Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 5, 2026

DEEP LEARNING-BASED TRAFFIC SIGN DETECTION IN DEVELOPING-COUNTRY ROAD CONDITIONS: A COMPARATIVE STUDY OF YOLOV8, YOLOV5, VISION TRANSFORMER AND RESNET18

Asad Ullah Gill[*1], Qamar Farooq[2], Haroon Noor[3], Muhammad Hamza Afzal[4], Qamar Ayyub[5]

[*1,2,3,4,5]*Department of Computer Science, the superior university Lahore (Faisalabad Campus)*

[*1]qfarooq506@gmail.com

DOI: https://doi.org/10.5281/zenodo.20374040

Keywords
Traffic sign detection; YOLOv8n; YOLOv5; Vision Transformer; ResNet18; intelligent transportation systems; road safety; developing countries

Article History
Received: 27 March 2026
Accepted: 07 May 2026
Published: 25 May 2026

Corresponding Author: *
Asad Ullah Gill

Abstract
Traffic sign detection is a safety-critical perception task for intelligent transportation systems, driver-assistance applications and road-infrastructure monitoring. In developing-country road environments, this task is complicated by faded or damaged signboards, inconsistent installation heights, dust, partial occlusion, motion blur, illumination variation and visually cluttered backgrounds. This paper presents a comparative deep learning study for traffic sign detection and recognition using YOLOv8n, YOLOv5nu, Vision Transformer (ViT-tiny) and ResNet18. A YOLO-format dataset containing 2,099 labelled images across 21 class identifiers was normalized into 1,679 training images and 420 validation images. YOLO models were trained at 640-pixel image size for 22 epochs using AdamW, while the image-level classifier branch was fine-tuned for 10 epochs. Experimental results show that YOLOv8 achieved 95.94% precision, 97.53% recall, 96.73% F1-score, 98.51% mAP@50 and 84.50% mAP@50-95. YOLOv5 obtained a slightly higher mAP@50 of 98.72%, whereas YOLOv8 provided stronger recall and marginally better mAP@50-95. For classification, ResNet18 reached 98.90% accuracy and weighted F1-score, while ViT-tiny achieved 62.91% accuracy, indicating that the transformer branch requires more data, stronger augmentation or hybrid local-global design before deployment. The findings support YOLOv8n as a practical real-time detection backbone for cost-aware traffic sign monitoring, while also showing that detector-classifier cascades must be validated end to end before being claimed as operationally superior.

1. Introduction
Road signs communicate regulatory, warning and guidance information to drivers, pedestrians and autonomous or semi-autonomous perception systems. Accurate automatic detection of these signs is therefore important for advanced driver-assistance systems, road safety auditing, intelligent surveillance, smart-city planning and autonomous mobility. Standard benchmark datasets have shown that traffic sign recognition can reach very high accuracy under controlled conditions, but field performance remains sensitive to data quality, scene variation and deployment hardware [1],[2].

Developing-country road environments introduce additional challenges that are not always represented in clean benchmark data. Signboards may be faded, bent, scratched, partially hidden by trees, placed at non-standard heights or mounted near visually similar advertisements. Camera

streams may include dust, rain, low illumination, motion blur, compression artifacts and complex backgrounds. These factors reduce the reliability of traditional hand-crafted algorithms and create a need for data-driven models that can generalize to non-ideal road conditions.

Classical traffic sign systems typically rely on color thresholding, contour extraction, Hough transforms, histogram of oriented gradients, template matching and support vector machines. Such approaches are interpretable and computationally light, but they depend heavily on stable color, shape and lighting assumptions. Deep learning offers a stronger alternative because convolutional and attention-based models learn discriminative features directly from annotated images. One-stage detectors such as YOLO are particularly attractive because they localize and classify objects in a single forward pass, which supports real-time road-scene processing [3], [4].

This study evaluates YOLOv8n and YOLOv5nu as object detectors and compares ViT-tiny and ResNet18 as recognition classifiers. The goal is not to claim a new network architecture; rather, the contribution is an experimentally grounded comparison and deployment interpretation for a traffic sign dataset prepared in YOLO annotation format. This careful positioning is important for journal submission because it avoids unsupported claims about an end-to-end hybrid model when the available evidence is component-level.

1.1 Research Problem

The central research problem is the selection of a reliable, lightweight and reproducible model configuration for detecting traffic signs in non-ideal road scenes. A detector deployed in traffic-safety applications should not only achieve high average precision but also maintain high recall, because missed signs may have a greater operational cost than occasional false detections. At the same time, a recognition module should be validated carefully before being added to a real-time cascade, because a weak classifier can reduce the reliability of an otherwise accurate detector.

The study therefore asks which evaluated model is most suitable as a practical traffic sign detection backbone, how detector and classifier performance differ on the available data, and what limitations must be addressed before real-world deployment.

1.2 Research Contributions

- A normalized YOLO-format traffic sign dataset of 2,099 images and 21 class identifiers is documented for detection and recognition experiments.
- YOLOv8n and YOLOv5nu are compared using precision, recall, F1-score, mAP@50 and mAP@50-95 under the same validation split.
- ViT-tiny and ResNet18 are evaluated to assess the feasibility of a recognition-refinement branch after sign localization.
- A deployment-oriented interpretation is provided for developing-country road monitoring, including class-map limitations, validation requirements and edge-device considerations.
- The manuscript distinguishes between component-level evidence and a fully validated end-to-end detector-classifier cascade, reducing the risk of overclaiming.

2. Related Work

2.1 Traditional Traffic Sign Detection

Traditional traffic sign detection methods exploit the distinctive color and geometric structure of road signs. Red, blue, yellow and white regions are often isolated in RGB, HSV or HSI color spaces and then processed using connected-component analysis, contour filtering or Hough transforms. Candidate regions are classified using hand-crafted descriptors such as HOG, SIFT or SURF combined with classifiers such as SVM or random forests.

Although these methods are fast and explainable, they are fragile in field conditions. Illumination changes alter color distribution, rain and dust reduce contrast, occlusion damages shape boundaries, and cluttered visual backgrounds increase false positives. These weaknesses are especially problematic when local traffic

infrastructure is not standardized or regularly maintained.

2.2 CNN-Based Recognition and Object Detection

Convolutional neural networks learn hierarchical representations from raw pixels. Early layers capture edges and textures, while deeper layers capture object parts and semantic patterns. Residual networks improved deep CNN training through skip connections, enabling high-performing image recognition models such as ResNet18 [5]. In traffic sign recognition, CNNs remain competitive because road signs contain strong local shape and symbol patterns.

YOLO-based detectors formulate detection as a single-stage regression and classification problem. YOLOv3 and YOLOv4 established strong speed-accuracy trade-offs for real-time detection [3], [4]. YOLOv8, distributed through the Ultralytics framework, provides a modern real-time detection interface and an optimized family of model sizes, including the lightweight YOLOv8n variant [6], [7]. Recent traffic sign studies have also shown interest in improved YOLOv8 variants for small or degraded traffic signs [8].

2.3 Vision Transformers for Traffic Sign Recognition

Transformers use self-attention to model relationships across tokens and were first popularized for sequence modelling [9]. Vision Transformer adapts this idea by dividing an image into patches and applying transformer encoders to the resulting patch embeddings [10]. This global context can be beneficial when traffic signs are partially degraded or embedded in complex scenes.

However, pure transformer models often require larger data volumes and stronger regularization than CNNs because they have weaker built-in local inductive bias. Recent traffic sign studies therefore frequently explore hybrid approaches that combine convolutional locality with transformer attention [11], [12], [13]. The weak ViT-tiny result in the present experiment is consistent with this broader observation: transformer-based recognition can be promising, but it must be trained and validated under data conditions that support attention-based learning.

3. Materials and Methods

3.1 Dataset Preparation

The experiment used a YOLO-format traffic sign dataset uploaded as a ZIP archive and normalized into a standard train/validation directory structure. Valid image-label pairs were identified and copied into the appropriate folders. The final dataset contained 2,099 images, divided into 1,679 training images and 420 validation images, corresponding to an approximately 80:20 split.

The annotation files contained 21 numeric class identifiers. Because the available dataset metadata did not include verified human-readable class names, the study reports class identifiers as class_0 to class_20. This reporting choice avoids assigning incorrect semantic names to classes. For final deployment, the numeric class map should be replaced with official traffic sign names supplied by the dataset owner or verified manually by domain experts.

Table 1. Dataset and experimental environment summary.

Item	Value
Total labelled images	2,099
Training images	1,679
Validation images	420
Number of class identifiers	21
Dataset split	Approximately 80% train / 20% validation
Annotation format	YOLO text labels
YOLO input image size	640 pixels
Classifier input image size	224 pixels

Training environment reported by notebook	Google Colab with NVIDIA A100 GPU

3.2 Preprocessing and Augmentation

Images and labels were reorganized into a format compatible with the Ultralytics YOLO training interface. The YOLO training configuration used standard augmentation mechanisms, including scale transformation, translation, mosaic augmentation, color-space augmentation and horizontal flipping where appropriate. These augmentations help expose the detector to variation in size, position and illumination.

For the classification branch, images were resized to 224 x 224 pixels and transformed using PyTorch/timm-compatible preprocessing. The detector and classifier experiments were treated as separate experimental components. This distinction is important because an end-to-end cascade score would require passing every detected bounding box to a classifier and then evaluating the final combined localization-recognition output.

3.3 Model Architectures and Training Configuration

Four model families were evaluated. YOLOv8n was selected as the primary lightweight detector because it is suitable for real-time deployment on resource-constrained platforms after export and optimization. YOLOv5nu was used as a detection baseline. ViT-tiny-patch16-224 and ResNet18 were fine-tuned as classifier baselines to study whether a separate recognition-refinement stage is feasible.

The YOLO models were trained for 22 epochs with image size 640, batch size 16, AdamW optimizer and learning rate 0.001. The classifier models were fine-tuned for 10 epochs with batch size 16, AdamW optimizer and learning rate 0.001. Table 2 summarizes the experimental settings.

Table 2. Model configurations used in the experiment.

Model	Experimental role	Training setting
YOLOv8n	Primary real-time traffic sign detector	22 epochs, image size 640, batch size 16, AdamW, learning rate 0.001
YOLOv5nu	Detection baseline	22 epochs, image size 640, batch size 16, AdamW, learning rate 0.001
ViT-tiny-patch16-224	Transformer-based recognition classifier	10 epochs, batch size 16, AdamW, learning rate 0.001
ResNet18	CNN recognition classifier baseline	10 epochs, batch size 16, AdamW, learning rate 0.001

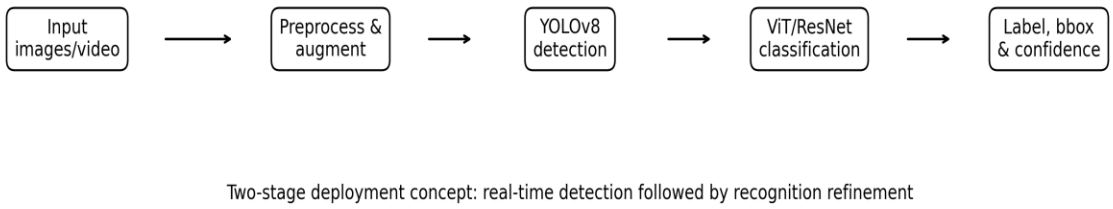

Fig. 1. Proposed two-stage deployment concept: real-time detection followed by optional recognition refinement.

3.4 Evaluation Metrics

Object detectors were evaluated using precision, recall, F1-score, mAP@50 and mAP@50-95. Precision measures the proportion of predicted detections that are correct, while recall measures the proportion of ground-truth signs that are detected. F1-score is the harmonic mean of precision and recall. mAP@50 reports mean average precision at an intersection-over-union threshold of 0.50, whereas mAP@50-95 averages performance across stricter thresholds and is therefore more informative about localization quality.

Classifier models were evaluated using accuracy, weighted precision, weighted recall and weighted F1-score. For YOLO models, mAP@50 is reported separately and should not be interpreted as image-level classification accuracy. This distinction prevents misleading comparison between object detection and classification tasks.

Precision = TP / (TP + FP), Recall = TP / (TP + FN), F1 = 2 x Precision x Recall / (Precision + Recall)

4. Experimental Results

4.1 YOLO Detector Performance

Both YOLO detectors produced strong validation performance. YOLOv5 obtained a slightly higher mAP@50 value, whereas YOLOv8 achieved higher recall and slightly better mAP@50-95. For traffic safety and infrastructure monitoring, recall is especially important because missed signs may create more serious downstream risk than a manageable number of false detections.

YOLOv8 achieved 95.94% precision, 97.53% recall, 96.73% F1-score, 98.51% mAP@50 and 84.50% mAP@50-95. YOLOv5 achieved 97.53% precision, 94.79% recall, 96.14% F1-score, 98.72% mAP@50 and 84.00% mAP@50-95. The difference is small, but YOLOv8 is preferred in this study because it provides the stronger recall and better localization robustness under the stricter mAP@50-95 metric.

Table 3. YOLO detector validation performance.

Detector	Precision (%)	Recall (%)	F1-score (%)	mAP@50 (%)	mAP@50-95 (%)
YOLOv8n	95.94	97.53	96.73	98.51	84.50
YOLOv5nu	97.53	94.79	96.14	98.72	84.00

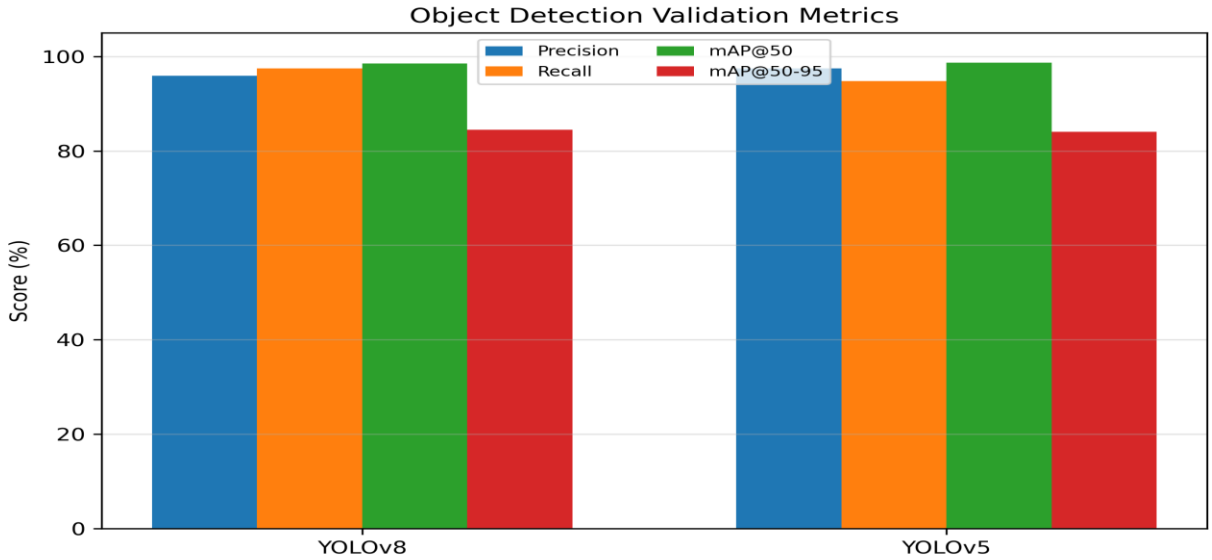

Fig. 2. Detection metric comparison between YOLOv8 and YOLOv5.

4.2 Classifier Performance

The classification results show a clear difference between the CNN and transformer branches. ResNet18 achieved 98.90% accuracy and 98.90% weighted F1-score, while ViT-tiny achieved 62.91% accuracy and 59.80% weighted F1-score. This gap indicates that the evaluated transformer configuration did not receive sufficient data diversity, training duration or regularization to compete with the CNN baseline.

The result does not imply that transformer-based recognition is unsuitable for traffic signs. Rather, it indicates that pure or small transformer variants are more sensitive to data volume and tuning. In moderate-sized traffic sign datasets, CNNs such as ResNet18 retain an advantage because their convolutional filters efficiently model local edges, shapes and symbols.

Table 4. Overall model performance summary. For detector rows, mAP@50 is not image-level classification accuracy.

Model	mAP@50 / Accuracy proxy (%)	Precision (%)	Recall (%)	F1-score (%)
YOLOv8n	98.51	95.94	97.53	96.73
YOLOv5nu	98.72	97.53	94.79	96.14
ViT-tiny	62.91	65.40	62.91	59.80
ResNet18	98.90	98.94	98.90	98.90

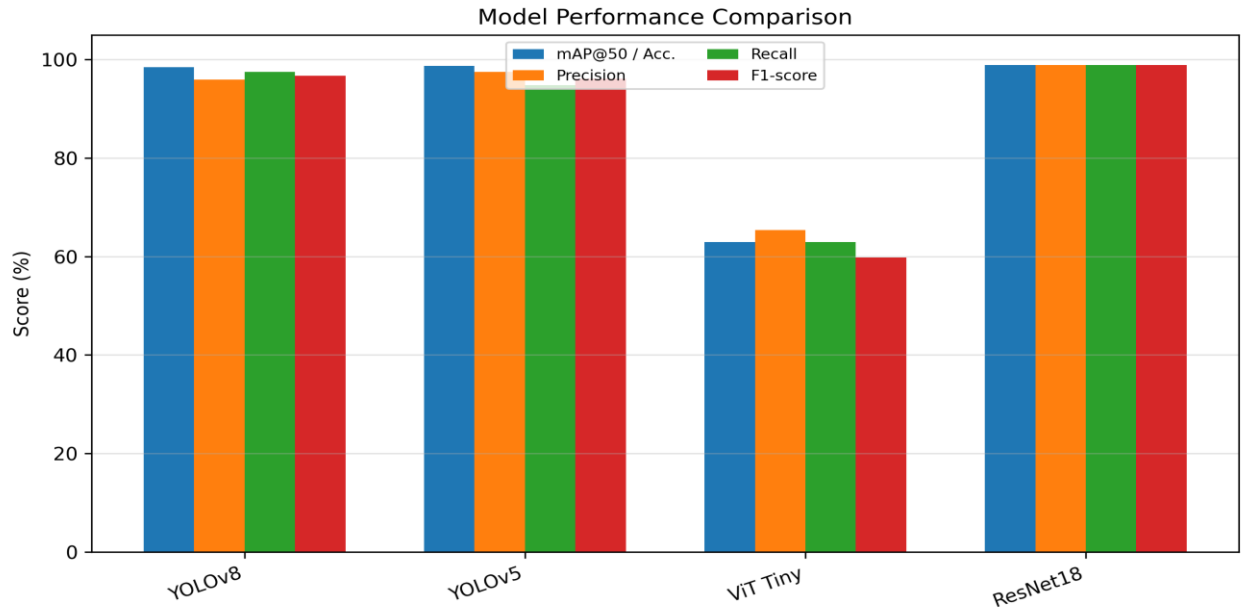

Fig. 3. Comparative performance of detection and classification models.

4.3 Qualitative Detection Output

A qualitative validation output is shown in Fig. 4. The predicted label is displayed using a numeric class identifier because verified human-readable class metadata were not available in the uploaded dataset. The output demonstrates that the detector localizes the sign region and returns a confidence score, but semantic interpretation should be improved by adding a validated class map before deployment.

Fig. 4. Example traffic sign localization output generated during validation.

5. Discussion

5.1 Interpretation of Detector Results

The detector results suggest that one-stage object detection is effective for traffic sign localization in the prepared dataset. YOLOv8n offers a favourable balance between high recall, strong mAP and lightweight deployment potential. While YOLOv5nu produced a marginally higher mAP@50 and precision, YOLOv8n achieved the strongest recall, which is preferable when the operational priority is to avoid missing signs.

The mAP@50-95 result is also important. A high mAP@50 can be achieved when predicted boxes overlap the ground truth at a relatively loose threshold, but mAP@50-95 requires accurate localization across multiple stricter thresholds. YOLOv8n obtained a slightly higher mAP@50-95 than YOLOv5nu, indicating that it provides a more reliable bounding-box quality under stricter evaluation.

5.2 Interpretation of Classifier Results

ResNet18 performed strongly in the classification branch, confirming that compact CNNs remain highly competitive for traffic sign recognition. Traffic signs are visually structured objects with distinctive local edges, symbols and colors, making them suitable for convolutional filters. This result supports the use of a CNN-based recognition model if a classifier refinement stage is required.

ViT-tiny underperformed substantially. This can be explained by limited training data, the absence of stronger transformer-specific augmentation, a short fine-tuning schedule, possible class imbalance and the weaker local inductive bias of pure transformer models. Future transformer experiments should consider hybrid CNN-transformer architectures, longer fine-tuning, balanced sampling, stronger augmentation, and pretraining on traffic-sign or road-scene data.

5.3 Relevance for Developing-Country Road Conditions

Developing countries often require cost-effective systems that can operate with limited local datasets and affordable computing infrastructure. A lightweight YOLOv8n-based detector is appropriate for such settings because it can be exported to ONNX or TensorRT and deployed on modest GPU-enabled or edge platforms after optimization. The model can support road-sign inventory generation, driver-assistance prototypes,

traffic-rule monitoring and infrastructure maintenance planning.

For municipal road monitoring, high recall is valuable because missed signs may allow damaged, faded or missing infrastructure to remain unnoticed. False positives can often be reviewed by a human operator, especially in offline road-audit workflows. In contrast, driver-assistance applications require stricter real-time validation, robust temporal stability and safety testing before deployment.

5.4 Threats to Validity

- Dataset scope: the validation set contains 420 images, which is useful for initial evaluation but not sufficient for broad deployment claims across all road and weather conditions.
- Class-map limitation: numeric class identifiers reduce interpretability and restrict per-sign-type error analysis.
- Component-level evaluation: detector and classifier results were evaluated separately, so the manuscript does not claim a completed end-to-end YOLOv8-ViT cascade.
- Hardware limitation: training was performed in a high-performance Colab environment; real-time performance must be measured on the actual target device.
- Environmental limitation: night-time, fog, rain, dust, occlusion and motion-blur subsets should be evaluated separately to quantify robustness.

Table 5. Practical limitations and recommended controls before field deployment.

Issue	Risk	Recommended control
Missing class names	Weak interpretability and limited reviewer confidence	Add verified class map and per-class metrics
Class imbalance	Rare signs may receive poor recall despite high average scores	Report per-class AP and collect targeted samples
Small signs	Localization errors and missed detections	Use higher-resolution testing and difficult-case subsets
Weather and lighting variation	Reduced field reliability	Create separate night, rain, fog, dust and blur tests
Unvalidated cascade	Recognition stage may reduce final system accuracy	Evaluate detector-classifier pipeline end to end

6. Deployment and Reproducibility Considerations

6.1 Proposed Deployment Workflow

A practical deployment workflow begins with frame acquisition from a dashboard camera, roadside camera or mobile inspection vehicle. Each frame is resized and passed through the YOLO detector. Detected signs are returned as bounding boxes, class identifiers and confidence scores. High-confidence detections can be stored directly, while low-confidence or ambiguous detections can be reviewed by a classifier or human operator.

For road-infrastructure auditing, outputs should be stored with timestamp, route segment, bounding-box coordinates, predicted class, confidence and optional GPS metadata. This allows municipalities to build a traffic-sign inventory and prioritize maintenance for damaged or missing signs. Human review should remain part of early deployment so that uncertain predictions can be corrected and added to the training dataset.

6.2 Reproducibility Checklist

Reproducibility is essential for journal review. The data split, training logs, class map, model weights, framework versions and random seeds should be preserved. Validation tables should be generated from raw logs rather than manually transcribed values. When possible, trained weights and inference scripts should be archived

as supplementary material subject to dataset license restrictions.

Table 6. Reproducibility checklist for the traffic sign detection experiment.

Step	Reproducibility action	Purpose
1	Store train/validation split and class map	Allows the same images and labels to be reused
2	Record framework versions and hyperparameters	Supports fair replication of all model runs
3	Save raw training and validation logs	Prevents metric transcription errors
4	Archive final weights and YAML configuration	Enables independent testing and deployment
5	Report hardware and inference latency	Separates model quality from hardware speed

7. Conclusion and Future Work

This paper presented a comparative deep learning study for traffic sign detection and recognition in non-ideal developing-country road conditions. YOLOv8n, YOLOv5nu, ViT-tiny and ResNet18 were evaluated on a normalized YOLO-format dataset containing 2,099 images and 21 class identifiers. YOLOv8n achieved 95.94% precision, 97.53% recall, 96.73% F1-score, 98.51% mAP@50 and 84.50% mAP@50-95, making it the preferred real-time detection backbone in this study. YOLOv5nu remained highly competitive and produced a marginally higher mAP@50, but YOLOv8n offered stronger recall and slightly better stricter localization performance.

The classifier comparison showed that ResNet18 was highly effective, achieving 98.90% accuracy and F1-score, while ViT-tiny underperformed with 62.91% accuracy. This finding suggests that CNN-based recognition remains reliable for moderate-sized traffic sign datasets, whereas transformer-based recognition requires more data, stronger augmentation, longer training or hybrid local-global design. The study recommends a YOLOv8n-first deployment strategy with optional CNN-based recognition refinement only after end-to-end validation.

Future work should add verified human-readable class names, report per-class average precision, expand the dataset with local road scenes, evaluate adverse-weather and night-time subsets, measure edge-device latency, and validate a complete detector-classifier cascade. Additional work should also investigate model compression, temporal smoothing for video, active learning with human review and integration with geographic information systems for road-infrastructure monitoring.

Declarations

Funding

No external funding was reported for the present manuscript.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," in *The 2011 International Joint Conference on Neural Networks*, San Jose, CA, USA: IEEE, Jul. 2011, pp. 1453–1460. doi: 10.1109/IJCNN.2011.6033395.

J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Netw.*, vol. 32, pp. 323–332, Aug. 2012, doi: 10.1016/j.neunet.2012.02.016.

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, *arXiv*. doi: 10.48550/ARXIV.2004.10934.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

"Ultralytics, 'YOLOv8 Documentation,' Ultralytics Docs. [Online]. Available: https://docs.ultralytics.com/models/yolov8/. Accessed: May 19, 2026."

M. Yaseen, "What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," 2024, *arXiv*. doi: 10.48550/ARXIV.2408.15857.

B. Ji, J. Xu, Y. Liu, P. Fan, and M. Wang, "Improved YOLOv8 for small traffic sign detection under complex environmental conditions," *Frankl. Open*, vol. 8, p. 100167, Sep. 2024, doi: 10.1016/j.fraope.2024.100167.

A. Vaswani *et al.*, "Attention Is All You Need," 2017, *arXiv*. doi: 10.48550/ARXIV.1706.03762.

A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 2020, *arXiv*. doi: 10.48550/ARXIV.2010.11929.

A. Farzipour, O. N. Manzari, and S. B. Shokouhi, "Traffic Sign Recognition Using Local Vision Transformer," in *2023 13th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, Islamic Republic of: IEEE, Nov. 2023, pp. 191–196. doi: 10.1109/ICCKE60553.2023.10326288.

G. Zeng, Z. Wu, L. Xu, and Y. Liang, "Efficient Vision Transformer YOLOv5 for Accurate and Fast Traffic Sign Detection," *Electronics*, vol. 13, no. 5, p. 880, Feb. 2024, doi: 10.3390/electronics13050880.

J. M. Kaleybar, H. Khaloo, and A. Naghipour, "Efficient Vision Transformer for Accurate Traffic Sign Detection," 2023, *arXiv*. doi: 10.48550/ARXIV.2311.01429.