Smart Detection Systems for Traffic Sign Recognition: Pioneering Road Safety Solutions with YOLO Models

Abhay Divakar Patil[a], Sakshi Sonoli[b], Sanjeet Kalasannavar[c], Shreyas Rajeev Patil[d]
and Prema T. Akkasaligar[e]

Department of Computer Science and Engineering, KLE Tech, University, Dr. MSSCET, Belagavi, India

Keywords: ADAS, YOLO, Self Attention Mechanisms, Preprocessing Techniques, Data Augmentation, Segmentation, Transformer Based Architectures.

Abstract: Traffic sign detection is crucial for autonomous vehicles (AVs) and advanced driver-assistance systems (ADAS). Detecting Indian traffic signs is challenging due to diverse designs, environmental conditions, and multilingual content. This paper proposes a method integrating self-attention mech- anisms with You Only Look Once (YOLO) models, such as YOLO11n,YOLO11s, and YOLOv8n. The self-attention modules enhance feature extraction and localization in complex scenarios like occlusions and low-light conditions. Preprocessing techniques, including segmentation and data augmentation, play a significant role in improving model performance. This research contributes to the advancement of traffic sign detection in real-world scenarios and provides a foundation for future innovations, such as transformer-based architectures.

1 INTRODUCTION

Traffic Sign Detection plays a crucial role in both autonomous vehicles and traffic management systems. By utilizing both visual and infrared technology, these systems can reliably recognize and interpret road signs even in complex conditions such as poor lighting, adverse weather, or cluttered environments. The autonomous vehicles and traffic systems become more prevalent, ensuring their ability to detect road signs accurately is essential for maintaining road safety, reducing accidents, and enhancing traffic flow. Failure to detect signs in time could lead to serious consequences, making reliable traffic sign detection a critical need for future transportation systems.

Apparently one person dies from a traffic accident every 24 seconds worldwide, which amounts to roughly 1.3 million fatalities per year. An additional 20-50 million individuals get serious injuries (Dewi, Chen, et al. 2024). Road accidents are the major cause of mortality for young people and children between the ages of 5 years and 29 years. It ranks as the eighth most common cause of death globally. Developing countries have a disproportionate share of the economic and social

costs, since traffic-related mishaps can cost them up to 3% of their GDP. Despite having fewer vehicles, they are responsible for 93% of road fatalities worldwide (Flores-Calero, Astudillo, et al. 2024)*. Over half of these fatalities are among vulnerable road users, including motorcyclists, cyclists, and pedestrians. Over 90% of deaths worldwide occur in lower- and middle-income nations, making them the ones that suffer the most.

Even though just 7% of crash deaths occur in highincome nations, issues including speeding, distracted driving, and a growing senior population still exist. This man-made epidemic can be considerably reduced by investing in smart traffic management and preventative measures, which will be far less expensive than the losses to society. With improved accuracy and resilience, the suggested model tackles traffic sign identification issues, makes it perfect for real-time applications such as autonomous electronic vehicle (EV) navigation. Its sophisticated design guarantees dependability, raising road safety standards and providing a scalable response to contemporary transportation demands.

Current state-of-the-art solutions use advanced computer vision techniques, including deep learning

[a] https://orcid.org/0009-0004-7651-905X
[b] https://orcid.org/0009-0000-7907-355X
[c] https://orcid.org/ 0009-0000-7056-1717
[d] https://orcid.org/0009-0002-5164-5186
[e] https://orcid.org/0000-0002-2214-9389

Patil, A. D., Sonoli, S., Kalasannavar, S., Patil, S. R. and Akkasaligar, P. T.
Smart Detection Systems for Traffic Sign Recognition: Pioneering Road Safety Solutions with YOLO Models.
DOI: 10.5220/0013608900004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 65-74
ISBN: 978-989-758-763-4

models like Convolutional Neural Networks (CNNs), combined with sensor fusion methods (e.g., LIDAR and infrared imaging) to identify and classify traffic signs under various conditions. These methods have shown success in controlled environments but face challenges in more complex real-world scenarios.

Despite of recent advances, existing solutions struggle in scenarios involving occlusion, varying weather conditions, and changes in lighting. Furthermore, realtime processing with high accuracy remains a challenge, particularly in highly dynamic environments such as highways or urban streets.

To address these challenges, our study focuses on developing a robust machine learning model for traffic sign detection, capable of handling diverse and complex scenarios. By leveraging advanced techniques and a comprehensive dataset, we explore variations of the You Only Look Once (YOLO) object detection framework, including YOLO11n, YOLO11s, and YOLOv8. YOLO11n and YOLO11s incorporate self-attention mechanisms to enhance feature extraction, while YOLOv8 serves as a baseline for performance comparison. These models are designed to improve the accuracy, speed, and reliability of traffic sign detection, making them suitable for real-time applications and contributing to enhanced road safety standards.

2 LITERATURE SURVEY

A key component of intelligent transportation systems, traffic sign detection and recognition (TSDR). It is essential for both unmanned vehicle operations and road safety. Early approaches has limits in terms of their adaptability in the real world because they depended on conventional computer vision techniques like color segmentation and analysis. TSDR is revolutionized by the advent of machine learning and deep learning, which allowed models to acquire intricate patterns and attain great accuracy in dynamic settings. Resilience under various circumstances still exist due to notwithstanding progress, issues including edgecase accuracy, computing effectiveness. This overview examines the development of TSDR, examining conventional methods, contemporary developments, and continuing chances for creativity in safe traffic sign recognition.

In (Mannan et al., 2019), propounded a method for bifurcating ordinary site visitors symptoms the usage of adaptive models and switch learning. They employed a modified Gaussian mixture model and CNNs to decorate detection and popularity accuracy. The system adapts to image distortion however calls for specialized hardware like GPUs for green processing. In(Lopez-Montiel et al., 2021), evaluated deep gaining knowledge of-based totally visitors detection the usage of MobileNet v1 and ResNet50 with SSD. TPUs significantly outperformed GPUs,implying faster processing and higher detection accuracy, although ResNet50 had high memory needs. In (Boujemaa et al., 2021), added the ATTICA dataset for Arabic visitors symptoms. R-FCN performed the first-class overall performance in detecting and spotting traffic symptoms, although demanding situations consisting of class conflicts and language-particular issues had been cited. In (Triki et al., 2024), assessed TSR structures on Raspberry Pi and Nvidia Jetson Nano. The Jetson Nano exhibited higher detection accuracy and velocity, but each structures faced limitations on low-give up gadgets. In (Greenhalgh and Mirmehdi, 2015), evolved a site visitors sign text detection machine employing MSER functions and HSV coloration space with OCR for textual content popularity. It performed high detection accuracy but multiplied complexity may want to have an effect on performance.

In (Youssef et al., 2016), devised a cutting-edge method for identifying traffic signs that utilizes color segmentation techniques along with HOG and CNNs. Advantage: This approach allows for rapid identification of road signs. Disadvantage: It necessitates preprocessing to effectively narrow down the search parameters. In (Avramović et al., 2020), developed a detection framework for traffic signs within high-definition imagery, employing several YOLO architectures. Advantage: This system provides high accuracy in detecting signs under HD conditions. Disadvantage: It demands considerable computational resources for efficient parallel processing. In (Fleyeh and Dougherty, 2006), examined a range of strategies for recognizing road signs, focusing on techniques that analyze both color and shape. Advantage: Their work encompasses a broad spectrum of detection techniques. Disadvantage: A lack of standardization for color extraction methods remains a challenge. In (Tabernik and Skočaj, 2020), introduced a sophisticated system for large-scale traffic sign detection using an enhanced version of Mask R-CNN. Advantage: This method achieves high levels of precision in sign recognition. Disadvantage: It faces difficulties in detecting smaller signs in intricate environments. In (Flores-Calero et al., 2024a), With its remarkable accuracy and real-time speed, the YOLO algorithm transforms traffic sign identification and is perfect for ADAS and autonomous driving applications. Its ability to effectively detect objects even with sparse data is its strength. But issues include trouble seeing tiny or hidden signals and performance lapses in bad weather or low-resolution situations point to areas that need improvement. Because of its accessibility and agility, YOLO serves as a foundation for safer, smarter roads and encourages continuous innovation to go over its practical constraints.

In (Flores-Calero et al., 2024b), a systematic review was conducted on the application of YOLO object detection algorithms for traffic sign detection and recognition. The study analyzed 115 research papers from 2016–2022, identifying three main applications: road safety, ADAS, and autonomous driving. The GTSDB and GTSRB datasets were frequently utilized for benchmarking. The most commonly used hardware included Nvidia RTX 2080 and Titan Tesla V100 GPUs, alongside Jetson Xavier NX for embedded systems. The study highlighted a wide mAP range from 73% to 89%, with YOLOv5 being the most efficient version. Challenges such as lighting variability, adverse weather, and partial occlusion were extensively discussed. However, addressing complex real-world scenarios remains a limitation. This review provides a foundational analysis for advancing YOLO-based systems for robust traffic environments. In (Dewi et al., 2024), developed a method to improve road marking visibility at night for autonomous vehicles using YOLO models. Combining CLAHE with YOLOv8 yielded 90 percent accuracy, precision, and recall. Advantage: Enhanced detection of road signs in low-light conditions for safer driving. Disadvantage: Real-time processing demands significant computational power. In (Gao et al., 2024), proposed a CNN-based system for detecting traffic signs under adverse conditions like rain and fog. The model, which includes VGG19, Enhance-Net, and YOLOv4, reached 95.03 percent accuracy and improved detection speed by 12.03 fps. Advantage: Faster and more resilient detection in harsh conditions. Disadvantage: Accuracy may drop slightly under specific challenging environments.

In (Cao et al., 2024), introduced YOLOv7-tiny-RCA, a lightweight traffic sign detection system for edge devices. Using ELAN-REP, CBAM, and AFPN modules, the model achieved 81.03 percent mAP with fewer parameters and faster inference speeds. Advantage: Ideal for real-time edge applications due to its efficiency. Disadvantage: Struggles with highly occluded or complex scenes. In (Khalid et al., 2024) employed the YOLOv5s model for detection, coupled with the MSER algorithm for text localization and OCR for text recognition, using the ASAYAR dataset. This approach demonstrated high accuracy and reduced false positives, though it faced challenges with occluded panels and highlighted the need for more robust handling of occlusions. In (Mahadshetti et al., 2024) proposed a YOLOv7-based model with SE blocks and attention mechanisms, utilizing the GTSDB dataset. This model achieved an impressive mAP of 99.10% and significantly reduced model size by 98%, yet struggled in complex road environments, indicating limited adaptability to varied conditions.

In (Valiente et al., 2023)introduced a system combining YOLO, OCR, and machine learning for detection, text extraction, and 3D orientation analysis, applied to a dataset of degraded traffic signs. Their model excelled in detecting degraded signs with high accuracy but required GPU support for real-time performance, while preprocessing steps added computational overhead. These studies underscore the effectiveness of YOLO-based approaches but also highlight recurring challenges such as occlusions, environmental variability, and computational demands.

As a result of deep learning, traffic sign identification and recognition made great progress, with high accuracy and real-time performance. This study shows resilience in difficult situations, including remote or blurry indicators, and improves detection accuracy. Enhancing TSDR's dependability and influence on intelligent transportation systems will need tackling the outstanding issues with edge-case handling and efficiency.

3 PROPOSED METHODOLOGY

Traffic sign detection and recognition (TSDR) is essential for autonomous vehicles and intelligent traffic systems, requiring reliable solutions to handle complex real-world conditions. The YOLO (You Only Look Once) framework is renowned for its real-time performance, combining region proposal and classification into a single network. The latest version, YOLO11, incorporates advancements such as refined loss functions for better accuracy, anchor-free detection for reduced complexity, and transformer-based attention mechanisms for enhanced feature extraction. These features address challenges like occlusion, variable lighting, and real-time processing, making YOLO11 ideal for TSDR applications.

This study utilizes the Indian Road Traffic Sign Dataset (IRTSD-Datasetv1), annotated with Roboflow, which includes 37 classes of traffic signs such as ”No Parking,” ”Pedestrian Crossing,” ”Speed Limit 40,” and so on. The dataset captures diverse scenarios, including varying weather conditions, lighting environments, and occasional occlusion, while assuming negligible motion blur and high-

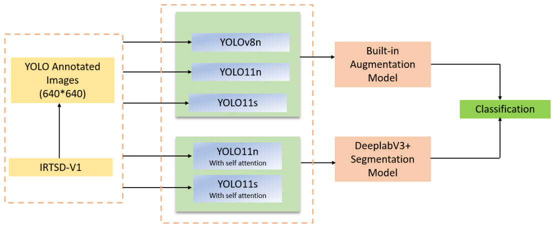

YOLO Annotated Images (640*640)
IRTSD-V1
YOLOv8n
YOLO11n
YOLO11s
YOLO11n With self attention
YOLO11s With self attention
Built-in Augmentation Model
DeeplabV3+ Segmentation Model
Classification

Figure 1: Workflow of the proposed method using the IRTSD-V1 dataset. YOLO-annotated images (640×640) are processed through two model streams: standard YOLO models (YOLOv8n, YOLO11n, YOLO11s) with built-in augmentation and enhanced YOLO models (YOLO11n, YOLO11s) integrated with self-attention. Outputs from these models are passed to a built-in augmentation module or DeepLabV3+ segmentation model for final classification

quality input feeds. The proposed work incorporates training and evaluvation of three YOLO variants YOLO11n, YOLO11s, and YOLOv8 as shown in Fig. 1. Along with this also trained YOLO11n and YOLO11s with self-attention mechanisms to improve feature extraction in complex scenes. YOLOv8, serving as a baseline, does not incorporate attention but provides a comparison for assessing the impact of attention mechanisms. These models are trained on the annotated dataset, emphasizing the detection of signs under challenging conditions such as partial occlusion and low contrast.

With the growing adoption of autonomous vehicles, TSDR systems must reliably handle real-world complexities. For example, autonomous vehicles in urban areas must accurately detect signs that may be obscured by sunlight or partially blocked by other vehicles. Inaccurate or delayed detection could lead to serious consequences, including traffic violations or accidents. By leveraging YOLO's advanced capabilities, this research seeks to improve TSDR performance for Indian roadways, contributing to safer and more efficient transportation systems.Objectives are as followed

Design and Implement a Robust Traffic Sign Detection and Recognition System: Develop an efficient system that accurately detects and recognizes traffic signs from real-time video frames or images, using the Indian traffic sign dataset (IRTSD-Datasetv1). This system should be capable of handling variations in sign size, orientation, and lighting

conditions.

Leverage the YOLO Architecture for Real-time Detection: Utilize the YOLO11n, YOLO11s, and YOLOv8 models for object detection tasks, chosen for their speed and accuracy in

detecting traffic signs in real-world environments. Implement these models with appropriate adaptations and fine-tuning to meet the requirements of traffic sign detection.

Enhance Model Performance with Attention Mechanisms: Integrate self-attention mechanisms in the YOLO11n and YOLO11s models to improve the detection of small and occluded traffic signs. This aims to enhance the model's focus on critical areas of the image, increasing detection accuracy and robustness.

Conduct Comparative Analysis Across Multiple Models: Perform a thorough evaluation of different model architectures, including YOLO11n, YOLO11s with and without attention, and YOLOv8. The goal is to assess the impact of attention mechanisms and compare the performance in terms of accuracy, inference speed, and robustness in real-world scenarios.

Optimize and Fine-tune Hyperparameters for Improved Accuracy: Experiment with various hyperparameters (e.g., learning rate, batch size, epochs) and techniques (such as data augmentation) to optimize model performance and ensure robust generalization across diverse traffic sign types and environments.

There are 4,553 annotated traffic sign images in IRTSD-v1 dataset, which are divided into 37 different classes. This dataset is annotated for YOLO classification by the means of Roboflow. The Annotated IRTSD-v1 is then fed to multiple YOLO architectures with all the requirements of YOLO being fulfilled.

The pairwise self-attention process is shown in Fig. 3 , where the attention weight $\alpha(f_i, f_j)$ is calculated by extracting relational information using transformations φ and ψ. The original input is added as a residual link, and the transformed feature $\beta(f_j)$ is coupled with the attention weight using the Hadamard product. A context-aware feature map is produced by refining the output using Batch Normalization and LeakyReLU (Wang et al., 2020).

Pairwise self-attention is defined as in equation (1)

$$f_i' = \sum_{j \in R(i)} \alpha(f_i, f_j) \odot \beta(f_j), \tag{1}$$

where $\odot$ represents the Hadamard product, i indicates the spatial index of the feature vector f_i, and $R(i)$ is the local footprint. The attention weight α is decomposed as:

$$\alpha(f_i, f_j) = \gamma(\delta(f_i, f_j)), \tag{2}$$

where $\delta(f_i, f_j)$ encodes the relation between f_i and f_j through trainable transformations ϕ and ψ. Position encoding normalizes coordinates to $[-1, 1]$, calculates differences $p_i - p_j$, and concatenates them with features before mapping.

Fig. 2 illustrates the implementation that utilized the IRTSD-Datasetv1, which contains 37 classes of traffic signs, and followed a series of preprocessing

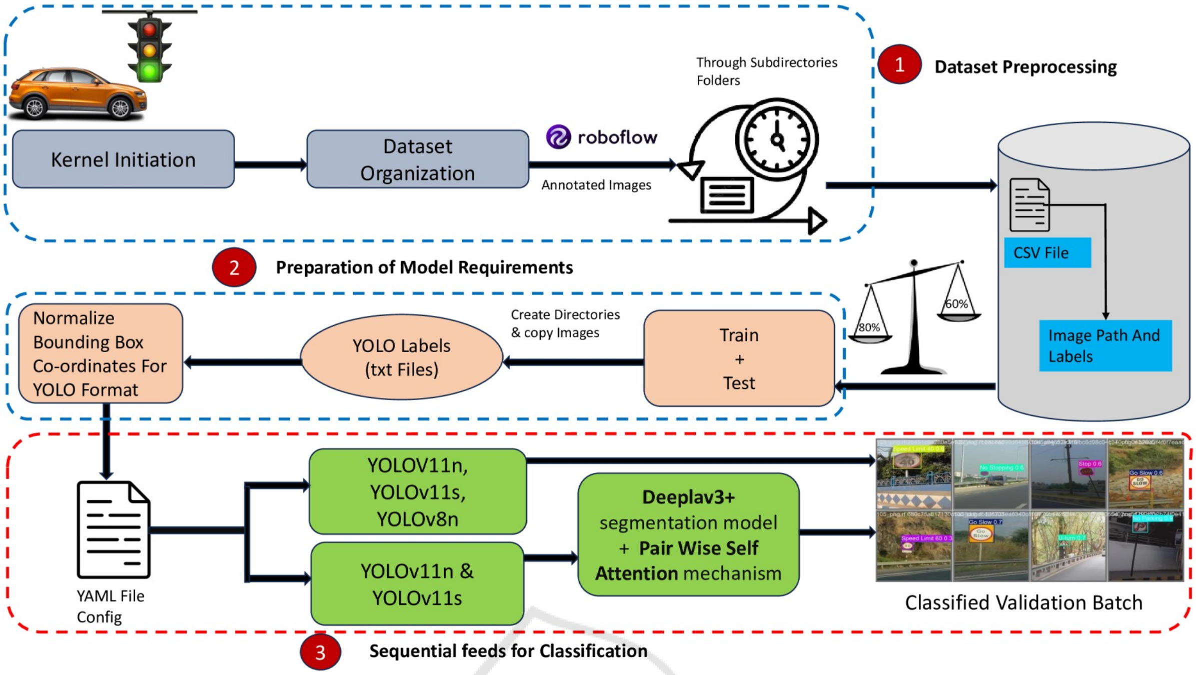

1 Dataset Preprocessing
Kernel Initiation
Dataset Organization
roboflow
Annotated Images
Through Subdirectories Folders
CSV File
Image Path And Labels
2 Preparation of Model Requirements
Normalize Bounding Box Co-ordinates For YOLO Format
YOLO Labels (txt Files)
Create Directories & copy Images
Train + Test
80%
60%
YAML File Config
YOLOV11n, YOLOv11s, YOLOv8n
YOLOv11n & YOLOv11s
Deeplav3+ segmentation model + Pair Wise Self Attention mechanism
Classified Validation Batch
3 Sequential feeds for Classification

Figure 2: The end-to-end pipeline of the traffic sign detection and recognition using the IRTSD-V1 dataset. The process is divided into three stages: (1) Dataset Preprocessing, where annotated images are organized and split into training and testing sets; (2) Preparation of Model Requirements, involving normalization of bounding box coordinates, generation of YOLO labels, and configuration setup for training; and (3) Sequential Feeds for Classification, utilizing YOLOv11n, YOLOv11s, YOLOv8, and DeepLabv3+ models with a Pair-Wise Self-Attention mechanism to produce accurately classified validation batches.

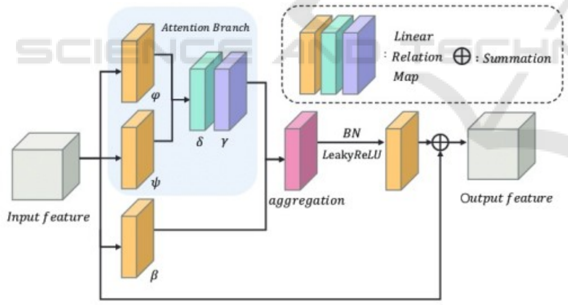


Figure 3: Illustration of the pairwise self-attention mechanism, demonstrating how input features are transformed through attention branches and aggregated to produce context-aware output features.

Table 1: Parameters of YOLOv8n, YOLO11n, and YOLO11s Models

Model	Size (pixels)	mAPval 50-95	Speed CPU (ms)	Speed T4 (ms)	Params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLO11n	640	39.5	56.1	1.5	2.6	6.5
YOLO11s	640	47.0	90.0	2.5	9.4	21.5

steps to prepare the data for training. The raw images are annotated using Roboflow, enabling the creation of bounding boxes and class labels for each object in the image. To enhance the dataset, data augmentation techniques, including rotation, flipping, zooming, and scaling, are applied. This helped improve the model's robustness and prevent overfitting. Additionally, pixel normalization is applied to scale all pixel values to a range of 0 to 1, aiding the convergence process during training. The dataset is split into training, validation, and test sets with a 70%, 15%, and 15% ratio, respectively.

In terms of model architecture, four variants of the YOLO11 and YOLOv8 models are explored. YOLO11n (without attention) is a lightweight version optimized for real-time object detection, focusing on speed and efficiency. While fast, it may not achieve the highest accuracy, especially in detecting small or overlapping objects. YOLO11s (with self-attention) incorporates a self-attention mechanism to enhance feature learning, improving detection accuracy, especially for complex and small traffic signs. YOLO11s (without self-attention) is a simpler version of YOLO11s, which trades off some accuracy for increased speed. YOLOv8n, the most advanced model, is designed to be faster and more accurate than its predecessors, but it does not incorporate the self-attention mechanism, which may lead to the loss of subtle features. The models are evaluated based on precision, recall, F1-score, and mean average precision (mAP). YOLOv8 provided the best overall performance in terms of speed and accuracy, while YOLO11s with self-attention showed superior results in handling complex detection scenarios, particularly with occlusions or similar-looking signs. The inclusion of data augmentation and preprocessing steps is crucial for improving the robustness of the models, allowing them to handle varying lighting conditions, angles, and occlusions.

In conclusion, YOLOv8 delivered the best performance for real-time traffic sign detection, surpassing YOLO11n. However, YOLO11s and YOLO11n with self-attention proved effective in complex detection tasks. Data preprocessing and augmentation significantly enhanced the models' ability to handle challenging detection environments.

The Algorithm 1 shows step by step process of Pair-Wise Self Attention mechanism of YOLO.

Algorithm 1 Pairwise Attention Mechanism for Feature Enhancement

```
Input: Feature map $F \in \mathbb{R}^{H\times W\times C}$
Output: Enhanced feature map $F_{\text{enhanced}} \in \mathbb{R}^{H\times W\times C}$
# Step 1: Input Transformation
Reshape the feature map $F$ into a sequence $F_{\text{seq}} \in \mathbb{R}^{(H\times W)\times C}$
Initialize the pairwise attention weights matrix $W_{\text{attn}} \in \mathbb{R}^{(H\times W)\times(H\times W)}$
# Step 2: Pairwise Attention Computation
for each pair $(i,j)$ where $i,j \in \{1,2,\dots,H\times W\}$ do
    Compute similarity: $S_{ij} = \text{softmax}(\text{dot}(F_{\text{seq}}[i], F_{\text{seq}}[j]))$
    Update attention weights: $W_{\text{attn}}[i,j] = S_{ij}$
end for
# Step 3: Weighted Feature Aggregation
for each feature vector $F_{\text{seq}}[i]$ where $i \in \{1,2,\dots,H\times W\}$ do
    Aggregate features:
        $F_{\text{seq}}[i]_{\text{enhanced}} = \sum_j W_{\text{attn}}[i,j] \cdot F_{\text{seq}}[j]$
end for
# Step 4: Output Transformation
Reshape $F_{\text{seq, enhanced}} \in \mathbb{R}^{(H\times W)\times C}$ back to $F_{\text{enhanced}} \in \mathbb{R}^{H\times W\times C}$
Return: $F_{\text{enhanced}}$
```

The Algorithm 2 shows step by step process of YOLO performing Object Detection.

Algorithm 2 YOLO-based Object Detection Algorithm

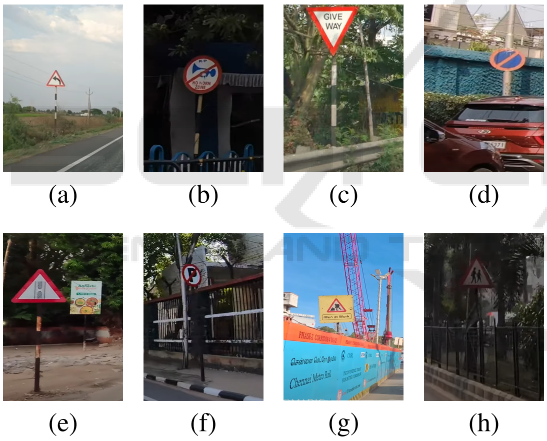
```
Input: Input image $I$ of size $640\times 640\times 3$, pretrained YOLO model weights, confidence threshold, IoU threshold
Output: Bounding boxes $B = \{b_1, b_2, \dots, b_n\}$ and class probabilities $P = \{p_1, p_2, \dots, p_n\}$
# Step 1: Image Preprocessing
Resize the input image $I$ to the YOLO model's required input size (e.g., $640\times 640$ pixels)
Normalize pixel values to the range $[0,1]$
Convert the image into a tensor format suitable for the YOLO model
# Step 2: Feature Extraction
Pass the preprocessed image through the YOLO backbone network
Extract multi-scale feature maps from the input image
# Step 3: Object Detection
Apply detection heads to the extracted feature maps
Predict bounding box coordinates, objectness scores, and class probabilities
Filter predictions using a confidence threshold (e.g., 0.5)
# Step 4: Non-Maximum Suppression (NMS)
for each detected class do
    Sort bounding boxes by their confidence scores
    Remove overlapping boxes based on the Intersection over Union (IoU) threshold (e.g., 0.5)
end for
# Step 5: Post-Processing
Map the filtered bounding boxes to the original image dimensions
Assign class labels and confidence scores to each detected object
Return: Bounding boxes $B$ and class probabilities $P$
```

4 RESULTS AND DISCUSSION

This study utilizes the YOLO11 and YOLOv8n architecture, optimized with self-attention and segmentation mechanisms, for real-time traffic sign detection and classification. The models YOLO11n and YOLO11s are trained on the IRTSD-Datasetv1, comprising 4,553 annotated images across 37 classes(Sample images are shown in figure 2), using PyTorch with data augmentation, early stopping, and the AdamW optimizer. Segmentation preprocessing with DeepLabV3+ and custom self-attention modules improved feature learning and detection performance. YOLO11s with self-attention achieved the best results, with an mAP@50 of 84.0%, Precision of 86.2%, and Recall of 86.4%, outperforming YOLOv8n and other variants in robustness and accuracy, especially under challenging conditions

Table 2: Performance Comparison of YOLOv11 Models with and without Attention, and YOLOv8

Model	Precision	Recall	mAP@50	mAP@50-95
YOLOv11n with attention	0.7622	0.7716	0.7855	0.4818
YOLOv11n without attention	0.7239	0.7127	0.7326	0.4501
YOLOv11s with attention	0.8093	0.8181	0.8401	0.5228
YOLOv11s without attention	0.7664	0.7754	0.7873	0.4780
YOLOv8	0.7112	0.7246	0.7312	0.4457



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: Sample images from the IRTSD-Datasetv1 depicting traffic signs in various conditions and settings. (a) Left Turn; (b) Horn Prohibited; (c) Give Way; (d) No Stopping; (e) Gap In Median; (f) No Parking; (g) Men at Work; (h) School Ahead.

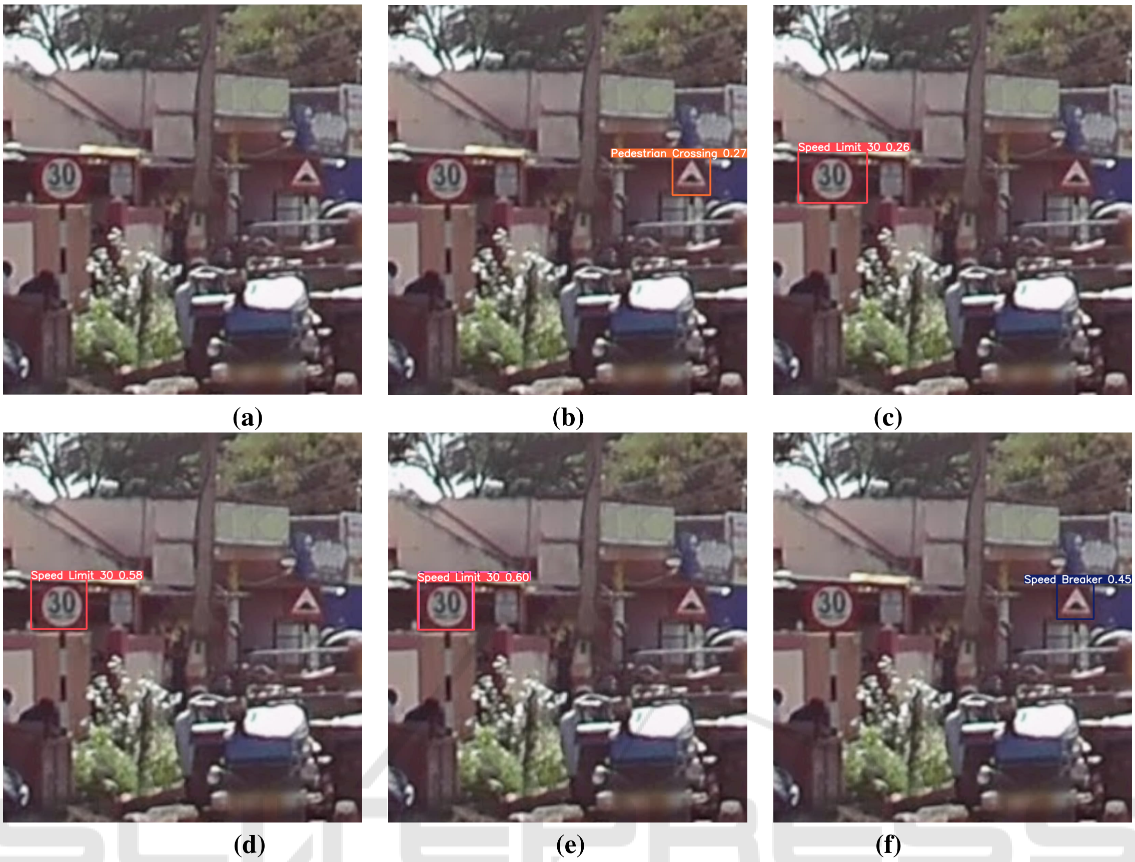
The performance comparison of different YOLO models, as presented in Table 2, demonstrates the efficacy of incorporating attention mechanisms in traffic sign detection and recognition. Among the tested models, YOLO11n and YOLO11s are evaluated with and without attention, alongside YOLOv8 as a baseline. The results highlight the significant improvements achieved by leveraging attention mechanisms, particularly in enhancing feature extraction and localization capabilities. For the YOLO11n models, the attention augmented version outperformed the model without attention across all metrics, achieving a precision of 0.76 and a recall of 0.77, compared to 0.72 and 0.71, respectively. Similarly, the mAP@50 and mAP@50-95 values showed a considerable increase from 0.73 and 0.45 to 0.785523 and 0.48.

The YOLO11s models exhibited even stronger performance, with YOLO11s with attention emerging as the best performing model. It achieved a precision of 0.80, a recall of 0.81, a mAP@50 of 0.84, and a mAP@50-95 of 0.52. This indicates that the larger architecture of YOLO11s effectively leverages attention mechanisms to deliver superior results. In contrast, YOLOv8, although efficient and widely used for real time applications, demonstrated lower accuracy, achieving a precision of 0.71, a recall of 0.72, a mAP@50 of 0.73, and a mAP@50-95 of 0.44. These metrics underscore the limitations of YOLOv8 in handling complex real-world challenges such as occlusions, varying lighting conditions, and cluttered backgrounds.

The inclusion of attention mechanisms significantly enhances the performance of YOLO models, particularly in challenging scenarios. YOLO11 models with attention consistently outperformed their non-attention counterparts, highlighting their robustness and reliability. The results affirm the potential of attention-augmented YOLO11 models, especially YOLO11s with attention, for real-time traffic sign detection in autonomous vehicle systems which is computed with the phenomena of feeding an complex and robust traffic sign image to all the models shown in Fig. 5.YOLO11s employed with self attention dominating the classification results in achieving astonishing precision and recall as shown in Fig. 6 and Fig. 7 respectively.High precision and recall, along with strong mAP values, the proposed models demonstrate their capability to address practical difficulties in traffic sign recognition, ensuring safe and efficient operation in real-world applications.

To verify the performance of the YOLO family on the IRTSD-Datasetv1, an illustration is provided in Fig. 8. It depicts the mAP (mean Average Precision) at 50% IoU for YOLO models over the first 60 epochs, highlighting the impact of self-attention mechanisms and architectural variations. Models integrated with attention mechanisms, such as YOLO11n with attention and YOLO11s with attention, consistently outperform their counterparts without attention, demonstrating improved feature representation and detection accuracy.

As shown in Figure 8, The graph shows the Comparison of mAP (metrics/mAP50(B)) vs Epoch for Different YOLO Models (First 60 Epochs) and the performance of five YOLO models, including



(a) (b) (c)

(d) (e) (f)

Figure 5: Classification performed by different YOLO models for a robust and occluded traffic sign image. (a) Cluttered and Robust Traffic Sign Image, (b) YOLOv8n classification on the image, (c) YOLO 11n classification on the image, (d) YOLO 11s classification on the image, (e) YOLO 11n With Self Attention classification, (f) YOLO 11s With Self Attention classification.

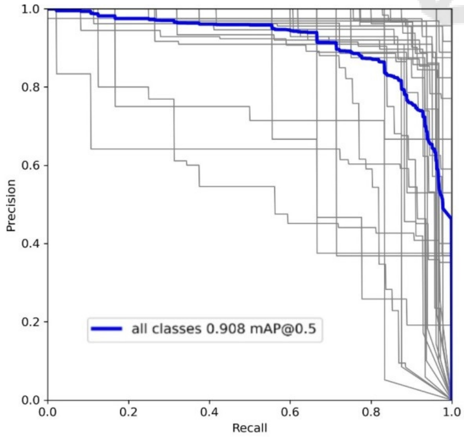


Figure 6: Precision-Recall Curve of YOLO11s employed with Self Attention

YOLOv11n and YOLOv11s with and without attention mechanisms, and YOLOv8, highlighting the impact of attention mechanisms on model accuracy over epochs.

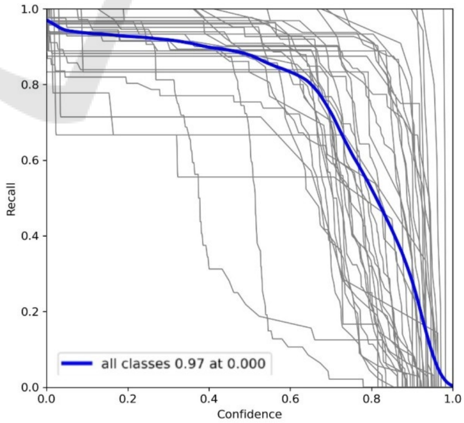


Figure 7: ROC Curve of YOLO11s employed with Self Attention

The YOLO11s variants (with and without attention) achieve slightly higher mAP values than YOLO11n, indicating that the larger model leverages its capacity to learn more intricate features effectively. Among all models, YOLO11s with at-

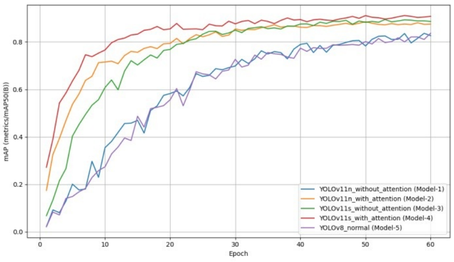


Figure 8: Comparison of mAP vs Epoch

tention converges the fastest, reaching higher mAP within the initial 20 epochs and maintaining stability after 40 epochs, showcasing its robustness. While YOLOv8n achieves competitive performance, it is outperformed by YOLO11 models with attention, particularly YOLO11s with attention. This validates the superiority of the YOLO11 architecture with integrated attention mechanisms for enhanced traffic sign detection.

5 CONCLUSION

The difficulty of real-time traffic sign detection and recognition, which is essential for the secure and effective operation of autonomous cars, is successfully addressed in this work, to sum up. In this work YOLO11s employed with Pair-Wise Self Attention mechanism obtaining Precision 80.93% , Recall 81.81% ,achieving mAP@50 of 84.01% and mAP@50-95 of 52.28% showcased ground breaking detection and classification results. The study emphasizes how crucial it is to incorporate attention strategies in order to improve detection performance by using sophisticated YOLO models, both with and without attentive mechanisms. Occlusions, changing lighting, and broken signage are just a few examples of the real-world difficulties that the models are refined to handle through careful dataset preparation, accurate annotation, and the application of reliable training techniques. The results show that attention-integrated models are suitable for real-time applications, since they perform noticeably better than their counterparts in terms of recognition and detection accuracy.By presenting a workable approach that improves autonomous vehicle navigation and establishes the foundation for future developments in the field, this research advances the field of traffic sign identification.

REFERENCES

Avramović, A., Sluga, D., Tabernik, D., Skočaj, D., Stojnić, V., and Ilc, N. (2020). Neural-network-based traffic sign detection and recognition in high-definition images using region focusing and parallelization. *IEEE Access*, 8:189855–189868.

Boujemaa, K. S., Akallouch, M., Berrada, I., Fardousse, K., and Bouhoute, A. (2021). Attica: A dataset for arabic text-based traffic panels detection. *IEEE Access*, 9:93937–93947.

Cao, X., Xu, Y., He, J., Liu, J., and Wang, Y. (2024). A lightweight traffic sign detection method with improved yolov7-tiny. *IEEE Access*, 12:105131–105147.

Dewi, C., Chen, R.-C., Zhuang, Y.-C., and Manongga, W. E. (2024). Image enhancement method utilizing yolo models to recognize road markings at night. *IEEE Access*, 12:131065–131081.

Fleyeh, H. and Dougherty, M. (2006). Road and traffic sign detection and recognition.

Flores-Calero, M., Astudillo, C. A., Guevara, D., Maza, J., Lita, B. S., Defaz, B., Ante, J. S., Zabala-Blanco, D., and Armingol Moreno, J. M. (2024a). Traffic sign detection and recognition using yolo object detection algorithm: A systematic review. *Mathematics*, 12(2).

Flores-Calero, M., Astudillo, C. A., Guevara, D., Maza, J., Lita, B. S., Defaz, B., Ante, J. S., Zabala-Blanco, D., and Armingol Moreno, J. M. (2024b). Traffic sign detection and recognition using yolo object detection algorithm: A systematic review. *Mathematics*, 12(2).

Gao, Q., Hu, H., and Liu, W. (2024). Traffic sign detection under adverse environmental conditions based on cnn. *IEEE Access*, 12:117572–117580.

Greenhalgh, J. and Mirmehdi, M. (2015). Recognizing text-based traffic signs. *IEEE Transactions on Intelligent Transportation Systems*, 16(3):1360–1369.

Khalid, S., Shah, J. H., Sharif, M., Dahan, F., Saleem, R., and Masood, A. (2024). A robust intelligent system for text-based traffic signs detection and recognition in challenging weather conditions. *IEEE Access*, 12:78261–78274.

Lopez-Montiel, M., Orozco-Rosas, U., Sánchez-Adame, M., Picos, K., and Ross, O. H. M. (2021). Evaluation method of deep learning-based embedded systems for traffic sign detection. *IEEE Access*, 9:101217–101238.

Mahadshetti, R., Kim, J., and Um, T.-W. (2024). Sign-yolo: Traffic sign detection using attention-based yolov7. *IEEE Access*, 12:132689–132700.

Mannan, A., Javed, K., Ur Rehman, A., Babri, H. A., and Noon, S. K. (2019). Classification of degraded traffic signs using flexible mixture model and transfer learning. *IEEE Access*, 7:148800–148813.

Tabernik, D. and Skočaj, D. (2020). Deep learning for large-scale traffic-sign detection and recognition. *IEEE Transactions on Intelligent Transportation Systems*, 21(4):1427–1440.

Triki, N., Karray, M., and Ksantini, M. (2024). A comprehensive survey and analysis of traffic sign recognition systems with hardware implementation. *IEEE Access*, 12:144069–144081.

Valiente, R., Chan, D., Perry, A., Lampkins, J., Strelnikoff, S., Xu, J., and Ashari, A. E. (2023). Robust perception and visual understanding of traffic signs in the wild. *IEEE Open Journal of Intelligent Transportation Systems*, 4:611–625.

Wang, C., Wu, Y., Su, Z., and Chen, J. (2020). Joint self-attention and scale-aggregation for self-calibrated deraining network. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2517–2525.

Youssef, A., Albani, D., Nardi, D., and Bloisi, D. D. (2016). Fast traffic sign recognition using color segmentation and deep convolutional networks. In Blanc-Talon, J., Distante, C., Philips, W., Popescu, D., and Scheunders, P., editors, *Advanced Concepts for Intelligent Vision Systems*, pages 205–216, Cham. Springer International Publishing.