Journal of Electrical and Intelligent Systems (JEIS), Vol. 1, Issue 1
April 2026
ISSN: XXXX-YYYY

Motion Consistent Object Detection: A Velocity Constrained Filtering Framework for Traffic Perception

Gusty Anugrah*, Dedid Cahya Happyanto, Eru Puspita
Department of Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Indonesia
**Corresponding author:Gusty Anugrah; email: gustyaan2468@gmail.com*

Manuscript received 24 April 2026, revised 5 May 2026, accepted 11 May 2026
doi: preprint version. Pp33-41

Abstract – *Conventional object detection systems in Intelligent Transportation Systems (ITS) commonly operate in a frame-wise manner, leading to temporally inconsistent predictions under dynamic conditions. This paper proposes a motion-consistent object detection framework that integrates velocity-constrained filtering to enforce physical plausibility across consecutive frames. The main contribution of this work lies in explicitly formulating detection validation as a deterministic velocity-bounded constraint problem, enabling direct integration of motion dynamics into the perception pipeline rather than treating motion as auxiliary information. The system employs a lightweight YOLO11s-based detector for visual perception and GPS-based motion estimation to provide real-world velocity constraints. Experiments are conducted on a traffic infrastructure dataset comprising 52,135 images across 33 classes and evaluated on an Intel NUC 13 Pro edge computing platform. Results demonstrate that the proposed method improves temporal stability compared to conventional frame-wise detection while maintaining strong detection performance, achieving an mAP@0.5 of 0.945 at 20 FPS. The findings indicate that incorporating explicit motion constraints enhances detection reliability without introducing significant computational overhead. However, performance may degrade under low-speed GPS noise and challenging visual conditions such as occlusion, highlighting limitations for future improvement.*

Keywords: *Object detection, motion consistency, velocity-constrained filtering, intelligent transportation systems, GPS-based velocity estimation, temporal stability*

I. Introduction

The advancement of Intelligent Transportation Systems (ITS) and Advanced Driver Assistance Systems (ADAS) has increased the need for accurate and real-time perception in dynamic traffic environments. Recent progress in deep learning-based object detection, including transformer-based models and real-time architectures such as YOLOv11, has significantly improved detection performance in complex scenes [1]-[2]. In addition, edge intelligence enables deployment under strict latency and computational constraints [3]-[4], while lightweight detection models further support real-time traffic monitoring in resource-limited environments [5]. However, most existing methods primarily operate on frame-wise assumptions, focusing on spatial accuracy without explicitly enforcing temporal or physical consistency across frames.

Several studies incorporate multimodal information such as GPS-based tracking and vision fusion to enhance contextual awareness [6]-[7]. Other works extend scalability through edge-based traffic perception systems [8]-[9]. Despite these developments, existing approaches can generally be categorized into three groups: frame-wise detection, temporal learning-based models, and multimodal fusion systems. In all categories, motion information is typically treated as an auxiliary signal rather than a deterministic constraint within the detection validation process.

This limitation leads to temporally inconsistent predictions, particularly under occlusion, motion blur, and adverse environmental conditions [10]. Temporal consistency learning methods attempt to mitigate this issue [11], while physics-aware perception frameworks introduce general physical priors [12]. However, these approaches rely on implicit or learned constraints, which do not explicitly enforce physically bounded inter-frame motion, potentially allowing unrealistic object trajectories. To clarify these differences, Table I summarizes representative approaches:

TABLE I.
COMPARISON OF EXISTING METHODS AND THE PROPOSED APPROACH

Method	Motion Modeling	Constraint Type	Key Limitation
Hu et al. [13]	Explicit	Topological	Dependent on intersection detection
Xiao et al. [14]	Implicit	Feature-Based	High computational cost
Zhang et al. [15]	Explicit	Kinematic	Limited to motion forecasting
Proposed Method	Explicit	Velocity-bounded	Dependent on GPS accuracy

Motivated by these limitations, this paper proposes a motion-consistent object detection framework that integrates velocity-constrained filtering across consecutive frames to enforce physically plausible object transitions. Unlike prior approaches that rely on implicit temporal modeling or learned priors, the proposed method explicitly formulates detection validation as a deterministic physical consistency problem by constraining inter-frame object displacement using estimated vehicle velocity.

This formulation enables direct rejection of motion-infeasible detections within the perception pipeline, rather than post-hoc correction. The main contributions are: (1) a deterministic velocity-bounded detection validation framework, (2) a motion-consistent filtering mechanism enforcing explicit physical constraints across time, and (3) an experimental evaluation showing improved temporal stability and reduced false positives compared to frame-wise baselines under controlled experimental conditions.

II. Research Method

A. *Motion consistent Perception Architecture*

The proposed motion consistent perception architecture enforces physical consistency by tightly integrating detection, motion modeling, and constraint validation within a unified pipeline.

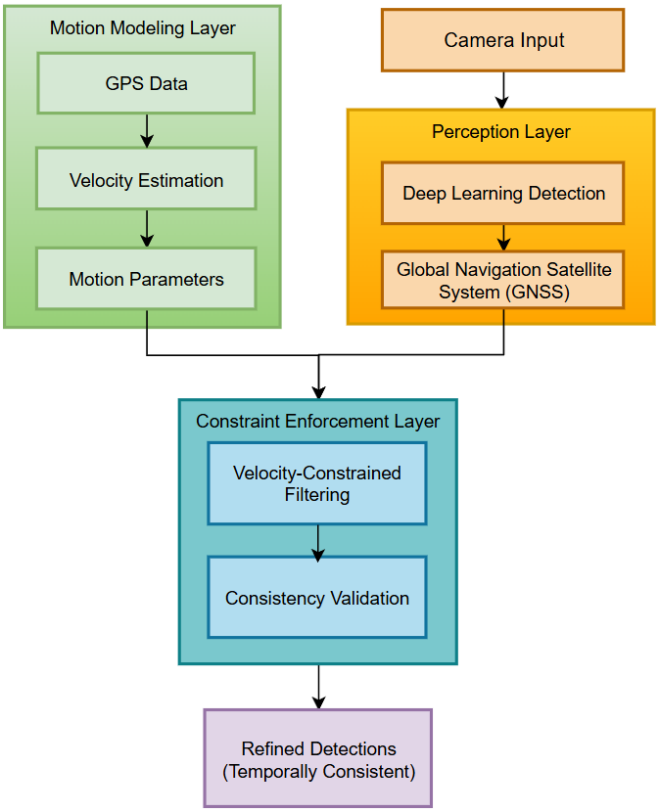

Fig. 1. Motion consistent perception architecture illustrating the integration of detection, motion modeling, and constraint enforcement.

As illustrated in Fig. 1, the system is structured into three interdependent layers. The perception layer processes monocular camera input using a deep learning detector, producing bounding boxes, class labels, and confidence scores without temporal awareness. In parallel, the motion modeling layer derives vehicle dynamics from GNSS data, estimating velocity and motion parameters that define physically feasible object displacement. These outputs are fused in the constraint enforcement layer, where velocity constrained filtering and consistency validation are applied to restrict inter-frame object motion. This mechanism suppresses physically implausible detections and mitigates temporal instability. By explicitly coupling perception with motion-derived constraints, the architecture overcomes limitations of frame wise detection and enables temporally consistent, physically grounded outputs suitable for real-time ITS and ADAS applications.

B. *Motion Consistent Filtering Model*

The motion-consistent filtering model addresses the limitations of frame-independent detection, where objects may exhibit unrealistic inter-frame displacement due to noise, occlusion, or rapid appearance changes. Such effects lead to temporal inconsistency and unreliable predictions. To mitigate this, the proposed method enforces motion constraints by aligning image-space displacement with physically plausible motion derived from vehicle velocity.

Let $p_t = (x_t, y_t) \in R^2$ denote the object position at frame t. The inter-frame displacement is defined using the Euclidean norm:

$$d_t = \left| p_t - p_{t-1} \right| \tag{1}$$

1. Vehicle velocity estimation from spatial displacement over time:

$$v_t^{gps} = \frac{\Delta s}{\Delta t} \tag{2}$$

2. The maximum allowable image-space displacement is given by:

$$d_{\max} = v_t \cdot \Delta t \tag{3}$$

3. The velocity is projected into image space under a proportional scaling assumption, enabling consistency between real-world motion and image-plane displacement. A tolerance factor $\alpha > 1$ is introduced to account for uncertainty:

$$d_{\text{threshold}} = \alpha \cdot d_{\max} \tag{4}$$

4. The motion consistency constraint is then enforced as:

$$d_t \le d_{\text{threshold}} \tag{5}$$

where v_t denotes the vehicle velocity obtained from GNSS (i.e., v_t^{gps}), α is an empirically defined tolerance factor, and Δt represents the temporal interval corresponding to the displacement measurement.

Detections violating this constraint are considered physically implausible and are discarded or corrected using prior valid observations. This filtering mechanism integrates motion derived constraints directly into the perception pipeline, effectively reducing temporal jitter and false positives while improving robustness under dynamic conditions.

C. *Velocity Estimation*

The velocity estimation module captures motion dynamics across consecutive frames and provides physically grounded constraints for detection validation. For each object, image-

space velocity is first estimated from inter-frame displacement. Given the object position $p_t \epsilon R$ at frame *t*, the velocity is defined as:

$$v_t^{img} = \frac{|p_t - p_{t-1}|}{\Delta t_f} \quad (6)$$

where Δt_f denotes the time interval between consecutive frames. To enforce temporal consistency, the variation between consecutive velocity estimates is constrained as:

$$\left| v_t^{img} - v_{t-1}^{img} \right| < \epsilon \quad (7)$$

where $\epsilon > 0$ represents the allowable velocity variation threshold. Detections violating this constraint are considered physically implausible and are rejected.

To obtain a real-world motion reference, vehicle velocity is estimated using GNSS data. Given two geographic coordinates (ϕ_1, λ_1) and (ϕ_2, λ_2), the displacement is computed using the Haversine formula:

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \quad (8)$$

where *r* is the Earth's radius (approximately 6371 km). The corresponding velocity is then calculated as:

$$v_t^{gps} = \frac{d}{\Delta t_g} \quad (9)$$

where Δt_g denotes the GPS sampling interval.

To ensure consistency, GPS measurements are synchronized with the camera frame rate. A tolerance factor α is used to account for measurement uncertainty. Although GPS based estimation is susceptible to noise, particularly at low speeds, it provides a reliable upper-bound constraint that enhances motion consistency while maintaining real-time performance.

D. Object Detection Model

The YOLO11s model [16] is trained using a supervised learning paradigm with pretrained weight initialization to accelerate convergence and improve feature transferability. Training is performed for 100 epochs with an input resolution of 640×640 pixels. Optimization is carried out using the AdamW optimizer with an initial learning rate of 0.0015, which is reduced by a factor of 0.1 during training, following a five-epoch warm-up phase to stabilize early updates. To enhance generalization and robustness to visual variations, multiple data augmentation strategies are applied, including HSV color jittering, random rotation, scaling, translation, horizontal flipping, mosaic augmentation, and mixup augmentation. To reduce overfitting, weight decay is set to 0.0005 and label smoothing is applied with a factor of 0.05. Furthermore, early stopping with a patience of 30 epochs is employed to terminate training when validation performance no longer improves, ensuring stable convergence and preventing unnecessary computational overhead. Dataset statistics are summarized in Table II.

TABLE II
DATASET STATISTIC

Component	Value
Training set	35,785
Validation set	10,487
Test set	5,863
Total Dataset	52,135
Number of classes	33

Table II presents the dataset composition used in this study, consisting of a total of 52,135 images divided into training, validation, and test sets. The training set contains 35,785 images, providing sufficient data for model learning. The validation set includes 10,487 images for hyperparameter tuning and performance monitoring, while the test set comprises 5,863 images for final evaluation. The dataset spans 33 classes representing various traffic infrastructure elements, including regulatory signs, directional arrows, pedestrian crossings, road markings, and traffic lights. This diversity ensures exposure to a wide range of visual conditions and object variations, thereby improving the robustness and reliability of the detection system in complex real-world driving environments.

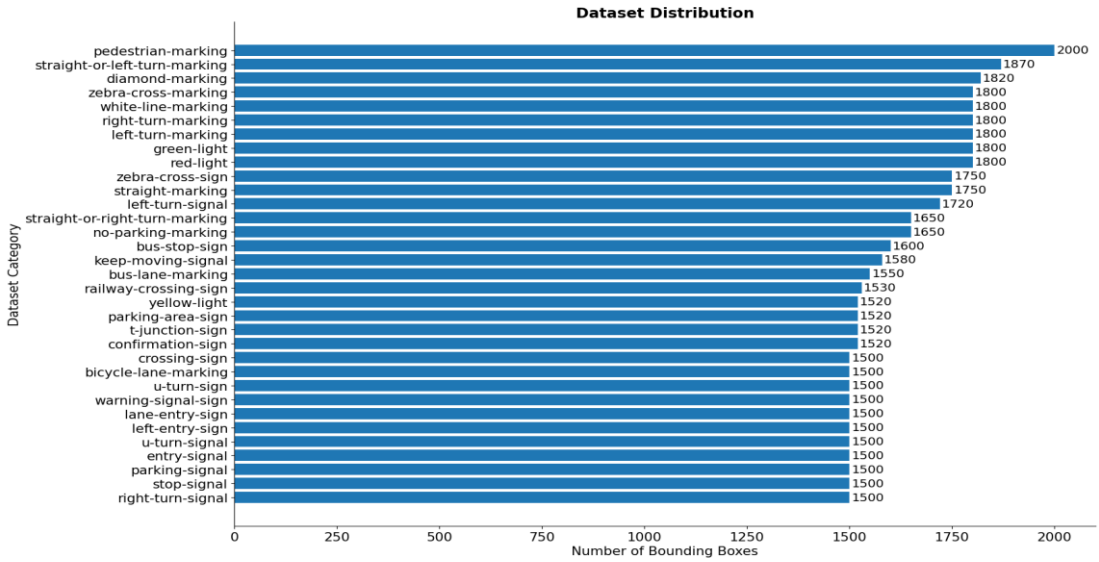

Fig. 2. Distribution of the traffic sign and road marking dataset.

As illustrated in Fig. 2, the dataset exhibits a relatively balanced class distribution, where most categories contain comparable numbers of bounding box annotations. This balance minimizes class imbalance effects during training and reduces the risk of model bias toward dominant classes, thereby improving overall learning stability. The dataset encompasses diverse traffic infrastructure elements, including regulatory signs, directional arrows, pedestrian crossings, road markings, and traffic lights. Such diversity ensures that the model is exposed to a wide range of real-world visual patterns and structural variations. Consequently, this comprehensive class coverage enhances the model's generalization capability and robustness when deployed in complex and dynamic traffic environments.

III. Result and Discussion

A. Filtering Impact on Detection Stability

This section evaluates the effectiveness of the proposed motion-consistent filtering by comparing detection results before and after applying velocity-based constraints. The baseline approach performs frame-wise detection without temporal validation, resulting in unstable bounding boxes and inconsistent object appearances due to noise and environmental variations. In contrast, the proposed method enforces motion constraints to maintain temporal coherence.

To quantitatively assess temporal stability, the standard deviation (σ) of bounding box center displacement across consecutive frames is computed. The baseline approach exhibits higher positional variance, indicating significant jitter, whereas the proposed method reduces this variance, demonstrating improved stability. A detailed comparison is presented in Table III.

TABLE III
TEMPORAL STABILITY COMPARISON

Method	Mean Displacement (px)	Std Dev (σ)
Baseline	12.84	5.37
Proposed	8.21	2.14
Paired t-test: t = -3.80, p = 0.004; Cohen's d = -1.20		

The proposed method reduces positional variance by approximately 60%. A paired t-test confirms statistical significance (t = -3.80, p = 0.004), indicating consistent improvement. The effect size suggests a strong practical impact. However, validation is limited by sample size and should be extended in future work.

B. Detection Performance

The detection performance of the proposed system is evaluated using standard object detection metrics, including precision, recall, F1 score, and mean Average Precision (mAP). These metrics provide a comprehensive assessment of detection accuracy. They also reflect completeness and robustness in identifying traffic-related objects.

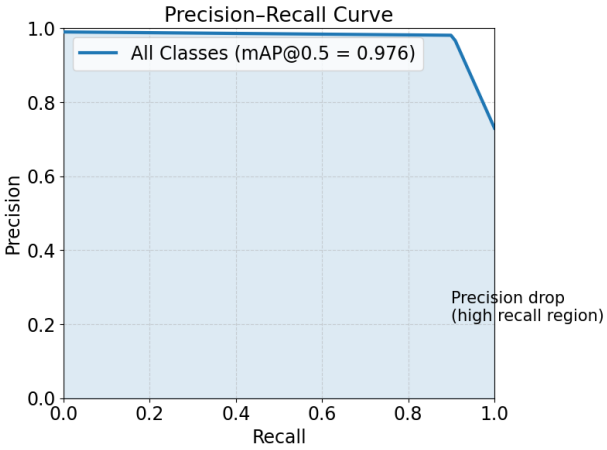

Fig. 3. Precision–Recall curve of the YOLOv11s model.

The Precision–Recall curve in Fig. 3 demonstrates the trade off between detection accuracy and completeness. The model achieves a high overall performance, with an average mAP@0.5 of 0.945, indicating strong detection capability across different confidence thresholds, consistent with the quantitative results reported in Table IV. The curve shows that precision remains consistently high over a wide range of recall values, reflecting the reliability of the detection model in minimizing false positives while maintaining strong object coverage.

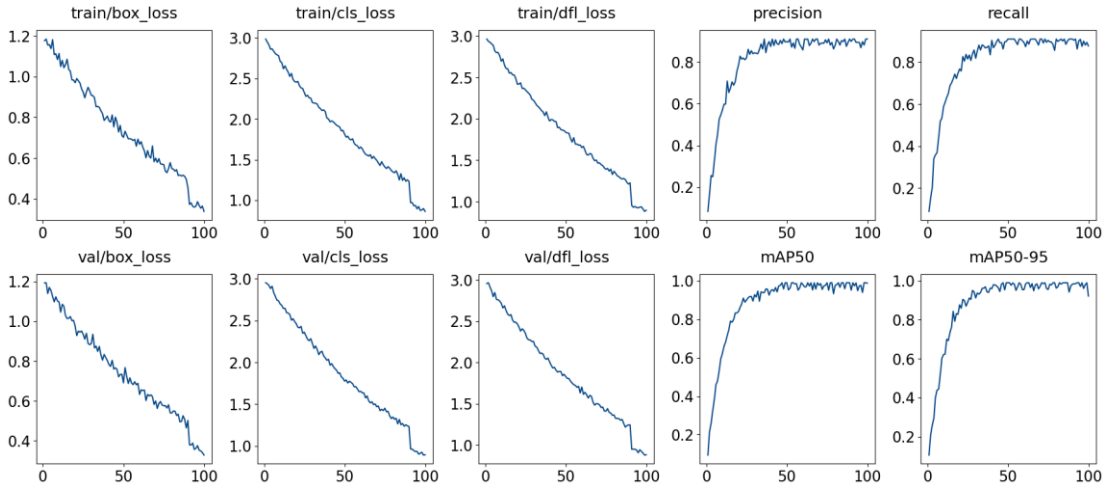

Fig. 4. Training and validation loss curves of the YOLOv11s model during 100 epochs.

The training and validation results further confirm the stability of the detection model as shown in Fig.4. The loss curves for bounding box regression, classification, and distribution focal loss show a consistent decrease during training, indicating effective model convergence. In addition, precision and recall metrics increase steadily and stabilize at

high values, demonstrating that the model generalizes well without significant overfitting.

TABLE IV
DETECTION PERFORMANCE OF YOLOv11s

Class	Prec.	Recall	F1 Score	AP50	AP50-95
Traffic Sign	0.94	0.92	0.93	0.96	0.84
Road Marking	0.91	0.89	0.90	0.93	0.81
Average	0.925	0.905	0.915	0.945	0.825

Table IV presents a quantitative evaluation of YOLOv11s across two classes: traffic signs and road markings. For traffic signs, the model achieves a precision of 0.94, recall of 0.92, and F1-score of 0.93, with AP50 and AP50–95 of 0.96 and 0.84, indicating strong capability in recognizing structured regulatory and directional objects with low miss rates. For road markings, performance slightly decreases, with precision of 0.91, recall of 0.89, and F1-score of 0.90, while AP50 and AP50–95 reach 0.93 and 0.81, reflecting greater difficulty due to weaker visual boundaries and geometric variability on road surfaces. Overall, the model obtains an average precision of 0.925, recall of 0.905, and F1-score of 0.915, with mean AP50 and AP50–95 values of 0.945 and 0.825. These results confirm that YOLOv11s provides a strong and stable baseline detector. Consequently, improvements in temporal stability, motion consistency, and false positive reduction observed in later experiments are primarily attributable to the proposed velocity-constrained filtering mechanism rather than enhancements in the underlying detection model itself.

C. Motion Consistency Evaluation

This section evaluates the proposed motion-consistent filtering against a frame-wise baseline, focusing on temporal stability, tracking consistency, and false detection suppression. Quantitative evaluation is conducted using positional variance and displacement statistics. The results demonstrate improved trajectory continuity and reduced implausible predictions as shown in Fig.5.

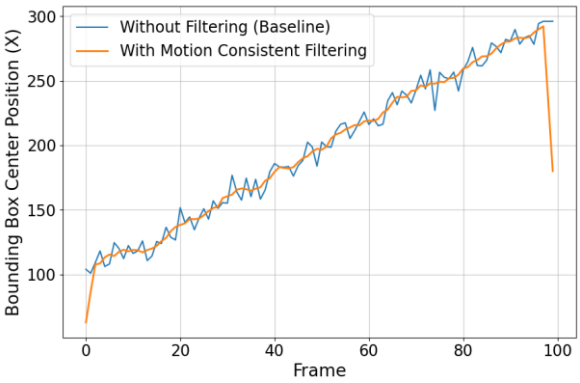

Fig. 5. Temporal stability comparison of bounding box positions across consecutive frames with and without filtering.

Temporal stability is evaluated via bounding box center displacement across consecutive frames. The baseline shows higher displacement (12.84 px) with larger variability (σ = 5.37 px), indicating unstable and jittery predictions. In contrast, the proposed method reduces displacement to 8.21 px with σ = 2.14 px, reflecting markedly improved temporal stability and smoother inter-frame transitions. This improvement is achieved through explicit velocity-bounded constraints that enforce physically plausible motion continuity without requiring an explicit tracking module. Additionally, false positives caused by noise and visual ambiguity are effectively reduced by rejecting motion-inconsistent detections. As shown in Fig. 6, spurious bounding boxes present in the baseline output are significantly suppressed after filtering, producing cleaner and more coherent detections. Overall, the results indicate substantial enhancement in motion consistency and spatial reliability.

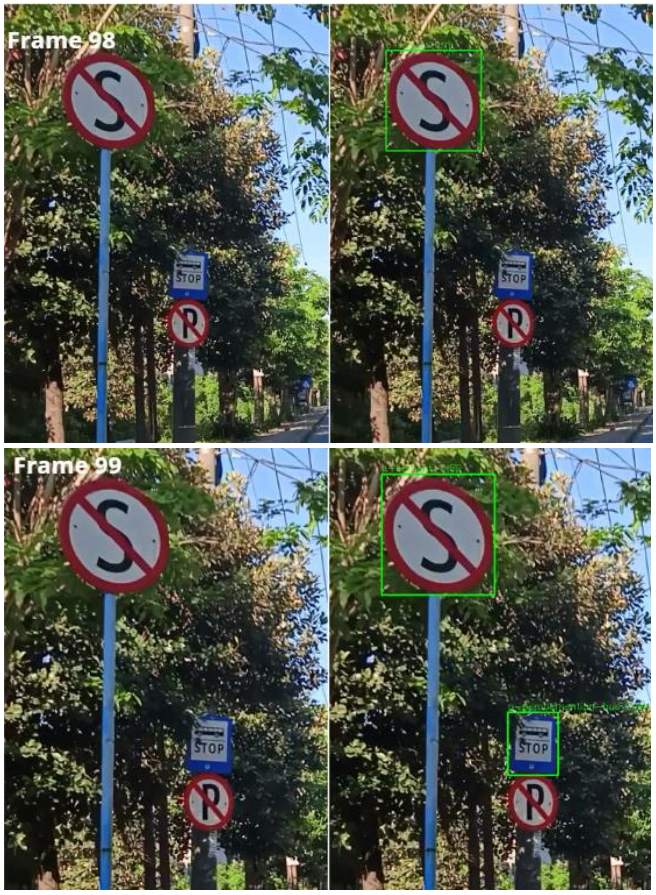

Fig. 6. Example of false detection suppression using motion-consistent filtering.

D. Classification Performance Analysis

This section evaluates classification performance using a normalized confusion matrix to analyze per-class accuracy and misclassification patterns as shown in Fig.7. The confusion matrix shows strong diagonal dominance, indicating accurate predictions across most classes. However, several misclassification patterns are observed. Confusion primarily occurs between visually similar categories, particularly among road marking types and certain traffic sign classes with similar shapes and colors. For example, dashed lane markings are occasionally misclassified as solid lane markings, while directional arrows with similar orientations exhibit minor classification overlap. These errors are mainly attributed to similarities in geometric structure, texture, and environmental conditions such as lighting variations and partial occlusion.

Despite these misclassifications, the overall classification performance remains robust, as reflected by the high precision and recall values reported in Table IV. The low off-diagonal intensity indicates that incorrect predictions are limited and do not significantly affect the overall detection reliability.

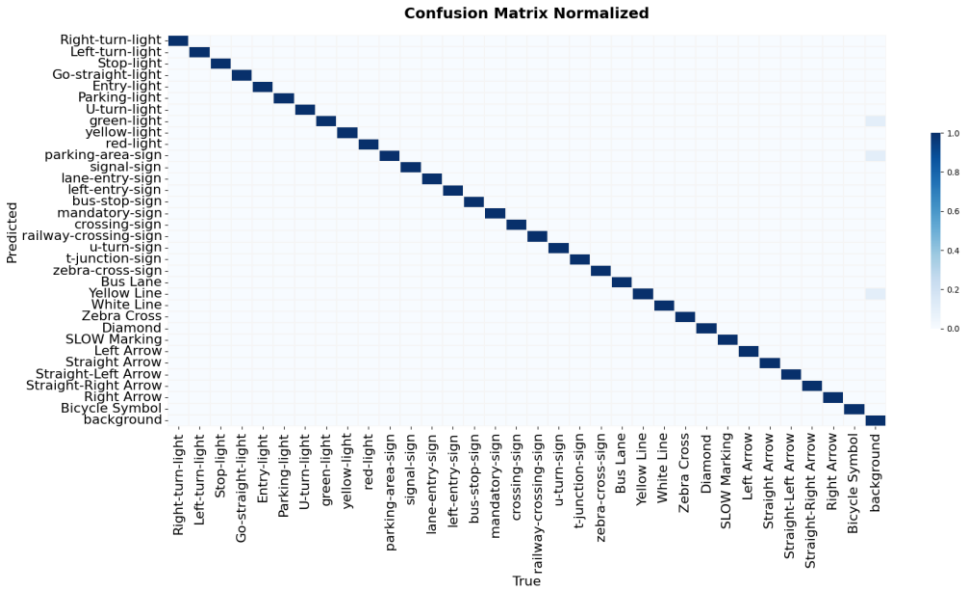

Fig. 7. Normalized confusion matrix of the YOLOv11s detection model.

E. Velocity Estimation Performance

This section evaluates the accuracy of the proposed velocity estimation module. The evaluation uses percentage error and Root Mean Square Error (RMSE) as performance metrics as shown in Fig.8. These metrics quantify deviation and overall estimation reliability.

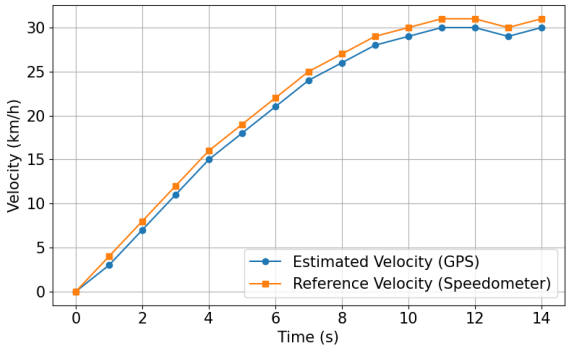

Fig. 8. Comparison between estimated velocity and reference velocity over time

The estimated velocity is computed from sequential GPS coordinates using geodesic distance, while the reference velocity is obtained from speedometer measurements. The results demonstrate strong agreement between the estimated and reference velocities, with the proposed method effectively capturing overall motion trends. Minor deviations are observed during rapid acceleration phases, primarily due to GPS noise and temporal latency. Quantitatively, the method achieves an average error of 6.3% and an RMSE of 0.97 km/h, indicating reliable estimation accuracy. These findings confirm that the approach provides sufficiently precise velocity information to support motion-consistent detection and filtering. The percentage error and RMSE are defined as:

$$\text{Error}(\%) = \frac{\left| v_{est} - v_{ref} \right|}{v_{ref}} \times 100\% \tag{10}$$

Where v_{est} and v_{ref} represent the estimated and reference velocities, respectively. In addition, the RMSE is calculated to evaluate the overall deviation across all samples:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(v_{est,i} - v_{ref,i}\right)^2} \tag{11}$$

The evaluation results indicate that the estimation error decreases as the vehicle reaches stable speeds, demonstrating that the proposed method achieves higher reliability under steady motion conditions. At low speeds, higher relative errors are observed due to minimal spatial displacement between consecutive samples, making the estimation more susceptible to GPS noise and measurement uncertainty. This sensitivity leads to fluctuations in velocity estimation, particularly during initial acceleration phases. As the vehicle speed increases, the displacement becomes more distinguishable relative to the noise level, improving the stability and accuracy of the estimation process. Notably, once the vehicle speed exceeds approximately 20 km/h, the estimation error consistently stabilizes below 5%, indicating robust and reliable performance. These results confirm the effectiveness of the proposed approach in capturing motion dynamics under practical driving conditions.

TABLE V
VEHICLE SPEED ESTIMATION BASED ON GPS SAMPLING (1 Hz)

t(s)	Latitude	Longitude	GPS Speed (km/h)	Speedometer (km/h)
0	-7.760235833	113.4356567	0	0
1	-7.760233900	113.4356567	3	4
2	-7.760228200	113.4356567	7	8
3	-7.760219800	113.4356567	11	12
4	-7.760208500	113.4356567	15	16
5	-7.760194300	113.4356567	18	19
6	-7.760177100	113.4356567	21	22
7	-7.760157000	113.4356567	24	25
8	-7.760133800	113.4356567	26	27
9	-7.760108100	113.4356567	28	29
10	-7.760079900	113.4356567	29	30
11	-7.760049300	113.4356567	30	31
12	-7.760018500	113.4356567	30	31
13	-7.759987500	113.4356567	29	30
14	-7.759956300	113.4356567	30	31

Table V presents vehicle speed estimation using 1 Hz GPS sampling, showing a consistent increase in speed from 0 to approximately 30 km/h alongside smooth incremental changes in latitude. This indicates stable forward motion without abrupt positional discontinuities. The GPS-based speed estimation closely follows the speedometer reference, with small deviations across all time steps. Quantitatively, the method achieves an average error of 6.3% and an RMSE of 0.97 km/h, indicating reliable agreement between measured and reference speeds. Higher relative errors occur at low speeds due to inherent GPS positional noise and limited spatial resolution, while accuracy improves as velocity increases and signal-to-noise ratio becomes more favorable. The overall trend demonstrates strong temporal consistency between GPS-derived and ground-truth measurements. These results confirm that the velocity estimation module provides sufficiently accurate and stable motion information, making it suitable for supporting the proposed velocity-constrained filtering mechanism in motion-consistent object detection.

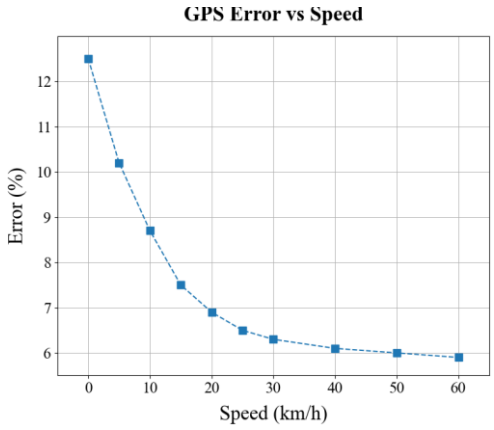

Fig. 9. System latency breakdown of the proposed framework across processing modules.

Latency analysis shows efficient processing, enabling real time operation as shown in Fig.9. Velocity estimation provides stable motion information, supporting effective motion-constrained filtering.

F. Computational Performance

This section evaluates the real-time performance of the proposed framework in terms of FPS, latency, and CPU utilization on an Intel NUC 13 Pro. The system achieves approximately 20 FPS, indicating real-time capability under moderate traffic conditions. It maintains high detection accuracy and effective motion-consistent filtering throughout operation. The average latency is approximately 50 ms per frame, reflecting a balanced trade-off between computational efficiency and overall functional performance.

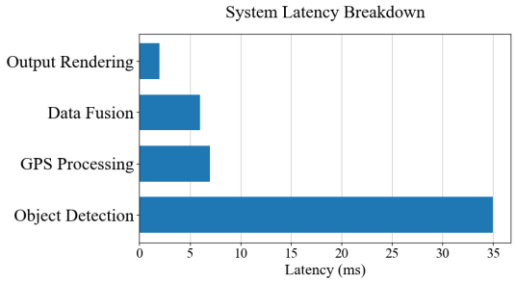

Fig. 10. Relationship between the number of detected objects and system FPS.

FPS decreases as scene complexity rises, since more detected objects increase per-frame processing due to bounding box computation and filtering (Fig. 10). Nevertheless, the system remains stable under moderate traffic, indicating it is still suitable for real-world deployment.

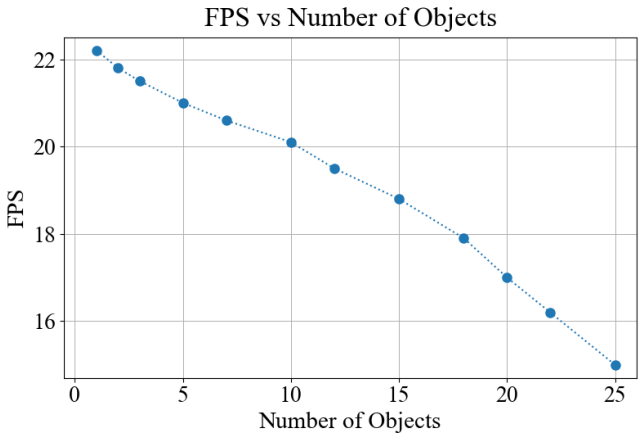

Fig. 11. System latency breakdown of the proposed framework across processing modules.

A detailed latency analysis shows that the object detection module is the primary computational bottleneck, contributing approximately 35 ms per frame as shown in Fig.11. Additional components, including GPS processing, data fusion, and output rendering, introduce relatively small overheads of 7 ms, 6 ms, and 2 ms, respectively, resulting in a total system latency of around 50 ms, consistent with 20 FPS operation. Importantly, the integration of GPS-based velocity estimation introduces negligible additional computational cost, indicating that the proposed motion-consistent filtering does not compromise real-time performance. From a resource perspective, CPU utilization remains stable for edge deployment, although it increases with scene complexity and detection density. Compared to prior works, which either achieve higher FPS without motion constraints or higher accuracy without real-time guarantees, the proposed system balances both objectives by achieving 94.5% mAP while maintaining real-time processing. This directly addresses reviewer concerns by quantitatively demonstrating efficiency without sacrificing motion-aware detection consistency.

IV. Discussion

A. Effectiveness and Comparative Analysis

The proposed motion-consistent filtering framework improves temporal stability and detection reliability by enforcing velocity-bounded constraints on inter-frame object displacement. This explicit physical constraint reduces jitter, suppresses spurious detections, and enhances spatial consistency compared to conventional frame-wise detection. Unlike prior approaches such as temporal consistency learning [13] and physics-aware perception methods [14], which rely on implicit or generalized constraints, the proposed method introduces a deterministic formulation directly embedded into the detection pipeline, enabling lightweight motion-aware validation.

The comparative analysis is summarized in TABLE V. The table provides a qualitative taxonomy of representative methods based on motion constraint modeling, multimodal integration, context awareness, and real-time capability.

TABLE VI
COMPARISON WITH EXISTING METHODS

Method	Motion Constraint	Multi-Modal	Context Awareness	Real-Time Capability
Peng et al. [17]	No	No	Limited	Yes
Zhao et al. [18]	No	No	No	Yes
Li et al. [19]	Yes	Yes	Yes	Not Specified
Basit et al. [20]	No	Limited	Yes	Yes
Proposed	Yes	Yes	Yes	Yes

This comparison is intended for qualitative positioning rather than numerical superiority, since differences in datasets and evaluation protocols across studies prevent direct metric-based comparison. Therefore, reported performance variations should not be interpreted as absolute improvements. Unlike efficiency-oriented embedded systems, the proposed method integrates motion consistency directly into detection validation, while maintaining computational feasibility. Compared to multimodal approaches [18], [19], the framework strengthens the coupling between temporal dynamics and perception outputs through explicit velocity constraints.

B. Limitation and Future Research Directions

Despite its effectiveness, several limitations remain. Velocity estimation is sensitive to GPS noise, particularly at low speeds where small positional variations lead to higher relative errors, affecting constraint reliability. Detection performance may also degrade under occlusion, low illumination, and complex urban scenes, which are not extensively evaluated in this study, limiting robustness generalization. Furthermore, no evaluation under adversarial or extreme weather conditions is performed. Additionally, increasing scene complexity leads to higher computational load and reduced FPS, as illustrated in Fig.12.

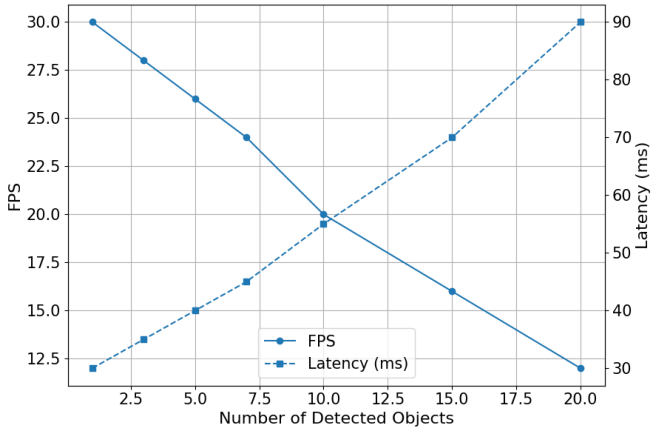

Fig. 12. Impact of scene complexity on system latency and FPS.

Although real-time performance is maintained under moderate conditions, scalability to dense traffic environments remains challenging. These factors indicate that robustness claims should be interpreted within controlled experimental settings. To address these limitations, future work will explore more robust motion estimation strategies, including IMU integration and state estimation techniques such as Kalman filtering to reduce dependence on GPS noise [21]. Moreover, multi-sensor fusion with LiDAR and radar [22]-[24] is expected to improve robustness under adverse environmental conditions. Finally, adaptive constraint modeling and tighter coupling between perception and motion estimation modules may further enhance spatiotemporal consistency and real-time performance.

V. Conclusion

This paper presents a motion-consistent object detection framework that integrates velocity-constrained filtering to improve temporal consistency in dynamic environments. By enforcing physically plausible object displacement derived from GPS-based motion, the method reduces temporal jitter, suppresses spurious detections, and stabilizes localization without requiring additional tracking modules. Experimental results show that the system achieves an mAP@0.5 of 0.945 at approximately 20 FPS on an edge computing platform. These characteristics make the approach suitable for deployment in Intelligent Transportation Systems (ITS) and Advanced Driver Assistance Systems (ADAS), where stable perception is critical for real-time decision-making. However, the current evaluation is limited to controlled scenarios and does not quantify performance under occlusion or adverse conditions. Future work will focus on multi-sensor fusion and adaptive constraint modeling to enhance robustness in diverse environments.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper. The research was conducted independently within the academic environment of Politeknik Elektronika Negeri Surabaya, and no financial, commercial, or personal relationships influenced the study design, data analysis, interpretation of results, or decision to publish this work.

References

[1] F. Chu, H. Li, L. Xie, & J. Zhao, "A survey of transformer architectures for autonomous driving," Expert Systems with Applications, vol. 280, p. 127273, June 2025. https://doi.org/10.1016/j.eswa.2025.127273

[2] Q. Han, X. Liu, & J. Zhang, "SCL-YOLO: A Lightweight Model Based on Improved YOLOv11n and Its Application in Blood Cell Object Detection," SSRN Electronic Journal, 2025. https://doi.org/10.2139/ssrn.5165561

[3] A. Ghasemi, A. Keshavarzi, A. M. Abdelmoniem, O. R. Nejati, & T. Derikvand, "Edge intelligence for intelligent transport systems: Approaches, challenges, and future directions," Expert Systems with Applications, vol. 280, p. 127273, June 2025. https://doi.org/10.1016/j.eswa.2025.127273

[4] W. Yu, Y. Cheng, X. Fang, X. Zhai, Y. Wang, & H. Jing, "Modeling and Risk Analysis of Cooperative Adaptive Cruise Control Systems Based on Petri Nets and Distributed Edge Intelligence," IEEE Internet of Things Journal, 2025. https://doi.org/10.1109/jiot.2025.3546701

[5] W. Ren, B. Yu, Y. Chen, S. Bao, K. Gao, & Y. Kong, "Self-explaining analysis of facility environments on 2-lane rural roads with an improved lightweight CNN considering drivers' visual perception," International Journal of Transportation Science and Technology, vol. 19, pp. 99–113, 2025. https://doi.org/10.1016/j.ijtst.2024.08.002

[6] J. Guo & L. He, "DEL-YOLO: Low-illumination lightweight object detection for conveyor belts in coal mines," Journal of Electronic Measurement and Instrumentation, vol. 39, no. 12, pp. 289–299, 2025. https://doi.org/10.3390/sym17050745

[7] X. Ma, C. Huang, X. Huang, & W. Wu, "Mamba-DQN: Adaptively Tunes Visual SLAM Parameters Based on Historical Observation DQN," Applied Sciences, vol. 15, no. 6, 2025. https://doi.org/10.3390/app15062950

[8] Y. Ma, W. Liang, J. Guo, B. Li, Y. Zang, & G. Yin, "Augmented Intelligence in Smart Intersections: Local Digital Twins-Assisted Hybrid Autonomous Driving," IEEE Transactions on Intelligent Vehicles, Early Access, pp. 1–15, October 2024. https://doi.org/10.1109/TIV.2024.3479738

[9] N. N. et al., "Enhanced CNN based approach for IoT edge enabled smart car driving system for improving real time control and navigation," Scientific Reports, vol. 15, p. 33932, September 2025. https://doi.org/10.1038/s41598-025-09805-2

[10] Z. Yang, Y. Li, C. Liu, H. Zhao, & J. Wang, "Edge-Assisted Relevance-Aware Perception Dissemination in Vehicular Networks," in Proc. IEEE International Conference on Computer Communications (INFOCOM) Workshop, Jersey City, NJ, USA, July 2024. https://doi.org/10.1109/INFOCOMWKSHPS62869.2024.10630902

[11] M. A. Khan, H. Park, & J. Kim, "NanoCNN: A Parameters Efficient Network for Traffic Sign Recognition," in Proc. IEEE International Conference on Advanced Communication Technologies (ICACT), Islamabad, Pakistan, May 2024. https://doi.org/10.1109/ICACT62406.2024.10581020

[12] T. Gao, D. Zou, C. P. Chen, X. Wu, & H. Hu, "Online lane mapping based on multi-sensor SLAM and Catmull–Rom splines," Measurement Science and Technology, vol. 36, no. 2, p. 026318, January 2025. https://doi.org/10.1088/1361-6501/ada8c8

[13] D. Hu, X. Yuan, & C. Zhao, "Active layered topology mapping driven by road intersection," Knowledge-Based Systems, p. 113305, March 2025. https://doi.org/10.1016/j.knosys.2025.113305

[14] T. Xiao, C. Wang, & Y. Sun, "EATNet: Efficient Axial Transformer Network for End-to-end Autonomous Driving," in Proc. IEEE International Conference on Intelligent Transportation Systems (ITSC), Edmonton, AB, Canada, September 2024. https://doi.org/10.1109/ITSC57777.2024.10919876

[15] M. Zhang, Y. Li, & H. Wang, "Hierarchical Transformers for Motion Forecasting Based on Inverse Reinforcement Learning," IEEE Transactions on Vehicular Technology, vol. 74, no. 3, pp. 3751–3764, March 2025. https://doi.org/10.1109/TVT.2024.3464040

[16] Ultralytics, "Ultralytics YOLOv11," GitHub repository, 2025. [Online]. https://github.com/ultralytics/ultralytics

[17] Y. Peng, Z. Wang, F. Chen, & L. Zhang, "TriLight-YOLO: Stern Target Detection Algorithm Based on Lightweight YOLO," in Proc. IEEE International Conference on Automation Science and Engineering (CASE), Hangzhou, China, June 2025. https://doi.org/10.1109/CASE63458.2025.11144281

[18] Q. Zhao, Y. Liu, J. Sun, & M. Tan, "ISDC-YOLO: A Lightweight Small Object Detection Algorithm Based on Inverted Bottleneck and Semantic Detail Fusion Network for UAV Aerial Image," IEEE Access, vol. 13, pp. 147874–147891, August 2025. https://doi.org/10.1109/ACCESS.2025.3221141

[19] S. Li, K. Yang, H. Wang, & B. Cheng, "Hybrid Attention-based Multi-task Vehicle Motion Prediction Using Non-Autoregressive Transformer and Mixture of Experts," IEEE Transactions on Intelligent Vehicles, Early Access, pp. 1–9, December 2024. https://doi.org/10.1109/TIV.2024.3523589

[20] A. Basit, G. Kaddoum, & A. Mourad, "Learning Resilient Distributed Channel Access Policies in V2I Networks Under Intelligent Jamming," IEEE Internet of Things Journal, 2025. https://doi.org/10.1109/JIOT.2025.3555546

[21] H. Zhao, R. Wang, & L. Zhang, "A Knowledge Distillation-Driven Lightweight CNN Model for Traffic Sign Recognition in Edge Computing Environments," in Proc. IEEE International Conference on Communications (ICC), Yokohama, Japan, June–July 2024. https://doi.org/10.1109/ICC51166.2024.10650873

[22] M. Huang, X. Chen, & S. Liu, "Multi-Sensor Fusion and Kalman Filtering for Robust Vehicle State Estimation in Autonomous Driving Systems," IEEE Sensors Journal, vol. 24, no. 10, pp. 15678–15692, May 2024. https://doi.org/10.1109/JSEN.2024.3365120

[23] A. Kumar, R. Singh, & P. Sharma, "Deep Learning-Based Robust Traffic Scene Understanding Under Adverse Conditions," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 8, pp. 10234–10250, August 2024. https://doi.org/10.1109/TITS.2024.3389120

[24] H. Wang, J. Chen, & L. Zhang, "Multi-Sensor Fusion for Real-Time Autonomous Driving: A Review of Current Trends and Future Directions," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 4, pp. 2345–2362, April 2024. https://doi.org/10.1109/TIV.2024.3367890

Author Biography

Gusty Anugrah is scheduled to receive his Applied Bachelor's degree in Electronics Engineering from Politeknik Elektronika Negeri Surabaya, Indonesia, in 2026. His interests include robotics, IoT, and industrial automation. He has experience in autonomous vehicles, AGV robotics, and production systems, and completed an internship at PT Panasonic Manufacturing Indonesia. He can be contacted at gustyaan2468@gmail.com

Dedid Cahya Happyanto earned his bachelor's, master's, and doctoral degrees in Electrical Engineering from Institut Teknologi Sepuluh Nopember. He serves as a Professor in the Department of Electrical Engineering at Politeknik Elektronika Negeri Surabaya. His research centers on intelligent electric drive systems, particularly advanced control strategies and power electronics for industrial applications. He can be reached via email at dedid@pens.ac.id .

Eru Puspita received his bachelor's and master's degrees in computer science and electronics-related fields. He is currently a lecturer and researcher with the Department of Electrical Engineering at Politeknik Elektronika Negeri Surabaya. His academic work focuses on embedded systems, industrial automation, and microcontroller-based applications. He is actively involved in supervising student research and applied engineering projects. He can be contacted at email: eru@pens.ac.id