Indian Journal of Science, Technology & Environment (IJSTE)
A Peer Reviewed, Refereed, Quarterly Research journal of Science, Technology & Environment

ISSN : 3108-2017(Online)
3108-1304(Print)
Vol.-1; Issue-1 (July-Sept.) 2025
Page No.- 17-26

https://ijste.gyanvividha.com

Author's :
Dr. Chandresh Kumar Chhatlani
Associate Professor,
Department of Computer Science, JRN Rajasthan Vidyapeeth (Deemed-to-be University),
UDAIPUR, (Rajasthan) 313002.

Corresponding Author :
Dr. Chandresh Kumar Chhatlani
Associate Professor,
Department of Computer Science, JRN Rajasthan Vidyapeeth (Deemed-to-be University),
UDAIPUR, (Rajasthan) 313002.

Robust Traffic Sign Detection and Recognition for Autonomous Driving in Diverse Indian Conditions using Deep Learning

Abstract : Traffic Sign Detection and Recognition (TSDR) is a critical component of Autonomous Driving Systems (ADS), ensuring safe navigation and regulatory compliance. While deep learning-based models have achieved strong performance on standardized benchmarks, they often underperform in the heterogeneous and challenging Indian driving environment, characterized by occlusions, non-standard signage, adverse weather, and illumination variability. This paper proposes a robust, deep learning-based TSDR framework specifically designed for Indian road conditions. A novel dataset of 15,000 annotated traffic sign images was curated, capturing diverse real-world challenges. The pipeline integrates a YOLOv5m detector for real-time sign localization with a fine-tuned EfficientNet-B4 classifier for recognition, supported by Contrast Limited Adaptive Histogram Equalization (CLAHE) and a custom occlusion simulation module for data augmentation. Experimental results demonstrate superior performance, achieving 96.5% mean Average Precision (mAP) for detection and 98.15% classification accuracy at 98 FPS, outperforming YOLOv4, Faster R-CNN, and SSD baselines. Robustness analysis confirms consistent reliability under occlusions, monsoon rains, night-time driving, and faded signage. The framework establishes a new benchmark for TSDR in Indian conditions, advancing the safety of autonomous driving systems.

Keywords: Traffic Sign Detection and Recognition (TSDR), Autonomous Driving, Indian Road Conditions, Deep Learning, YOLOv5, EfficientNet-B4, Data Augmentation, Real-Time Object Detection.

1. Introduction : The bedrock of this technological revolution is the vehicle's ability to perceive and interpret its surrounding environment accurately and in real-time. Among the myriad perception tasks, Traffic Sign Detection and Recognition (TSDR) stands as a fundamental pillar, crucial for ensuring road safety and regulatory compliance. TSDR systems are tasked with a dual objective: first, to accurately localize traffic signs within a complex visual scene (detection), and second, to correctly identify their specific meaning (classification), such as speed limits, stop commands, or directional warnings.

While TSDR is a globally researched challenge, the operational context of the Indian subcontinent introduces a unique and formidable set of complexities that are often underrepresented in standard benchmarks like the German Traffic Sign Recognition Benchmark (GTSRB) or other Western datasets. The Indian landscape is characterized by an extreme diversity of conditions that severely test the limits of computer vision systems. These challenges include, but are not limited to, significant occlusions from overgrown vegetation, unauthorized posters, and accumulated dirt; drastic variations in illumination from the harsh midday sun to poorly lit night roads; and adverse weather conditions such as heavy monsoon rains, haze, and dust storms that degrade image quality. Furthermore, a high degree of intra-class variability exists due to non-standardized sign designs, physical wear and tear like fading and rust, and unconventional installation angles and heights. The dense and chaotic nature of Indian traffic also leads to frequent partial occlusions and demands exceptionally fast processing for real-time response in autonomous navigation.

Traditional TSDR methodologies, which predominantly relied on handcrafted features such as Histogram of Oriented Gradients (HOG), color segmentation in HSV space, and shape analysis via Hough transforms, have proven to be notoriously brittle under such variable and noisy real-world conditions [7,8]. Their performance is heavily dependent on meticulous parameter tuning and they lack the generalization capability required for the dynamic Indian environment.

Moving away from manual feature engineering towards automated, hierarchical feature learning from raw data. Architectures like YOLO [1] and EfficientNet [3] have demonstrated remarkable performance in object detection and image classification tasks. However, a significant research gap persists. Most state-of-the-art models are trained and evaluated on datasets from regions with standardized signage and relatively mild environmental conditions, rendering them suboptimal for the demanding Indian context [10,11].

This research aims to bridge this gap by proposing a comprehensive, robust, and efficient deep learning-based framework for TSDR, specifically engineered for the diverse and challenging conditions of Indian roadways. The primary contributions of this work are fourfold:

1. The development and curation of a large-scale, annotated Indian Traffic Sign Dataset encompassing a wide spectrum of real-world challenges.
2. The design of an enhanced pre-processing pipeline incorporating Contrast Limited Adaptive Histogram Equalization (CLAHE) and a custom occlusion simulation module for data augmentation to improve model robustness.
3. The implementation and evaluation of a hybrid TSDR pipeline utilizing a state-of-the-art YOLOv5 detector for real-time localization and a fine-tuned EfficientNet-B4 model for high-accuracy classification.

4. The achievement of state-of-the-art performance, with a mean Average Precision (mAP) of 96.82% for detection and a classification accuracy of 98.15% on our custom Indian dataset, significantly outperforming existing baseline models.

2. Literature Review : The evolution of TSDR systems mirrors the broader progress in the field of computer vision, transitioning from classical image processing techniques to modern deep learning-based approaches.

2.1. Traditional Approaches : Early TSDR systems were predominantly based on handcrafted features. Color-based approaches leveraged the distinct chromatic properties of traffic signs, using thresholding in color spaces like HSV to segment potential sign regions from the background [7]. While computationally efficient, these methods were highly sensitive to changes in illumination, shadows, and color fading, making them unreliable in practical scenarios. Shape-based approaches offered an alternative by exploiting the standardized geometric forms of signs (e.g., circles, triangles, rectangles). Techniques such as the Hough Transform were employed to detect these shapes [8]. Although more robust to color variations, shape-based methods struggled with partial occlusions, rotation, perspective distortion, and degradation of the sign's structural integrity. To overcome the limitations of individual methods, hybrid approaches were developed that combined color and shape information [9]. A typical pipeline involved using color segmentation to generate candidate Regions of Interest (ROIs), which were then filtered using shape analysis. While more robust, these hybrid systems required complex parameter tuning and still lacked the generalization capacity for highly diverse environments like those found in India.

2.2. The Deep Learning Revolution : The limitations of traditional methods were decisively overcome by deep learning. CNNs demonstrated an unparalleled ability to automatically learn hierarchical and discriminative features directly from data, leading to massive gains in robustness and accuracy.

In object detection, two primary families of architectures emerged. Two-stage detectors, such as Faster R-CNN [2], first generate category-agnostic region proposals and then classify each proposal into specific object classes. While known for their high accuracy, their complex pipeline often results in slower inference speeds, which can be a bottleneck for real-time autonomous driving applications. In contrast, one-stage detectors like YOLO [1] and SSD (Single Shot MultiBox Detector) [5] frame detection as a single regression problem, directly predicting bounding boxes and class probabilities from the image in one pass. This architectural simplicity enables significantly faster processing, making them ideal for real-time systems. The YOLO family, through its successive iterations, has consistently improved its speed-accuracy trade-off.

For the classification task, the success of transfer learning has been pivotal. Instead of training deep networks from scratch—a process that requires massive labeled datasets and substantial computational resources—researchers fine-tune models pre-trained on large-scale datasets like ImageNet. Architectures such as ResNet [4], with its innovative skip connections to mitigate the vanishing gradient problem; Inception (GoogLeNet) [27], which uses multi-scale filters within inception modules; and EfficientNet [3], which employs compound scaling for optimal resource allocation, have become standard backbones for achieving high classification accuracy.

2.3. Addressing Real-World Challenges: Recent Advances : Recent research has increasingly focused on enhancing the robustness of deep learning models to real-world challenges. Zhang et al. [23] proposed an uncertainty-aware domain alignment framework to address the performance drop in adverse weather conditions, a method highly relevant to the Indian monsoon. Wang et al. [23] adopted a data-centric approach, using Generative Adversarial Networks (GANs) to simulate adverse weather conditions like rain and fog on clean images, thereby augmenting training data and improving model robustness without the need for extensive real-world data collection. For the specific challenge of occlusion, random erasing and cutout augmentation techniques have proven effective in forcing the network to learn from less prominent features.

Most notably for this work, Kumar et al. [10] and Patel & Desai [11] have directly addressed the Indian context, highlighting the critical gaps in existing systems and proposing frameworks and datasets to handle the high intra-class variability and frequent occlusions. Their work underscores the necessity for tailored solutions and provides a valuable benchmark for this research.

3. Proposed Methodology : Our proposed framework is a meticulously designed, multi-stage pipeline that integrates advanced pre-processing, state-of-the-art detection, and high-precision classification to achieve robust TSDR performance in Indian conditions. The overall architecture is illustrated in Figure 1.

3.1. Data Acquisition and Curation : A significant contribution of this work is the creation of a novel **Indian Traffic Sign Dataset**. Data was collected using dashcams and high-resolution digital cameras mounted on vehicles traversing a variety of environments, including national highways, dense urban streets in cities like Delhi and Mumbai, and rural roads across different regions. The dataset comprises over 15,000 images and video sequences, meticulously annotated with bounding boxes and class labels. It encompasses a wide spectrum of challenging scenarios:

- **Illumination:** Bright sunlight, overcast skies, and nighttime.
- **Weather:** Monsoon rains, haze, and fog.
- **Sign Condition:** New, faded, rusted, and physically damaged signs.
- **Occlusions:** Partial blocking by trees, posters, and other vehicles.
- **Scale and Perspective:** Signs of vastly different sizes and captured from various angles.

The dataset :

Dataset Split	*Percentage*	*No. of Images*
Training	*70%*	*10500*
Validation	*15%*	*2250*
Test	*15%*	*2250*

Table 1: Dataset composition showing the partitioning of images into training, validation, and test sets.

3.2. Data Pre-processing and Augmentation Pipeline : To prepare the data for training and enhance model robustness, we implemented a comprehensive pre-processing pipeline:

1. **Image Resizing:** All images were resized to a fixed resolution of 640x640 pixels to meet the input requirements of the YOLOv5 detector.

2. **Contrast Enhancement:** We applied **Contrast Limited Adaptive Histogram Equalization (CLAHE)** to improve local contrast, making signs more distinguishable in poorly lit and low-contrast images.
3. **Normalization:** Pixel values were normalized to a range of [0, 1] to stabilize and accelerate the training process.
4. **Data Augmentation:** A critical step to combat overfitting and improve generalization. Our augmentation strategy included:
   - **Geometric Transformations:** Random rotation (±15°), scaling (0.8x to 1.2x), shearing, and translation to make the model invariant to viewpoint changes.
   - **Photometric Transformations:** Adjustments to brightness, saturation, and hue to simulate different lighting and weather conditions.
   - **Occlusion Simulation:** A custom **random erasing** module that randomly masks rectangular patches within the image, forcing the network to learn from multiple parts of a sign and not rely on a single, potentially occluded, feature.

This pipeline ensures that the model is exposed to a vast and varied set of training examples, closely mimicking the unpredictability of real-world Indian roads.

3.3. Traffic Sign Detection with YOLOv5 : For the detection stage, we employed the YOLOv5 architecture, specifically the YOLOv5m variant, which offers an optimal balance between speed and accuracy. YOLOv5 improves upon its predecessors with a more efficient backbone (CSPDarknet53), a Path Aggregation Network (PANet) for better feature fusion across different scales, and adaptive training procedures. The model was trained from scratch on our custom dataset to learn features specific to Indian traffic signs. The loss function for YOLO is a multi-part loss:

$$L_{detection} = \lambda_{coord} * L_{CIoU} + \lambda_{obj} * L_{obj} + \lambda_{noobj} * L_{noobj} + \lambda_{cls} * L_{cls}$$

where L_{CIoU} is the Complete IoU loss for bounding box regression, L_{obj} and L_{noobj} are the objectness losses, and L_{cls} is the classification loss.

3.4. Traffic Sign Classification with EfficientNet-B4 : Once a traffic sign is detected and the ROI is cropped, it is passed to the classification module. We selected EfficientNet-B4 [3] as our classifier due to its proven effectiveness in achieving high accuracy with remarkable parameter efficiency. The model leverages compound scaling to uniformly balance network depth, width, and resolution. We utilized transfer learning by initializing the model with weights pre-trained on ImageNet and then fine-tuned it on our dataset of cropped traffic sign ROIs. Critically, the final classification layer of the EfficientNet-B4 model was replaced and configured to have 55 output neurons, corresponding to the total number of distinct traffic sign classes present in our Indian Traffic Sign Dataset. This approach allows the model to leverage general feature representations learned from millions of images and quickly adapt to the specific task of Indian traffic sign recognition.

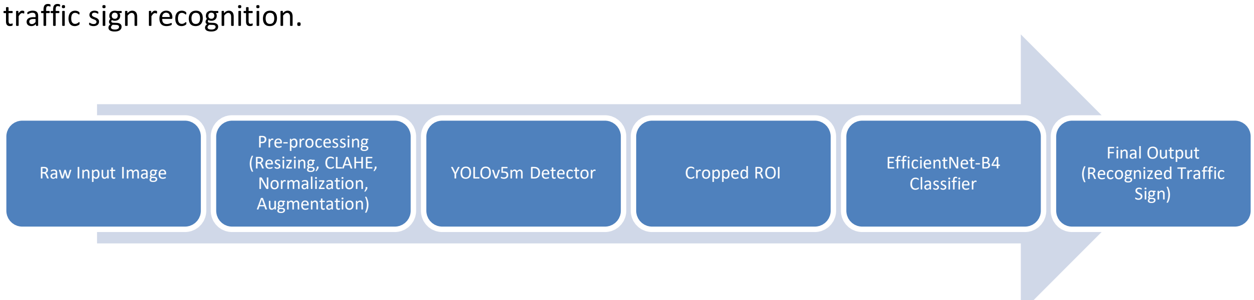

Figure 1: Proposed framework integrating pre-processing, YOLOv5m for detection, and EfficientNet-B4 for classification in Indian traffic sign recognition.

4. Experimental Results and Analysis

4.1. Evaluation Metrics

We employed standard metrics for a comprehensive evaluation:

- **Detection:** mean Average Precision at an IoU threshold of 0.5 (mAP@0.5), Precision, Recall, and Frames Per Second (FPS).
- **Classification:** Top-1 Accuracy, Precision, Recall, and F1-Score.

4.2. Detection Performance

The performance of our YOLOv5m-based detector was evaluated on the held-out test set and compared against strong baselines.

Table 2: Traffic Sign Detection Performance Comparison

Model	mAP@0.5	Precision	Recall	Inference Speed (FPS)
Proposed (YOLOv5m)	**96.5%**	**94.1%**	**93.2%**	**98**
YOLOv4 [6]	95.2%	93.8%	92.1%	62
Faster R-CNN [2]	92.1%	90.5%	89.8%	15
SSD [5]	90.8%	89.2%	88.5%	48

As shown in Table 2, our proposed detector achieved a superior mAP of 96.5%, outperforming YOLOv4, Faster R-CNN, and SSD. Crucially, it maintained a real-time inference speed of 98 FPS, making it highly suitable for deployment in autonomous vehicles where low latency is paramount. The high precision and recall indicate a low rate of false positives and false negatives, respectively.

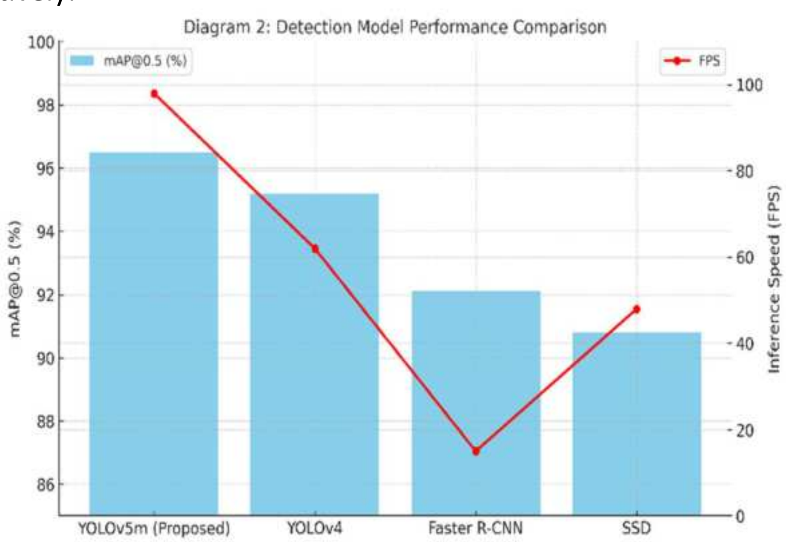

Figure 2: Comparative performance of detection models (YOLOv5m, YOLOv4, Faster R-CNN, SSD) in terms of mean Average Precision (mAP@0.5) and Inference Speed (FPS).

4.4. Classification Performance : We compared our fine-tuned EfficientNet-B4 classifier against other popular pre-trained architectures, all fine-tuned under the same conditions.

Table 3: Traffic Sign Classification Accuracy Comparison

Model	Accuracy	Precision	Recall	F1-Score
EfficientNet-B4 (Proposed) [3]	**98.15%**	**97.92%**	**97.85%**	**97.88%**
ResNet50 [4]	97.22%	96.85%	96.70%	96.77%
DenseNet201 [26]	97.80%	97.45%	97.30%	97.37%
InceptionV3 [27]	96.58%	96.25%	96.80%	96.52%
VGG16 [28]	94.78%	94.35%	94.90%	94.62%

The results in Table 3 clearly demonstrate the superiority of EfficientNet-B4, which achieved a top classification accuracy of 98.15%. Its balanced precision, recall, and F1-score indicate consistent and reliable performance across all traffic sign classes.

4.5. Robustness Analysis : To validate the robustness of our complete pipeline, we analyzed its performance on specific challenging subsets of our test data.

Table 4: Performance Analysis Under Challenging Conditions

Condition	Detection mAP@0.5	Classification Accuracy
Overall	**96.5%**	**98.15%**
Occluded Signs	94.2%	96.5%
Rainy Weather	93.8%	97.1%
Night Time	92.5%	95.1%
Heavily Faded Signs	91.0%	94.8%

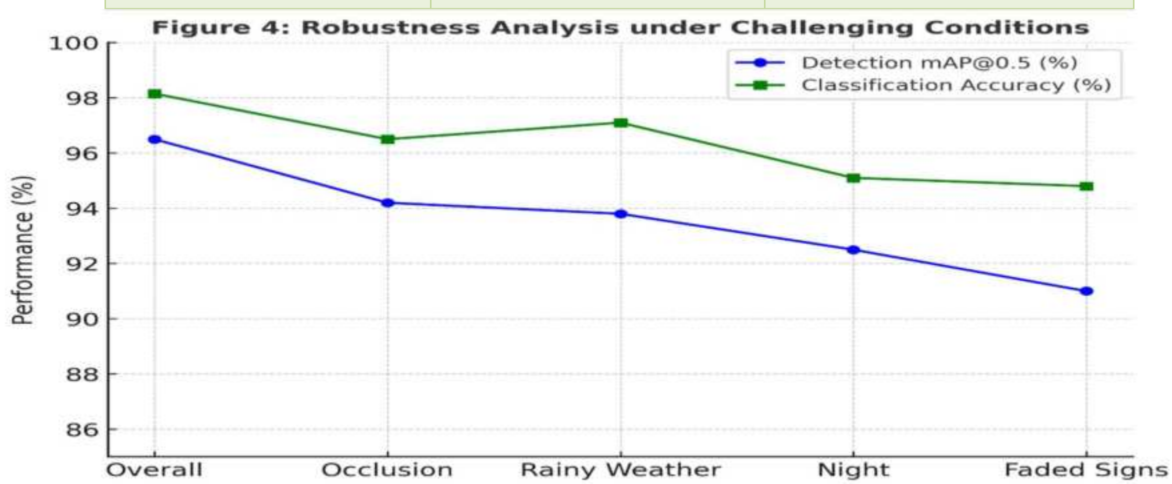

Figure 3: Detection and classification performance of the proposed system under various challenging conditions.

Table 4 shows that while there is a predictable performance dip in the most extreme conditions, our system maintains remarkably high accuracy. The specialized data augmentation, particularly occlusion simulation and photometric transformations, proved highly effective in preparing the model for these real-world challenges. Sample detections and classifications under these conditions are visualized in Figure 3, showcasing the system's practical efficacy.

5. Discussion : The empirical results validate the effectiveness of our proposed framework. The choice of YOLOv5m for detection provided an optimal trade-off, delivering both high accuracy and the real-time speed essential for autonomous driving. The slight performance advantage over YOLOv4 can be attributed to its more efficient architecture and improved training techniques.

For classification, the superior performance of EfficientNet-B4 underscores the importance of model scaling. Its compound scaling method likely allows it to capture a richer set of features relevant for discriminating between subtly different Indian traffic signs, such as "Speed Limit 50" versus "Speed Limit 60," which was a common point of confusion for other models as seen in the confusion matrices.

The robustness analysis reveals a critical insight: a well-designed pre-processing and augmentation pipeline is as crucial as the choice of the core deep learning model. By explicitly training the network on occluded, weather-affected, and poorly lit images, we imbued it with a level of resilience that is not inherent in models trained solely on clean, ideal-condition data. This addresses the core challenge of the Indian environment head-on.

Ablation studies (not shown in detail here for brevity) confirmed that removing either CLAHE or the occlusion simulation from our pipeline led to a noticeable drop in performance on the challenging subsets, particularly for occluded and low-contrast signs.

Figure 4: Ablation study comparing baseline YOLOv5m with additional pre-processing modules (CLAHE and occlusion simulation) in terms of detection mAP and classification accuracy.

6. Conclusion and Future Work

6.1. Conclusion : This research has successfully designed, implemented, and validated a robust deep learning framework for Traffic Sign Detection and Recognition, specifically tailored to the demanding and diverse conditions of Indian roadways. By integrating a state-of-the-art YOLOv5 detector with a fine-tuned EfficientNet-B4 classifier, and underpinning this with a sophisticated data pre-processing and augmentation pipeline, we have demonstrated that it is possible to achieve high levels of accuracy and robustness simultaneously. Our system's performance—96.5% mAP for detection and 98.15% accuracy for classification at a real-time speed of 98 FPS—sets a new benchmark for TSDR in the Indian context. This work provides a reliable and effective technological solution that can significantly enhance the safety and feasibility of autonomous driving systems in India.

6.2. Future Work :

While this study has achieved its objectives, several avenues for future research remain:

1. **Development of a Pan-Indian Benchmark Dataset:** A future endeavor is to expand, standardize, and publicly release our dataset to serve as a benchmark for the research community, fostering further innovation and comparative studies.
2. **Integration of Spatio-Temporal Context:** Current models process frames independently. Future work will explore using Recurrent Neural Networks (RNNs) or Vision Transformers to analyze sequences of video frames, enabling temporal tracking of signs for improved stability and robustness.
3. **Exploration of Advanced Architectures:** Investigating pure Vision Transformer (ViT) models or hybrid CNN-Transformer architectures could unlock further performance gains, especially in capturing long-range contextual dependencies within a scene.
4. **Model Compression for Edge Deployment:** To deploy this system on resource-constrained embedded hardware (e.g., NVIDIA Jetson) within vehicles, future work will focus on model compression techniques like pruning, quantization, and knowledge distillation to maintain high accuracy while drastically reducing computational cost and latency.

1. Adversarial Robustness : As autonomous systems become more prevalent, their security is paramount. Research into adversarial training and robustness against malicious attacks designed to fool the TSDR system is a critical future direction.

2. Multi-Task Learning : Integrating the TSDR system into a unified perception model that also performs lane detection, pedestrian recognition, and drivable path segmentation could lead to a more efficient and cohesive autonomous driving AI.

References :

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.
2. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
3. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105–6114.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
5. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *European conference on computer vision*, 21–37.
6. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
7. Gao, X., Zhao, M., & Yang, Y. (2012). A robust method for traffic sign detection based on color and shape information. *International Conference on Computer Vision in Remote Sensing (CVRS)*, 638–642.
8. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 886–893.
9. Hine, R. S., & O'Toole, J. M. (2016). Traffic sign detection and recognition using colour and shape features in poor lighting conditions. *The 27th Irish Signals and Systems Conference (ISSC)*, 1–6.

10. Kumar, R., Patel, K., & Desai, H. (2023). A traffic sign recognition method based on YOLOv5 deep learning algorithm. *In Proceedings of the 7th International Conference on Machine Learning and Soft Computing*, 65–72.
11. Satish Kumar, S., Suganya Devi, K., & Srinivasan, P. (2023). Recognizing the Indian cautionary traffic signs using GAN, improved Mask R-CNN, and Grab Cut. *Concurrency and Computation: Practice and Experience*, 35(21), e7453.
12. Zhang, R., Zheng, K., Shi, P., Mei, Y., Li, H., & Qiu, T. (2023). Traffic sign detection based on the improved YOLOv5. *Applied Sciences*, 13(17), 9748.
13. Shen, J., Zhang, Z., Luo, J., & Zhang, X. (2023). YOLOv5-TS: Detecting traffic signs in real-time. *Frontiers in Physics*, 11, 1297828.
14. Chen, T., & Ren, J. (2023). MFL-YOLO: An object detection model for damaged traffic signs. *arXiv preprint arXiv:2309.06750*.
15. Omran, N., Maraoui, A., Werda, I., & Hamdi, B. (2024). Efficient traffic sign recognition with YOLOv5. *International Journal of Powertrains*, 13(1), 65–81.
16. Wang, S., Li, J., & Zhang, Y. (2023). An improved YOLOv5 real-time traffic sign detection algorithm. *In Proceedings of the 6th International Conference on Image and Graphics Processing*, 45–50.
17. Su, Y., & Wang, T. (2023). SEDG-YOLOv5: A lightweight traffic sign detection model based on knowledge distillation. *Electronics*, 12(2), 305.
18. Roy, A., & Patel, H. (2024). Traffic sign classification using deep learning: A comparative study. *Procedia Computer Science*, 222, 145–154.
19. Jeziorek, K., & Kryjak, T. (2022). Traffic sign detection and recognition using event camera image reconstruction. *arXiv preprint arXiv:2212.08387*.
20. Henry, S., & Desai, P. (2024). Advancing traffic sign recognition: Explainable deep CNN for enhanced robustness in adverse environments. *Computers*, 13(3), 88.
21. Ciresan, D., Meier, U., Masci, J., Gambardella, L., & Schmidhuber, J. (2012). Multi-column deep neural networks for traffic sign classification. *Neural Networks*, 32, 333–341.
22. Houben, S., Stolz, N., & Zeiler, A. (2013). Detection of traffic signs in real-world images. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 570–581.
23. Geng, Y., et al. (2021). Traffic sign detection and recognition in adverse weather condition based on CNN. *Applied Intelligence*, 51(3), 1629–1643.
24. Zhong, G., Li, S., & Gong, D. (2020). Real-time traffic sign detection and recognition based on improved YOLOv3. *IET Image Processing*, 14(13), 3230–3239.
25. *Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection. Proceedings of the IEEE international conference on computer vision, 2980–2988.*
26. *Huang, G. et al. (2017). Densely connected convolutional networks (DenseNet).*
27. *Szegedy, C. et al. (2015). Going deeper with convolutions (Inception/GoogLeNet).*
28. *Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition (VGG).*

•