RT-DETR with Attention-Free Mechanism: A Step towards Scalable and Generalizable Traffic Sign Recognition

Shiva Shankar Reddy[1], Midhunchakkaravarthy Janarthanan[2], Inam Ullah Khan[3]

[1]Faculty of Engineering and Built Science, Lincoln University of College, Malaysia.
[2]Faculty of AI Computing and Multimedia, Lincoln University of College, Malaysia.
[3]Lincoln University of College, Malaysia
[3]Multimedia University, Cyberjaya, Malaysia

Email: [1]pdf.shivareddy@lincoln.edu.my, [1]shiva.shankar591@gmail.com
[2]midhun@lincoln.edu.my,[3]inamullahkhan@mmu.edu.my

ABSTRACT

Computer vision-based road sign detection enhances intelligent transportation systems and driver assistance technology. Existing object detection models are insufficient for ensuring high accuracy under changing environmental conditions such as fog, low illumination, and rain. This research suggests improving road sign detection by utilising the proposed MLP + RT-DETR (Multilayer Perceptron + Real-Time Detection Transformer) with a modified backbone that replaces conventional attention mechanisms with an attention-free architecture. The aim is to enhance real-time detection efficiency while minimising computational burdens. To evaluate the performance of this approach, a test bed of road sign images was created under four weather conditions: normal, fog, low light, and rain. The new model compared favourably with baseline RT-DETR, demonstrating greater precision, recall, and detection resilience under adverse conditions. Performance metrics, such as the F1-Confidence Curve and Precision-Recall Curve, indicate that the attention-free backbone achieves a mean Average Precision (mAP) of 0.934, accompanied by a high F1-score of 0.93, demonstrating its ability to effectively balance precision and recall. Qualitative analysis confirms that the adopted model achieves better detection consistency, particularly under challenging visibility situations. These results demonstrate that eliminating attention mechanisms can enhance real-time object detection performance without compromising accuracy, thereby suggesting a potential solution for real-world intelligent transportation environments.

Keywords: Road Sign Recognition, RT-DETR, MLP, Attention-Free Architecture, Real-Time Object Detection

I. INTRODUCTION

Road infrastructure is the pillar of contemporary transportation, serving as the backbone for ensuring vehicle safety and free flow. It enables vehicles to detect and interpret traffic signs, helping drivers make informed decisions while enhancing overall road safety. Road signs follow a universal nomenclature and serve as an international language for all road users by communicating speed limits, changing lanes, crossing pedestrians, and dangers. Although crucial, the lack of detection of road signs is a significant problem that leads to traffic violations, congestion, and accidents.

The World Health Organization (WHO) states that more than 1.3 million people die each year globally as a result of road accidents, which are mainly caused by inadequate road infrastructure and absent or unreadable traffic signs [1]. In the United States, the Federal Highway Administration (FHWA) reports that approximately 22% of road accidents are caused by erroneous or missing road signs [2]. Additionally, road sign-related errors in India account for approximately 15% of road fatalities, resulting

in more than 9,000 casualties annually [3]. However, RSR (Road Sign Recognition) faces critical challenges, including environmental factors, physical barriers, and antiquated infrastructure. Poor weather conditions, such as fog, rain, and low lighting in the evening, significantly reduce the visibility of signs and recognition accuracy.

Additionally, degraded, ruined, or defaced road signs become illegible to both human motorists and AI-driven detection systems. Varying luminosity conditions, such as direct sunlight or shade, complicate sign detection and result in issues of overexposure or underexposure. Research has shown that nearly 30% of highway wrong-way driving accidents are caused by obstructed or poorly placed road signs, highlighting the importance of advanced maintenance methods and new-generation detection systems [4]. The worn-out Road Signs are shown in Figure 1.

Figure 1: Worn-Out Road Signs

Classic RSR techniques were based on rule-based image processing methods and manual inspections, which were time-consuming, inefficient, and prone to human error. As vehicles and road complexity increase, conventional methodologies are no longer sufficient [5]. Speed limit, the severity of the injury, time of the accident, drunk driving, month, weather conditions, road surface, and light conditions can predict accidents [6]. Helmetless motorcycle riders have their plates registered [7]. Face recognition is widely used due to its applications in image processing and biometrics [8], [9]. Manual pothole detection is time-consuming and prone to error [10]. The Motor Vehicle Safety Division reports an increase in the use of automobiles over the past 5-10 years [11]. A driver distraction detection system presents attention detection methods [12]. Although successful in laboratory settings, these approaches faltered in real-world applications due to occlusion, varying illumination, and sign degradation. Developing machine learning algorithms, such as Support Vector Machines (SVMs) and Random Forest (RF) classifiers, has enhanced classification performance; however, it still requires substantial feature engineering [13].

Convolutional Neural Networks (CNNs) are now the core of current RSR systems as they can capture spatial hierarchies of features, yielding high recognition rates [14]. LeNet, AlexNet, VGGNet (Visual Geometry Group Network), and DenseNet201 are deep learning architectures that have proven robust in traffic signs. For real-time applications, the You Only Look Once (YOLO) object detection framework has gained popularity due to its speed and efficiency [15]. Variants such as YOLOv4 and YOLOv5 have found wide application for high-frame-rate purposes, particularly in autonomous cars [16]. Further developments have been made in YOLOv8, improving accuracy in challenging settings [17]. Automated detection is a method to enhance traffic surveillance efficiency while reducing human labour [18]. It helps find the right spare component within an appropriate time frame and avoid avoidable delays associated with typical human-operated services, as opposed to unavoidable delays such as customs [19]. Object identification is one of the most common uses of computer vision. It is a way to find and shape things in the actual world. Even if there are many ways to find things, they aren't accurate or efficient enough [20]. For image processing-based apps to work well, they require precise components [21]. Transformer-based models such as the Swin Transformer have recently shown

promise in RSR by effectively handling long-range image dependencies. These models have been coupled with CNNs to enhance recognition accuracy without reducing computational efficiency [22]. Hybrid strategies, such as a combination of CNN with Optical Character Recognition (OCR), have also been suggested to enhance the detection of text traffic signs [23].

This paper introduces an advanced road sign recognition system that combines RT-DETR with an attention-free MLP-Mixer backbone to overcome the current methods. The main contributions of this work are environmental robustness—proposed model has high detection accuracy under different environmental scenarios, such as normal weather, fog, low lighting, and rain; degraded sign handling—the integration of RT-DETR's anchor-free detection with MLP-Mixer's holistic feature extraction enhances foreground-background separation and eliminates false detections; computational efficiency—by avoiding standard self-attention mechanisms, our model dramatically eliminates computational Overhead while ensuring high accuracy; real-time performance—RT-DETR's lightweight, fast detection structure, tuned with training parameters (batch size = 16, image size = 640), ensures efficient real-world execution; and scalability and adaptability—the presented model is flexible for different sign classes and geographic locations with minimal retraining. These developments render our method appropriate for autonomous driving applications and smart traffic monitoring systems, solving some of the most significant challenges in current RSR research.

II. RELATED WORK

RSR has advanced significantly due to the development of deep learning, computer vision, and artificial intelligence, which play a crucial role in autonomous vehicle driving. Conventional RSR methods have utilised colour and shape for feature extraction, which can fail in situations such as varying light conditions and weather changes. Zhao et al. [24] introduced TSD-YOLO, a hybrid model that integrated YOLO v8 with the Mamba framework to improve model performance and generalise across diverse datasets. Similarly, Ge et al. [25] proposed MambaTSR, a lightweight model that reduces computational complexity and maintains high accuracy in TSR tasks. Simultaneously, Han et al. [26] proposed EDN-YOLO (Efficient Dense Network-YOLO), a multi-scale detection method developed for complex environments by integrating EfficientVit (a Vision Transformer) into the backbone and decoupling the detection heads for improved performance. Benchmarking studies, such as Assemlali and Sael [27], have performed comparative analyses among various models and explained that, beyond detection and classification, there is a need for optimisation.

Yu et al. [28] utilised the diffusion model to optimise rural road environments, focusing on the role of deep learning in enhancing road safety. Le et al. [29] assessed traffic knowledge among teenagers by introducing e-learning platforms with games to introduce road safety awareness. Chen et al. [30] categorised the detection methods into conventional, deep learning, and hybrid models and concluded that hybrid models are better for real-time detection. Another detailed survey carried out by Triki et al. [31] has performed a comparative analysis with Raspberry Pi and Jetson Nano for efficient real-time processing. The work done in the Road Sign Detection domain is summarised in Table 1.

Table 1: Literature Review

REF	YEAR	MODELS USED	RESULTS	LIMITATIONS
[32]	2024	CNN with Attention Mechanism	The model has achieved better accuracy, reducing complexity.	Limited to a particular dataset may not generalise well.
[33]	2024	Learning-Assisted OCR (LAOCRP), Generic OCR (GOCR)	Works better for text-based signs	Struggles for degraded signs.
[34]	2024	YOLOv8	High precision, recall, and	Complex backgrounds reduce

			mAP	performance
[35]	2024	Gabor Wavelet Transform, HSV	High accuracy and precision at nights	It might not work better in extreme lighting conditions
[36]	2024	YOLOv8, EasyOCR	mAP of 0.824	Dataset-specific cannot generalise
[37]	2024	YOLOv5s, MobileNet	87.34% true prediction	Computational complexity due to ensemble learning
[38]	2024	CNN	Better classification and voice alerts.	Performance is highly dependent on the dataset.
[39]	2024	CNN(5 layer)	High accuracy	Cannot work on instances with different regions
[40]	2024	Reinforcement Learning (RLHF) with LeNet-5	Fine-tuned LeNet-5 has performed better than the original.	Requires high computational power and extensive feedback
[41]	2024	Multi-task Learning Architecture (TSI-arch)	Successfully generated texts for traffic signs	Works only on Chinese road signs
[42]	2024	YOLO v5	Achieved 90.09% F1-score and 87.55% mAP.	Limited to 15 sign classes, performance may vary in diverse conditions.
[43]	2024	CNN, DenseNet201	DenseNet201 performed better than CNN.	Real-time application needs optimisation.
[44]	2024	YOLOv8m	It performed better than YOLO v8.	May struggle in extreme weather conditions.
[45]	2025	CNN, YOLOv4	Gave accurate voice alerts	It may be confined to predefined signs that cannot be generalised.
[46]	2025	ERF-YOLO	84.2% mAP	Struggles with extreme weather conditions.

The next sections of this paper are framed as follows. Section 3 outlines the proposed model architecture, including modifications to improve Generalization across diverse environmental conditions, and describes the dataset in Section 4. The performance of the proposed model is assessed and compared with existing approaches using key evaluation metrics. The findings are discussed, highlighting accuracy, robustness, and improvements in computational efficiency. Finally, Section 5 concludes the paper by summarising key contributions, discussing potential limitations, and suggesting future research directions to further enhance road sign detection and recognition systems.

GAPS Identified

1. **Limited Generalisation:** From the above literature, it can be observed that existing models struggle to adapt across different environmental conditions and regions; therefore, we need to create a model that generalises even in adverse environmental conditions.
2. **Performance degradation:** Due to poor visibility and complex backgrounds, model performance is reduced, and the main challenges arise from worn-out signs, adverse weather conditions, and occlusions.
3. **High computational cost:** Many models require high processing power, making real-time detection difficult. Lack of optimisation leads to delays, limiting deployment in autonomous systems and intelligent traffic monitoring.

III. METHODOLOGY

OBJECTIVES

1. Enhance the model's adaptability and resilience for reliable road sign recognition across diverse geographic regions and challenging environmental conditions.
2. Ensure better detection of road signs, even in adverse weather conditions, with worn-out signs and complex backgrounds**.**
3. Maximise computational efficiency to allow real-time detection of road signs with low latency while maintaining high accuracy and reliability for intelligent transportation systems.

DATASET

The dataset is a private dataset generated from videos taken. The dataset comprises a total of 19,880 images, divided into:

- **Training Set**: 14,912 images with corresponding labels.
- **Validation Set**: 4,968 images with corresponding labels.

The images available in the dataset at various conditions are shown in Figure 2.

Normal Image

Image with Fog

Image with Low Light

Image with Rain

Figure 2. Images in different conditions

Rain, fog, and low-light conditions were simulated in digital images using OpenCV transformation techniques to expand the dataset. The rain effect was generated by adding a layer of randomly oriented streaks to simulate falling raindrops. Gaussian blur was applied to soften the streaks, and the overall brightness was reduced to enhance the perception of a rainy atmosphere. The fog effect was introduced by overlaying a semi-transparent white layer onto the image, mimicking natural fog dispersion and reducing visibility. The low-light condition was simulated by adjusting the image brightness using contrast scaling, making the scene appear dimmer. These modifications enable controlled weather simulations, which are valuable for training computer vision models and testing image processing algorithms in adverse conditions, as shown in Figure 2. The various classes in the dataset are shown in Table 2.

Table 2. Dataset Classes

 Cross Road Ahead		 Danger Ahead	
 Double Curve Ahead	 A gap in Median Ahead	 Go Slow	 Horn Prohibition
 Keep Left	 Left Curve Ahead	 Narrow Bridge Ahead	 National Speed Limit
 Pedestrian Crossing Ahead	 Railway Track Ahead	 Right Curve Ahead	 Road Work Ahead
 School Ahead	 Side Road Ahead	 Sound Horn	 Speed Breaker Ahead
 Speed Limit 20	 Speed Limit 30	 Speed Limit 40	 Speed Limit 60
 Speed Limit 80		 T Intersection Ahead	 U Turn Ahead

	 STOP		

PROPOSED MODEL

The proposed MLP + RT-DETR (Multilayer Perceptron + Real-Time Detection Transformer) is designed to enhance real-time road sign recognition, addressing limitations such as high computational complexity, attention bottlenecks, and environmental variability in the standard RT-DETR. The modifications involved replacing a Transformer-Based Backbone with an MLP Mixer to eliminate self-attention and reduce computational Overhead and integrating an efficient hybrid encoder to facilitate multi-scale feature fusion. These modifications enable faster, more accurate, and computationally efficient detection of road signs. The modified architecture is shown in Figure 3.

The convolutional feature extraction modules illustrated in Figure 4 are used in the RT-DETR CCFM module shown in Figure 3. Each module has a convolutional layer, followed by batch normalisation (BN) and Sigmoid Linear Unit (SiLU) activation function, as shown in Figure 4. The modules aid in the hierarchical extraction of features from the input and increase the model's ability to detect road signs irrespective of the environment.

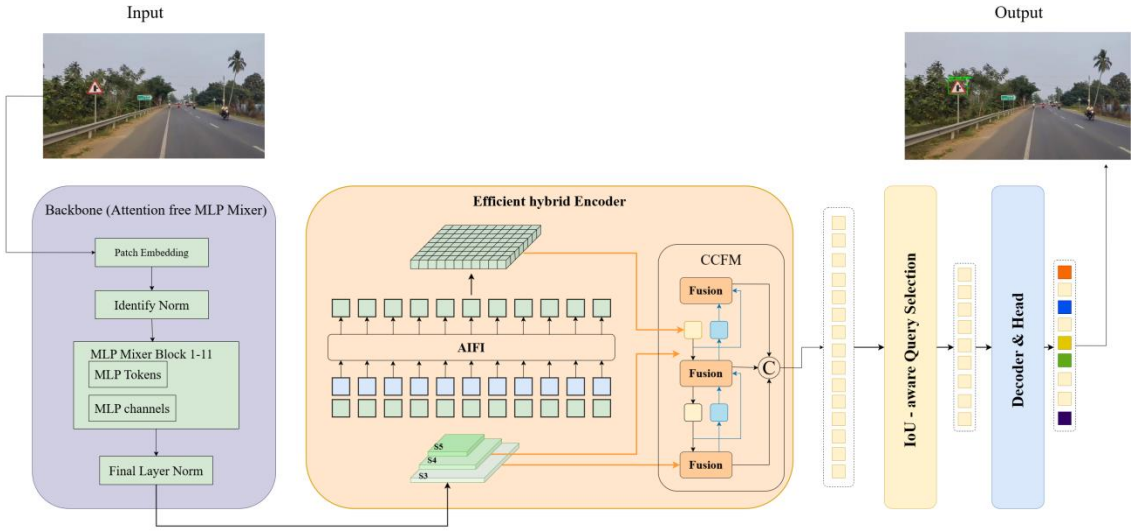


Figure 3. MLP + RT-DETR Architecture

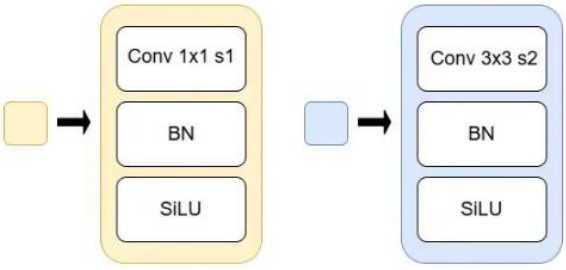


Figure 4. Convolutional Feature Extraction Modules in RT-DETR

The various blocks in Figure 3 are explained below.

A. Backbone: Attention-Free MLP Mixer for Feature Extraction

The original RT-DETR utilises a CNN-based backbone with self-attention layers, which increases memory usage and computation time, the primary gap identified. Also, transformers require global attention mechanisms that are not feasible for real-time object detection.

1. Patch Embedding

Before processing the input, the image is split into non-overlapping patches, which are then embedded into a lower-dimensional representation. A standard approach is using a convolutional layer:

$$\mathrm{Xp} = \mathrm{Conv2D}(\mathrm{X}, \mathrm{Wp}) \qquad Equation\ 1$$

Here, X represents the input image of shape B × H × W × C. Where B → Batch size, H → Height, W → Width, C → Number of channels, Wpis the convolutional kernel for patch embedding, and Xpis the resulting **patch token** of shape B×N×D. Where N is the number of patches, and D is the embedding dimension,

2. Identify Norm (Normalisation Layer)

Normalisation is applied to stabilise training and improve convergence. A common choice is Layer Normalization (LN)

$$LN(X) = (\mathrm{X} - \mu) \,/\, (\sigma + \varepsilon) * \gamma + \beta \qquad Equation\ 2$$

Where μ and σ are the mean and standard deviation per feature, (γ and β) are learnable affine parameters.

It helps to keep activations well-scaled.

3. MLP Mixer Blocks (1 to 11)

The MLP Mixer replaces self-attention with MLP-based transformations. Each block consists of two fully connected layers. Token Mixing MLP (operates along spatial dimensions) and Channel Mixing MLP (operates along feature dimensions).

Token Mixing MLP

Processes each feature independently across spatial tokens:

$$\mathrm{X}' = \mathrm{X} + \mathrm{W2} * \sigma(\mathrm{W1} * \mathrm{XT}) \qquad Equation\ 3$$

Where X is the input token embeddings, (W1 and W2) are learnable weight matrices, and σ is an activation function (e.g., GELU or ReLU).

Channel Mixing MLP

Applies to an MLP on the feature channels:

$$Y = Y + w4 * \sigma(\mathrm{w3} * \mathrm{Y}) \qquad Equation\ 4$$

Where Y = LayerNorm(X') is the input from the token-mixing step, (W3 and W4) are learnable parameters.

4. Final Layer Normalisation

A final normalisation is applied to the Mixer output before passing it to the encoder.

$$\mathrm{X_{final}} = \mathrm{LN}(\mathrm{X_{MLP}Mixer\ Output}) \qquad Equation\ 5$$

This block reduces computational complexity by removing self-attention and spatial components. A feature-level dependency was captured due to improved token and channel mixing, ensuring that feature representations remain stable before passing them to the next stage.

B. Efficient Hybrid Encoder

In this module, spatial and semantic information were combined to enhance feature extraction, thereby increasing the model's robustness.

1. Multi-Scale Feature processing

The encoder takes multi-scale feature maps. S_3,S_4 and S_5they need to be processed for transformation using the convolution layer from the backbone.

$$F_i = Conv(S_i, W_i) \quad Equation\ 6$$

where S_i different scale feature maps, W_i represents the weights of convolutions used for transformations, F_i is the feature processed at each scale.

2. Adaptive Interactive Feature Integration (AIFI)

Merges multi-scale features obtained in the previous step dynamically, allowing information to flow between different spatial scales. These transformations can be denoted as

$$AIFI(F) = \sigma(W_a F + b_a) \quad Equation\ 7$$

WhereW_a and b_a are parameters that are learned, σ is the activation function (non-linear), $AIFI(F)$ is the updated feature representation after integration.

3. Cross-Scale Complementary Fusion Module (CCFM)

In this block, Multi-scale integrated features from the previous step are taken through multiple fusion stages and enhanced with feature fusion between different spatial and channel-wise components. Cross-channel modulation is used to refine the interactions between features.

$$F_{fusion} = W_f.\left(F_i + \sum_{j \neq i} \alpha_{ij} F_j\right) \quad Equation\ 8$$

where, α_{ij}represents attention weights for different feature scales, W_f Represents fusion weight (learnable), F_{fusion} represents feature representation after final fusion.

C. **IoU-Aware Query Selection**

The IoU (Intersection over Union)-Aware Query Selection module selects the most relevant object queries based on IoU scores, measuring the overlap between predicted and ground-truth bounding boxes. This module ensures that the most confident queries are used for decoding, improving detection performance.

The IoU calculation is done using.

$$\mathrm{IoU} = \frac{|\mathrm{B_p} \cap \mathrm{B_g}|}{|\mathrm{B_p} \cup \mathrm{B_g}|} \quad Equation\ 9$$

where $\mathrm{B_p}$ is predicting bounding box and $\mathrm{B_g}$ is ground truth.

The IoU prediction for each query is done using

$$IoU_q = \sigma(W_q.F_q + b_q) \quad Equation\ 10$$

where F_q is feature expression for query q, (W_q and b_q) are the parameters that can be trained.

In an Efficient Hybrid Encoder, the multi-scale features from the backbone are sent to various blocks, where refined multi-scale feature representations that encode spatial and contextual information are generated, making them suitable for accurate object query selection and detection in subsequent stages.

D. Decoder and Head

The decoder & head module refine the queries selected in the previous stage and generate the final object detection outputs. The decoder updates the feature interaction using the MLP mixer instead of self-attention, which reduces computational cost. The refined query is then passed to the detection head.

The class prediction is calculated using the equation.

$$\hat{c} = Softmax(W_c Q^n) \qquad Equation\ 11$$

The Bounding Box regression is calculated using the equation

$$\hat{b} = \sigma(W_b Q^n) + b_0 \qquad Equaution\ 12$$

The Confidence score for each detection is calculated using the equation

$$s = Sigmoid(W_s Q^n) \qquad Equation\ 13$$

After all, Non-Maximum Suppression (NMS) filters the redundant detections and ensures accurate localisation and classification of objects.

ALGORITHM: **MLP + RT-DETR**	
INPUT: Raw image III OUTPUT: Bounding box with the road sign name	
Step 1:	The raw image III (an RGB image with road signs and background) is converted into non-overlapping patches using Equation 1.
Step 2:	Normalise the input patches and perform Token Mixing and Channel Mixing. Once again, the final normalisation layer is applied before sending it into the encoder using equations 2-5.
Step 3:	Multi-scale feature maps S3, S4, and S5 are extracted from the backbone network using equation 6.
Step 4:	The extracted features are passed into the **AIFI module**, which ensures an effective fusion of multi-scale features in equation 7.
Step 5:	The output from AIFI is sent to the **CCFM**, where multi-level feature interactions are performed using equation 8.
Step 6:	The final refined features from the **CCFM** are sent forward for **IoU-aware query selection**.
Step 7:	Feature maps are encoded into query representations. IoU scores are calculated for potential object queries. Queries are filtered, and irrelevant ones are discarded using equations 9-10.
Step 8:	The decoder processes the IoU-selected queries, which predict object class probabilities and bounding box coordinates using equations 11-13.
Step 9:	The final detection boxes and labels are produced.

The process begins by splitting the raw image into non-overlapping patches, and then normalisation and token-channel mixing are applied to enhance the features. Multi-scale features are extracted from the backbone and combined with the AIFI module for effective representation. These enhanced features are then processed in the CCFM for multi-level feature interactions. IoU-aware query selection eliminates irrelevant queries, and the remaining queries are decoded to estimate object class probabilities and bounding box coordinates, yielding precise road sign detection.

IV. RESULT ANALYSIS

Table 3 compares the detection performance of objects between the standard RT-DETR model and the proposed RT-DETR with an attention-free backbone mechanism for road sign recognition. The input image contains various environmental conditions, including images with normal lighting, low light, fog,

and rain. In Table 3, each row corresponds to a particular light condition. The first column displays the original input image, the second column shows the detection results using RT-DETR, and the third column presents the results obtained using RT-DETR with the attention-free backbone.

Table 3: Comparison of RT-DETR and RT-DETR with Attention-Free Backbone

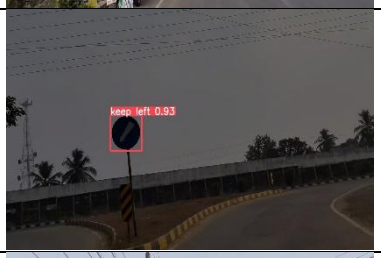
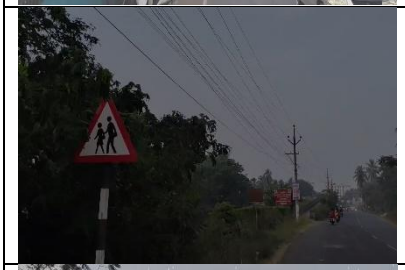
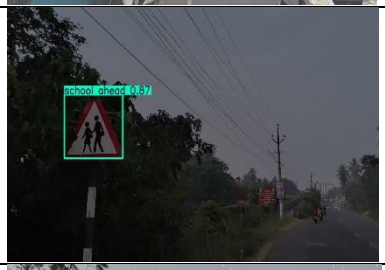
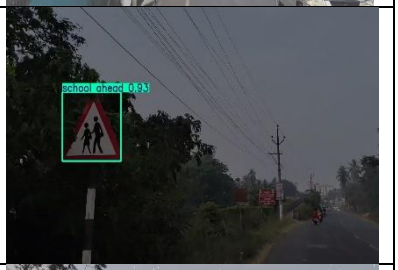
Input	RT-DETR	RT-DETR with AttentionFree Backbone
		
		
		
		
		

Table 3 shows that both models have successfully identified the road signs, but differences in detection accuracy and consistency are noticeable. The RT-DETR with an attention-free backbone performs better, especially under low-light and foggy scenarios, where the baseline RT-DETR often fails. The bounding boxes from the attention-free model look more accurate and robust under different environmental conditions. Moreover, the attention-free backbone offers stable detection quality in

adverse conditions, such as rain and darkness, indicating an improved feature extraction ability. These findings demonstrate that eliminating attention mechanisms in the backbone can enhance efficiency without necessarily affecting detection performance. Figure 5, F1-Confidence Curve, illustrates the trade-off between the F1-score and the confidence threshold for all identified classes. The F1 Score is an accuracy-recall measure that provides a good estimate of model performance. It is used to find the best confidence Threshold at which the F1-Score is maximised, and there is the most effective trade-off between false positives and false negatives.

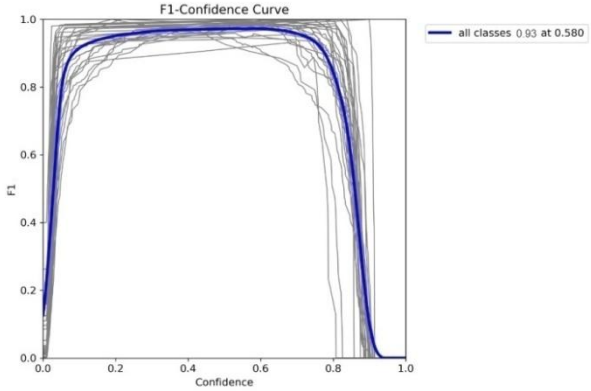


Figure 5. F1-Confidence Curves

The curve shows the F1-score is uniformly high for a vast range of confidence values before decreasing rapidly as confidence approaches 1.0. The best F1-score is 0.93 at a threshold of 0.58, which provides the optimal trade-off between recall and precision for both classes. The number of grey curves provides a hint of class-wise differences in performance, but the overall model achieves excellent detection accuracy with the optimal threshold. In Figure 6, the Precision-Recall (PR) Curve below illustrates the trade-off between precision and recall at various confidence levels. The PR curve illustrates the model has high precision at almost all recall values, with a slight decrease in the extreme recall area. The mean Average Precision (mAP@0.5) of 0.934 indicates excellent overall detection performance. The fact that there are several gray curves shows per-class differences in precision and recall, but the overall trend (in blue) implies stable and reliable detection. In Figure 7, the Recall-Confidence Curve is a significant model that predicts confidence reliability measurement metrics. It is applied to analyse whether the model is well-balanced between recall and confidence scores, particularly in object detection. Greater recall at a particular confidence level indicates improved performance in detecting all objects of interest and minimising false detections.

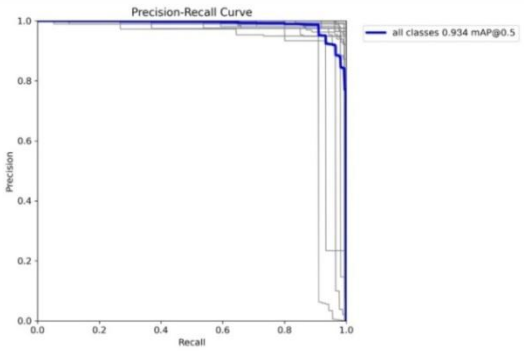


Figure 6. Precision-Recall (PR) Curves

The Recall-Confidence Curve of the model under consideration has a high recall value of 0.91 at a confidence level of 0.83, indicating that the model can correctly label road signs without compromising recall and precision. The curve is steeply decreasing with high confidence values, indicating that the

model is highly calibrated with few false negatives. Demonstrates the model's strength under poor environmental conditions, degraded signs, and occluded backgrounds, as required by this research.

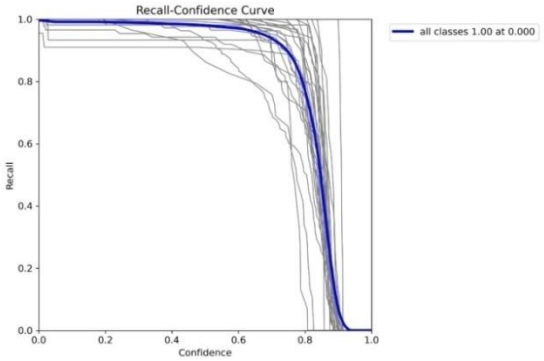


Figure 7. Recall-Confidence Curve

In Figure 8, the precision-confidence curve is a measure of performance where precision is conveyed through confidence levels. That kind of excellent model would be one where the performance remains good even with very low confidence levels, and hence, few false positives are implied. Precision is the ratio of properly labelled positive instances to the total, and recall is the model's capacity to identify all the corresponding cases. The curve is most important for object detection models, in which precision and recall must be in equipoise at all costs. The mean Average Precision (mAP) at an Intersection over Union (IoU) of 0.5 is also available, providing a general idea of how accurately the model identifies road signs.

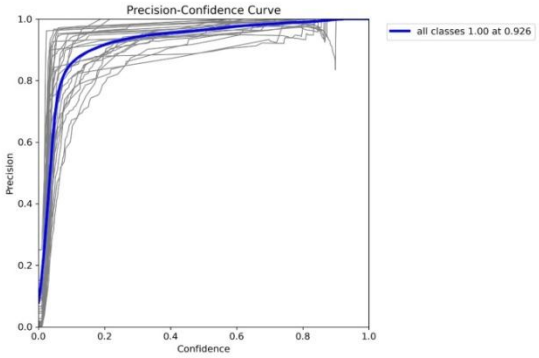


Figure 8. Precision-Confidence Curve

The model's Precision-Confidence Curve is 0.92 at a confidence level of 0.85, which verifies the model's accuracy in detecting road signs with a low number of false positives. The steeply increasing slope, followed by a flat, high-precision rate, shows that the model is well-calibrated to provide consistent results even in poor environmental conditions and cluttered scenes. These results align with the study's objectives, justifying the model's efficiency and effectiveness.

To compare the efficacy of the presented model, we compare its performance with previous models of road sign detection. A comparison is made against major performance measures like mean Average Precision (mAP) and F1-score, as shown in Table 4.

Table 4: Comparison with previous models

Ref	Model Used	Results
[24]	YOLOv8, EasyOCR	mAP of 0.824
[25]	YOLOv5s, MobileNet	87.34% true prediction
[30]	YOLO v5	Achieved 90.09% F1-score and

		87.55% mAP.
[34]	ERF-YOLO	84.2% mAP
[47]	YOLOv8, CNN, EasyOCR	mAP: 82.4%
[48]	Deep CNN	mAP@.5: 65%
[49]	YOLOv5s6, YOLOv8s	(YOLOv5s6), 76.2%
[50]	R-CNN	Accuracy: 90%
	Proposed Model	**0.93 mAP**

The results shown in Table 4 indicate that the present model is superior to existing models, as it boasts an improved mAP (0.93) while maintaining real-time efficiency. In contrast to YOLO-based approaches, the present model offers improved accuracy and stability, with favourable performance in poor environmental conditions. The improvement is attributed to the simplified architecture and removal of attention mechanisms that conserve computational Overhead while improving detection performance. These results verify the efficiency of the RT-DETR with an attention-free backbone as a real-time road sign detection solution under different conditions.

V. CONCLUSION

The proposed RT-DETR model, with an attention-free backbone, best aligns with the primary objectives outlined in this study. We have improved its ability to generalise over different regions and weather situations by removing attention mechanisms and streamlining the model structure. The output indicates that the model is highly effective in detecting objects under various conditions, including normal, foggy, low-light, and rainy conditions, demonstrating its resilience to real-world environmental variability. Furthermore, the enhanced model is better equipped to detect road signs in challenging circumstances, including adverse weather conditions, damaged signs, and complex backgrounds. The qualitative and quantitative evaluations confirm the new backbone gains robustness at fewer misdetections and false negatives. One of this paper's key contributions is balancing computational complexity without losing precision. The Precision-Recall Curve and F1-Confidence Curve indicate that the model achieves a very high mAP of 0.934 and an F1-score of 0.93, which effectively demonstrates its precision and recall balance. Removing attention mechanisms significantly enhances real-time performance, and the model is more likely to be deployed in real-time traffic monitoring and intelligent transportation systems. In short, the proposed modifications significantly enhance efficiency, robustness, and Generalization, enabling accurate and real-time detection of road signs in dynamic and challenging situations. The research adds to smart transport by compelling road sign detection systems to become faster, more robust, and ready for real-world implementation.

REFERENCES

[1]. World Health Organization. Global status report on road safety 2023: summary. World Health Organization; 2023 Dec 7.

[2]. Federal Highway Administration. Traffic Control Devices and Road Safety Report. FHWA, 2024.

[3]. Ministry of Road Transport and Highways, India. Road Accidents in India 2023. Government of India, 2024.

[4]. Liu Q, Huang J, Zhao X, Li J, Chen Y, Wu C. Study on optimisation design of guide signs in dense interchange sections of eight-lane freeway. Accident Analysis & Prevention. 2025 Jan 1;209:107828.

[5]. Triki N, Karray M, Ksantini M. A comprehensive survey and analysis of traffic sign recognition systems with hardware implementation. IEEE Access. 2024 Sep 12.
[6]. Yasaswini L, Mahesh G, Shankar RS, Srinivas LV. Identifying road accidents severity using convolutional neural networks. International Journal of Computer Sciences and Engineering. 2018 Jul 6;6(7):354.
[7]. Sahnkar RS, Raminaidu CH, Ravibabu D, Gupta VM. A survey to raise the awareness of road accidents due to not-wearing helmet. Int. J. Ind. Eng. Prod. Res. 2020 Sep 10;31(3):367-77.
[8]. Reddy SS, Gupta VM, Srinivas LV, Swaroop CR. Methodology for eliminating plain regions from captured images. Int J Artif Intell ISSN.;2252(8938):1359.
[9]. Shankar RS, Raminaidu C, Rajanikanth J, Raghaveni J. Frames extracted from video streaming to recognition of face: LBPH, FF and CNN. InAIP Conference Proceedings 2023 Dec 15 (Vol. 2901, No. 1). AIP Publishing.
[10]. Mahesh G, Shankar RS, Rao VM. An object detection framework and deep learning models used to detect the potholes on the streets. In2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) 2024 May 16 (pp. 1-7). IEEE.
[11]. Reddy SS, Rao VR, Voosala P, Nrusimhadri S. You only look once model-based object identification in computer vision. IAES International Journal of Artificial Intelligence (II-AI). 2024 Mar;13(1):827-38.
[12]. Shankar RS, Neelima P, Priyadarshini V, Chigurupati SR. An approach to classify distraction driver detection system by using mining techniques. Indonesian Journal of Electrical Engineering and Computer Science. 2022 Sep;27(3):1670-80.
[13]. Taki Y, Zemmouri E. A Lightweight Model for Traffic Sign Recognition Based on Attention Mechanism. In2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS) 2024 Oct 23 (pp. 1-6). IEEE.
[14]. Yin K. Application of deep learning in traffic sign detection and recognition system. In2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE) 2023 Aug 18 (pp. 1036-1042). IEEE.
[15]. Luo S, Wu C, Li L. Detection and recognition of obscured traffic signs during vehicle movement. Ieee Access. 2023 Nov 1;11:122516-25.
[16]. Yang C, Zhuang K, Chen M, Ma H, Han X, Han T, Guo C, Han H, Zhao B, Wang Q. Traffic sign interpretation via natural language description. IEEE Transactions on Intelligent Transportation Systems. 2024 Jul 25.
[17]. Wicaksono I, Negara MA, Laagu MA, Herdiyanto DW, Setiabudi D, Rahardi GA. Enhancing Traffic Signs Recognition Systems Through Gabor Feature Extraction Techniques. In2024 IEEE 2nd International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT) 2024 Nov 22 (pp. 284-289). IEEE.
[18]. Reddy SS, Khan IU. Weather and Monsoon Resilient Pothole Detection: A YOLO Based Real-Time Application for Diverse Road Conditions. SGS-Engineering & Sciences. 2025 May 7;1(1).
[19]. Shiva Shankar R, Devareddi R, Mahesh G, MNSSVKR Gupta V. Develop a smart data warehouse for auto spare parts autonomous dispensing and rack restoration by using iot with dds protocol. InComputer Networks, Big Data and IoT: Proceedings of ICCBI 2021 2022 May 22 (pp. 879-895). Singapore: Springer Nature Singapore.
[20]. Shankar RS, Srinivas LV, Neelima P, Mahesh G. A framework to enhance object detection performance by using YOLO algorithm. In2022 international conference on sustainable computing and data communication systems (ICSCDS) 2022 Apr 7 (pp. 1591-1600). IEEE.
[21]. Reddy SS, Rao VV, Sravani K, Nrusimhadri S. Image quality evaluation: evaluation of the image quality of actual images by using machine learning models. Bulletin of Electrical Engineering and Informatics. 2024 Apr 1;13(2):1172-82.

[22]. Kirubakaran D, Lakshmisridevi S, Ramal PJ, Rajeswari M, Rani AJ, Pandi VS. Driving into the Future: Artificial Intelligence based Traffic Sign Recognition using Learning Assisted OCR Principle. In2024 International Conference on Automation and Computation (AUTOCOM) 2024 Mar 14 (pp. 262-266). IEEE.
[23]. Vimalambikaipakan G, Amarasinghe C, Rajapaksha T, Thayananthan T, Mariyathas J. Traffic sign recognition and auditory alert system for sri lankan drivers using deep-learning. In2024 International Research Conference on Smart Computing and Systems Engineering (SCSE) 2024 Apr 4 (Vol. 7, pp. 1-5). IEEE.
[24]. Zhao R, Tang SH, Shen J, Supeni EE, Rahim SA. Enhancing autonomous driving safety: a robust traffic sign detection and recognition model TSD-YOLO. Signal Processing. 2024 Dec 1;225:109619.
[25]. Ge Y, Chen Z, Yu M, Yue Q, You R, Zhu L. MambaTSR: You only need 90k parameters for traffic sign recognition. Neurocomputing. 2024 Sep 28; 599:128104.
[26]. Han Y, Wang F, Wang W, Zhang X, Li X. EDN-YOLO: Multi-scale traffic sign detection method in complex scenes. Digital Signal Processing. 2024 Oct 1; 153:104615.
[27]. Hamza AS, Nawal SA. Traffic sign classification using deep learning comparative study. Procedia Computer Science. 2024 Jan 1;233:939-49.
[28]. Yu B, Zhu Z, Chen Y, Wang J, Gao K, Qian X. A Diffusion Model-based Intelligent Optimization Method of Rural Road Environments. International Journal of Transportation Science and Technology. 2025 Feb 3.
[29]. Le HN, Cuenen A, Trinh TA, Janssens D, Wets G, Brijs K. Investigating the immediate and mid-term effect of a gamified e-learning platform for the enhancement of traffic knowledge and skills among Vietnamese adolescents operating powered two-wheelers. Journal of Safety Research. 2024 Sep 1; 90:62-72.
[30]. Chen H, Ali MA, Nukman Y, Abd Razak B, Turaev S, Chen Y, Zhang S, Huang Z, Wang Z, Abdulghafor R. Computational Methods for Automatic Traffic Signs Recognition in Autonomous Driving on Road: A Systematic Review. Results in Engineering. 2024 Dec 7:103553.
[31]. Triki N, Karray M, Ksantini M. A comprehensive survey and analysis of traffic sign recognition systems with hardware implementation. IEEE Access. 2024 Sep 12.
[32]. Taki Y, Zemmouri E. A Lightweight Model for Traffic Sign Recognition Based on Attention Mechanism. In2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS) 2024 Oct 23 (pp. 1-6). IEEE.
[33]. Kirubakaran D, Lakshmisridevi S, Ramal PJ, Rajeswari M, Rani AJ, Pandi VS. Driving into the Future: Artificial Intelligence based Traffic Sign Recognition using Learning Assisted OCR Principle. In2024 International Conference on Automation and Computation (AUTOCOM) 2024 Mar 14 (pp. 262-266). IEEE.
[34]. Choudhary N, Sharma R, Upadhyay D, Verma A, Jain V. Enhanced Traffic Sign Recognition Using Advanced YOLOv8 Model. In2023 4th International Conference on Intelligent Technologies (CONIT) 2024 Jun 21 (pp. 1-4). IEEE.
[35]. Wicaksono I, Negara MA, Laagu MA, Herdiyanto DW, Setiabudi D, Rahardi GA. Enhancing Traffic Signs Recognition Systems Through Gabor Feature Extraction Techniques. In2024 IEEE 2nd International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT) 2024 Nov 22 (pp. 284-289). IEEE.
[36]. Xian CL, Sheikh UU, Bakar SA. Malaysia Traffic Sign Recognition for Autonomous Vehicles with Textual Information Using Computer Vision. In2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA) 2024 Sep 3 (pp. 1-6). IEEE.
[37]. Wang LJ, Suwattanapunkul T, Thalauy J, Jansengrat P. Research on Traffic Sign Detection and Recognition System Using Deep Ensemble Learning. In2024 6th International Conference on Computer Communication and the Internet (ICCCI) 2024 Jun 14 (pp. 61-66). IEEE.

[38]. Priya DD, Chinnasamy P, Ayyasamy RK, Kiran A, Jalil NB, Sangodiah A. Smart Drive Safe: Harnessing CNN for Enhanced Traffic Sign Recognition and Voice Alerts. In2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC) 2024 Aug 7 (pp. 1915-1920). IEEE.
[39]. PP HK, Ravindran S, Vijean V. Traffic Sign Classification for Road Safety using CNN. In2024 International Conference on Emerging Systems and Intelligent Computing (ESIC) 2024 Feb 9 (pp. 462-466). IEEE.
[40]. Dai S, Chen Z, Hu Z, Zhou L. Traffic Sign Classification with Reinforcement Learning from Human Feedback. In2024 5th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC) 2024 Nov 8 (pp. 169-173). IEEE.
[41]. Yang C, Zhuang K, Chen M, Ma H, Han X, Han T, Guo C, Han H, Zhao B, Wang Q. Traffic sign interpretation via natural language description. IEEE Transactions on Intelligent Transportation Systems. 2024 Jul 25.
[42]. Vimalambikaipakan G, Amarasinghe C, Rajapaksha T, Thayananthan T, Mariyathas J. Traffic sign recognition and auditory alert system for sri lankan drivers using deep-learning. In2024 International Research Conference on Smart Computing and Systems Engineering (SCSE) 2024 Apr 4 (Vol. 7, pp. 1-5). IEEE.
[43]. Sneha P, Nagarathna N. Traffic Sign Recognition for Automatic Vehicles. In2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE) 2024 Feb 16 (pp. 534-538). IEEE.
[44]. Zhou X, Huang J, Zhou T. Traffic Sign Recognition under Autonomous Driving Based on YOLOv8 Enhancement. In2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI) 2024 May 24 (pp. 697-703). IEEE.
[45]. Bhuvaneshwari B, Abbijeet R, Singh S, Priyadharshini CS. Traffic Sign Classification And Voice Alert System Using Convolutional Neural Network. Procedia Computer Science. 2025 Jan 1;252:995-1004.\
[46]. Sun Y, Li X, Zhao D, Wang QG. Evolving Traffic Sign Detection via Multi-scale Feature Enhancement, Reconstruction and Fusion. Digital Signal Processing. 2025 Jan 27:105028.
[47]. Xian CL, Sheikh UU, Bakar SA. Malaysia Traffic Sign Recognition for Autonomous Vehicles with Textual Information Using Computer Vision. In2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA) 2024 Sep 3 (pp. 1-6). IEEE.
[48]. HD NU, Aravind BN, Yashwanth N, Ruj M. Real-Time Traffic Sign Detection and Recognition for Enhanced Road Safety in Autonomous Driving: A Deep CNN-Based Approach. In2023 Fourth International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) 2023 Dec 8 (pp. 1-4). IEEE.
[49]. Suwattanapunkul T, Wang LJ. The efficient traffic sign detection and recognition for taiwan road using yolo model with hybrid dataset. In2023 9th International Conference on Applied System Innovation (ICASI) 2023 Apr 21 (pp. 160-162). IEEE.
[50]. Gade PS, Shinde SL, Kolapkar SB, Walte AV, Sharma SK, Chintamani RD. Traffic Sign Board Recognition and Alert System Using R-CNN. In2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM) 2023 Dec 12 (pp. 01-05). IEEE.