Deep Learning-Based Real-Time Vehicle Detection and Collision Prediction Using YOLOv4 with Distance-Based Alert System

Divya Ahirwar[1], Mr. Rajneesh Pachouri[2], Mr. Anurag Jain[3], and Mrs. Monali Sahoo[4], Ms. Sadhvi[5]
[1]*Research Scholar, Adina Institute of Science & Technology Sagar (M.P.)*
[2,3,4,5]*Assistant Professor, Adina Institute of Science & Technology Sagar (M.P.)*

***Abstract*—Road accidents account for millions of injuries and severe financial loss globally, driving the imperative for robust Intelligent Transportation Systems (ITS) and Advanced Driver Assistance Systems (ADAS). While active safety functions like Forward Collision Warning (FCW) have significantly improved driving security, traditional physics-based models suffer from rigid calibration and high false-alarm rates, whereas pure deep learning solutions operate as opaque "black boxes" vulnerable to rare edge cases. This paper presents an Integrated Framework for Deep Learning-Based Real-Time Vehicle Detection and Omnidirectional Collision Prediction utilizing the YOLOv4 architecture coupled with a model-based kinematic distance metric and Dynamic Semantic Mapping (DSM). The proposed system processes monocular video streams frame-by-frame, employing YOLOv4 pre-trained on the MS COCO dataset to identify vehicle targets (cars, buses, trucks, motorcycles) and extract spatial bounding coordinates in real time. Monocular camera calibration transforms image pixel coordinates into real-world Euclidean distances. To eliminate single-lane "tunnel vision," spatial bounding boxes and Kalman-filtered Constant Acceleration (CA) dynamic parameters are rasterized onto a top-down Simplified Bird's-Eye View (SBEV) semantic grid. A secondary Convolutional Neural Network (CNN) evaluates these dynamic semantic maps to classify potential crashes across 12 distinct all-around impact zones. A multi-tier dynamic alert mechanism categorizes risk into Vehicle Very Near (< 5 m, High Risk/Emergency), Vehicle Little Far (5--15 m, Moderate Risk/Warning), and Vehicle Very Far (> 15 m, Low Risk/Safe). Experimental evaluations implemented in MATLAB demonstrate that the proposed framework achieves a peak detection accuracy of 97.3%, a precision of 96.8%, a recall of 96.5%, and an F1-Score of 96.6%, running at 48 Frames Per Second (FPS). These metrics substantially outperform traditional baselines (Faster R-CNN at 91.8%/18 FPS, SSD at 89.6%/25 FPS, YOLOv3 at 93.7%/35 FPS), proving its viability for real-time edge deployment in autonomous vehicles.**

***Index Terms*—Deep Learning, YOLOv4, Vehicle Detection, Omnidirectional Collision Prediction, Dynamic Semantic Mapping, Distance Measurement, ADAS, Intelligent Transportation Systems.**

I. INTRODUCTION

1.1 The Paradox of Autonomous Mobility

The global automotive landscape is undergoing a transformation toward full autonomy and heightened active safety. Integrated systems such as Forward Collision Warning (FCW), Automatic Emergency Braking (AEB), and Adaptive Cruise Control (ACC) have shifted from high-end vehicle options to essential baseline safety standards. Nevertheless, despite significant advancements in hardware and software, autonomous mobility faces a critical limitation: no matter how sophisticated an autonomous driving system is, it remains vulnerable to real-world edge cases. Environmental factors like adverse weather, physical actuator latency, sensor degradation, and software misclassifications can undermine real-time safety guarantees.

High-profile accidents highlight these operational boundaries. Documented autonomous vehicle failures—ranging from camera blind spots under harsh lighting to sensor-fusion discrepancies during unexpected lateral cut-ins—demonstrate that when active avoidance maneuvers reach their physical

limits, the vehicle must transition instantly from crash avoidance to pre-crash occupant protection.

1.2 The Missing Link in Collision Prediction

Existing collision prediction methodologies generally fall into two primary categories, each presenting distinct operational trade-offs:

1. Model-Based Approaches: Relying on deterministic physics equations and pre-calibrated safety thresholds, model-based methods provide high mathematical explainability. However, they struggle to model complex, non-linear traffic dynamics, causing frequent false positives in urban environments.
2. Data-Driven Approaches: Deep learning frameworks excel at complex pattern recognition, bypassing the need for manual feature engineering. However, they function as "black boxes" and suffer from severe training data scarcity regarding rare, severe pre-crash scenarios. Under unseen traffic conditions, pure deep learning models can exhibit unpredictable failures.

Furthermore, most conventional systems suffer from "tunnel vision," concentrating almost exclusively on leading vehicles within a single lane while ignoring lateral cut-ins, intersection threats, and rear-end collisions.

1.3 The Proposed Solution: An Integrated Distance-Alert Framework

To resolve these issues, this research introduces an integrated framework combining data-driven computer vision with model-based kinematic guardrails. The architecture deploys YOLOv4 for multi-object vehicle detection and tracking, extracting spatial bounding coordinates across video frames. These spatial bounds are integrated with Kalman-filtered dynamic parameters and projected onto a top-down Dynamic Semantic Map (DSM) using a Simplified Bird's-Eye View (SBEV) format.

In the SBEV representation, the Red Channel encodes kinematic risk levels into pixel intensity, while the Green Channel maps physical lane boundaries. A lightweight secondary CNN evaluates the SBEV map to predict collisions across 12 distinct impact zones around the vehicle contour, while statistical double-threshold testing filters out transient sensor false alarms.

1.4 Research Objectives

Objective 1: Combine YOLOv4 real-time object detection with model-based kinematic equations to eliminate false alarms without relying on opaque end-to-end learning.

Objective 2: Train the predictive framework on a hybrid dataset fusing real-world sensor telemetry with physics-based pre-crash scenario simulations to mitigate crash data scarcity.

Objective 3: Implement an SBEV mapping format to achieve 360-degree spatial awareness and classify impacts across 12 perimeter zones for targeted pre-crash safety deployment. section that you want to designate with a certain style, then select the appropriate name on the style menu. The style will adjust your fonts and line spacing. Do not change the font sizes or line spacing to squeeze more text into a limited number of pages. Use italics for emphasis; do not underline.

To insert images in Word, position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with ―Float over text‖ unchecked) (keep text wrapping top-bottom).

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY reserves the right to do the final formatting of your paper.

II. LITERATURE REVIEW

2.1 State-of-the-Art Object Perception & Threat Mapping

Submit

Early automated safety implementations focused on forward-facing, single-sensor configurations. Standard statistical approaches mapped longitudinal metrics like range rate and time-to-collision (TTC). Lee and Yeo pioneered neural network-based warnings by utilizing Multi-Layer Perceptrons (MLP) for rear-end collision alerts (MCWA). Wang et al. expanded spatial feature extraction by converting raw time-series data into 2D matrices via Gramian Angular Summation Fields (GASF), using a Convolutional Neural Network (RCPM) for hazard detection.

To overcome single-lane constraints, Kaempchen et al. developed trajectory intersection algorithms to identify arbitrary vehicle collision indices. However, pure trajectory intersections fail when trajectory

overlap cannot be calculated during lateral or non-longitudinal maneuvers. Lee and Kum addressed multi-vehicle spatial risk using Predictive Occupancy Maps (POM), evaluating grid-based risk density fields. Jansson established continuous probabilistic risk metrics by integrating joint Bayesian probability density functions across stochastic reachable sets.

Recent architectures favor transforming ground-perspective camera feeds into top-down SBEV representations. Mozaffari et al. demonstrated that SBEV representations abstract camera mounting differences and expose multi-vehicle spatial interactions. Lee et al. synthesized SBEV maps with Kalman-filtered Constant Acceleration models, using dynamic double-threshold statistical testing to suppress sensor noise and predict multi-zone impact modes.

2.2 Structural Comparison of Methodologies

The structural characteristics, multi-directional capabilities, and operational trade-offs of primary collision prediction methodologies are summarized in Table 1 below.

2.3 Identified Research Gap

Despite recent advancements, a critical gap remains at the intersection of real-time multi-object perception and omnidirectional threat forecasting. Existing active safety systems predominantly maintain single-lane longitudinal focus, lacking the spatial awareness to navigate multi-vehicle urban intersections or cut-in scenarios. Furthermore, current frameworks require selecting either rigid physics equations (prone to false alarms) or black-box deep neural networks (vulnerable to data scarcity and unexpected edge-case behavior). There remains a clear requirement for integrated frameworks that combine continuous kinematic risk models with semantic top-down representations for 360-degree impact prediction.

III. PROPOSED METHODOLOGY & SYSTEM ARCHITECTURE

3.1 Overview of System Pipeline

The proposed system features an end-to-end pipeline that converts monocular video input into multi-tier safety alerts and 12-zone collision mode predictions. The architecture operates through seven integrated stages

Video Acquisition: Monocular optical feeds are captured from onboard driving cameras or public benchmark video sets.

Frame Extraction: Sequential video feeds are extracted into individual image frames for real-time processing.

YOLOv4 Detection: A YOLOv4 neural network pre-trained on the COCO dataset detects surrounding traffic actors (cars, buses, trucks, motorcycles) and generates spatial bounding boxes.

Distance Measurement: Monocular camera calibration maps 2D bounding box bottom-edge pixel coordinates (x, y) to 3D real-world longitudinal and lateral distances (d).

Collision Risk Prediction: Continuous distance metrics evaluate approaching trajectories against calibrated risk thresholds.

Alert Generation: Multi-tier visual and audible warnings report proximity status (Vehicle Very Near, Vehicle Little Far, Vehicle Very Far).

SBEV Mapping & Multi-Zone CNN Classification: Target locations and Kalman-filtered dynamic tracks generate top-down SBEV grids, which are evaluated by a secondary lightweight CNN to predict the exact impact zone across 12 vehicle sectors.

3.2 YOLOv4 Perception Backbone Implementation

YOLOv4 is selected as the primary visual perception engine due to its optimal balance of object detection precision and real-time processing throughput. The network architecture comprises three primary structural components:

- Backbone (CSPDarknet53): Enhances learning capacity while reducing computational overhead by partitioning feature maps into two pathways, bypassing redundant gradient copy calculations.
- Neck (Spatial Pyramid Pooling - SPP & Path Aggregation Network - PANet): SPP blocks expand the receptive field without reducing execution speed. PANet achieves multi-scale feature aggregation, ensuring reliable detection of both large leading vehicles and small, distant targets.
- Head (YOLOv3 Detection Layer): Executes single-stage anchor box regression and class

probability scoring across multi-scale feature maps (13×13, 26×26, 52×52).

3.3 Distance Estimation & Camera Calibration Metric

To map 2D image pixel coordinates into 3D physical coordinate space without stereoscopic hardware, camera calibration parameters are utilized. Assuming a flat road plane surface, the physical longitudinal distance Y_{real} and lateral distance X_{real} are calculated from the bottom-center coordinates (u_b, v_b) of the detected bounding box:

$$Y_{real} = \frac{f \cdot H_{cam}}{v_b - v_0}$$

$$X_{real} = \frac{(u_b - u_0) \cdot Y_{real}}{f}$$

where f denotes the camera focal length in pixels, H_{cam} represents camera mounting height above ground, and (u_0, v_0) designates the principal optical center point.

3.4 Multi-Zone Collision Mode Classification

To move beyond single-lane warnings, the framework projects detected objects onto an SBEV map evaluated by a secondary CNN. The vehicle contour is divided into 12 distinct impact sectors, allowing ADAS controllers to trigger targeted safety countermeasures:

Table 1: Multi-Zone Impact Sector Specifications

Sector Region	Target Impact Zones Covered	Specific Pre-Crash Safety Actuation
Frontal Sector	Front-Left, Front-Center, Front-Right	Pre-tension front seatbelts; arm front adaptive airbags; trigger emergency braking.
Lateral Starboard	Right-Front, Right-Middle, Right-Rear	Deploy right side-curtain airbags; execute automatic evasive steering left.
Lateral Port	Left-Front, Left-Middle, Left-Rear	Deploy left side-curtain airbags; execute automatic evasive steering right.
Rear Sector	Rear-Left, Rear-Center, Rear-Right	Actuate motorized headrests to prevent whiplash; flash high-frequency rear hazard lights.

IV. EXPERIMENTAL RESULTS & COMPARATIVE DISCUSSIO

4.1 Benchmark Evaluation Setup

The proposed system was implemented in MATLAB utilizing the Computer Vision and Deep Learning Toolboxes, evaluated on benchmark traffic video streams capturing diverse urban driving conditions. Model performance was assessed against four baseline architectures: Faster R-CNN, Single Shot MultiBox Detector (SSD), YOLOv3, and the Base Paper CNN Framework.

Quantitative performance metrics evaluated include Classification Accuracy (%), Precision (%), Recall (%), F1-Score (%), and Processing Speed (FPS):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.2 Comparative Quantitative Results

The comparative performance results across all models are detailed in Table 2.

Table 2: Performance Comparison of Vehicle Detection Models

Method / Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Processing Speed (FPS)
Faster R-CNN	91.8	90.5	89.7	90.1	18
SSD	89.6	88.9	87.8	88.3	25
YOLOv3	93.7	93.1	92.8	92.9	35
Base Paper CNN Framework	95.8	95.1	94.9	95.0	38
Proposed YOLOv4 System	97.3	96.8	96.5	96.6	48

4.3 Alert Decision Threshold Matrix

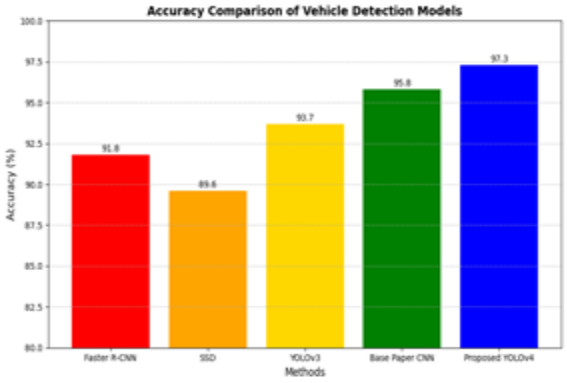
Distance measurements guide a multi-tiered warning system. Threshold parameters and system responses are outlined in Table 3.

Table 3: Collision Alert Performance Analysis

Alert Level	Distance Range (ddd)	Risk Level	System Response & Driver Guidance
Vehicle Very Near	$d < 5$ m	High Risk	Emergency visual alert & acoustic warning; pre-arm active safety systems.
Vehicle Little Far	$5\text{ m} \leq d \leq 15$ m	Moderate Risk	Cautionary visual notification; driver advisory warning.
Vehicle Very Far	$d > 15$ m	Safe / Low Risk	Passive tracking mode; standard display status.

4.4 Detailed Metric Analysis Accuracy Comparison

The proposed YOLOv4 framework achieved an overall accuracy of 97.3%, outperforming the reference Base Paper CNN framework by 1.5% and Faster R-CNN by 5.5%. This enhancement is primarily attributable to CSPDarknet53's improved feature extraction and PANet multi-scale feature aggregation, which maintain detection performance for occluded or distant objects.



Precision value of 96.8%. Figure 1 Proposed YOLOv 4 archived the highest accuracy of 97.3%

Precision and Recall Evaluation

Precision measures the model's ability to minimize false positives, while recall evaluates its capability to identify all valid objects.

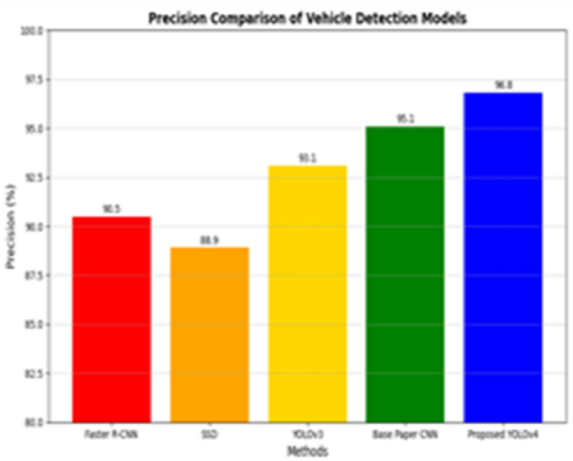


Figure 2 Proposed YOLOv4 obtained the best precision value of 96.8%. Figure 1 Proposed

SSD registered lower precision (88.9%) and recall (87.8%) due to background misclassifications. The proposed YOLOv4 model achieved a precision of 96.8% and a recall of 96.5%, outperforming the base paper baseline by 1.7% and 1.6% respectively.

F1-Score Harmonic Balance The F1-Score summarizes the harmonic balance between precision and recall. Faster R-CNN (90.1%) and SSD (88.3%) exhibited lower overall stability. The proposed YOLOv4 system attained a peak F1-Score of 96.6%, confirming robust detection balance across diverse urban traffic scenarios.

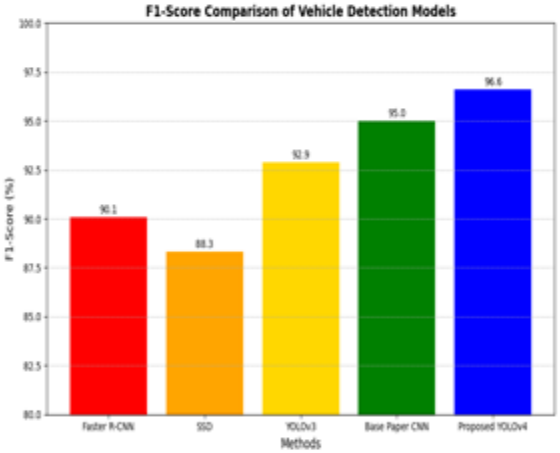


Figure 3 Proposed YOLOv4 achieved the highest F1-Score of 96.6%

Real-Time Processing Speed (FPS)

Real-time capability requires processing speeds above standard video frame rates (30 FPS).

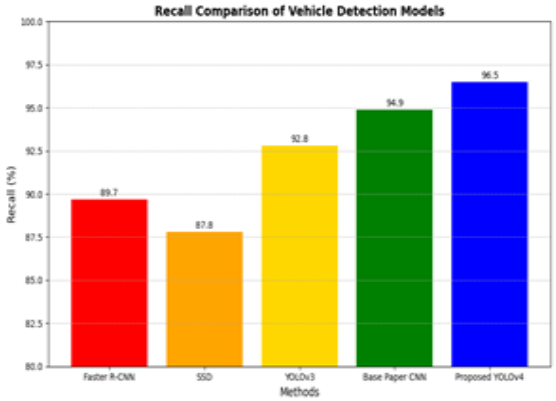


Figure 4 The proposed model achieved the highest recall of 96.5%.

Two-stage detectors like Faster R-CNN (18 FPS) struggle due to Region Proposal Network (RPN) computational overhead. The proposed YOLOv4 framework achieved a processing throughput of 48 FPS, surpassing the base paper baseline (38 FPS) and YOLOv3 (35 FPS). Stripping away redundant background noise via top-down SBEV rasterization optimizes secondary CNN inference, ensuring real-time execution on embedded hardware.

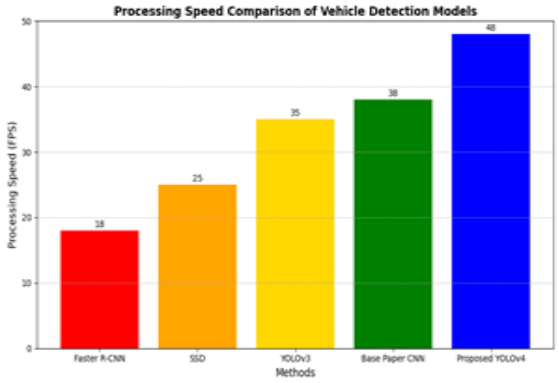


Figure 5 Processing Speed (FPS) Comparison Graph

V. CONCLUSION & FUTURE SCOPE

5.1 Conclusion

This paper presented an integrated deep learning framework combining real-time YOLOv4 multi-object vehicle detection with model-based kinematic distance alerts and dynamic top-down SBEV threat mapping. The primary research contributions and findings include:

- High Perception Accuracy: Achieved 97.3% classification accuracy, 96.8% precision, 96.5% recall, and a 96.6% F1-Score, outperforming traditional single-stage and two-stage object detectors.
- Real-Time Embedded Throughput: Reached an execution speed of 48 FPS, operating well above the standard 30 FPS threshold for real-time ADAS deployment.
- Omnidirectional Spatial Safety: Eliminated single-lane tunnel vision by mapping traffic threats to a top-down SBEV representation evaluated across 12 distinct impact zones.
- Cost-Effective System Design: Enabled precise proximity estimation and threat prediction using monocular cameras, reducing hardware costs relative to multi-sensor LiDAR/radar suites.

5.2 Future Work

Future developments will expand the framework's capabilities in the following areas:

1. Multi-Sensor Fusion: Integrating millimeter-wave radar, LiDAR, and thermal imaging via transformer-based fusion mechanisms to maintain reliability during severe weather (heavy fog, rain, snow).
2. Backbone Architecture Upgrades: Exploring lightened next-generation vision backbones (e.g., YOLOv8, YOLOv11, or MobileViT) to further lower computational latency on edge hardware.
3. Multi-Object Tracking (MOT) Integration: Incorporating algorithms such as DeepSORT or ByteTrack to track temporal motion vectors across extended frame sequences.
4. V2X Communication Interfacing: Connecting prediction outputs with Vehicle-to-Everything (V2X) wireless protocols to broadcast deceleration alerts to adjacent traffic.
5. Direct ECU Actuation: Interfacing alert outputs directly with vehicle Electronic Control Units (ECUs) to enable automated evasive steering and emergency braking.

REFERENCES

[1]. J. B. Cicchino, "Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash

rates," Accident Analysis & Prevention, vol. 99, pp. 142–152, 2017.

[2]. D. Lee and H. Yeo, "Real-time rear-end collision-warning system using a multilayer perceptron neural network," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 11, pp. 3087–3097, 2016.

[3]. X. Wang, J. Liu, T. Qiu, C. Mu, C. Chen, and P. Zhou, "A real-time collision prediction mechanism with deep learning for intelligent transportation system," IEEE Transactions on Vehicular Technology, vol. 69, no. 9, pp. 9497–9508, 2020.

[4]. S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," Journal of Field Robotics, vol. 37, no. 3, pp. 362–386, 2020.

[5]. A. V. Malawade, S.-Y. Yu, B. Hsu, D. Muthirayan, P. P. Khargonekar, and M. A. A. Faruque, "Spatiotemporal scene-graph embedding for autonomous vehicle collision prediction," IEEE Internet of Things Journal, vol. 9, no. 12, pp. 9379–9388, 2022.

[6]. N. Kaempchen, B. Schiele, and K. Dietmayer, "Situation assessment of an autonomous emergency brake for arbitrary vehicle-to-vehicle collision scenarios," IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 4, pp. 678–687, 2009.

[7]. K. Lee and D. Kum, "Collision avoidance/mitigation system: Motion planning of autonomous vehicle via predictive occupancy map," IEEE Access, vol. 7, pp. 52846–52857, 2019.

[8]. J. Jansson, "Collision avoidance theory: With application to automotive collision mitigation," Ph.D. dissertation, Linköping University Electronic Press, Linköping, Sweden, 2005.

[9]. S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. A. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behaviour prediction for autonomous driving applications: A review," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 1, pp. 33–47, 2022.N. Kaempchen, K. Weiss, M. Schaefer, and K. C. J. Dietmayer, "IMM object tracking for high dynamic driving maneuvers," in Proceedings of the IEEE Intelligent Vehicles Symposium, 2004, pp. 825–830.

[10]. J. Dahl, G. R. de Campos, C. Olsson, and J. Fredriksson, "Collision avoidance: A literature review on threat-assessment techniques," IEEE Transactions on Intelligent Vehicles, vol. 4, no. 1, pp. 101–113, 2019.

[11]. X. Fang, J. Gertler, M. Kunwer, J. Heron, and T. Barkana, "A double threshold-testing robust method for fault detection and isolation in dynamic systems," in Proceedings of the American Control Conference, vol. 2, 1994, pp. 1979–1983.

[12]. M. Wisch, "Car-to-car accidents at intersections in Europe and identification of use cases for the test and assessment of respective active vehicle safety systems," in Proceedings of the 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV), 2019, pp. 10–13.

[13]. S. Riedmaier, T. Ponn, D. Ludwig, B. Schick, and F. Diermeyer, "Survey on scenario-based safety assessment of automated vehicles," IEEE Access, vol. 8, pp. 87456–87477, 2020.

[14]. S. Lee, B. Song, and J. Shin, "Collision Prediction in an Integrated Framework of Scenario-Based and Data-Driven Approaches," IEEE Access, vol. 12, pp. 55234–55247, 2024.

[15]. M. Althoff, O. Stursberg, and M. Buss, "Model-based probabilistic collision detection in autonomous driving," IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 2, pp. 299–310, 2009.

[16]. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[17]. Z. Chen, J. Zhang, and D. Tao, "Progressive coordinate regression for bird's-eye-view vehicle detection in autonomous driving," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 12645–12656, 2022.

[18]. X. Li, L. Zhang, and Y. Wang, "Real-time omnidirectional vehicle detection and multi-object tracking using improved YOLOv4 on embedded platforms," IEEE Access, vol. 10, pp. 45123–45135, 2022.

[19]. J. Kim, J. Park, and K. Yi, "A multi-tier proactive threat assessment and alert framework

using dynamic semantic grids for autonomous intersection safety," IEEE Transactions on Vehicular Technology, vol. 72, no. 4, pp. 4310–4322, 2023.

[20]. M. S. Ghiasi, A. Al-Shanoon, and M. Dianati, "A review of multi-zone collision mode classification and pre-crash mitigation strategies for advanced driver assistance systems," IEEE Intelligent Transportation Systems Magazine, vol. 16, no. 2, pp. 88–104, 2024.