Scene Matching-Assisted Adaptive Control of Autonomous Vehicles in CARLA Simulator

Yeonsu Jeong, Jiwon Lee and Minseok Choi
Department of Electronic Engineering, Kyung Hee University, Yongin, South Korea
E-mails: jys3049@khu.ac.kr, jiwonn612@khu.ac.kr, choims@khu.ac.kr

***Abstract*—This paper presents a practical and lightweight system designed to support robust autonomous driving in adverse weather, especially dense fog. Instead of relying on expensive LiDAR or fragile camera-only setups, our approach combines YOLOv5-based object detection with Vehicle-to-Vehicle (V2V) communication. A key feature of the system is a scene matching method that uses ROI Align to compare feature maps between vehicles and verify visual consistency. This enables a two-stage distance estimation process—starting with an estimate based on V2V velocity and timing data, then refining it using the relative sizes of matched objects. We tested the system in the CARLA simulator under varying fog levels, showing stable performance with over 90% accuracy and real-time processing speeds under 0.3 seconds. The system also includes a fail-safe mechanism that ensures safety in extreme conditions. Our results show that combining camera input with V2V communication can provide reliable cooperative driving even when visibility is severely limited.**

***Index Terms*—Autonomous Driving, Adverse Weather Perception, V2V Communication, Scene Matching, Distance Estimation.**

I. INTRODUCTION

Autonomous driving technology has the potential to significantly improve road safety and traffic efficiency. To operate reliably in the real world, autonomous vehicles must be capable of perceiving and responding to their surroundings in real time—even under challenging conditions such as fog, heavy rain, or snow. However, verifying such systems in real environments is often expensive, time-consuming, and risky. To address this, simulation platforms like CARLA have become essential tools for safely testing diverse driving scenarios, and are used extensively in this study.

Sensor reliability is a major concern in adverse weather. LiDAR and radar, while widely used, suffer from reduced performance in fog or rain due to signal scattering and attenuation [4], [6]. LiDAR, in particular, is sensitive to fog, and although radar offers better robustness, its low resolution often limits object recognition and distance accuracy [2]. On the other hand, cameras provide rich visual information at a lower cost but are easily affected by poor visibility, which can critically degrade perception and control performance. While many studies have explored solutions such as sensor fusion, advanced image enhancement, or deep learning improvements [5], these approaches often add complexity, cost, or fail to fully address the fundamental limitations of the individual sensors in extreme conditions.

In this work, we propose a camera-based system enhanced by Vehicle-to-Vehicle (V2V) communication to support safe and stable driving in foggy conditions. Several studies have employed V2V communications to support task-specific [10] or delay-sensitive services [11]. While YOLO-based object detectors have shown promise in adverse weather, their effectiveness drops significantly when visual information is severely limited [3], [12]. Our approach addresses this gap by combining camera perception with cooperative V2V communication, offering a cost-effective alternative to expensive sensor suites like LiDAR and radar.

The main contributions of this paper are as follows:

- **Fog-specific dataset generation and training:** We collected a custom dataset under dense fog in the CARLA simulator and trained a YOLOv5 model optimized for these conditions.
- **Scene matching and distance estimation:** We introduce a novel scene matching method using intermediate features from YOLOv5, enabling a two-stage distance estimation process based on V2V data and visual similarity.
- **Experimental validation in CARLA:** We demonstrate the effectiveness of our system in simulated foggy environments, showing accurate distance estimation and stable vehicle control performance.

II. SYSTEM MODEL IN CARLA SIMULATOR

To evaluate the proposed system, we use the open-source CARLA simulator (v0.9.15) [8], which offers a realistic driving environment and fine-grained control over vehicle behavior, sensors, and weather. This section describes the simulation setup and provides an overview of the system architecture.

A. CARLA Simulator Environment Configuration

All simulations were conducted on CARLA's Town04 map, chosen for its long, straight highway sections well suited for testing vehicle following and lane-change scenarios. As shown in Fig. 1, the test section supports a range of lane configurations for evaluating various driving maneuvers. The ego vehicle (EV) is equipped with a forward-facing RGB camera (800×600 resolution, 90-degree FOV) capturing images at 20 FPS. Lead vehicles (LVs) are placed ahead of the EV and follow predefined paths using CARLA's BasicAgent module. All vehicles use the Tesla Model 3 blueprint.

To emulate real-world road conditions, the following settings were applied:

- **Lane Configuration**: The selected highway segments contain 1 to 4 lanes per direction, allowing tests under

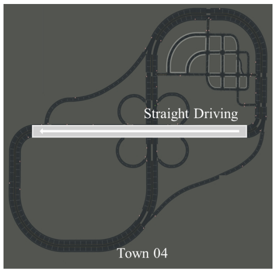


Fig. 1: The Town04 highway environment in the CARLA simulator, highlighting the selected long highway sections for driving scenarios.

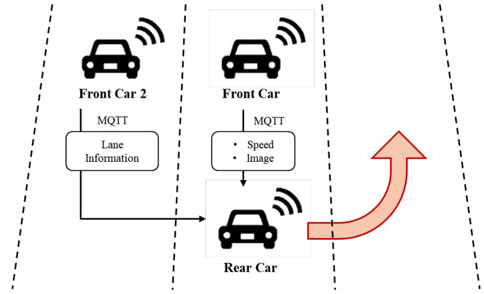


Fig. 2: Driving Scenarios in Adverse Weather.

diverse traffic scenarios including lane keeping and lane changes.

- **Environmental objects**: Standard CARLA roadside elements such as speed limit signs, warning signs, streetlamps, trees, and billboards were deployed. These were also used to build the object detection dataset through manual labeling on the Roboflow platform.
- **Vehicle speed and arrangement**: In the main scenario, the LV drives at 50 km/h in the second lane, while the EV follows at 57 km/h in the same lane. An additional lead vehicle (LV2) is placed in either the adjacent first or third lane, traveling at 55 km/h. This setup allows the system to be tested in multi-vehicle and lane-change situations.

All vehicles are virtually equipped with GNSS and IMU sensors, providing ground truth data for position and velocity. These values are used to simulate V2V communication between the vehicles.

B. Driving Scenarios in Adverse Weather

To evaluate system performance under poor visibility, we configured various fog conditions using CARLA's built-in weather engine. The fog density parameter was gradually increased from 0 (clear) to 100 (dense), with primary experiments conducted in the range of 20 to 100 to simulate realistic adverse scenarios.

In our main scenario, the ego vehicle (EV) follows a lead vehicle (LV) in the same lane while maintaining a safe distance. A second lead vehicle (LV2) travels in an adjacent lane and plays a supporting role by sharing lane occupancy information. This setup is intended to test whether the EV can maintain safe driving even when the LV becomes visually indistinct due to fog. To support this, V2V communication allows the LV to transmit its velocity and front camera image to the EV, while LV2 shares information about its current lane and whether it is occupied. This shared data enables the EV to estimate inter-vehicle distance and make informed decisions even when its own visual perception is impaired. Fig. 2 shows the complete driving scenario, including the relative positions of all vehicles, lane layout, and the direction of data exchanged via V2V links.

C. System Architecture

As shown in Fig. 3, the system is composed of four modules: perception, communication, decision-making, and control. The perception module performs object detection, feature extraction, scene matching, and distance estimation using camera input and V2V images. The communication module exchanges real-time data such as velocity, camera images, and lane status with nearby vehicles. Based on this information, the decision-making module selects appropriate actions, such as maintaining distance, braking, or changing lanes, which are executed by the control module via low-level commands. To enhance safety, a Fail-Safe mechanism is triggered under severe perception degradation, prompting the vehicle to slow down or stop as needed.

III. Scene Matching and Adaptive Control Methodology

This section outlines the core components of our system: a perception and a control module. We describe how each component contributes to achieving robust performance under foggy conditions, using only camera input and V2V communication.

A. Object Detection and Custom Dataset Generation

To enable reliable object detection under adverse weather, we trained a YOLOv5s model on a custom dataset collected in the CARLA simulator. We selected YOLOv5s for its balance between accuracy and speed, leveraging its CSPDarknet backbone and PANet/FPN architecture for multi-scale detection [9]. The dataset was built by capturing images from the EV's front-facing RGB camera while driving at approximately 60 km/h on the Town04 highway map. Fog conditions were applied to simulate realistic low-visibility environments.

Objects were categorized into two types: base classes (e.g., tree, streetlamp) representing general road features, and event classes (e.g., warning signs, advertisements, accidents) conveying critical road information. These were randomly placed at regular intervals (about 20 meters) along the road to ensure spatial diversity. A total of 663 images were collected and labeled using the Roboflow platform, with a 70:30 train-validation split. This dataset enabled the model to detect both static environmental features and dynamic events in foggy conditions, forming the foundation for downstream scene

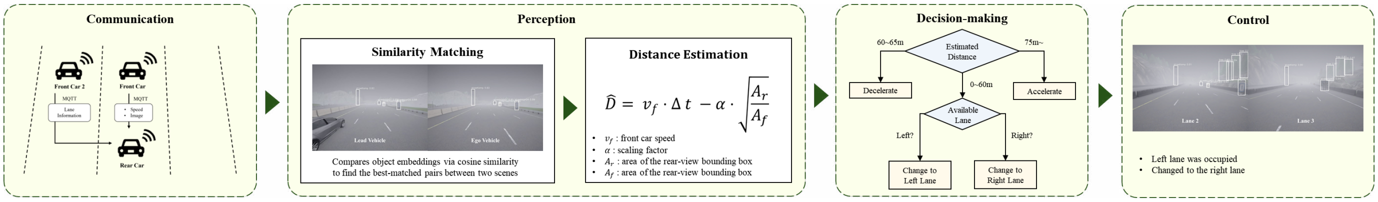


Fig. 3: Overall System Architecture Diagram.

Fig. 4: Objects placed and utilized in the CARLA simulator for dataset generation.

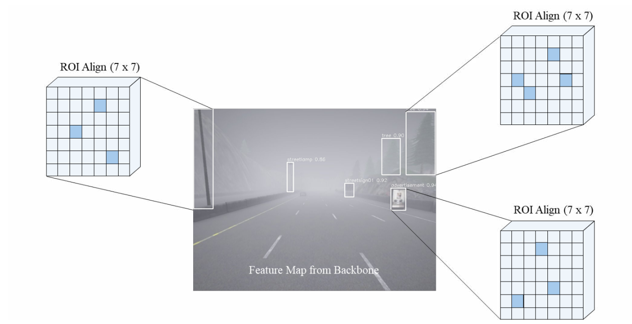


Fig. 5: ROI Align Architecture for Feature Extraction.

matching and distance estimation. Examples of labeled objects used in training are shown in Fig. 4.

B. V2V Communication Module

To support cooperative perception, we implemented a lightweight V2V communication module based on MQTT, a publish–subscribe protocol suitable for low-bandwidth, high-frequency data exchange.

In our setup, the lead vehicle (LV) continuously publishes its velocity and front camera images to a local MQTT broker. The second lead vehicle (LV2), driving in an adjacent lane, publishes lane-related information such as its current lane ID and occupancy status. All data is encoded in JSON format and includes timestamps to ensure synchronization.

The ego vehicle (EV) subscribes to both data streams. While LV images are transmitted at 20 Hz, the EV stores them at a reduced rate of 2 Hz to balance performance and memory usage. This asynchronous sharing of velocity, vision, and lane status enables the EV to supplement its own visual input with cooperative information, especially in situations where visibility is degraded due to fog. This communication scheme is essential for enabling downstream modules, such as scene matching and adaptive control, to function reliably under adverse conditions.

C. Scene Matching Algorithm

To establish visual consistency between vehicles, we apply a scene matching pipeline that compares detected event objects from the EV and the LV using intermediate features extracted from YOLOv5.

First, both EV and LV images undergo object detection using the trained YOLOv5s model. We focus on detecting two types of objects: base classes (e.g., trees, streetlamps) and event classes (e.g., traffic signs, warning signs, advertisements). However, only event objects are used for scene matching, as they are more distinct and informative across different viewpoints.

For each detected event object, we extract features using region-of-interest (ROI) Align from the fourth layer of YOLOv5's backbone. ROI Align generates fixed-size 7×7 feature maps that preserve spatial precision, enabling reliable comparisons even when objects appear at different scales or angles. The feature extraction process is illustrated in Fig. 5.

We then compute the cosine similarity between corresponding event object features from the EV and LV. If the similarity exceeds a predefined threshold, denoted as τ_{sim}, the pair is considered a match. Scene similarity is quantified by the match ratio—the proportion of matched event objects relative to total detections. This metric serves as the basis for both distance estimation and fail-safe activation.

D. Enhanced Multi-Stage Distance Estimation

We estimate the distance between the EV and the LV using a two-stage method that combines V2V data with visual correction based on scene matching. In the first stage, we compute an initial distance estimate D based on the LV's velocity v_f (received via V2V) and the time difference Δt between when the LV and EV captured matching scenes:

$$D = v_f \cdot \Delta t. \tag{1}$$

Here, Δt is calculated using timestamps from the LV's transmitted image and the EV's current frame. This gives a rough approximation of the distance based on temporal separation.

In the second stage, this estimate is refined using visual information. Specifically, we compare the bounding box areas

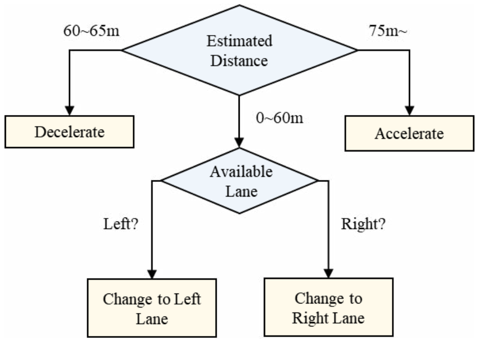


Fig. 6: Adaptive vehicle control decision-making process based on estimated inter-vehicle distance and V2V lane information.

of matched event objects between the two vehicles. Since the apparent size of an object in an image is inversely related to its distance, the area ratio provides a cue for correction:

$$\overline{D} = v_f \cdot \Delta t - \alpha \cdot \left\{ \sqrt{\left(\frac{1}{n} \sum_{i=1}^{n} \frac{A_i^r}{A_i^f} \right)} - 1 \right\}, \qquad (2)$$

where A_i^r and A_i^f are the areas of the i-th matched object in the EV and LV images, respectively, and n is the number of matched objects. The correction coefficient α is empirically set to 0.15 to ensure stable adjustment within $\pm 5\%$. This two-stage approach enables accurate distance estimation even when visual input is partially degraded, by combining temporal V2V data with spatial visual cues.

E. Vehicle Control Execution and Fail-Safe Mechanism

The final distance estimate $\overline{D}$ computed from scene matching and V2V communication feeds directly into the control logic, which determines safe driving actions based on real-time conditions. A custom controller then executes these actions through throttle, brake, and steering commands, extending CARLA's BasicAgent functionality.

Fig. 7 illustrates this adaptive decision-making process. The system classifies the inter-vehicle distance into several ranges to determine the appropriate response:

- When the estimated distance is below a critical threshold, denoted as D_m, it is considered a close proximity risk, triggering emergency braking or a possible lane change.
- When the distance exceeds a safe driving threshold, denoted as D_s, the system accelerates to optimize driving efficiency while remaining cautious.
- Within a maintenance range, i.e., $D_m \leq \overline{D} \leq D_s$, the vehicle adjusts its speed to maintain a steady distance.

To support safe lane changes, the system uses real-time lane occupancy data received from LV2 via V2V communication. Before initiating a maneuver, it checks adjacent lanes for availability and only proceeds when a safe option is confirmed.

In addition, a Fail-Safe mechanism is integrated to handle perception failure or system instability. If the scene similarity score falls below a preset threshold, denoted as τ_{fail}, or object

Fig. 7: Ego vehicle and lead vehicle camera view in fog with object detection.

Fig. 8: Example of an adaptive lane change maneuver performed due to close-proximity risk detection, leveraging V2V lane occupancy information.

detection becomes unreliable due to extreme fog, the vehicle switches to a safe fallback mode. In this mode, the EV slows to a minimum speed, maintains a conservative following distance, or activates hazard lights. This mechanism ensures that the vehicle remains operational and safe even in severely degraded conditions, and can prompt driver intervention if necessary.

IV. Experimental Results

We evaluated our system in the CARLA simulator under various fog densities, focusing on four key aspects: object detection performance, distance estimation accuracy, adaptive control behavior, and real-time processing capability. Each component was tested using controlled driving scenarios with synthetic fog ranging from light to dense, as detailed in Section II-B. The results demonstrate how our system maintains robust perception and safe vehicle control using only cameras and V2V communication, even when visibility is severely degraded.

A. Simulation Results Visualization

Fig. 7 shows an example of object detection and scene matching in dense fog. Despite low visibility, both the EV and LV successfully detect common road elements such as trees and traffic signs. Matched objects yield high cosine similarity scores (typically ¿ 0.6), indicating strong visual consistency between vehicle perspectives.

Fig. 8 presents a successful lane change scenario triggered by a close-proximity event. When the estimated distance to the LV dropped below the safety threshold, and lane occupancy information from LV2 confirmed that the adjacent third lane was available, the EV executed a smooth and safe lane change. This illustrates the system's ability to combine visual and V2V information for real-time adaptive control.

TABLE I: Object Detection Performance (AP and F1 Score) in Fog Density 60%.

Index	Tree	Streetlamp	Streetsign01	Streetsign04	Advertisement	Warning
AP	0.939	0.967	0.972	0.978	0.967	0.946
F1	0.915	0.936	0.942	0.935	0.890	0.903

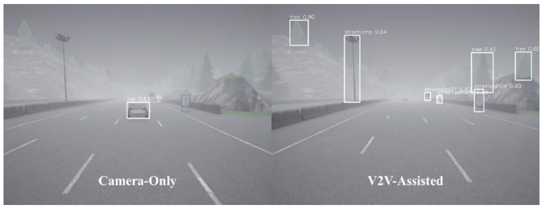


Fig. 9: Comparison of perception range and safe following distance between a Camera-Only YOLO system and the proposed V2V-Assisted system in foggy conditions.

(a) Fog Density 0% (b) Fog Density 20% (c) Fog Density 30%

(d) Fog Density 40% (e) Fog Density 60% (f) Fog Density 100%

Fig. 10: Visual representation of varying fog densities (0% to 100%) as simulated in the CARLA environment with object detections overlaid. Each subfigure illustrates the visual conditions at a specific fog density level.

B. Object Detection Performance

We evaluated the trained YOLOv5s model under foggy conditions to assess its ability to detect both base and event objects. Table I shows the average precision (AP) and F1 scores for each class at a representative fog density of 60%. The model achieved high performance across most object types, with AP values exceeding 0.93 and F1 scores above 0.89. Notably, structured objects like traffic signs and streetlamps showed slightly higher detection accuracy than more variable or less distinct objects such as advertisements and warning signs.

While increasing fog density generally led to a minor drop in accuracy—especially for small or visually ambiguous objects—the overall detection quality remained sufficient for downstream processing. In particular, the stable detection of event-class objects provided a reliable foundation for scene matching, which depends heavily on accurate localization of such features. Furthermore, the use of V2V-shared images from the lead vehicle helped compensate for occasional detection failures on the EV side, enhancing the overall robustness of the perception pipeline under adverse conditions.

C. Comparison with Camera-Only System: Perception Range and Safe Distance

To highlight the benefits of integrating V2V communication, we compared our system with a conventional camera-only YOLOv5 setup under foggy conditions. Fig. 9 illustrates the difference in perception range and safe following distance. In the camera-only system, stable object detection was possible only within approximately 13.6 meters in dense fog. At a driving speed of 60 km/h, this distance is traversed in just 0.8 seconds—insufficient time for reliable braking or control, posing a serious safety risk.

By contrast, the proposed system significantly extended the perception horizon. Through V2V communication and scene matching, the ego vehicle was able to maintain consistent tracking of the lead vehicle and preserve a safe following distance, even in dense fog. The system maintained inter-vehicle spacing within the defined safety thresholds ($D_m = 40--60$ m, $D_s = 60$ m), which would not have been possible with visual input alone. These results demonstrate the critical role of V2V assistance in adverse weather, where reliance on camera input alone leads to unsafe proximity and delayed reactions.

D. Distance Estimation Accuracy vs. Fog Density

To evaluate the reliability of our distance estimation algorithm, we tested its performance under varying fog densities, using both the initial estimate (from V2V time offset) and the corrected value (using visual area ratios). Fig. 10 visually represents these varying fog conditions. Table II summarizes the results across five levels of fog density.

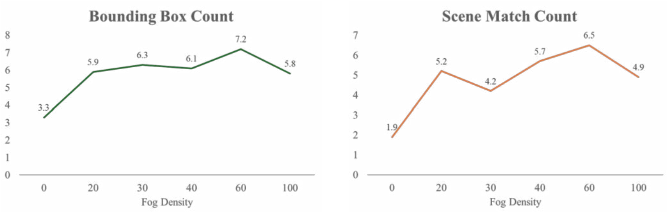
The initial estimate achieved accuracy between 85% and 91%, depending on visibility. After applying the visual correction stage, accuracy consistently improved, exceeding 90% in all cases and reaching up to 94.18% at 100% fog density. The correction coefficient $\alpha = 0.15$, introduced in Equation (2), was empirically chosen to ensure that the average correction ratio remained within $\pm 5\%$. This setting provided a stable balance between under- and overestimation, even when the number of matched objects varied due to visual occlusion. These results demonstrate that the two-stage estimation process—starting with V2V-based timing and refining via visual similarity—offers robust and adaptive distance tracking across a wide range of visibility conditions.

E. Scene Matching Performance and Real-Time Capability

Fig. 11 illustrates the performance of our scene matching algorithm across various fog densities, detailing both the initial object detection capability and subsequent matching accuracy. As shown in Fig. 11(a), the bounding box count remains high at moderate fog levels, indicating robust initial object detection by YOLOv5s. However, at 100% fog density, the number of

TABLE II: Distance Estimation Accuracy vs. Fog Density.

Fog Density (%)	Initial Accuracy (%)	Corrected Accuracy (%)
20	85.86	91.69
30	87.94	91.17
40	89.65	93.58
60	88.98	93.58
100	90.72	94.18



(a) Bounding Box Count

(b) Scene Match Count

Fig. 11: Scene Matching Performance across various fog densities. (a) illustrates the total number of bounding boxes detected by YOLOv5s, and (b) shows the count of successfully scene-matched objects. Both metrics are presented as a function of increasing fog density.

detected bounding boxes sharply decreases due to the severe visual obscuration of objects. Correspondingly, Fig. 11(b) shows that the scene matching count also exhibits a similar trend, directly impacted by the initial detection performance. At 100% fog density, the sharp decrease in scene matching accuracy is primarily attributed to the reduced bounding box count, as insufficient initial detections make feature vector comparison impossible.

The entire perception to control loop, including the scene matching algorithm implemented with ThreadPoolExecutor for parallel processing, executed in an average of 0.3 seconds. This satisfies the real-time requirements for autonomous driving systems. The Fail-Safe system was also confirmed to operate as designed in all tested conditions, switching to a safe mode upon scene matching failures or system instability, thus enhancing overall system reliability.

V. Conclusion

We presented a practical and cost-effective system that combines YOLOv5-based object detection with V2V communication to address the challenges of autonomous driving in foggy conditions. By integrating scene matching using ROI Align and a two-stage distance estimation algorithm, the system achieved robust perception and safe inter-vehicle control without relying on expensive sensors like LiDAR. Extensive simulation in the CARLA environment demonstrated that our approach maintains over 90% distance estimation accuracy and real-time processing performance (¡0.3 seconds), even under dense fog. The system also incorporates adaptive lane control and a Fail-Safe mechanism, ensuring safe operation in degraded visual environments. Our results suggest that combining vision-based detection with lightweight V2V communication is a viable alternative for reliable cooperative driving under adverse weather.

Acknowledgment

This work was supported in part by Institute of Information & communications Technology Planning & Evaluation (IITP) under YKCS Open RAN Global Collaboration Center (IITP-2025-RS-2024-00434743) grant funded by the Korea government (MSIT), and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00521295).

References

[1] A. Haider, M. Pigniczki, S. Koyama, M. H. Köhler, L. Haas, M. Fink, M. Schardt, K. Nagase, T. Zeh, A. Eryildirim, et al., "A Methodology to Model the Rain and Fog Effect on the Performance of Automotive LiDAR Sensors," *Sensors*, vol. 23, p. 6891, 2023.

[2] S. Chaudhary, S. Khichar, Y. Meng, and A. Sharma, "Multiple target detection using photonic radar for autonomous vehicles under atmospheric rain conditions," *PLoS ONE*, vol. 20, no. 5, p. e0322693, 2025.

[3] T. Kim, H. Jeon, and Y. Lim, "Challenges of YOLO Series for Object Detection in Extremely Heavy Rain: CALRA Simulator based Synthetic Evaluation Dataset," *arXiv preprint*, arXiv:2312.07976, 2023.

[4] D. J. Yeong, G. Velasco Hernandez, J. Barry, and J. Walsh, "Sensor and Sensor Fusion Technology in autonomous vehicles: A Review," *Sensors*, vol. 21, no. 6, p. 2140, 2021.

[5] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Autonomous Driving in Adverse Weather Conditions: A Survey," *arXiv preprint*, arXiv:2112.08936, 2021.

[6] S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, and M. A. Kaafar, "The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self driving car," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 103-111, 2019.

[7] R. Walambe, A. Marathe, K. Kotecha, and G. Ghinea, "Lightweight Object Detection Ensemble Framework for autonomous vehicles in challenging weather conditions," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1-12, 2021.

[8] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. Conf. Robot Learning (CoRL)*, PMLR, pp. 1-16, 2017.

[9] J. Glenn, "YOLOV5 Release v6.1," Ultralytics, 2022. [Online]. Available: https://github.com/ultralytics/yolov5/releases/tag/v6.1

[10] M. Choi, J. Kim and J. Moon, "Adaptive Detector Selection for Queue-Stable Word Error Rate Minimization in Connected Vehicle Receiver Design," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3635-3639, April 2018.

[11] W. J. Yun, D. Kwon, M. Choi, J. Kim, G. Caire and A. F. Molisch, "Quality-Aware Deep Reinforcement Learning for Streaming in Infrastructure-Assisted Connected Vehicles," in IEEE Transactions on Vehicular Technology, vol. 71, no. 2, pp. 2002-2017, Feb. 2022.

[12] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image adaptive YOLO for object detection in adverse weather conditions," in Proc. AAAI Conf. Artif. Intell., vol. 36, no. 2, pp. 1792-1800, 2022.

[13] T. Sharma, B. Debaque, N. Duclos, A. Chehri, B. Kinder, and P. Fortier, "Deep learning based object detection and scene perception under bad weather conditions," Electronics, vol. 11, no. 4, p. 563, 2022.

[14] H. Jeon, Y. O. Kim, M. Choi, D. Park, S. Son, J. Lee, G. Choi, and Y. Lim, "CARLA Simulator based Evaluation Framework Development of Lane Detection Accuracy Performance under Sensor Blockage Caused by Heavy Rain for Autonomous Vehicle," IEEE Robot. Autom. Lett., vol. 7, no. 4, pp. 9977-9984, 2022.