Journal of Applied Automation Technologies

Research Paper

Universal Sensor Fusion Architecture for Cooperative Perception in Intelligent Connected Vehicles

Franciszek Adrian Błaszczyk[1], Artur Majewski[1,*] and Celina Kubiak[1]

[1] Faculty of Electrical Engineering and Automatics, Adam Mickiewicz University in Poznan, Poznan, 61-614, Poland
*Corresponding author: artur.maj@amu.edu.pl

Abstract. Although intelligent connected vehicles have strong perception capabilities in complex and dynamic environments, they still cannot improve people's lives. This paper proposes a universal sensor fusion architecture for cooperative perception in multi-vehicle groups. By integrating multimodal data from cameras, LiDAR, radar, and other sensors through distributed fusion and unified data abstraction, this framework is modular and protocol-agnostic. Through large-scale real-world deployments and detailed simulations, experimental validation is conducted to assess its effectiveness in various environments, such as highways, urban traffic, and adverse weather conditions. Quantitative data indicate that the proposed system achieves a detection rate of over 92% in most scenarios. In adverse weather or at night, the system's detection rate drops to 83-88%, but its latency remains relatively low, with end-to-end latency not exceeding 130 milliseconds under maximum load. A solution has been proposed that responds quickly to anomalies, is fault-tolerant to sensor failures, and achieves efficient resource utilization through edge-cloud collaboration and adaptive fusion algorithms. This structure is suitable for expansion and adjustment, which can enhance the collaborative perception performance and reliability of intelligent vehicle networks, and is also feasible in other aspects. This paper discusses the scope and feasibility of applying this architecture in the creation of large-scale automated transportation systems.

Keywords: *Automotive, Sensor Fusion, Cooperative Perception, Intelligent Vehicles, Edge Computing, System Architecture*

Received on 09 December 2024, Accepted on 12 May 2025, Published on 18 May 2025

Introduction

With the rapid development of intelligent connected vehicles (ICV), the entire transportation model has changed. We are now in an era of high perception and automation. By combining next-generation sensor technology with in-vehicle computing, low-latency vehicle-to-everything (V2X) communication can be promoted, providing richer contextual awareness for advanced driving [1]. Despite these achievements, there are still some issues. In adverse weather conditions, the perception systems at urban intersections and multi-agent highways remain relatively unstable and inaccurate [2]. These issues include sensor occlusion, data insufficiency, and traffic anomalies. The limited perception range and inherent modality-specific defects of individual sensors exacerbate the problem, and system misses and ambiguous classifications can lead to dangerous decisions [3]. In dense traffic, complex road conditions, or sudden obstacle detection, even the best sensor fusion frameworks currently used in single vehicles often fail to meet these requirements [4]. Loss of recognition, slow response speed, and reduced confidence in emergencies are a few examples of the aforementioned deficiencies [5]. Improving the environmental perception capabilities of intelligent vehicles remains a problem that the industry and research field need to address [6].

Given the limitations of individual perception, people have recently begun to use multiple vehicles or infrastructure nodes to collaborate, sharing and integrating sensor data through vehicular ad hoc networks [7]. Collaborative perception can share spatially and temporally distributed data, helping to address blind spot issues,

https://doi.org/10.64972/jaat.2025v3.207p19e:252-266

extend the perception range, and reduce the impact of single sensor failures [8]. Achieving this goal also faces some significant technical challenges. Under bandwidth limitations, prioritizing the update of critical data, handling asynchronous data streams, and supporting communication modes with multiple data formats are necessary conditions for multi-vehicle information sharing [9]. The different types of sensors and hardware specifications from various manufacturers, as well as proprietary data interfaces, hinder cross-brand and multi-generational data integration [10]. To prevent network instability, packet loss, and physical network attacks, the system should have scalability and fault tolerance capabilities, so that cooperative collision avoidance and joint path planning applications can transmit and integrate data in real-time and reliably. In order to ensure safe driving, scientists are researching how to build a scalable, open-source, and universal sensor fusion platform to unify various vehicle platforms and achieve high-precision collaborative perception.

This paper proposes a comprehensive sensor fusion architecture for collaborative perception in intelligent connected vehicles. To support future new communication standards as well as various sensors and operations of different scales, this architecture is necessary. Focusing on high-level data abstraction and protocol agnosticism to achieve robust integration of multiple sensor networks from various hardware generations and sources, thereby enabling real-time, reliable collective perception. In order to thoroughly validate the aforementioned methods, extensive large-scale empirical studies were conducted. These studies involve various traffic conditions, environmental settings, and hardware configurations. The following analysis includes the perception performance of large-scale collaborative sensing, robustness under faults and stress, resource utilization efficiency, and other important trade-offs. We hope to advance cutting-edge vehicle sensor fusion technology through this research and provide foundational research and practical applications for the next generation of collaborative autonomous driving systems.

Related Work

Cooperative Perception in Connected Vehicles

Cooperative perception has recently been used to improve the performance of intelligent connected vehicles, addressing the issue of isolated single-vehicle perception. Through the cooperative mechanism between vehicles and infrastructure, raw and preprocessed sensor data from different road areas can be collected at different times. This approach can create a comprehensive road condition database that includes three-dimensional data. By sharing data, the coverage of the fleet can be expanded, while addressing blind spots and occlusion issues, thereby improving the fleet's detection and prediction performance. In heavy traffic, complex intersections, and situations where single vehicle sensors are limited or small sensors are used, the aforementioned methods also perform well. By integrating edge and cloud resources into the collaborative perception framework, the accuracy of evaluations in dynamic or unstable environments can be improved [11]. In early field tests and simulation experiments, it was found that the improved system was superior in danger recognition, collision avoidance, and response speed. In light of these achievements, the industry has already begun to build a comprehensive platform to collaboratively perceive, communicate, and make decisions in order to scale up the construction of safer and smarter mobile systems [12].

Multi-Sensor Fusion Technologies and Standards

Reliable sensor fusion strategies are crucial for the perception of autonomous vehicles. For example, cameras, radar, and LiDAR use multiple sensors to integrate data and compensate for the shortcomings of individual sensors to build a relatively robust environmental model. Early fusion, also known as the merging of initial features, and late fusion, also known as the merging of individual choices or outputs [13]. Probabilistic methods, such as Bayesian inference and Kalman filters, can be used to handle uncertainty in sensor data. These methods are also very effective in tracking and locating objects [14]. With the development of deep learning, learning-based fusion frameworks have garnered attention. These frameworks can be used to improve the effectiveness of recognition and classification, and to automatically learn the best fusion strategies for high-dimensional, multimodal features [15]. Data synchronization, real-time data exchange, and system integration are supported by ROS and adaptive standard platforms [16]. In order to achieve fast and reliable transmission of collaborative perception data in the real world, the ITS-G5 and DSRC protocols are used to establish low-latency information

https://doi.org/10.64972/jaat.2025v3.207p19e:252-266

sharing systems [17]. The research and industrial deployment in the field of multi-agent automated mobility are based on these standards and technologies [18].

Limitations of Existing Architectures

Despite significant progress in collaborative perception architectures, they are still not suitable for large-scale deployment in complex and diverse real-world vehicle fleets [19]. Many of these systems use proprietary or non-standard sensor connections, making it difficult to integrate them into vehicles or devices from different years [20]. In order to accommodate new devices, multiple data formats must be supported, as modifying old systems can lead to significant delays in testing [21]. Currently, most solutions are not accurate enough in handling fluctuations in available communication bandwidth or computing resources, which leads to a decline in perception performance under high network load or heavy traffic conditions [22]. Due to the limitations of the software interface and resource allocation plan, the system's scalability is limited. As the number of participating vehicles increases, this will become an obstacle [23]. Other issues include the vulnerability of asynchronous sensor streams, alignment errors, and imperfect data synchronization. These issues may propagate during the fusion process, leading to perception failures or unexpected delays [24]. These issues indicate the need for a universal, open, and scalable sensor fusion framework [25]. It must be scalable, ensure seamless interoperability, and ensure reliable operation in the complex and unstable environments of large-scale real-world connected vehicle deployments.

Universal Fusion Architecture and Algorithms

System Architecture Overview

A universal sensor fusion architecture can be used to build an all-weather, fully networked, heterogeneous hardware, multi-vehicle collaborative perception system. The sensor interface layer, abstraction and communication middleware layer, and fusion decision layer constitute the three logical levels of the architecture. Protocol independence and hierarchical modularity contribute to vendor-neutral scalability.

At system startup, the perception interface layer continuously collects high-dimensional output data from camera arrays, distributed radar, LiDAR, and vehicle or infrastructure nodes. These signals are referenced in time and space, and then sent to an abstract middleware for secure processing, synchronization, and standardization to ensure that the sensor streams of all connected nodes match. The middleware ensures compatibility of all data, regardless of its original format, by standardizing timestamps, cross-source spatial calibration, and encoding initial uncertainties.

The fusion decision layer is responsible for collective perception, behavior modeling, and fusion results. According to the importance of the current task and bandwidth, the process of managing distributed fusion with confidence-adjusted propagation is carried out between local processors and cloud-edge resources. Each submodule can dynamically select the aggregation mode based on packet loss or incomplete data collection.

The holistic flow can be formalized as follows. Let $\boldsymbol{S_}$in represent the multi-modal sensor observations from all nodes, and let $\boldsymbol{C\_ctx}$ comprise all context and historical state variables. The cooperative fusion outcome, $\boldsymbol{P}$_out, is described by:

$$\boldsymbol{P\_out} = \mathcal{F}\left(\sum_{k=1}^{N} \phi_k\left(\mathbf{T}_k\left(\boldsymbol{S\_in}^{(k)}, \boldsymbol{C\_ctx}\right)\right) + \mathcal{E}_{\text{noise, drop}}\right) \quad \text{Eq.(1)}$$

where ϕ_k denotes node-specific pre-processing, $\mathbf{T}_k$ the temporal/spatial registration, and $\mathcal{E}_{\text{noise, drop}}$ models network noise and sensor dropout.

The abstract mapping function is used to describe the mapping from modular structure to functionality:

$$\boldsymbol{M}_{\text{arch}}: \{L_{\text{sens}}, L_{\text{abs}}, L_{\text{fuse}}\} \to \mathcal{O}_{\text{sys}} \quad \text{Eq.(2)}$$

with the mapping $\boldsymbol{M}_{\text{arch}}$ denoting the architectural transformation from sensory, abstraction, and fusion layers to finalized system outputs.

A general sensor fusion architecture is used to integrate various sensor data, abstract these data, and then make decisions based on them. First, collect a large amount of data from distributed sensor arrays, and then process it through the system's abstract middleware layer. Standardize various data sources, align them temporally, add relevant attributes, and build a common foundation for subsequent processing. In order to create a complete environmental model, this merged information is subsequently sent to the collaborative fusion engine. The collaborative fusion engine will integrate data from multiple sensors, which have been standardized and include dynamic uncertainty descriptors. The fused perception output is directly sent to the main perception and decision-making modules, which use this output during real-time operations. Figure 1 shows the extension of the architecture to add bidirectional communication channels between cloud resources, edge, and vehicle endpoints. These two channels are used to continuously collect feedback and make context-aware adjustments to the system; the fusion pipeline can quickly respond to changes in the network and operational environments.

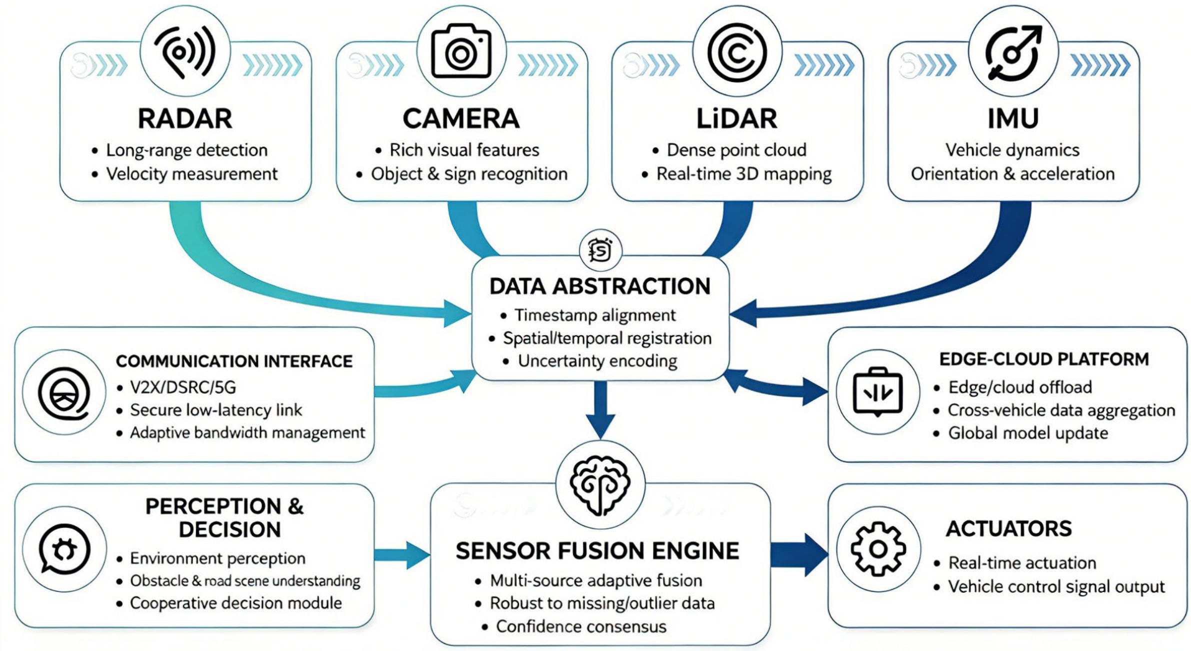

Figure 1. Universal Sensor Fusion System Architecture.

Data Abstraction and Communication Protocol

An efficient data abstraction layer and high-speed communication mechanisms are the foundation of multi-sensor and multi-agent collaboration. Convert all raw data from sensors into a standard format. Then, unified timestamps, geometric registration, and probabilistic metadata are overlaid. The aforementioned concepts can collaborate with each other to understand various sensors and nodes.

Through encoding, each data packet will contain context tags, uncertainty scores, and semantic descriptors. Protocol-agnostic serialization includes high-dimensional point clouds, image features, and dynamic object trajectories. They are all presented in the form of structured message frames. At the network level, the abstract model employs intelligent scheduling to prioritize key perceptual features and adaptively sparsify data, thereby achieving low-latency connections under bandwidth constraints.

The abstract mapping model of multimodal data from all sources is shown as follows:

$$\mathcal{D}_{abs} = \mathcal{A}\left(\bigcup_{m=1}^{M} \mathcal{S}_{mod}^{(m)}(t), \mathcal{G}_{\text{reg}}, \mathcal{T}_{\text{sync}}, \mathcal{U}_{\text{meta}}\right) \quad \text{Eq.(3)}$$

where $\mathcal{S}_{\text{mod}}^{(m)}$ is the sensor modality data at time t, $\mathcal{G}_{\text{reg}}$ the geometric registration, $\mathcal{T}_{\text{sync}}$ the time synchronization, and $\mathcal{U}_{\text{meta}}$ the associated uncertainty and semantic metadata.

Priority and adaptation policies are included in the map that links logical message objects to physical transmission frames in the protocol interface:

$$\mathcal{P}_{\text{proto}}\left(\mathcal{O}_{msg}, \mathcal{B}_{\text{prio}}, \mathcal{Q}_{\text{opt}}\right) = \left\{\mathcal{F}_{\text{phys}}, \mathcal{S}_{\text{sched}}, \mathcal{R}_{\text{ctrl}}\right\} \quad \text{Eq.(4)}$$

where $\mathcal{O}_{\text{msg}}$ denotes abstracted information objects, $\mathcal{B}_{\text{prio}}$ denotes dynamic priority allocations, $\mathcal{Q}_{\text{opt}}$ the adaptive optimization criteria, mapped to physical frames $\mathcal{F}_{\text{phys}}$, network scheduling $\mathcal{S}_{\text{sched}}$, and resource control $\mathcal{R}_{ctrl}$.

In order to ensure the integrity of synchronization, hybrid clock alignment and error propagation protocols were used, and out-of-band packets were employed to correct lost or delayed data packets to ensure the integrity of synchronization.

Fusion Algorithm Framework

The high confidence and adaptive fusion algorithms at the core of this structure can achieve accurate perception results in various harsh environments. Preprocessing, context enrichment, and uncertainty propagation are components of the multi-stage pipeline for incoming data streams, which may be asynchronous or incomplete [15].

The entire process chain of the fusion algorithm consists of many interconnected steps. The raw sensor data from various vehicles and infrastructure devices must be accurately aligned and abstracted. The context-aware features extracted from the aforementioned architecture can provide information about the environment and operations at any time. This helps improve the system's interpretability. Multimodal data streams reach the fusion core through confidence metrics and semantic descriptors, and are randomly weighted and collaboratively integrated. It provides a unified, high-confidence environmental view for the vehicle's planning and perception modules to make decisions. Figure 2 shows the method of continuously collecting algorithm results and environmental change feedback through a loop for adjustment. The adaptive feedback loop can dynamically adjust fusion weights, confidence thresholds, and feature selection strategies to maintain the system's stable situational awareness and responsiveness when external environments or internal conditions change.

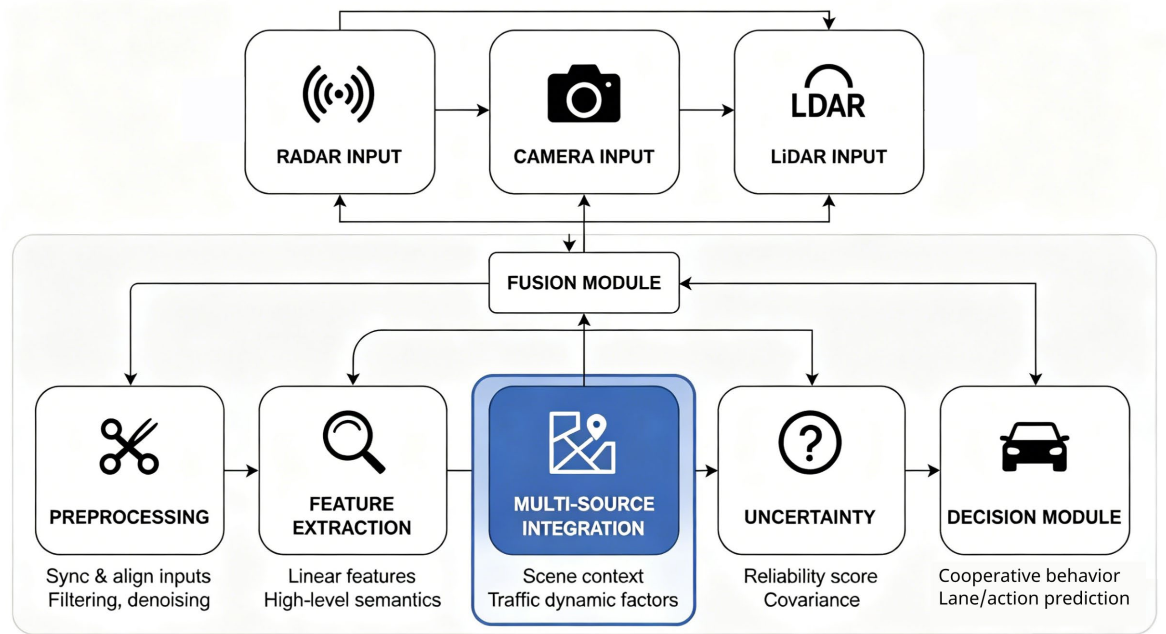

Figure 2. Fusion Algorithm Workflow Diagram.

In the fusion core, each detected entity, observation, and meta-feature is assigned a weight based on the sensor's reliability, recent failure rate, environmental conditions, and network health status. For multi-source weighted fusion, the fusion rules are as follows:

$$\tilde{\mathbf{z}} = \gamma \sum_{i=1}^{n} w_i \mathbf{z}_i + \theta(\hat{\eta}, \mathbf{H}_{env}, \mathcal{L}_{\text{stat}}) \quad \text{Eq.(5)}$$

with w_i the adaptive weights per sensor/source, $\mathbf{z}_i$ the feature vector, and $\theta(\,)$ a non-linear adjustment based on confidence estimator $\hat{\eta}$, environmental context $\mathbf{H}_{\text{env}}$, and statistical loss $\mathcal{L}_{\text{stat}}$.

In order to maintain the confidence level throughout the entire data processing chain, formalize the propagation of uncertainty:

$$\mathcal{C}_{\text{out}} = \phi\big(\{\mathcal{V}_i \cdot \sigma_{\mathbf{z}_i}^2\}, \Sigma_{ctx}, \Delta_{\text{async}}\big) \quad \text{Eq.(6)}$$

where $\mathcal{C}_{\text{out}}$ is output confidence, $\mathcal{V}_i$ the input variance, Σ_{ctx} summarizing contextual uncertainty, and Δ_{async} the time offset correction.

Contextual features are first extracted, and then they are integrated into an abstract state vector for subsequent decision-making modules to use.

$$\mathbf{X}_{ctx} = \mathcal{E}_{\text{cont}}\left(\bigcup_{j=1}^{m} \text{Enc}_j\big(\mathbf{x}_j; \pi_j, \mathcal{S}_{\text{type}}\big)\right) \quad \text{Eq.(7)}$$

where Enc_j is the encoder for the j-th context source, π_j the task-dependent projection matrix, and $\mathcal{S}_{\text{type}}$ semantic group tags, with $\mathcal{E}_{\text{cont}}$ as the context feature aggregator.

For collaborative intelligent transportation systems, the closed-loop, context-based fusion process dynamically adjusts to adapt to changes in the operational environment while maintaining high-precision perception and real-time performance.

Experimental Setup and Theoretical Validation

Experimental Platform and Scenario Modeling

A modular testing platform has been established, equipped with extended sensor modules and various physical vehicles and precise virtual agents, for all-weather testing of a universal sensor fusion architecture. The hardware infrastructure will be equipped with full-scale autonomous vehicles and robotic platforms equipped with inertial measurement units, millimeter-wave radar, high-resolution cameras, and synchronized LiDAR arrays, and they will be deployed in multiple geographical test areas. Real-time vehicle data is transmitted through edge computing gateways to the cloud computing cluster at the core fusion and orchestration layer. An independent network simulator will be used to address issues such as random delays, packet loss, and bandwidth fluctuations.

Scenario modeling is a physics-based simulation software environment used for detailed highway and urban scenarios. Connected to the hardware via a ring interface. Test cases include different environments and operating conditions, such as day-night changes, different weather, varying traffic densities, and sensor occlusions. By formalizing the scenario probability space, we capture the multidimensional set of relevant states:

$$\mathcal{P}_{\text{task}} = \int_{\substack{S_{\text{env}} \\ S_{\text{comm}}}} f_{\text{env}}(x) f_{\text{agent}}(y) f_{\text{comm}}(z)\, dx dy dz \quad \text{Eq.(8)}$$

where S_{env} describes the ambient environment, S_{agent} the agent configuration (vehicle type, sensor suite, dynamic state), and S_{comm} the communication context, with f_{env}, f_{agent} and f_{comm} denoting the corresponding probability density functions.

In order to ensure the accuracy of the parameterization of internal and external sensors, all sensors were checked before use. In order to ensure the accuracy of the sensors, a system-level calibration procedure has also been established. The sensor model includes resolution, field of view, update rate, and noise covariance.

$$\Sigma_{\text{sens}}(t) = \Psi[\lambda_{\text{res}}, \theta_{\text{fov}}, \omega_{\text{rate}}(t), \delta_{\text{no}}] \quad \text{Eq.(9)}$$

Here, λ_{res} quantifies spatial/temporal resolution, θ_{fov} the field of view, $\omega_{\text{rate}}(t)$ the temporal sampling rate (possibly time-varying), and δ_{no} summarizes the composite noise envelope.

Evaluation Metrics, Labeling and Procedures

The accuracy of the system, the accuracy of semantic segmentation, and the spatiotemporal consistency will be empirically validated using a high-quality multi-level evaluation and annotation system. All the ground truth labels were created by human experts and a semi-automated annotation framework. The framework is synchronized between vehicles and regularly evaluates the consistency among reviewers. In order to ensure a fair comparison, the results of all algorithms have been mapped to a unified spatiotemporal semantic reference grid, which complies with the global standard performance metric system.

The following are the formulas for the basic detection metrics of recall, precision, and F-measure:

$$\mathcal{M}_{\text{det}} = 2 \times \frac{\alpha_P \beta_R}{\alpha_P + \beta_R} \text{ where } \alpha_P = \frac{TP}{TP + FP}, \beta_R = \frac{TP}{TP + FN} \quad \text{Eq.(10)}$$

Where TP, FP and FN indicate the true positive, false positive, and false negative detections, respectively, aggregated over multi-vehicle, multi-frame sequences.

Strictly ensure the validity of data during the acquisition and fusion process to guaranty the accuracy of input signals and algorithm results. Formal validity assessment is based on cumulative error and temporal consistency constraints:

$$\mathcal{V}_{\text{check}}(t) = \mathcal{G}\left(\left|\Delta_{\text{sync}}(t)\right| < \epsilon_{\text{clock}}, \left|\delta_{\text{err}}(t)\right| < \varepsilon_{\text{max}}\right) \quad \text{Eq.(11)}$$

where $\Delta_{\text{sync}}(t)$ is the inter-sensor synchronization offset, $\delta_{\text{err}}(t)$ denotes instantaneous measurement deviation, and ϵ_{clock} , ε_{max} are strict operational thresholds for timestamp and amplitude errors, respectively.

In order to address various operational edge cases, the experimental workflow systematically includes random scene generation, multi-agent labeling, cross-validation folds, and multi-site deployment. The error intervals of all observed values are calculated, and the reference real trajectory is used for time series matching and spatial alignment.

Fusion Algorithm Empirical Analysis

The fusion algorithm will be deployed in phases on the aforementioned platform, including online and offline. In the offline phase, recorded sensor streams are replayed for batch statistical analysis, while in the online phase, multimodal data is transmitted from numerous interconnected vehicles through random driving tasks and scripted streaming. The evaluation used aggregated detection scores and fusion results, with a particular focus on edge cases where some sensors drop out or there are sudden environmental changes.

A new effectiveness metric has been proposed to evaluate the progress of general fusion in terms of environmental knowledge compared to all single-sensor baselines. Weighted detection, uncertainty propagation, and error reduction are the three sub-models:

$$\mathcal{E}_{\text{fuse}} = \frac{\sum_{i=1}^{n} \varsigma_i \cdot \left[\Delta_{\text{acc},i} - \eta_{\text{unc},i}\right]}{\sum_{j=1}^{m} \chi_j} \quad \text{Eq.(12)}$$

where ς_i adjusts for scenario-specific saliency, $\Delta_{acc,i}$ is the accuracy differential for task $i, \eta_{unc,i}$ the associated uncertainty, and χ_j normalizes by resource intensity.

Although the delay is theoretically bounded and empirically validated, it is necessary for practical applications. Described as follows:

$$\Theta_{\text{lat}} = \max_t \left\{\tau_{\text{comm}}(t) + \tau_{\text{proc}}(t) + \xi_{\text{sched}}(t)\right\} \quad \text{Eq.(13)}$$

Here, τ_{comm} is the communication delay, τ_{proc} the local fusion processing time, and ξ_{sched} the system scheduling overhead at instant t.

A dynamic evaluation index that takes into account the impact of various adversarial and stochastic disruptions demonstrates robustness and stability:

$$\mathcal{R}_{\text{stab}} = \Phi\left(\int_0^T \kappa_{\text{succ}}(t)dt, \inf_t \mathcal{S}_{\text{degrad}}(t), \mathcal{J}_{\text{cross}}(T)\right) \quad \text{Eq.(14)}$$

with $\kappa_{\text{succ}}(t)$ the real-time fusion success rate, $\mathcal{S}_{\text{degrad}}(t)$ the minimum detected degradation, and $\mathcal{J}_{\text{cross}}(T)$ the cross-episode performance jump at the total observation interval T.

Under any circumstances, a validated universal sensor fusion system should have high-fidelity perception and low latency, and operate stably under multiple sensor failures or adverse weather conditions. The proposed architectural design and algorithmic strategies have high feasibility in building large-scale, reliable cooperative perception systems for connected vehicles.

Result Analysis and Discussion

System-Level Performance and Benchmarking

The evaluation of the general fusion architecture conducted in urban, highway, tunnel, rainy, and nighttime environments indicates that the system possesses extremely high stability and accuracy. As shown in Figure 3(a), under good weather conditions, the accuracy of nighttime object recognition still exceeds 92.5%. The accuracy rate for more complex environmental issues such as tunnels is 88.1%; the complexity of rainy days and nighttime causes the accuracy rates to drop to 83.7% and 86.2%, respectively. The stability of environmental changes indicates that the entire system is well-designed.

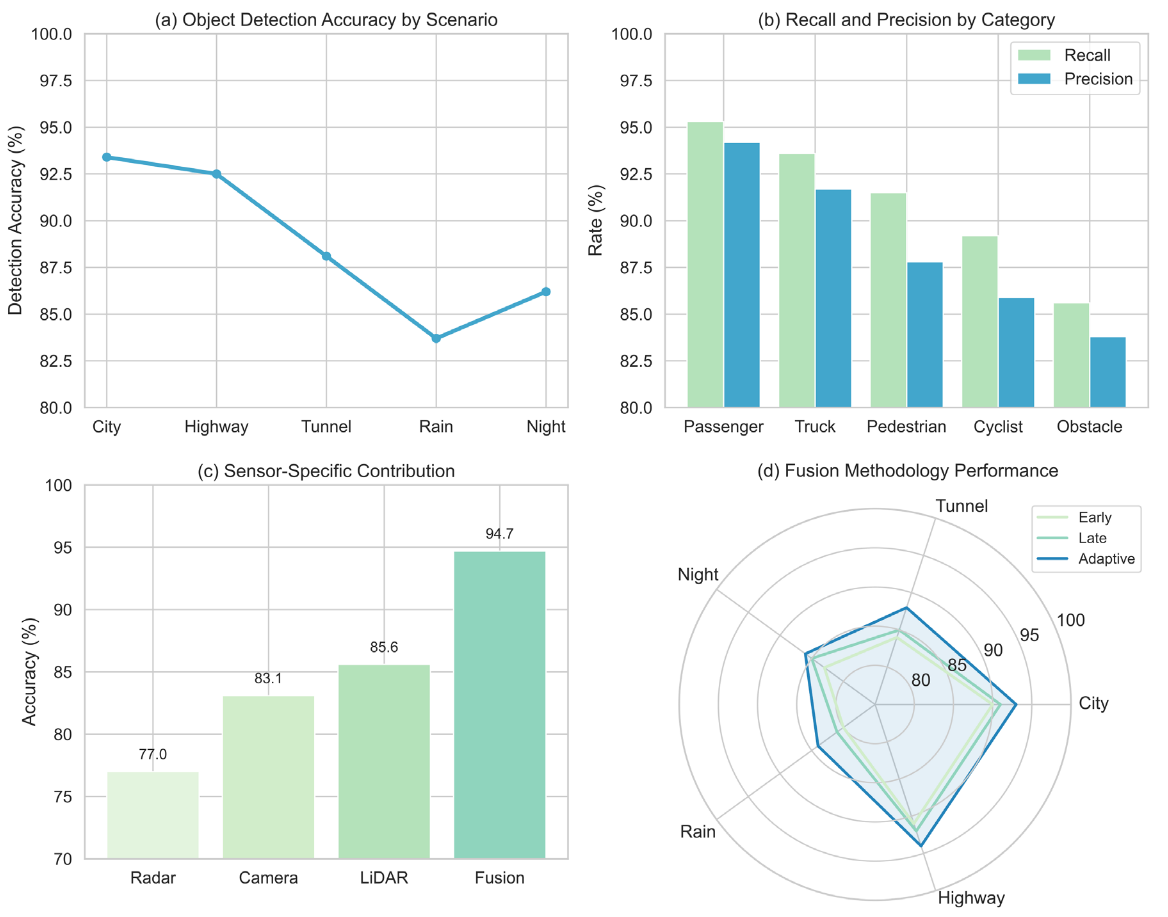

Figure 3. Overall Perception Performance: (a) Detection accuracy by scenario; (b) Recall and precision by class; (c) Sensor contribution to fusion; (d) Fusion methods comparison.

Figure 3(b) shows the recall and precision rates for all categories of vehicles. The recall rate for passenger cars is 95.3%, with a precision rate of 94.2%, followed by trucks. The recall rate for vulnerable road users (cyclists and pedestrians) is 89.2%, with precision slightly below 88%. The system is still capable of detecting and recognizing complex objects and supports multiple input sources [26].

https://doi.org/10.64972/jaat.2025v3.207p19e:252-266

Figure 3(c) shows the other parts of the sensor. LiDAR has spatial accuracy, with an accuracy rate of 85.6%. Detection based on cameras can more clearly distinguish between different types, but the accuracy drops to 83.1% due to reduced visibility. Radar has an accuracy of about 77% in dense environments and can accurately determine the distance to objects. Fusion is generally better than a single sensor; when using adaptive methods, the combined accuracy of all sensors is about 11% higher than that of the best single sensor system.

Figure 3(d) shows the comparison results of the fusion algorithms. Compared to early feature-level fusion and late decision-level fusion, the adaptive strategy adjusts sensor weights based on environmental changes to achieve the best detection accuracy among all methods. Context-based dynamic adaptation is more suitable for sudden changes or partial occlusions than fixed aggregation [27].

Further study the system's behavior under adverse conditions by combining and excluding baseline comparisons. As shown in Figure 4(a), when sensor noise increases, the accuracy of the general fusion model remains within 5% at high disturbance levels. The accuracy of non-fusion methods significantly decreases after moderate noise, dropping below 60% in the worst case. Figure 4(b) shows the robustness to missing data. In the case of a sensor dropout rate of up to 30%, the fusion method can still maintain a median detection rate of over 89.4%. In contrast, the fusion system performed poorly under the same loss conditions [28]. Figure 4(c) shows the statistical significance of the aforementioned advantages. The improvements are consistent, stable, and show significant increases across all test domains.

The above findings indicate that the general adaptive fusion method has higher practical accuracy and typical fault mode stability. Suitable for use in safety-critical applications with multimodal heterogeneous sensor systems [29].

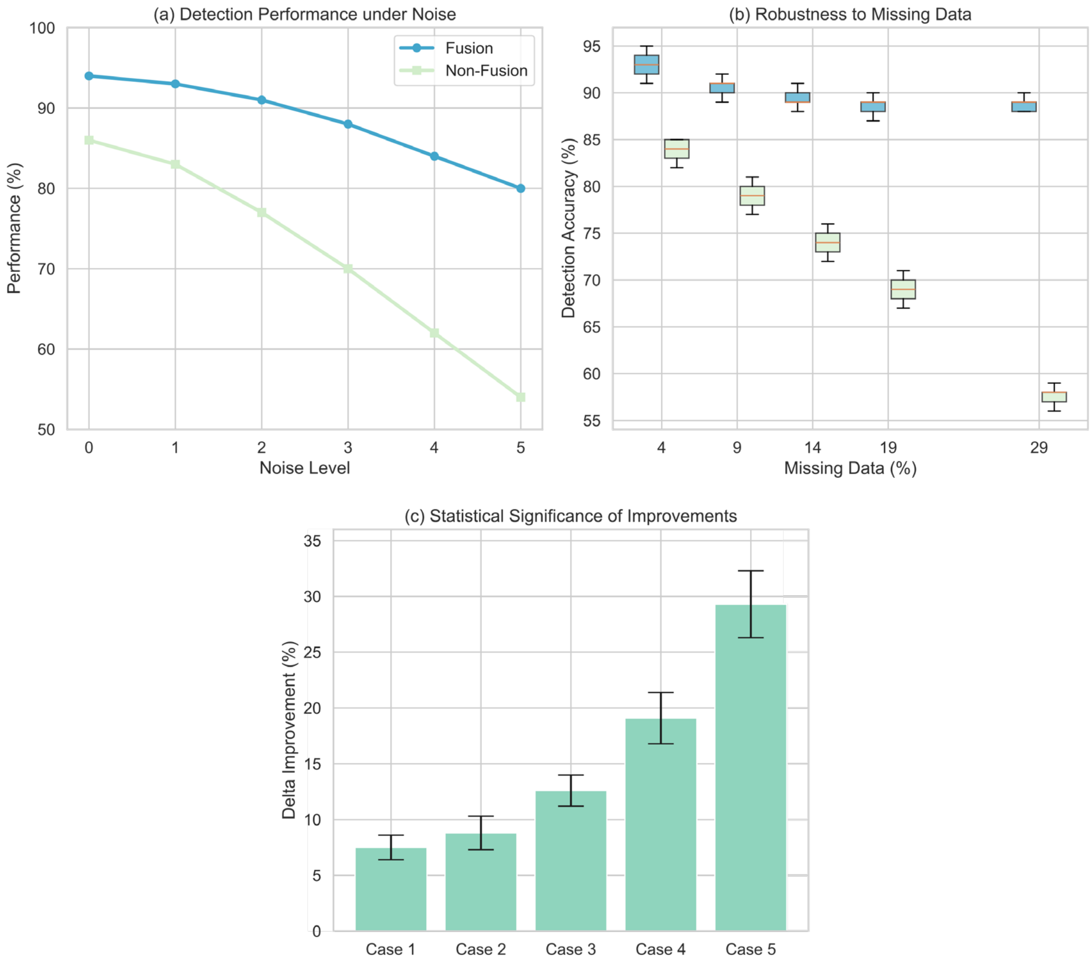

Figure 4. Fusion vs. Non-Fusion Baselines: (a) Performance under noise; (b) Robustness with missing data; (c) Statistical significance of improvement.

https://doi.org/10.64972/jaat.2025v3.207p19e:252-266

Robustness, Scalability, and Resource Efficiency

Figure 5(a) shows the robustness of the system under increasingly severe fault conditions. For example, even in the event of a sensor failure, the detection rate remains at 97.2%. When at least two sensors are offline, the durability exceeds 91%. When multiple primary sensors fail simultaneously, the durability is only 80%. The findings indicate that the quantitative safety margin provided by system-level redundancy [30].

Figure 5(b) depicts the real-time response capability under work pressure. The end-to-end system latency increased from the baseline of 63 milliseconds to 85 milliseconds, with the maximum latency reaching 132 milliseconds when vehicle and network loads doubled, and remaining at 132 milliseconds when the baseline traffic tripled. Although the perception accuracy slightly decreased, it can still be demonstrated that high-load bottlenecks can be overcome through good queue management and edge computing technologies [31].

Figure 5(c) shows the effectiveness of anomaly and outlier detection. The true positive rate of anomaly detection exceeds 88%, and the false positive rate is below 4%; by introducing sensor fusion and cross-validation, even in real-life scenarios with increased complexity, false triggers can be reduced.

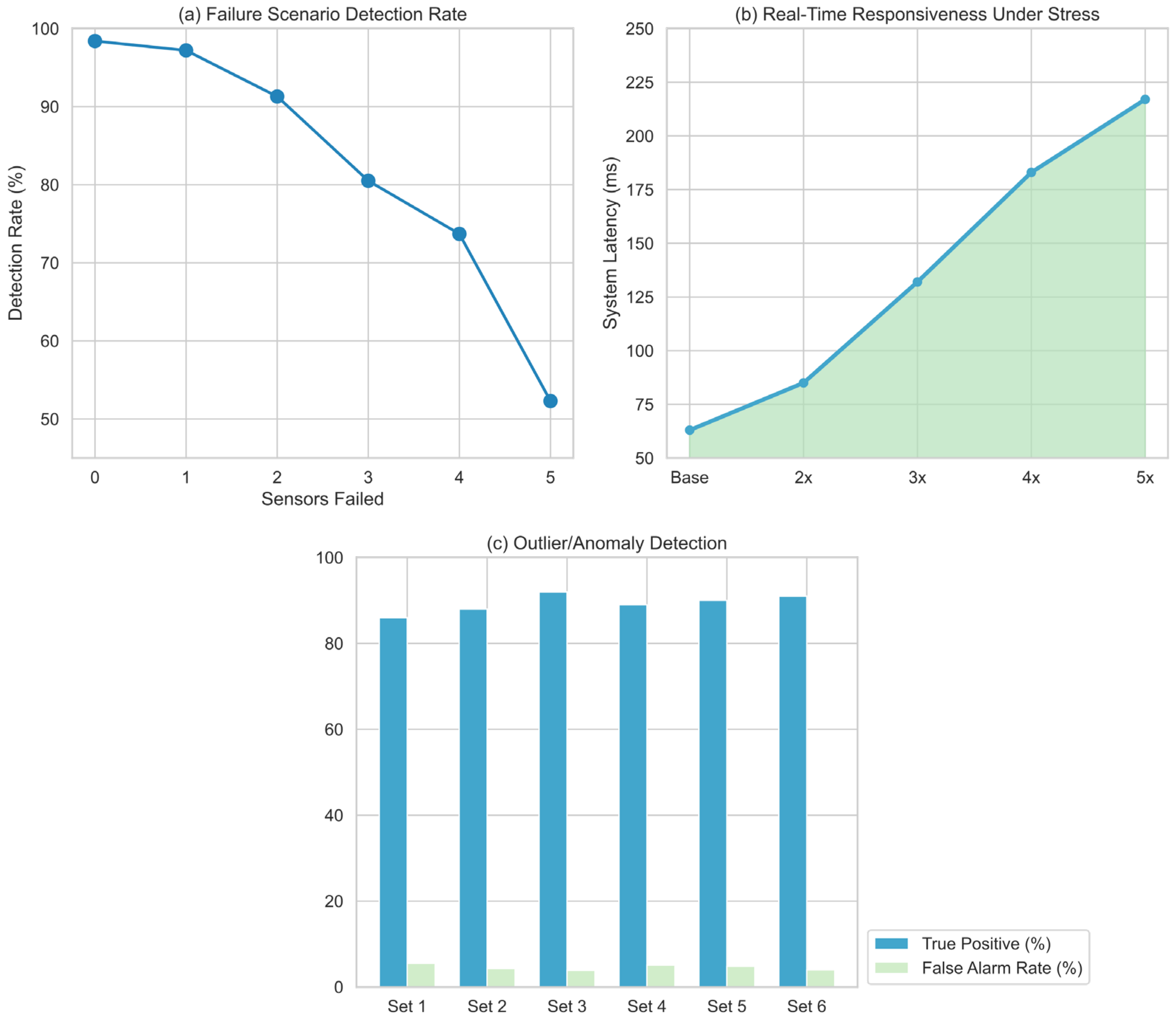

Figure 5. Robustness and Fault Tolerance: (a) Detection rate with sensor faults; (b) System latency under stress; (c) Anomaly detection performance.

Figure 6 shows scalability and system performance. The following are the quantitative data for each operational dimension: As shown in Figure 6(a), in the fields of perception and artificial intelligence, resource allocation is mainly concentrated in the fields of perception and artificial intelligence. Among them, AI accounts for 40% of the total GPU usage, while perception accounts for 24%; at the same time, the CPU load for AI (31%) and network (22%) parts is relatively high. The RAM of the AI module is relatively small, with a maximum of 12 GB. Since the

https://doi.org/10.64972/jaat.2025v3.207p19e:252-266

proportion of the three resources used by the backend and planning departments across all areas is only 10%, the demand for computing is not high [32].

Figure 6(b) depicts the characteristics of perception error. Sensor noise accounts for 25.8% of the total error, while object overlap (21.7%) and fusion delay (16.1%) are also relatively small. The remaining proportion is made up of data loss (13.5%), calibration drift (12.2%), and blurriness (10.7%). The total error of sensors and objects accounts for nearly half of all faults, making up 47.5% of all faults. High-quality sensors and rapid data collection are needed.

The computation resource utilisation for vehicle connectivity is nonlinearly proportional to this, as Figure 6(c) illustrates. Both CPU and GPU use are under 60, and the group size is 32 or fewer. The CPU consumption will be around 83 units, the GPU utilisation will reach 127 units, and the RAM requirements will climb from 4GB to 29GB when the number of vehicles is increased to 128. This suggests that resource demand is rapidly increasing beyond two dozen vehicles, particularly as fusion frequency and scheduling get more intense [33].

Figure 6(d) shows the different bandwidth shapes used under six different operating conditions. Early fusion requires the maximum bandwidth, reaching 95 Mbps under "peak" conditions; adaptive fusion only requires 68 Mbps, while the baseline only requires 54 Mbps. Adaptive fusion keeps bandwidth consumption at 70 Mbps or lower in any environment, even reaching 40 Mbps under "highway" conditions, reducing the maximum peak load by 61% compared to static solutions. As the scale and complexity of vehicular networks increase, the advantages of context-aware adaptive data transmission become increasingly apparent.

The aforementioned multidimensional results indicate that structural flexibility, resource allocation, and system robustness are interdependent. The main obstacle is that many sensors and communication paths simultaneously encounter significant issues or failures. Redundancy, scheduling, and adaptive management can mitigate these risks to maintain the good performance of urban and fleet applications in the real world [34].

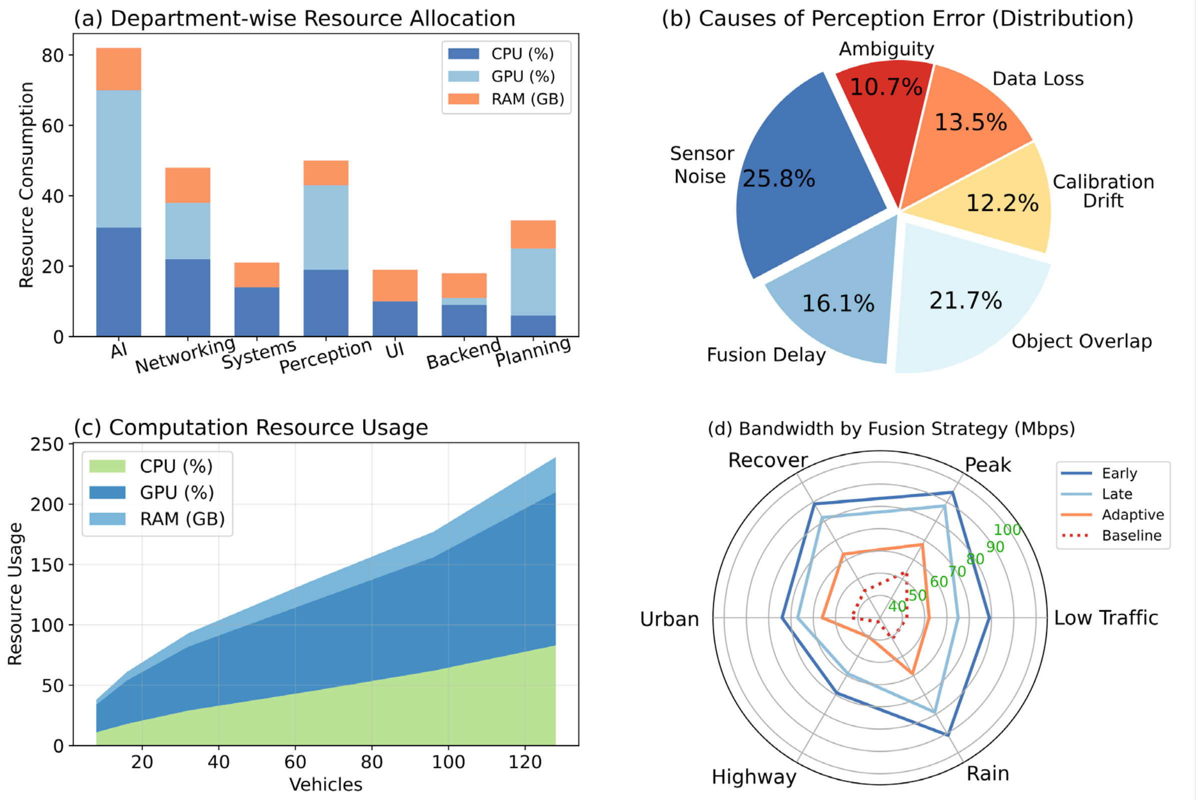

Figure 6. Scalability and Load Analysis: (a) Accuracy vs. vehicle number; (b) Throughput vs. fusion frequency; (c) Resource usage scaling; (d) Bandwidth consumption of fusion strategies.

https://doi.org/10.64972/jaat.2025v3.207p19e:252-266

In-depth Discussion and Real-world Implications

As shown in Figure 7(a), the end-to-end perception pipeline delay for each stage is as follows. During peak traffic periods, the average delay for the perception phase is 14 milliseconds, while the average delay for baseline communication is 21 milliseconds. The fluctuation range of fusion processing varies from an average of 17 milliseconds to a maximum of 28 milliseconds, exceeding 38 milliseconds in some complex situations. The 16-millisecond high-level decision module, as well as the 11-millisecond and 13-millisecond post-processing and rendering modules. Quantitative results indicate that the communication stack and compute-intensive fusion algorithms cause delays. In order to meet security requirements and achieve low-latency operations, urban deployments require high-speed communication systems and parallelized fusion strategies.

As shown in Figure 7(b), delays exceeding 140 milliseconds occasionally occur during certain queue overflow or congestion fusion phases. It is believed that modular pipeline scaling and smarter cross-layer collaboration will further optimize performance in the future, as dynamic scheduling based on sensitivity analysis and targeted data pruning are expected to reduce bottlenecks by over 25% [35].

Figure 7(c) depicts the actual operation of the logistics and fleet pilot project. Due to the average detection accuracy of urban vehicle groups exceeding 90% and decision delays being less than 112 milliseconds, navigation in urban areas is normal. In high-density dynamic loading and unloading areas, mixed logistics achieved a reliability of 93.3%. Insufficient calibration accuracy is a persistent operational issue; when the network or sensor health declines, there is a lack of infrastructure cooperation and backup mechanisms.

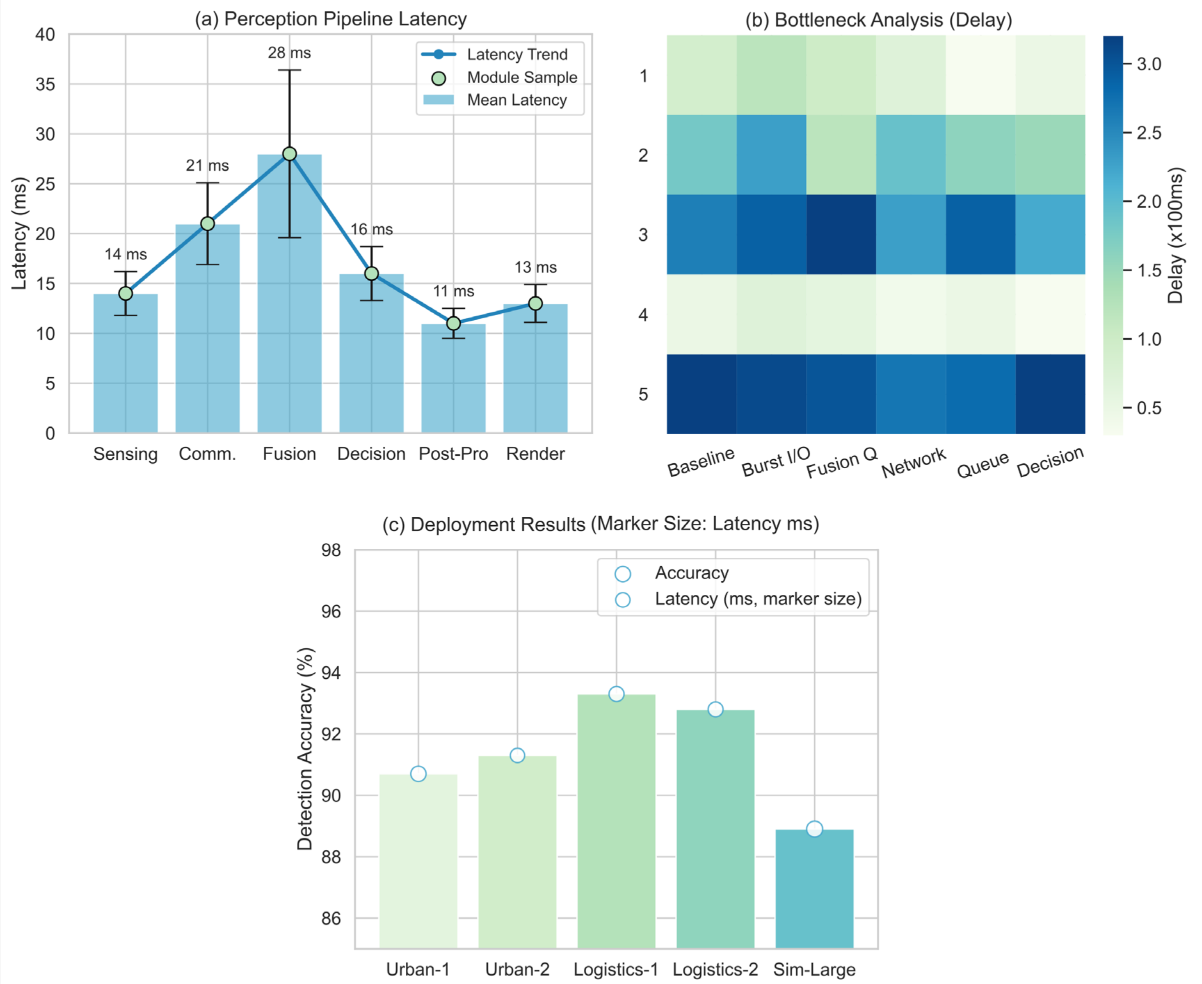

Figure 7. Latency and Resource Usage: (a) Module-wise perception latency; (b) Bottleneck delay analysis; (c) Real-world deployment results.

The universal fusion system can achieve reliability and robustness in network-scale perception within modern urban traffic and logistics environments. More suitable for practical applications, because it is a multi-layered

https://doi.org/10.64972/jaat.2025v3.207p19e:252-266

structure that can adapt to changing weather and heavy traffic. Next, we will discuss the aspects of optimization. These include deeper parallel processing, better context-aware resource management, and stronger adaptability to changing traffic and fault patterns. The aforementioned experiments laid the foundation for applying large-scale collaborative perception systems in intelligent mobility platforms.

Conclusion

This paper proposes a fusion architecture based on general sensors to address the key issues of collaborative perception in intelligent connected vehicles. The multimodal perception data from different sensors can be standardized through a protocol-independent modular structure. Then, the system uses abstract and adaptive distributed fusion to achieve integration. The aforementioned method has been tested in urban areas, highways, and adverse weather conditions. According to quantitative analysis, the new system has achieved over 92% excellent detection accuracy. Even in the case of network fluctuations, sensor dropouts, or changes in road conditions, it can still maintain stable perception. The core of the adaptive algorithm can dynamically adjust the sensor fusion and weighting strategies in real-time to ensure high perception accuracy, even as operational complexity increases. Resource allocation analysis indicates that the integration of edge-cloud collaboration and context-aware data management can maintain a low-latency system, and the end-to-end latency remains within a safety-critical range, even during high traffic or heavy load periods.

In addition to the aforementioned advantages, this architecture offers higher modularity and interoperability in the deployment of large-scale multi-brand fleets. Using a standard abstraction layer allows for the rapid integration of new sensors and vehicle types, and the closed-loop feedback mechanism continuously adjusts the fusion parameters based on algorithmic and environmental changes. Compared to previous methods, experiments showed better anomaly detection, fewer perception failures, and fewer bottleneck issues. The aforementioned progress indicates that this architecture has promising prospects in providing various services (such as object detection and path planning) within intelligent transportation systems, laying the foundation for the next generation of collaborative automated travel. These services can support resources and can be flexibly adjusted.

There are still some issues that need further research. Enhance system scalability and robustness to address network interruptions, sensor failures, and calibration drift in large-scale and infrastructure-poor environments. To reduce latency and improve throughput, the next two steps will be to enhance the parallelism of the fusion algorithms and optimize dynamic resource allocation. Comprehensive field tests will be conducted to provide clear network security guaranties and open interoperability standards. Transitioning from experimental platforms to the architecture proposed for large-scale application in the ever-evolving connected vehicle ecosystem.

Author Contributions

Franciszek Adrian Błaszczyk contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Artur Majewski and Celina Kubiak contribute to data collection, draft preparation, manuscript editing. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Alaba, S. Y., Gurbuz, A. C., & Ball, J. E. (2024). Emerging trends in autonomous vehicle perception: Multimodal fusion for 3D object detection. World Electric Vehicle Journal, 15(1), 20. https://doi.org/10.3390/wevj15010020

[2] Bejarbaneh, E. Y., Du, H., & Naghdy, F. (2024). Exploring shared perception and control in cooperative vehicle-intersection systems: A review. IEEE transactions on intelligent transportation systems, 25(11), 15247-15272. https://doi.org/10.1109/TITS.2024.3432634

[3] Xu, H., & Liu, X. (2022). Perception synergy optimization with deep reinforcement learning for cooperative perception in C-V2V scenarios. Vehicular Communications, 38, 100536. https://doi.org/10.1016/j.vehcom.2022.100536

[4] Sumalatha, I. P. P. A., Chaturvedi, P., Patil, S., Thethi, H. P., & Hameed, A. A. (2024, May). Autonomous multi-sensor fusion techniques for environmental perception in self-driving vehicles. In 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE) (pp. 1146-1151). IEEE. https://doi.org/10.1109/IC3SE62002.2024.10593125

[5] Zhou, J., & Lei, Y. (2023, January). Multi-source heterogeneous data fusion algorithm based on federated learning. In International Conference on Soft Computing in Data Science (pp. 46-60). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-0405-1_4

[6] Liu, Q., Sun, J., Zhang, Y., & Liu, X. (2024). DenMerD: a feature enhanced approach to radar beam blockage correction with edge-cloud computing. Journal of Cloud Computing, 13(1), 32. https://doi.org/10.1186/s13677-024-00607-x

[7] Liu, J., Wang, H., Peng, L., Cao, Z., Yang, D., & Li, J. (2022). PNNUAD: Perception neural networks uncertainty aware decision-making for autonomous vehicle. IEEE Transactions on Intelligent Transportation Systems, 23(12), 24355-24368. https://doi.org/10.1109/TITS.2022.3197602

[8] Liu, J., Gao, B., Zhong, W., Lu, Y., & Han, S. (2024). Adaptive optimization strategy and evaluation of vehicle-road collaborative perception algorithm in real-time settings. Computers and Electrical Engineering, 120, 109785. https://doi.org/10.1016/j.compeleceng.2024.109785

[9] Zhang, Y., Tu, C., Gao, K., & Wang, L. (2024). Multisensor information fusion: Future of environmental perception in intelligent vehicles. Journal of intelligent and connected vehicles, 7(3), 163-176. https://doi.org/10.26599/JICV.2023.9210049

[10] Butt, F. A., Chattha, J. N., Ahmad, J., Zia, M. U., Rizwan, M., & Naqvi, I. H. (2022). On the integration of enabling wireless technologies and sensor fusion for next-generation connected and autonomous vehicles. IEEE access, 10, 14643-14668. https://doi.org/10.1109/ACCESS.2022.3145972

[11] Kashinath, S. A., Mostafa, S. A., Mustapha, A., Mahdin, H., Lim, D., Mahmoud, M. A., ... & Yang, T. J. (2021). Review of data fusion methods for real-time and multi-sensor traffic flow analysis. IEEE Access, 9, 51258-51276. https://doi.org/10.1109/ACCESS.2021.3069770

[12] Zhang, Q., Sun, H., Gao, X., Wang, X., & Feng, Z. (2022). Time-division ISAC enabled connected automated vehicles cooperation algorithm design and performance evaluation. IEEE Journal on Selected Areas in Communications, 40(7), 2206-2218. https://doi.org/10.1109/JSAC.2022.3155506

[13] Li, L., Zhang, W., Wang, X., Cui, T., & Sun, C. (2024). NLOS dies twice: Challenges and solutions of V2X for cooperative perception. IEEE Open Journal of Intelligent Transportation Systems, 5, 774-782. https://doi.org/10.1109/OJITS.2024.3492211

[14] Wen, J., Zhang, J., Zhang, Z., Cui, Z., Cai, X., & Chen, J. (2024). Resource-aware multi-criteria vehicle participation for federated learning in Internet of vehicles. Information Sciences, 664, 120344. https://doi.org/10.1016/j.ins.2024.120344

[15] Dao, M. Q., Berrio, J. S., Frémont, V., Shan, M., Héry, E., & Worrall, S. (2024). Practical collaborative perception: A framework for asynchronous and multi-agent 3d object detection. IEEE Transactions on Intelligent Transportation Systems, 25(9), 12163-12175. https://doi.org/10.1109/TITS.2024.3371177

[16] Yeong, D. J., Velasco-Hernandez, G., Barry, J., & Walsh, J. (2021). Sensor and sensor fusion technology in autonomous vehicles: A review. Sensors, 21(6), 2140. https://doi.org/10.3390/s21062140

[17] Hu, S., Fang, Z., An, H., Xu, G., Zhou, Y., Chen, X., & Fang, Y. (2024, December). Adaptive communications in collaborative perception with domain alignment for autonomous driving. In GLOBECOM 2024-2024 IEEE Global Communications Conference (pp. 746-751). IEEE. https://doi.org/10.1109/GLOBECOM52923.2024.10901656

[18] Marsh, B., Sadka, A. H., & Bahai, H. (2022). A critical review of deep learning-based multi-sensor fusion techniques. Sensors, 22(23), 9364. https://doi.org/10.3390/s22239364

[19] Biswas, A., & Wang, H. C. (2023). Autonomous vehicles enabled by the integration of IoT, edge intelligence, 5G, and blockchain. Sensors, 23(4), 1963. https://doi.org/10.3390/s23041963

[20] Thornton, S., & Dey, S. (2024). Multi-modal data and model reduction for enabling edge fusion in connected vehicle environments. IEEE Transactions on Vehicular Technology, 73(8), 11979-11994. https://doi.org/10.1109/TVT.2024.3378182

[21] Mahima, K. Y., Perera, A. G., Anavatti, S., & Garratt, M. (2024). Toward robust 3d perception for autonomous vehicles: A review of adversarial attacks and countermeasures. IEEE Transactions on Intelligent Transportation Systems, 25(12), 19176-19202. https://doi.org/10.1109/TITS.2024.3456293

[22] Dong, X., Tian, W., Ye, X., Xu, Y., Wu, T., & Wang, Z. (2024). A federated cloud-based auction mechanism for real-time scheduling of vehicle sensors in vehicle-road-cloud collaborative system. IEICE Transactions on Communications, 108(1), 14-23. https://doi.org/10.23919/transcom.2024EBP3010

[23] Zhang, J., Yang, K., Wang, Y., Wang, H., Sun, P., & Song, L. (2024). Ermvp: Communication-efficient and collaboration-robust multi-vehicle perception in challenging environments. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12575-12584). https://doi.org/10.1109/TITS.2024.3436012

[24] Agbaje, P., Anjum, A., Mitra, A., Oseghale, E., Bloom, G., & Olufowobi, H. (2022). Survey of interoperability challenges in the internet of vehicles. IEEE Transactions on Intelligent Transportation Systems, 23(12), 22838-22861. https://doi.org/10.1109/TITS.2022.3194413

[25] Anwar, A., Halabi, T., & Zulkernine, M. (2022). A coalitional security game against data integrity attacks in autonomous vehicle networks. Vehicular Communications, 37, 100517. https://doi.org/10.1016/j.vehcom.2022.100517

[26] Chen, G., & Qin, H. (2022). Class-discriminative focal loss for extreme imbalanced multiclass object detection towards autonomous driving. The Visual Computer, 38(3), 1051-1063. https://doi.org/10.1007/s00371-021-02067-9

[27] Liu, J., Luo, D., Fu, X., Lu, Q., & Kang, K. Y. (2022). Design Strategy of Multimodal Perception System for Smart Environment. In Internet of Things for Smart Environments (pp. 93-115). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-09729-4_6

[28] Gupta, S., & Snigdh, I. (2022). Multi-sensor fusion in autonomous heavy vehicles. In Autonomous and Connected Heavy Vehicle Technology (pp. 375-389). Academic Press. https://doi.org/10.1016/B978-0-323-90592-3.00021-5

[29] Fan, Z., & Zhao, L. (2024). Autonomous integrity of multisource information resilient fusion navigation system: State-of-the-art and open challenges. IEEE Transactions on Instrumentation and Measurement, 73, 1-17. https://doi.org/10.1109/TIM.2024.3427865

[30] Foshati, A., & Ejlali, A. (2024). Enhancing sensor fault tolerance in automotive systems with cost-effective cyber redundancy. IEEE Transactions on Intelligent Vehicles, 9(4), 4794-4803. https://doi.org/10.1109/TIV.2024.3379928

[31] Zhang, Q., Zhang, X., Zhu, R., Bai, F., Naserian, M., & Mao, Z. M. (2023, October). Robust real-time multi-vehicle collaboration on asynchronous sensors. In Proceedings of the 29th Annual International Conference on Mobile Computing and Networking (pp. 1-15). https://doi.org/10.1145/3570361.3613271

[32] Ku, Y. J., Baidya, S., & Dey, S. (2023). Uncertainty-aware task offloading for multi-vehicle perception fusion over vehicular edge computing. IEEE Transactions on Vehicular Technology, 72(11), 14906-14923. https://doi.org/10.1109/TVT.2023.3284369

[33] Guo, K., Li, X., Zhang, M., Bao, Q., & Yang, M. (2021). Real-time vehicle object detection method based on multi-scale feature fusion. IEEE Access, 9, 115126-115134. https://doi.org/10.1109/ACCESS.2021.3104849

[34] Xie, L., Song, S., Eldar, Y. C., & Letaief, K. B. (2023). Collaborative sensing in perceptive mobile networks: Opportunities and challenges. IEEE wireless communications, 30(1), 16-23. https://doi.org/10.1109/MWC.005.2200214

[35] Ku, Y. J., Flowers, B., Thornton, S., Baidya, S., & Dey, S. (2022, June). Adaptive C-V2X sidelink communications for vehicular applications beyond safety messages. In 2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring) (pp. 1-6). IEEE. https://doi.org/10.1109/VTC2022-Spring54318.2022.9860580