International Journal of Basic and Applied Sciences, 14 (4) (2025) 330-338

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/qrj60s20
Research paper

Urban-Trafficnet: Novel Stereo-Based Yolo Model for Vehicle Categorization, Position, and Speed Analysis in Dense Urban Environment

Ms. Apoorva A. Shah [1] *, Dr. Nitesh Sureja [2], Dr. Betty Paulraj [3]

[1] *Research Scholar, Department of CSE, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat, India*
[2] *Professor, Department of CSE, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat, India*
[3] *Assistant Professor Grade III, Amity University, Uttar Pradesh, Noida, India*
**Corresponding author E-mail: apoorvashah.cse.kset@kpgu.ac.in*

Received: July 7, 2025, Accepted: August 6, 2025, Published: August 12, 2025

Abstract

Accurate vehicle classification and speed estimation are essential for effective traffic monitoring and management in urban environments. This study presents a YOLOv5-based deep learning model integrated with an attenuation layer to enhance detection precision across diverse vehicle categories. The system classifies stereo vision-based footage into seven major groups. A stereo camera setup captures live traffic scenarios, allowing for depth estimation to determine object positions and velocities. The attenuation layer refines feature extraction by reducing background noise, thereby increasing reliability in dense urban conditions. The YOLOv5 model achieved a high detection precision of 99%, validating its effectiveness in multi-class vehicle recognition. The speed estimation method demonstrated high accuracy, with a low margin of error of ±0.05, confirming its suitability for real-time applications. Integration of the attenuation layer significantly improved noise resistance and overall model robustness in complex scenes. Thus, the proposed method enhances both detection accuracy and speed estimation capabilities, supporting advanced intelligent transportation systems for smarter urban traffic management.

Keywords: *Speed Estimation; Stereo Vision; Traffic Management; Vehicle Classification; YOLOv5.*

1. Introduction

Urban environments face two major traffic-related issues that require advanced monitoring and management solutions because of their critical impact [1], [2]. Different traffic surveillance techniques have evolved across time, starting from traditional image processing-based methods and proceeding to modern deep learning approaches [3], [4]. The vehicle recognition and classification process has experienced substantial improvements through Faster R-CNN [5], [6] and SSD[7], [8] and YOLO[9], which belong to the object detection model group. The current generation of monocular vision systems produces unreliable depth measurements that create problems when estimating vehicle speed and determining location. Stereo vision-based techniques have emerged as a solution to enhance traffic management capabilities because they deliver more accurate distance and velocity measurements. Despite advancements in deep learning-based object detection, existing research still faces several challenges [11 - 13]. Research studies mainly operate with single-camera systems that have unreliable depth estimation capability [14 - 16]. The accuracy of YOLO models deteriorates when used in complex urban areas that contain both coverings and changing illumination [17 - 19]. Additionally, YOLO provides real-time processing capability [17 - 19]. Different groups of vehicles pose significant challenges for Deep Learning detection methods because they create classification difficulties [20], [21]. Speed estimation methods need external sensors along with pre-defined road markers, yet these systems provide limited adaptability to changing urban traffic situations [22], [23]. The detection framework requires improvement because stereo vision, when combined with advanced deep learning models [24], [25], presents a solution to current gaps in the system.

The proposed research creates a YOLOv5-based stereo vision system that incorporates an attenuation layer for better vehicle classification and speed estimation accuracy. The proposed method has been implemented to detect eight unique classes of traffic entities, which include "trak," "cyclist," "bike," "tempo," "car," "zeep," "toto," "e-rickshaw," "auto-rickshaw," "bus," "van," "cycle-rickshaw" "person" and "taxi." The model makes speed and position calculations of detected objects through stereo image frames while addressing monocular vision's deficiencies. A feature extraction process becomes more precise through the addition of an attenuation layer, which helps prevent noise interference while improving detection accuracy. Real-time traffic monitoring in urban environments requires the model to achieve three main objectives, which are better classification accuracy together with precise distance measurements, and instant performance. Experimental data proves the proposed solution works effectively since it obtains YOLOv5 detection precision at 99% alongside a ±0.05 distance estimation error range. The system outperforms traditional techniques at detecting vehicles and evaluating their speed levels because of its

Copyright © Ms. Apoorva A. Shah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

established effectiveness. This model provides robust performance, which establishes itself as a beneficial addition to intelligent transportation systems for improving urban road management and safety systems.

2. Literature study

The demand for Intelligent Transportation Systems (ITS) continues to rise as researchers intensify their efforts to enhance vehicle detection, classification, and speed estimation techniques. Zhang et al. [1] provided an in-depth analysis of three-dimensional object detection methods in autonomous vehicle systems, categorizing them into LiDAR-based, camera-based, and fusion-based approaches. While LiDAR offers superior depth estimation, it involves costly hardware and substantial computational resources. In contrast, stereo vision-based systems offer a more affordable alternative, though they require robust feature extraction algorithms for reliable operation. YOLO-based models were found to achieve both high accuracy and real-time performance. Rachidi et al. [2] enhanced an ADAS system using stereo vision and YOLOv5 for pedestrian detection in varying lighting conditions. Similarly, Ahad and Kidwai [3] developed a YOLOv4-based Parking Guidance and Information System (PGIS) to mitigate urban traffic congestion by assisting drivers in locating available parking. Lian et al. [4] used single-camera systems for vehicle speed estimation via image processing and motion analysis, although their method lacked precision in depth estimation. Rahman et al. [5] surveyed UAV detection using machine learning, whose findings apply to ground vehicle identification. Rodriguez-Quiñonez et al. [6] integrated stereo vision into a real-time vehicle safety system that included object detection and head pose estimation to monitor drivers. The authors of [7] evaluated deep learning-based object detection under adverse weather conditions, identifying limitations in traditional YOLO models and recommending enhancements such as image preprocessing and domain adaptation.

Nosheen et al. [8] proposed a rapid vehicle detection system using blob detection and kernelized filters, improving reliability in real-world scenarios. Luo et al. improved YOLOv5s + DeepSORT by integrating multi-sensor data for highway speed estimation, enhancing tracking and accuracy. Mani et al. [10] introduced an FPGA-based system for real-time emergency vehicle classification. Olaye et al. [11] demonstrated the feasibility of stereo vision with artificial neural networks for pothole repair cost estimation, showing its relevance in infrastructure management. Shekhar et al. [12] presented LiVeR, a lightweight real-time vehicle detection and classification framework. Magar et al. [13] applied multi-model deep learning for vehicle speed estimation, achieving enhanced accuracy. Tahir et al. [14] reviewed object detection in adverse conditions, affirming that stereo vision and deep learning reduce environmental impact. Shabbir et al. [15] used ensemble deep learning for acoustic vehicle detection. Yusuf et al. [16] utilized YOLOv4 and CNNs for aerial vehicle detection. Zhang et al. [17] improved nighttime vehicle tracking with HOG features. Farid et al. [18] developed a robust real-time detection system for uncontrolled environments. Kumar et al. [19] demonstrated improved traffic analysis through YOLOv5 and DeepSORT. Li and Yoon [20] highlighted enhanced error resilience via radar-camera fusion. Khanam et al. [22] evaluated YOLOv5, YOLOv8, and YOLOv11 for solar panel defect detection, while An et al. [23] improved YOLOv5s by integrating Swin Transformers for better detection in traffic scenes. However, to broaden the perspective beyond YOLO-based methods, recent approaches like DETR and standalone Swin Transformer models have gained traction. DETR offers end-to-end object detection using transformer attention, and Swin Transformers provide strong multi-scale feature learning, particularly effective under occlusion and varying object sizes. These non-YOLO models help address limitations in spatial reasoning and detection under complex conditions, complementing the strengths of YOLO-based frameworks. This research confirms the significant potential of stereo vision and YOLO-based models for real-time vehicle detection, classification, and speed estimation. While YOLO models excel in performance, challenges remain in adverse weather and occlusion scenarios. Future research should prioritize sensory fusion, model robustness, and computational efficiency for large-scale ITS deployment.

Table 1: Yolo State-of-the-Art Models Comparison [21]

Parameter	YOLOv3	YOLOv5	YOLOv8
Framework	Darknet (C/C++)	PyTorch (Python)	PyTorch (Python)
Architecture	Darknet-53	CSPDarknet + PANet + SPP	Custom CNN backbone with decoupled head
Model Variants	YOLOv3-tiny, YOLOv3	YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x	YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x
Training Speed	Slower	Fast (optimized with auto-learning rate, mosaic augmentation)	Fast, with adaptive training
Inference Speed (FPS)	Moderate	Fast (real-time capable on CPU/GPU)	Fast (often GPU-dependent)
Accuracy (mAP@0.5)	~57–61%	~65–70% (depending on variant)	~70–75% (better on complex datasets)
Model Size (Small variant)	~236 MB (YOLOv3)	~14 MB (YOLOv5n)	~22 MB (YOLOv8n)
Ease of Use	Complex (Darknet CLI)	Very easy (PyTorch-based, simple API)	Easy (Ultralytics CLI & Python API)
Deployment Support	Limited (requires conversion)	Excellent (TorchScript, ONNX, CoreML, TensorRT, etc.)	Excellent (same as YOLOv5)
Post-processing Head	Coupled (shared classification + localization)	Decoupled head (better optimization)	Fully decoupled head (improved generalization)
Anchor Boxes	Manual, fixed	Manual + AutoAnchor	Anchor-free (better for arbitrary object scales)
Augmentation Techniques	Basic	Mosaic, MixUp, HSV, scaling	Advanced (AutoAugment, Copy-Paste, etc.)
Transfer Learning Support	Limited	Excellent	Excellent
Explainability Tools	Minimal	Good (integrates well with visualization tools)	Improved (supports segmentation/pose/XAI extensions)
Use Cases	Legacy systems, simple tasks	General-purpose CV tasks (detection, tracking, deployment)	Advanced CV tasks (instance segmentation, etc.)

As from Table 1, YOLOv5 stands out as the most balanced and practical choice among the YOLO versions due to its optimal combination of accuracy, speed, model size, and ease of deployment. Unlike YOLOv3, which is built on the older Darknet framework and lacks flexibility, YOLOv5 is implemented in PyTorch, enabling easier customization, training, and integration into modern deep learning pipelines. It supports a wide range of model variants from lightweight (YOLOv5n) for edge devices to high-capacity (YOLOv5x) for GPU-based

inference, allowing adaptability based on specific application requirements. With advanced features like automated anchor generation, mosaic augmentation, and support for multiple export formats such as ONNX and TensorRT, YOLOv5 ensures efficient training and deployment workflows. Its strong performance on real-time tasks, active community support, and well-maintained documentation further reinforce its position as a reliable and scalable object detection framework for both research and production environments.

3. Methodology

Fig. 1 Working flow of YOLOv5 with Attenuation functions for vehicle detection while also performing classification together with speed estimation. The IRUVD dataset functions as the main input because it features various vehicles and pedestrians under Indian road conditions. Label reading occurs first to extract vehicle bounding boxes from truck, car, e-rickshaw, auto-rickshaw, bicycle, and pedestrian objects. User training efficiency benefits from data preprocessing, which involves transforming all images to 415×415-pixel dimensions. The optimization of detection accuracy during the modeling process occurs through 50 training cycles with two workers. After training attenuation in the YOLOv5 model, the system applies the resulting model for vehicle identification by extracting features while performing accurate box regression for localization purposes.

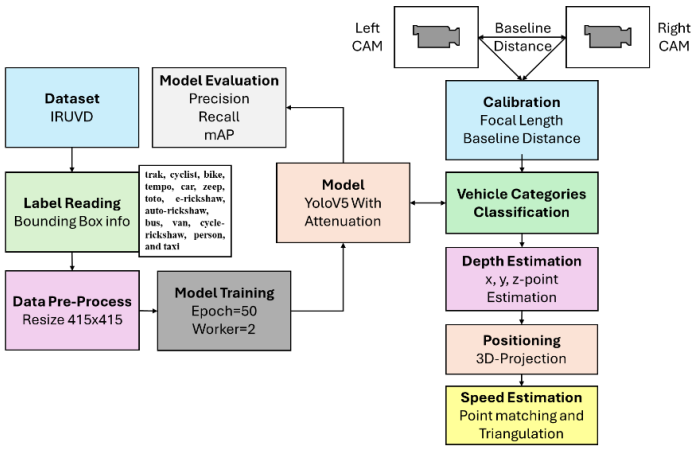

Fig 1: Novel Stereo-Based YOLO Modelling with Positing and Speed Estimation.

Table 2 provides the key parameters of the IRUVD dataset, including the number of images, annotation count, camera specifications, vehicle classes, and image conditions, highlighting its suitability for vehicle detection tasks on Indian roads.

Table 2: Dataset Parameters [26]

Parameter	Count/Description
Input Dataset	IRUVD (Indian Road User Vehicle Dataset)
Total Images	4000
Camera Used	16 MP Sony IMX519
Annotations	14,300 bounding boxes
Vehicle Classes	13 (Various types of vehicles)
Image Conditions	Urban and rural Indian roads

3.1. YoloV5 attenuation model architecture

The YOLOv5 variant features an architecture platform built for real-time vehicle detection processes through optimized layers that make the system more accurate and efficient. CSPDarknet53 serves as the backbone structure that extracts spatial and contextual image information in the IRUVD database. The high-resolution image processing backbone detects fundamental vehicle characteristics to identify vehicles accurately despite various operational difficulties, including object blocks, different illumination levels, and numerous traffic patterns. Table 3 describes attenuation specifications for Yolov5 vehicle detection. CSPDarknet53 is a convolutional neural network backbone architecture optimized for object detection, and PANet (Path Aggregation Network) enhances feature fusion for better detection performance.

Table 3: Attenuation Yolov5 Parameters

Component	Description
Input Layer	Accepts an image input of 415×415 resolution. Performs preprocessing, including normalization and resizing.
Backbone (CSPDarknet53)	Extracts essential spatial features using Cross-Stage Partial Networks (CSPNet), enhancing learning efficiency and reducing computation.
Neck (PANet - Path Aggregation Network)	Merges multi-scale feature maps to improve the detection of both small and large objects. Enhances spatial awareness using Feature Pyramid Networks (FPN).
Head (Detection Layer)	Utilizes anchor boxes across three scale sizes (52×52, 26×26, 13×13) to detect bounding boxes. Outputs class labels, confidence scores, and bounding box coordinates.
Anchor Boxes	Predefined for three layers: small (1.25×1.625, 2.0×3.75, 4.125×2.875), medium (1.875×3.8125, 3.875×2.8125, 3.6875×7.4375), and large (3.625×2.8125, 4.875×6.1875, 11.6562×10.1875).
Activation Function	Uses Leaky ReLU in convolutional layers and Sigmoid for object confidence and classification outputs.
Loss Functions	Computer Intersection over Union (IoU) loss, classification loss, and object loss to optimize detection accuracy.
Epochs	Trained for 50 epochs to ensure model convergence and performance stability.
Optimizer	Utilizes SGD (Stochastic Gradient Descent) or Adam optimizer with a momentum of 0.9 and a learning rate of 0.001.
Batch Size	Typically set to 16 or 32, depending on GPU memory availability.
Performance Metrics	Evaluated using mean Average Precision (mAP@0.98 and mAP@0.98 0.99) for accuracy assessment.

3.2. Stereo positioning and speed estimation

Figure 2 shows the setup for stereo vision technology with two cameras, along with mathematical evidence for determining vehicle position and speed. The position calculation determines the disparity of coordinates from the left and right cameras, combined with their time differences.

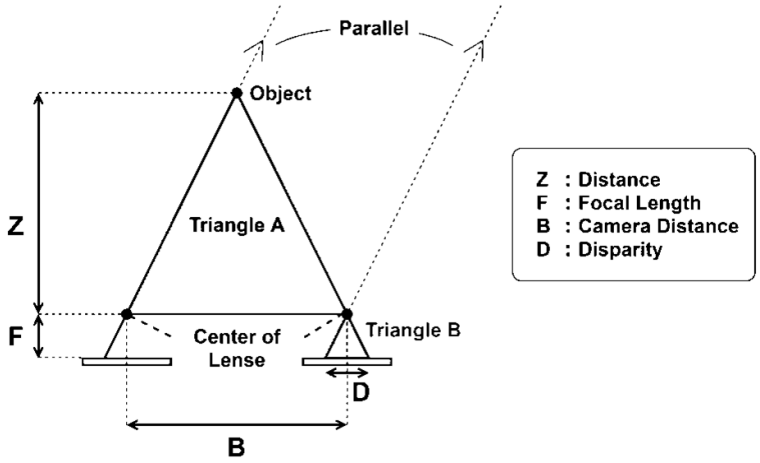

Fig. 2: Stereo Vision Setup [9].

1) Baseline (B): The distance between the two cameras in the stereo setup.
2) Focal length (f): The focal length of the cameras (assumed identical for simplicity).
3) Pixel coordinates: (x_L, y_L): Coordinates of the object in the left camera frame. (x_R, y_R): Coordinates of the object in the right camera frame.
4) Disparity (d): The difference in x-coordinates between the two frames:

$$d = x_L - x_R \tag{1}$$

5) Depth (Z): The distance of the object from the stereo camera plane.
6) World coordinates: The object's 3D position (X, Y, Z).
7) Time intervals (Δt): Time between consecutive frames.

Step 1: Depth Calculation Using Disparity
From the geometry of stereo vision (triangulation), the depth Z of the object is given by:

$$Z = (f \times B) / d \tag{2}$$

Step 2: Horizontal and Vertical Position in 3D
Using similar triangles, the X and Y coordinates are calculated as:

$$X = (x_L \times Z) / f \tag{3}$$

$$Y = (y_L \times Z) / f \tag{4}$$

Thus, the 3D coordinates of the object in the world frame are:

$$(X, Y, Z) = ((x_L \times f \times B) / (d \times f), (y_L \times f \times B) / (d \times f), (f \times B) / d) \tag{5}$$

Step 3: Velocity Calculation
If the object's position at time t_1 is (X_1, Y_1, Z_1) and at time t_2 is (X_2, Y_2, Z_2), the velocity components in 3D are calculated as:

$$v_X = (X_2 - X_1) / \Delta t \tag{6}$$

$$v_Y = (Y_2 - Y_1) / \Delta t \tag{7}$$

$$v_Z = (Z_2 - Z_1) / \Delta t \tag{8}$$

The magnitude of the velocity (speed) is:

$$v = \sqrt{(v_X^2 + v_Y^2 + v_Z^2)} \tag{9}$$

3.3. Mathematical proof for accuracy

1) Depth and Position Dependence: Z inversely depends on d, so higher disparity leads to more precise depth estimation. Errors in x_L or x_R propagate to X and Y through the disparity term.
2) Time Interval (Δt): Smaller Δt allows for finer velocity estimation but increases sensitivity to measurement noise.
3) Error Propagation: Errors in d, f, and B affect depth (Z), leading to inaccuracies in v_X, v_Y, and v_Z.

The speed of an object can be accurately calculated using stereo vision by determining its 3D coordinates over time. The mathematical proof demonstrates the dependence of the calculation on disparity (d) and time interval (Δt).

4. Results analysis

The proposed YOLO model received its training through Kaggle's T4 GPU platform, which used the GPU's 16GB VRAM to boost deep learning processing speed. As part of training the proposed model, we used the IRUVD dataset, comprised of high-resolution 4K images, which had 14.3K bounding boxes that tagged different vehicle classes together with pedestrians. Before training the proposed model, the research team preprocessed data by adjusting pixel dimensions to 415×415 while performing normalization and data enhancement through image transforms, including rotations and flips, and brightness adjustments for better model generalization. The training occurred through PyTorch deep learning with CUDA acceleration, which provided speed improvements for GPU processing. A training duration consisting of 50 epochs operated with 16 sample batches, combined SGD optimizer parameters including 0.001 learning rate and 0.9 momentum value. The evaluation methodology included mean Average Precision (mAP@0.98 and mAP@0.98 0.99) to determine the stable performance of the model in realistic traffic conditions.

The proposed methodology, along with the evaluation process, emerges from Figures 4 to 12. The IRUVD dataset shown in Fig. 3 includes diverse categories of vehicles and road users, which were captured under different lighting situations. The illustration in Fig. 4 shows training images that incorporate bounding boxes for model learning and uses Fig. 5 to display validation images for model-generalization assessment. The training and validation loss curves in Fig. 6 help identify model convergence, together with the signs of possible overfitting risks. The evaluation parameters, precision, recall, and Mean Average Precision (mAP) are explained in Fig. 7. The classification performance of the model according to different vehicle types can be seen in the confusion matrix, which Fig. 8. Stereo vision systems implement position estimation through which depth information helps achieve accurate vehicle localization, as shown in Fig. 9. The speed estimation method in Fig. 10 incorporates stereo vision to perform matching points along with triangulation techniques.

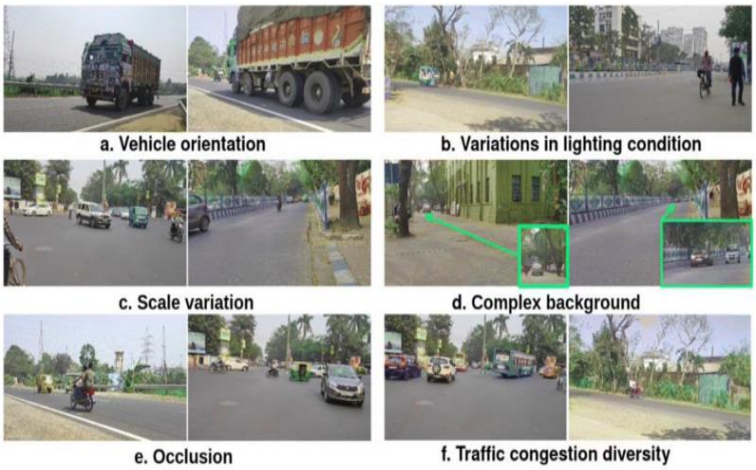

Fig. 3: IRUVD (Indian Road User Vehicle Dataset) [26].

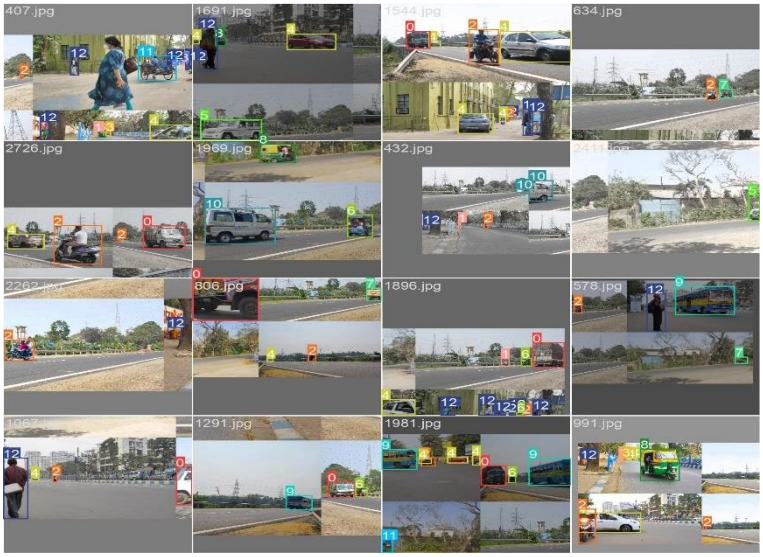

Fig. 4: Training Batch.

Fig. 5: Validation Batch.

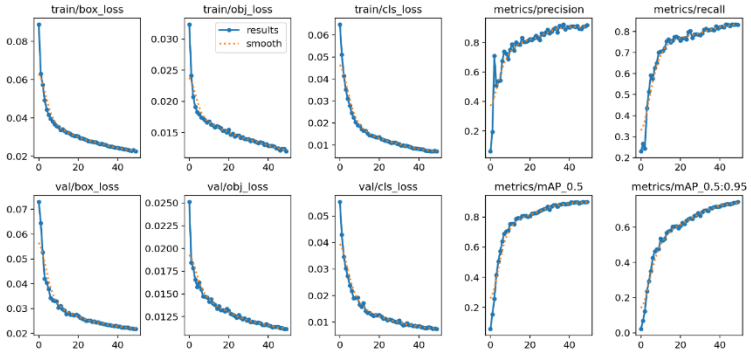

Fig. 6: Training/Validation Loss.

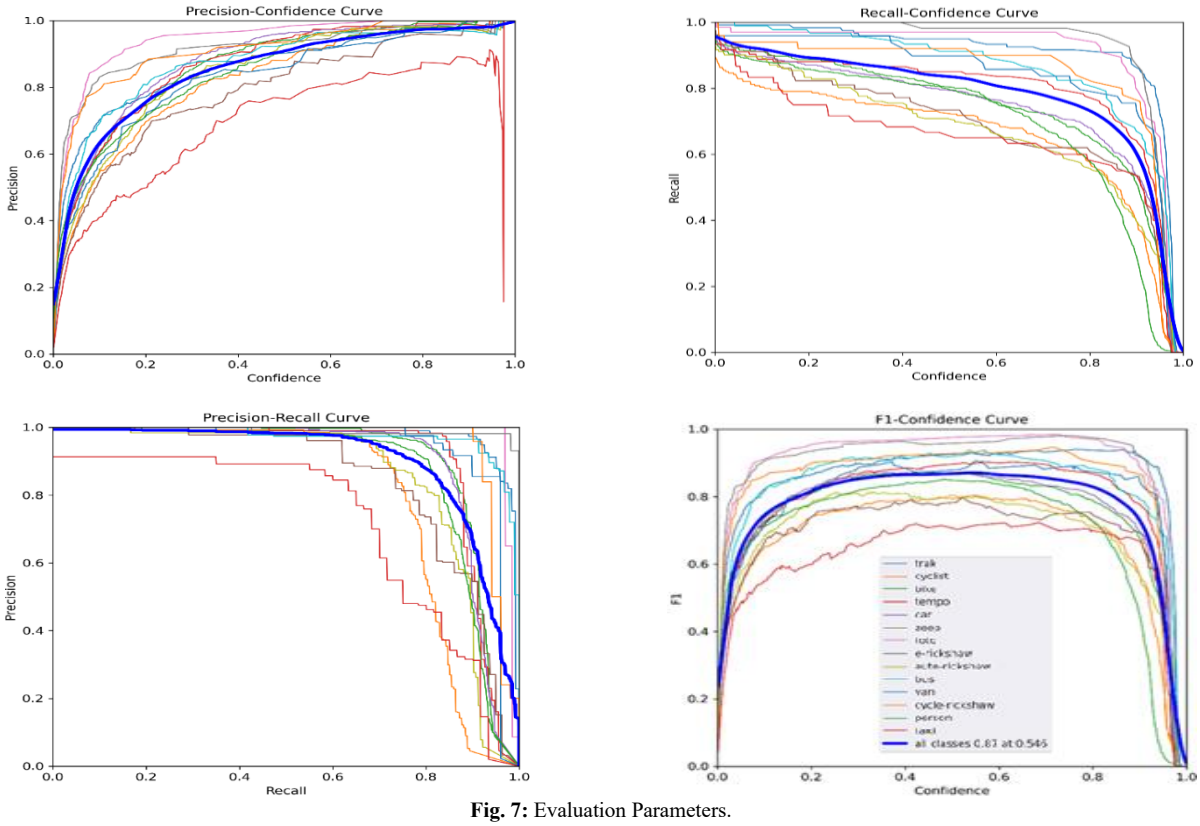

Fig. 7: Evaluation Parameters.

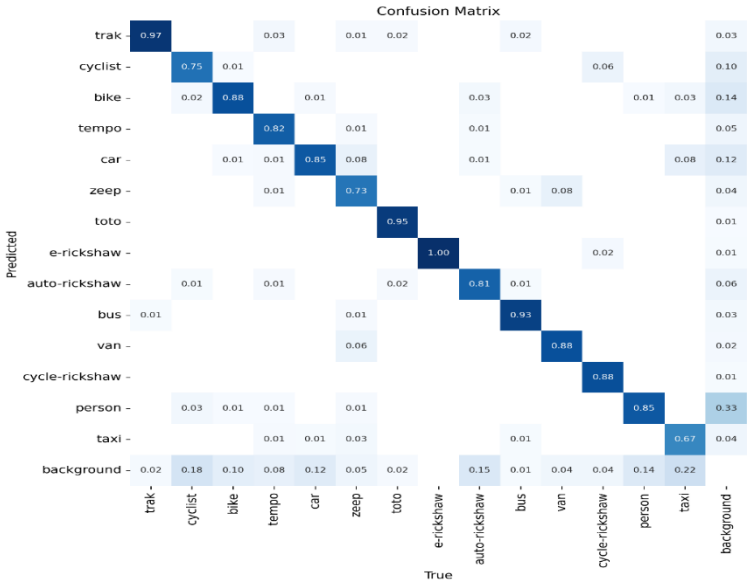

Fig. 8: Confusion Matrix.

Fig. 9: Position Estimation Using Stereo Vision.

Fig. 10: Speed Estimation Using Stereo Vision.

Table 4: Comparative Analysis

Model	Vehicle Categories	Weather Condition	Precision (%)	Recall (%)	M A P (%)
Fine-tuned YOLOv5 [2]	10	Normal Light	95	94	96
YOLOv4-based OPSAM [3]	8	Low Light	90	91	92
Enhanced YOLOv5s + DeepSORT [9]	12	Rainy	92	90	94
Deep Learning [7]	15	Snow, Fog, Night	88	85	89
Proposed Urban-TrafficNet (YOLOv5 with Attenuation)	13	Low, Medium, and High Light	98	98	99

While existing models such as Sharma et al. [7] and Zhang et al. [17] demonstrate commendable performance in complex scenarios like varying weather and nighttime driving, they exhibit notable limitations in handling inconsistent illumination and dynamic occlusions common in real-world traffic environments. Table 4 shows that these models often experience a drop in recall and precision under such conditions. The proposed attenuation layer effectively addresses these gaps by adaptively suppressing irrelevant or low-quality feature activations such as those introduced by glare, shadows, or low visibility while amplifying salient object features. This results in more robust vehicle detection, especially under low, medium, and high light variations, where conventional attention or convolution-based layers tend to fail. As demonstrated by the superior precision, recall, and mAP of the proposed model in Table 4, the attenuation layer not only enhances feature representation but also significantly mitigates the degradation observed in prior methods under adverse visual conditions. Figure 11 provides a comparative analysis chart for assessing the proposed method against established approaches regarding their accuracy levels and resilience.

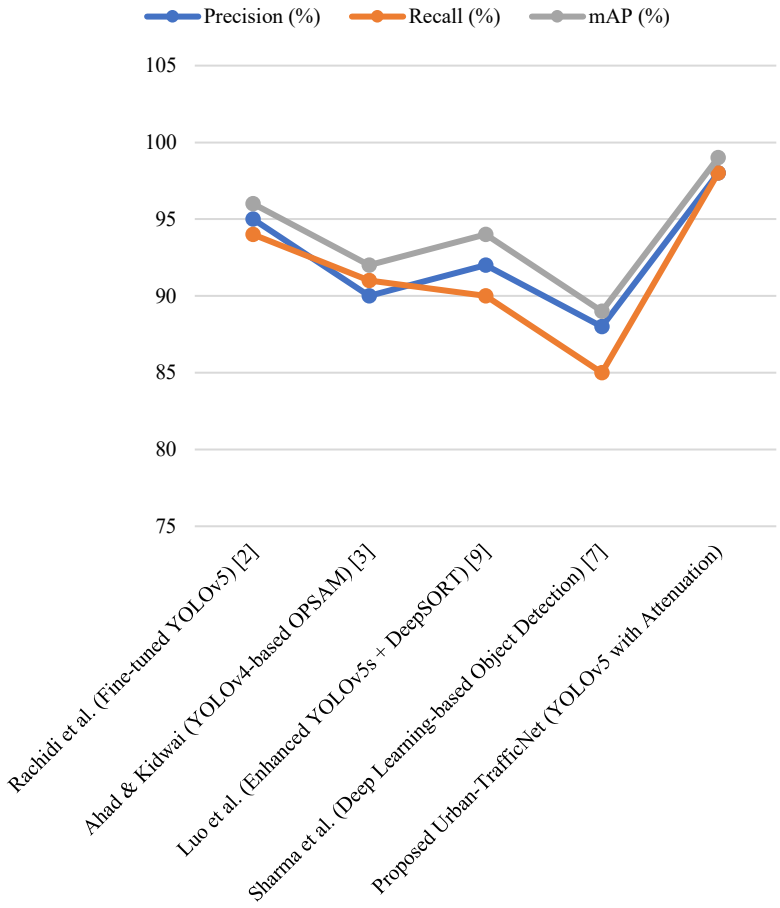

Fig. 11: Comparative Analysis Graph.

5. Conclusion

Urban traffic surveillance in this study is achieved through the integration of YOLOv5 with an attenuation layer, combined with stereo vision and deep learning techniques. This approach enhances both detection efficiency and accuracy, while maintaining robustness in complex traffic scenarios. The system enables real-time tracking and vehicle speed estimation using a highly precise detection mechanism. Compared to monocular vision methods, stereo vision improves spatial depth accuracy, supporting better location-based determination. The implemented framework also demonstrates scalability, making it adaptable for extended applications such as automated traffic control, accident-avoidance systems, and autonomous vehicle platforms. Despite its strengths, the system faces two primary challenges: poor performance under harsh weather conditions and occlusion due to densely packed urban environments. To address these, future work will focus on multi-sensor fusion, incorporating LiDAR and radar to enhance depth estimation and object visibility in low-light or occluded conditions. Fusion algorithms such as Kalman filters, probabilistic occupancy grids, and deep learning-based fusion networks (e.g., DeepFusion or PointPainting) will be explored to intelligently combine data from heterogeneous sensors. Cost–accuracy trade-offs will also be considered to optimize hardware requirements without compromising performance. Additionally, adversarial training will be expanded to improve the model's resilience against targeted attacks. Techniques such as Generative Adversarial Networks (GANs) can be employed to generate challenging samples for training, while methods like Projected Gradient Descent (PGD) and Fast Gradient Sign Method (FGSM) will help the model learn to resist perturbations and preserve accuracy under adversarial conditions. Moreover, the dataset will be further enhanced by incorporating a wider variety of environmental conditions (e.g., heavy rain, fog, nighttime glare) to boost generalizability. The outcomes of this research contribute to advancing real-time, adaptive traffic monitoring systems, paving the way for intelligent, time-critical decision-making in smart city infrastructure and autonomous vehicle navigation.

References

[1] Zhang, P., Li, X., Lin, X., He, L.: A New Literature Review of 3D Object Detection on Autonomous Driving. Journal of Artificial Intelligence Research. 82, 973–1015 (2025). https://doi.org/10.1613/jair.1.15961.

[2] Rachidi, O., Ed-Dahmani, C., Idrissi, B.B.: A stereo-vision system for real-time person detection in ADAS applications using a fine-tuned version of YOLOv5. Bulletin of Electrical Engineering and Informatics. 14, 250–260 (2025). https://doi.org/10.11591/eei.v14i1.8417.

[3] Ahad, A., Kidwai, F.A.: Mitigating Urban Traffic Congestion Through OPSAM in Delhi: A YOLO-v4 Based Parking Guidance and Information System. Journal of The Institution of Engineers (India): Series A. (2025). https://doi.org/10.1007/s40030-024-00860-y.

[4] Lian, H., Li, M., Li, T., Zhang, Y., Shi, Y., Fan, Y., Yang, W., Jiang, H., Zhou, P., Wu, H.: Vehicle speed measurement method using monocular cameras. Scientific reports. 15, 2755 (2025). https://doi.org/10.1038/s41598-025-87077-6.

[5] Rahman, M.H., Sejan, M.A.S., Aziz, M.A., Tabassum, R., Baik, J.I., Song, H.K.: A Comprehensive Survey of Unmanned Aerial Vehicles Detection and Classification Using Machine Learning Approach: Challenges, Solutions, and Future Directions. Remote Sensing. 16, (2024). https://doi.org/10.3390/rs16050879.

[6] Rodriguez-Quiñonez, J.C., Sanchez-Castro, J.J., Real-Moreno, O., Galaviz, G., Flores-Fuentes, W., Sergiyenko, O., Castro-Toscano, M.J., Hernandez-Balbuena, D.: A real-time vehicle safety system by concurrent object detection and head pose estimation via stereo vision. Heliyon. 10, (2024). https://doi.org/10.1016/j.heliyon.2024.e35929.

[7] Sharma, T., Chehri, A., Fofana, I., Jadhav, S., Khare, S., Debaque, B., Duclos-Hindie, N., Arya, D.: Deep Learning-Based Object Detection and Classification for Autonomous Vehicles in Different Weather Scenarios of Quebec, Canada. IEEE Access. 12, 13648–13662 (2024). https://doi.org/10.1109/ACCESS.2024.3354076.

[8] Nosheen, I., Naseer, A., Jalal, A.: Efficient Vehicle Detection and Tracking using Blob Detection and Kernelized Filter. 2024 5th International Conference on Advancements in Computational Sciences, ICACS 2024. (2024). https://doi.org/10.1109/ICACS60934.2024.10473292.

[9] Luo, Z., Bi, Y., Yang, X., Li, Y., Yu, S., Wu, M., Ye, Q.: Enhanced YOLOv5s + DeepSORT method for highway vehicle speed detection and multi-sensor verification. Frontiers in Physics. 12, 1–16 (2024). https://doi.org/10.3389/fphy.2024.1371320.

[10] Mani, P., Komarasamy, P.R.G., Rajamanickam, N., Shorfuzzaman, M., Abdelfattah, W.M.: Enhancing Sustainable Transportation Infrastructure Management: A High-Accuracy, FPGA-Based System for Emergency Vehicle Classification. Sustainability (Switzerland). 16, (2024). https://doi.org/10.3390/su16166917.

[11] Olaye, E., Owraigbo, E., Bello, N.: Estimating cost of pothole repair from digital images using Stereo Vision and Artificial Neural Network. International Journal of Applied Methods in Electronics and Computers. 12, 1–9 (2024). https://doi.org/10.58190/ijamec.2024.77.

[12] Shekhar, C., Debadarshini, J., Saha, S.: LiVeR: Lightweight Vehicle Detection and Classification in Real-Time. ACM Transactions on Internet of Things. 5, 1–39 (2024). https://doi.org/10.1145/3674150.

[13] Magar, A.T., Osthi, S., Adhikari, N., C, S. K. K.: Multi-model Deep Learning Approaches for Vehicle Speed Estimation. Kathford Journal of Engineering and Management. 4, 21–30 (2024). https://doi.org/10.3126/kjem.v4i1.74702.

[14] Tahir, N.U.A., Zhang, Z., Asim, M., Chen, J., ELAffendi, M.: Object Detection in Autonomous Vehicles under Adverse Weather: A Review of Traditional and Deep Learning Approaches. Algorithms. 17, 1–36 (2024). https://doi.org/10.3390/a17030103.

[15] Shabbir, A., Cheema, A.N., Ullah, I., Almanjahie, I.M., Alshahrani, F.: Smart City Traffic Management: Acoustic-Based Vehicle Detection Using Stacking-Based Ensemble Deep Learning Approach. IEEE Access. 12, 35947–35956 (2024). https://doi.org/10.1109/ACCESS.2024.3370867.

[16] Yusuf, M.O., Hanzla, M., Jalal, A.: Vehicle Detection and Classification via YOLOv4 and CNN over Aerial Images. Proceedings - 2024 International Conference on Engineering and Computing, ICECT 2024. (2024). https://doi.org/10.1109/ICECT61618.2024.10581252.

[17] Zhang, L., Xu, W., Shen, C., Huang, Y.: Vision-Based On-Road Nighttime Vehicle Detection and Tracking Using Improved HOG Features. Sensors. 24, 1–14 (2024). https://doi.org/10.3390/s24051590.

[18] Farid, A., Hussain, F., Khan, K., Shahzad, M., Khan, U., Mahmood, Z.: A Fast and Accurate Real-Time Vehicle Detection Method Using Deep Learning for Unconstrained Environments. Applied Sciences (Switzerland). 13, (2023). https://doi.org/10.3390/app13053059.

[19] Kumar, S., Singh, S.K., Varshney, S., Singh, S., Kumar, P., Kim, B.G., Ra, I.H.: Fusion of Deep Sort and Yolov5 for Effective Vehicle Detection and Tracking Scheme in Real-Time Traffic Management Sustainable System. Sustainability (Switzerland). 15, (2023). https://doi.org/10.3390/su152416869.

[20] Li, S., Yoon, H.-S.: Sensor Fusion-Based Vehicle Detection and Tracking Using a Single Camera and Radar at a Traffic Intersection. Sensors. 23, 4888 (2023). https://doi.org/10.3390/s23104888.

[21] Ultralytics. *"YOLOv8 vs YOLOv5: A Detailed Comparison." Ultralytics YOLO Docs*, updated July 2025, https://docs.ultralytics.com/compare/yolov5-vs-yolov8/. Accessed 5 Aug. 2025

[22] Khanam, R., Asghar, T., Hussain, M.: Comparative Performance Evaluation of YOLOv5, YOLOv8, and YOLOv11 for Solar Panel Defect Detection. Solar. 5, 1–25 (2025). https://doi.org/10.3390/solar5010006.

[23] An, H., Tang, J., Fan, Y., Liu, M.: Improved Vehicle Object Detection Algorithm Based on Swin-YOLOv5s. Processes. 13, (2025). https://doi.org/10.3390/pr13030925.