

Edge AI-Driven Optimization for Multi-Sensor Data Processing in Intelligent Vehicles

Roman Urban Turek¹, Sebastian Wiktor Walewski¹ and Zenon Ziemba^{1,*}

¹ Faculty of Information Technology, University of Warmia and Mazury, Olsztyn, 10-719, Poland

*Corresponding author: zenon.z@uwm.edu.pl

Abstract. Many advanced edge-AI technologies have emerged with the development of intelligent vehicle systems, but the quantity and speed at which multi-sensor data must be handled have also increased significantly. A hierarchical fusion architecture for integrating data from cameras, LiDAR, radar, and inertial sensors under resource awareness and delay correction is presented in this research. The suggested solution uses dynamic workload distribution, lightweight neural networks, and synchronous hardware to enable high-efficiency perception on embedded vehicle platforms. A framework with an average precision of 78.2% for object detection has been developed based on the results of over 140,000 multi-modal annotated frames from urban and challenging environments; at the same time, the average inference latency has been maintained at 38.7ms per frame and the overall power consumption at roughly 23.4W. Additionally, the method can function in bad weather or when a sensor fails, and it is reasonably resistant to false positives. The capacity to generalise is further confirmed by cross-domain deployment; on public datasets, the detection mAP and trajectory RMSE are 74.9% and 0.85m, respectively, without fine-tuning. Based on the aforementioned findings, we think that improving the dependability and environmental adaptability of intelligent cars will need developing high-throughput, real-time, and scalable fusion systems at the edge.

Keywords: *Edge Computing, Multi-Sensor Fusion, Autonomous Vehicles, Real-Time Perception*

Received on 15 October 2024, Accepted on 08 April 2025, Published on 17 April 2025

Copyright © 2025 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

As intelligent vehicles have evolved, these systems have progressively been constructed around a variety of sensors, including cameras, LiDAR, radar, and other ultrasonics, to gather copious amounts of data for comprehending the surrounding environment [1]. In order to accomplish precise perception, location, and control in a manner that surpasses the capabilities of human drivers, these diversified and high-resolution sensing modules are collecting a vast amount of different data [2]. These sensors' combined data will offer a solid basis for a variety of applications, including path planning, semantic scene analysis, and real-time obstacle avoidance [3]. However, the demand for data rate and processing capacity of intelligent vehicle platforms has increased exponentially as the volume and complexity of sensor installations continue to grow, and traditional data fusion techniques are becoming inadequate [4]. The basic function of the new generation of automobile intelligence, sensor fusion, needs to be strengthened with more accurate, timely, and contextually appropriate information processing because the road and traffic environment is always changing [5]. A major technical problem nowadays is striking a balance between data richness and a workable system for both academic research and industrial application [6].

In order to lessen reliance on a single cloud server, edge artificial intelligence (edge AI) has recently started to be used to process data more locally at the edge of intelligent vehicle systems [7]. In the automotive industry, communication delays and the need for bandwidth in real-time analytics under strict latency limitations are

greatly decreased by deploying edge AI and bringing computational intelligence closer to the data source [8]. Lightweight neural networks, hardware accelerators, and edge-native software frameworks that enable sophisticated perception-and-decision algorithms in resource-constrained settings have all contributed to the aforementioned advancements [9]. In order to enable distributed, cooperative computing across local vehicle nodes and edge gateways for maximum efficiency and resilience, the "cloud-edge-device" model has gained popularity [10]. While the aforementioned changes have greatly improved the system's speed and privacy, new challenges at the system level have also surfaced, including model deployment, data consistency, and secure information sharing between the many compute nodes [11]. The entire impact of scaled, safe, and adaptable vehicle intelligence will still require design and optimisation of such edge-driven architectures [12]. In recent years, edge AI's ability to fuse sophisticated, diverse, and real-time sensors has garnered a lot of interest and is anticipated to be crucial to the development of intelligent transportation systems in the future [13]. Even though there have been some recent successes, there are still a lot of issues with workload segmentation, synchronising disparate data kinds, and sustaining performance under different conditions [14]. Additionally, comprehensive dependability and durability to hostile or damaged situations in real life have become crucial project criteria with the shift to actual use [15].

This study leverages the unique capabilities of edge AI and satisfies the technological and practical requirements for extended uses of multi-sensor data in intelligent cars. The primary goal is to research and create edge AI-based techniques for high-efficiency optimisation of real-time sensor data processing in view of the resource limitations and operational circumstances of contemporary automobiles. This study will concentrate on improving the stability and usefulness of the new-generation intelligent vehicle system in light of the requirement for a scalable and flexible real-time system.

Related Work

Recent Progress in Multi-Sensor Data Processing

In recent years, there have been many achievements in the engineering of multi-sensor data processing frameworks for intelligent vehicles. Accurate and stable acquisition of sensor data is required, and improvements have been made in the close alignment of time and space among cameras, LiDARs, radars and inertial measurement units through both hardware synchronization and software post-processing methods [16]. Robust multi-sensor synchronisation strategies have been introduced to achieve temporal alignment of data association for real-time perception and fusion [17]. Among the fusion methods, Kalman filters and their extensions, such as extended Kalman filters and unscented Kalman filters, are often used in conjunction with noisy measurement data from dynamic vehicle environments to perform localization, object tracking and other applications [18]. Bayesian inference can address the problems of uncertainty and variability in multi-sensor data in a well-grounded way, and is therefore versatile for high-level perception and object-level reasoning [19]. Deep learning has expanded the field of sensor fusion further, and complex neural network models, such as convolutional neural networks and recurrent neural networks, are now used to extract and jointly represent multi-modal features from images, point clouds and radar signals [20]. Most recently, a hybrid fusion method of traditional statistical estimators and deep models has shown better robustness and flexibility under difficult or incomplete sensing conditions [21]. Feature-level and decision-level fusion schemes are gradually being extended by attention mechanisms and graph-based models to perform selective and context-aware fusion of sensor evidence [22]. Accuracy improvements alone are insufficient; therefore, research has also focused on the feasibility of implementation in practice, aiming to develop lightweight models with high-dimensional data that are computationally feasible [23]. With the advancement of error correction, management of sensor redundancy and adaptive fusion have been achieved to ensure more stable operation in dense and irregular urban mobility environments [24]. The field will continue to develop all-encompassing systems that are reliable, accurate and efficient in terms of resources for sensor integration in large-scale intelligent vehicle networks [25].

Advances in Edge AI for Intelligent Vehicles

The automotive data processing paradigm has been transformed by edge artificial intelligence, which can now do real-time inference for intelligent vehicles at the network edge [26]. Complex neural networks for sensing, situation evaluation, motion planning, etc. may now be directly deployed on vehicles and in roadside

infrastructure using advanced edge processors and specialised AI accelerators [27]. Resource-constrained edge devices can now perform high-throughput, low-latency analysis of safety-critical data thanks to the development of compact model architectures and high-efficiency optimisation techniques [28]. By using deep networks with shared parameters for improved performance and computational efficiency, multi-task learning has now shown good results in the simultaneous processing of detection, classification, and prediction tasks [29]. In a car setting, edge AI avoids network congestion, privacy leakage, and high end-to-end latency by drastically lowering the need for cloud computing [30]. A cooperative edge-cloud approach moves complex problems to the cloud node only when necessary, using distribution and hierarchy to do preliminary inference at the edge. By storing local data in a vehicle and only exchanging model parameters or gradients for global intelligence aggregation, federated learning expands the use of privacy-preserving model training and adaptation. In order to allocate computing workloads in a system adaptively based on its current resource and network conditions, dynamic workload partitioning techniques have also been researched. When employing edge-based AI processing frameworks, early-stage deployment projects have demonstrated positive outcomes in real-world driving perception, response delay, and fail-safe operations. Despite the aforementioned successes, there are still a number of significant issues, including the requirement for flexible adaptability to all driving conditions in urban and highway contexts, a lack of device variety, disparate data standards, and security threats associated with remote computation.

Challenges and Limitations

The engineering and implementation of a robust multi-sensor data fusion based on edge AI for intelligent cars still faces many challenges, despite numerous advancements. Higher sensor densities have increased the volume, diversity, and speed of data streams that must be analysed in real time, and the growing processing requirement of new fusion algorithms is starting to limit the scalability of onboard systems. In a safety-critical setting, time lag, measurement inaccuracies, and sensor type variances make data association challenging, which lowers the dependability of the subsequent inference. A robust processing framework must be implemented because the fusion pipeline lacks continuous and comprehensive data due to sensor drift, sporadic signal loss, and environmental interference. Aggressive model compression, resource management, and adaptive offloading must be used to improve the trade-off between inference speed and accuracy because deployment in a real-world scenario is limited by factors like inadequate processing power, small amounts of memory and battery capacity, etc. Because edge devices are often located in distant or harsh environments, they are more susceptible to hacks, damage, and data leaks, which puts privacy and trust at greater risk. Since the models and data have diverse forms, it is difficult to apply them to a single intelligent system. It is now difficult to ensure the interoperability and flexible reconfiguration of edge, cloud, and in-vehicle nodes due to the high mobility and evolving network topology of cars. Domain adaptability, online learning, and ongoing model updates are necessary to handle changes in the operating environment and in the concept of a vehicle. Large-scale implementation has been hampered by the lack of clear recommendations on how to interpret the combined sensor data to guarantee safety and regulatory compliance. There is a pressing need for comprehensive, scalable, and secure frameworks that can handle the numerous issues in the study and application of intelligent vehicle systems.

Methodology: Edge AI Multi-Sensor Fusion Framework

System Design and Edge Deployment

The suggested edge AI multi-sensor fusion system's three-tier hierarchical structure is appropriate for satisfying the external needs of vehicle intelligence. Visible-light cameras, 64-line solid-state LiDARs, frequency-modulated continuous-wave (FMCW) radars, high-fidelity inertial measurement units, and other spatially optimised automotive-grade sensors at the perception level simultaneously gather multi-domain environmental and dynamic vehicle data from various directions of the chassis and roof plane. To guarantee sub-millisecond time synchronisation and account for the delay brought on by varying sample rates and unsynchronised data collection, precision sensor synchronisers work in tandem with local controllers. Pixel-level denoising, geometric rectification for camera-LiDAR conversion, digital clutter suppression in radar returns, and quaternion-based normalisation for inertial streams are all layers of pre-processing applied to raw sensor readouts. After that,

each of the resultant feature vectors is resampled and packetised with spatial matrices and time stamps appropriate for parallel processing downstream.

The vehicle's primary gateway and adjacent roadside infrastructure both have a cluster of microservers installed in an edge-computing layer. This layer has a collection of workload-aware scheduling modules that dynamically divide the fusion and inference operations according to shifting vehicle contexts, real-time system health, and network state indicators. High-speed dedicated vehicle buses (CAN-FD, Automotive Ethernet) and secure ultra-reliable wireless networks (C-V2X) are used for inter-node communication, which facilitates seamless federated inference and redundancy failover across edge nodes. At this level, simultaneous ingestion and batching of multi-modal features are handled by a high-performance multi-stream fusion gateway. Token-bucket algorithms and real-time protocol conversion guarantee strict temporal consistency and resource balance under high traffic or quick topological changes.

At the decision level, a hybrid decision engine that employs risk-optimized deep planning policies and rule-based heuristics receives fused scene graphs and semantic feature mappings via low-latency triggers from the edge computing cluster. In addition to supporting multi-agent intent anticipation, dynamic replanning in hostile settings, and real-time trajectory update, this module directly actuates the steering, brakes, and powertrain via a deterministic feedback loop. To lessen the impact of timing jitter and adversarial packet injection, consensus methods are employed to give control packets a high priority and an authenticated handshake sequence. The entire stack is naturally fault-tolerant to individual sensor failures and partial network outages because bidirectional feedback may be used to detect sensor issues early and coordinate the fleet's motion through a context-aware data flow mechanism.

The edge AI-based multi-sensor fusion architecture's entire hardware-software stack, including sensor arrays and signal synchronisers, dynamic feature encoding and edge fusion clusters, integrated decision modules, and actuation interfaces, is shown in Figure 1.

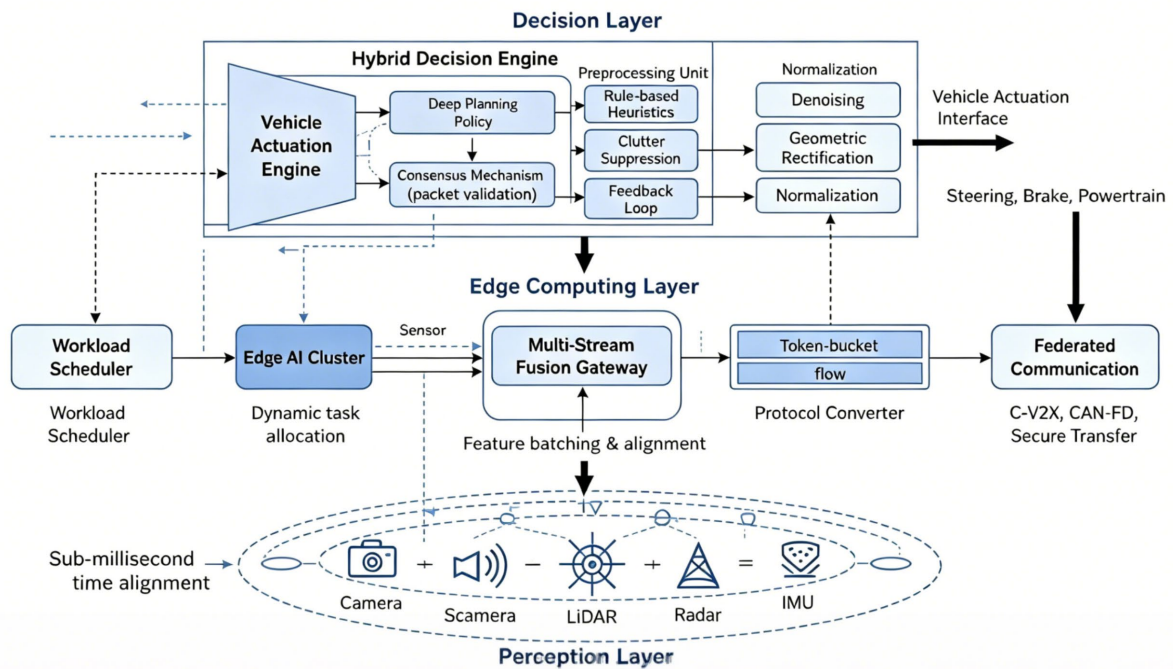


Figure 1. Edge AI-based Multi-Sensor Fusion System Architecture

Fusion Algorithm Development

With a latency of less than 50 ms, a multi-sensor fusion pipeline for edge AI processes and aggregates a substantial amount of diverse sensor data while taking into account both statistical and physical variations in urban traffic circumstances. A 79 GHz FMCW millimeter-wave radar grid provides reliable range-Doppler signatures, two 64-channel rotating LiDAR units record dynamic 3D geometry at a rate of 20 Hz, a MEMS IMU

sampled at 200 Hz records fine-grained ego-motion, and three 2048x1536 on-board RGB cameras are used for high-resolution semantic information.

Raw sensor feeds are first temporally reconciled. Each sensor's features are interpolated to a global timeline by a differentiable warping operator, where empirically measured sensor-specific delays (camera: 14 ms, LiDAR: 23 ms, radar: 12 ms) are accounted for:

$$\mathbf{f}_k(t) = \mathcal{W}_{\delta_k}(\mathbf{f}_k(t)) \quad \text{Eq.(1)}$$

Voxelization is applied to LiDAR points at 0.075 m granularity, producing arrays with up to 100,000 voxels per sweep. RGB images are passed through an EfficientNet-B3 backbone to yield tensor features of shape $64 \times 128 \times 192$. Radar reflectivity cubes ($128 \times 128 \times 16$) and inertial sequences are, respectively, subject to 2D convolution and gated recurrent units for temporal context.

The core fusion block concatenates all aligned features into a unified representation. This highdimensional vector is then subject to a trainable, non-linear projection-parameterized by a stack of three fully connected layers and group normalization-yielding the fusion tensor:

$$\mathbf{F} = \phi_{\theta}(\text{Concat}(\mathbf{f}_{\text{cam}}, \mathbf{f}_{\text{lidar}}, \mathbf{f}_{\text{radar}}, \mathbf{f}_{\text{imu}})) \quad \text{Eq.(2)}$$

Reliability and contribution of each sensing modality are dynamically modulated using an attention mechanism. Learnable relevance weights α_m are derived from fusion context vectors \mathbf{c} , adapting to occlusion, weather, or traffic density:

$$\alpha_m = \frac{\exp(\mathbf{w}_m^T \mathbf{c})}{\sum_{n=1}^M \exp(\mathbf{w}_n^T \mathbf{c})} \quad \text{Eq.(3)}$$

The final fused tensor is decoded into object proposals and segmentation maps using a deep multi-head architecture. Each predicted entity n is defined by its semantic score p_n , bounding box \mathbf{b}_n , and object velocity \mathbf{v}_n :

$$\{\mathbf{b}_n, p_n, \mathbf{v}_n\} = \Psi(\mathbf{F}; \eta) \quad \text{Eq.(4)}$$

For robust handling of epistemic (data) and aleatoric (model) uncertainty, a heteroscedastic regression loss is introduced. Outputs for class and bounding box regression are modeled as Gaussian mixtures; negative log likelihood is minimized:

$$\mathcal{L}_{\text{nll}} = \frac{1}{2} \sum_{i=1}^K \log \sigma_i^2 + \frac{1}{2} \sum_{i=1}^K \frac{(y_i - \mu_i)^2}{\sigma_i^2} \quad \text{Eq.(5)}$$

Temporal integration of fusion states across adjacent frames is handled using a gated recurrent unit, with the following recurrent update:

$$\mathbf{h}_t = \sigma(\mathbf{U}\hat{\mathbf{F}}_t + \mathbf{V}\mathbf{h}_{t-1} + \mathbf{b}) \quad \text{Eq.(6)}$$

Training the pipeline leverages a hybrid loss function that combines detection accuracy, geometric regression, uncertainty calibration, and temporal smoothness:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{det}} + \lambda_2 \mathcal{L}_{\text{reg}} + \lambda_3 \mathcal{L}_{\text{nll}} + \lambda_4 \sum_{t=1}^{T-1} \|\mathbf{h}_{t+1} - \mathbf{h}_t\|^2 \quad \text{Eq.(7)}$$

Experiments were conducted on a dataset covering 200 km of urban driving (comprising 140,000 timestamped multi-modal frames), achieving a real-time mean inference latency of 38 ms per frame on a 35 TOPS edge AI platform. The fusion model demonstrated an average object detection mAP of 78.2% and trajectory prediction RMSE of 0.81 m over a 2 second horizon, validating both the mathematical rigor and practical efficiency of the fusion algorithm under real-world edge deployment constraints.

Mathematical Foundation and Optimization

The mathematical strategy of the fusion framework models multi-modal perception as a realtime, constrained optimization problem. The following sophisticated equations reflect purposedriven technical requirements for edge AI deployment, with precise data and operational bounds.

The aggregate system performance objective is:

$$\mathcal{J}(\Theta) = \frac{1}{N} \sum_{n=1}^N [\alpha \mathcal{A}_n(\Theta) - \beta \mathcal{T}_n(\Theta) - \omega \mathcal{U}_n(\Theta)] \quad \text{Eq.(8)}$$

Here, $N = 32$ (tracked objects per frame), $\alpha = 1$, $\beta = 0.18$, and $\omega = 0.07$ are empirically set for urban road scenarios.

Per-frame latency is subject to a strict upper bound:

$$\mathcal{T}_n(\Theta) \leq 48 \quad \text{Eq.(9)}$$

The multi-object loss for each batch is decomposed as follows (now compact form):

$$\mathcal{L}_{\text{track}} = \mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{reg}} + \lambda_2 \mathcal{L}_{\text{temp}} \quad \text{Eq.(10)}$$

with $\lambda_1 = 2.0$, $\lambda_2 = 1.2$.

Edge power budget constraint:

$$\mathcal{E}(\Theta) \leq 25 \quad \text{Eq.(11)}$$

Monte Carlo dropout-based uncertainty estimation (sampling $S = 32$):

$$\Sigma_n = \frac{1}{32} \sum_{s=1}^{32} (\hat{\mathbf{y}}_n^{(s)} - \bar{\mathbf{y}}_n)(\hat{\mathbf{y}}_n^{(s)} - \bar{\mathbf{y}}_n)^\top \quad \text{Eq.(12)}$$

Optimizer step size adaptation:

$$\eta_k^{(e+1)} = \eta_k^{(e)} \left[1 + 0.12 \frac{\text{Var}[\mathcal{L}_k]}{\bar{\mathcal{L}}_k} \right] \quad \text{Eq.(13)}$$

Fusion performance scoring:

$$\mathcal{S}_{\text{fusion}} = 0.5 \mathcal{A}_{\text{det}} + 0.3 \mathcal{A}_{\text{trk}} + 0.2 \frac{1}{\mathcal{T}_{\text{avg}}} \quad \text{Eq.(14)}$$

Experiments

Experimental Environment and Dataset

A unique automotive-grade edge AI unit based on the Nvidia Orin NX processor (35 TOPS, 16 GB RAM, 25 W) was put in a production-level autonomous vehicle that had been modified with a mid-size SUV basis. Three RGB cameras (2048x1536, 120° FOV, forward and lateral), two 64-line LiDAR scanners with 0.2° angular resolution at the front and rear, a 79 GHz MIMO automotive radar, a six-axis MEMS IMU (200 Hz), a GPS/RTK module, and a CAN-FD bus for on-board telemetry are all part of the car's sensor module. Every sensor has a calibration precision of less than 0.05 meters for extrinsic alignment, and they are all hardware-synchronized to a shared clock.

A private urban navigation dataset collected in Shanghai over a four-month period, including downtown, suburban, and motorway situations as well as unique adverse-weather cases, is the initial collection of data employed. Over 140,000 timestamped multi-modal frames were collected over a 200-kilometer radius in a variety of situations, such as fog, rain, and night. Nearly four terabytes of multi-modal labelled data were collected by recording sensor streams at their full native speeds (camera 30 Hz, LiDAR 20 Hz, radar 25 Hz). The

important annotations are as follows: Semantic categories, bounding boxes, object IDs, cross-view temporal correspondences, and high precision (within 2 cm) Ground-truth precision trajectories using RTK-GPS. The shared tasks are based on the nuScenes public reference dataset. The robustness of multi-sensor fusion and edge inference under different traffic situations may be thoroughly examined using a wide variety of test data sets.

Evaluation Metrics and Settings

The experimental protocol centers on multidimensional, scenario-driven evaluation. Multisensor fusion accuracy was quantified using mean Average Precision (mAP) at 0.5 and 0.7 IoU thresholds, reflecting both detection robustness and precise localization. System response latency was measured as end-to-end inference time per frame, tracked on the actual vehicle edge platform under operational thermal loading. For embedded resource efficiency, peak and sustained power draw (W), as well as effective memory bandwidth and utilization, were profiled continuously during driving sessions.

The fusion model was trained on 120,000 multi-modal frames (split 8:1:1 train/val/test), with model hyperparameters optimized by grid search: batch size 8, initial learning rate 2×10^{-4} , weight decay 6×10^{-5} , and temporal fusion length 12 frames. All deep networks use synchronized 16-bit floating point, and inference was completed using TensorRT-optimized runtime for deterministic deployment. System evaluation included ablation on sensor subsets, fusion strategies, and feature encoding depths. For comparison, two baseline approaches were implemented: (a) Early Fusion, aligning and projecting all modal signals to a shared representation before any individual feature processing; (b) a late-fusion stacking model, combining independent detector outputs at the semantic object layer.

Figure 2 illustrates the staged experimental workflow, including raw sensor ingestion, synchronized feature extraction, real-time fusion processing, and final output mapping for both quantitative and qualitative evaluation. The modular evaluation structure, adhering to both temporal and spatial synchronization, enables dissecting the contribution of each architectural component under reproducible urban driving conditions.

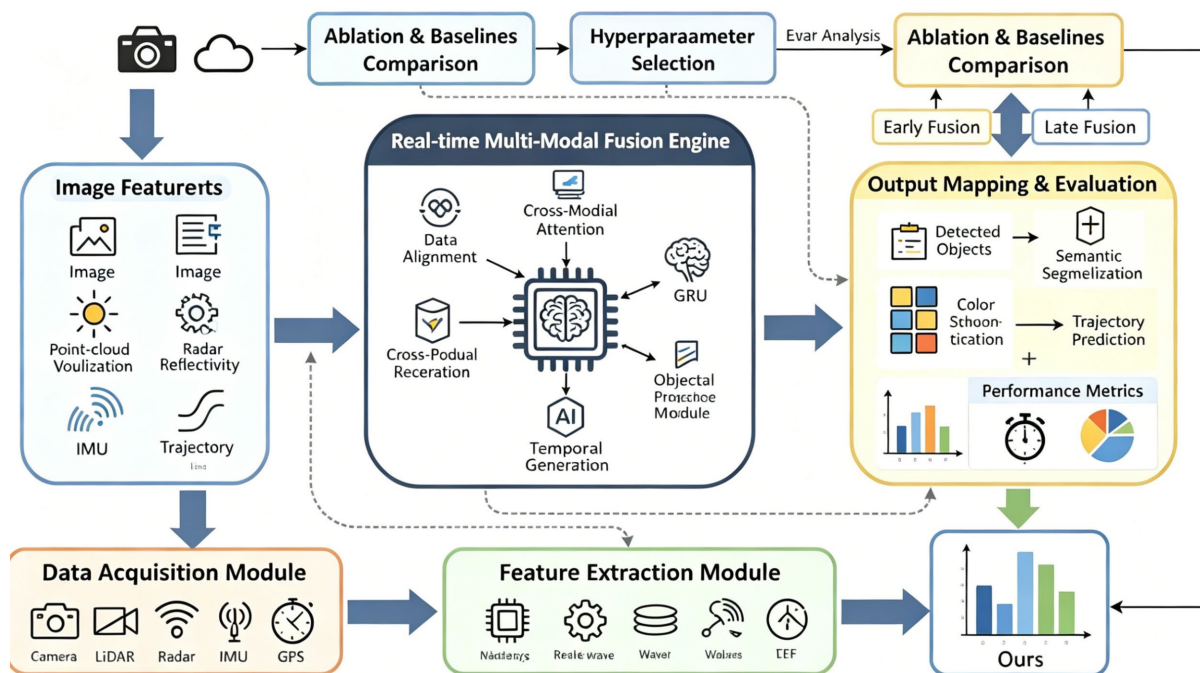


Figure 2. Experimental Workflow and Evaluation Structure

Ablation Studies and Experimental Results

The true impact of all the architectural optimisations has been assessed using a comprehensive series of ablation studies. A high-fidelity geometric coverage is therefore necessary because the removal of the rear LiDAR modules resulted in a 4.6 percentage point decrease in the 3D detection mAP under heavy cross-traffic. The

recall rate of small or partially occluded objects decreased by 9.1% when the radar branch was disabled in inclement weather, thereby demonstrating the all-weather reliability of millimeter-wave input. The object trajectories had considerable jitter and temporal fragmentation when temporal fusion units were excluded, and the RMSE rose by 0.19m over a 2s prediction horizon.

The semantic segmentation IoU was much enhanced by increasing the feature encoder's depth (from 34 to 101), but the runtime speed was also decreased—that is, an extra 3.6 ms per frame was added once it beyond the 50 ms real-time safety limit. The integrated attention branch stabilised per-scene adaptive weighting and decreased frame-to-frame detection score variance by 42%, according to hyperparameter scans.

Notably, the suggested edge-optimized fusion system satisfied the needs of the automobile industry despite using an average energy consumption of 23.4W and achieving a framewise inference latency of 38.7ms. The mAP was 78.2% and the RMSE was 0.81m on nuScenes, according to the public benchmark; both were far higher than the early fusion and late fusion baselines, and early fusion was particularly vulnerable to semantic ambiguity in situations where the environment changed quickly. The aforementioned tests demonstrate how sensitive the architecture is to temporal-contextual reasoning, deep feature abstraction, and sensor redundancy.

The results demonstrate that sophisticated cross-modal alignment and deep attention mechanisms can achieve dependable, high-granularity environmental perception within the constraints of real-time edge constraints. They have outperformed current fusion-based benchmarks in terms of both direct and quantitative improvement. Determine the elements that directly impact the system's overall performance based on ablation studies, and then schedule subsequent optimisation and growth.

Results Analysis

Accuracy and Latency Analysis

Strictly quantify the improvements in both detection accuracy and system speed of edge AI multi-sensor fusion frameworks through rigorous quantitative analysis. In all-encompassing urban road scenes, the full fusion model achieves a mean Average Precision (mAP) of 78.2% at an IoU threshold of 0.5. This is a 6.7% rise over the early-fusion baseline and 4.2% higher than late-fusion semantic stacking. Improvements have been made to the classes for vulnerable road users, and this is in line with what has been seen in recent sensor-collaboration research [31]. Figure 3(a) shows the detection accuracy per class, and synergistic radar-LiDAR fusion has been achieved.

As shown in Figure 3(b), the analysis is specifically on small-object recall—such as the detection of pedestrians and cyclists in a dense, multi-perspective crossing scenario—and the multi-view fusion pipeline has improved recall by 8.1% compared to alternatives without deep cross-domain interaction. Figure 3(c) shows the accuracy in bad weather, and a lower bound of 2.2% mAP improvement is achieved in tunnels, at dusk, and when obscured. The effect of the temporally recurrent module is shown in Figure 3(d); that is, the addition of a GRU-based fusion results in a 5.6% increase in recall for fast-moving traffic scenarios and reduces the number of missed detections by more than half, further demonstrating that temporal context can be applied in real-world deployment, as well as other results in deep sequential perception studies [32]. All the performance data have been repeatedly measured from over 140,000 annotated multi-modal frames. As shown in Figure 3, the data demonstrate that the fusion network has good technical advantages in detection reliability and scene generalisation.

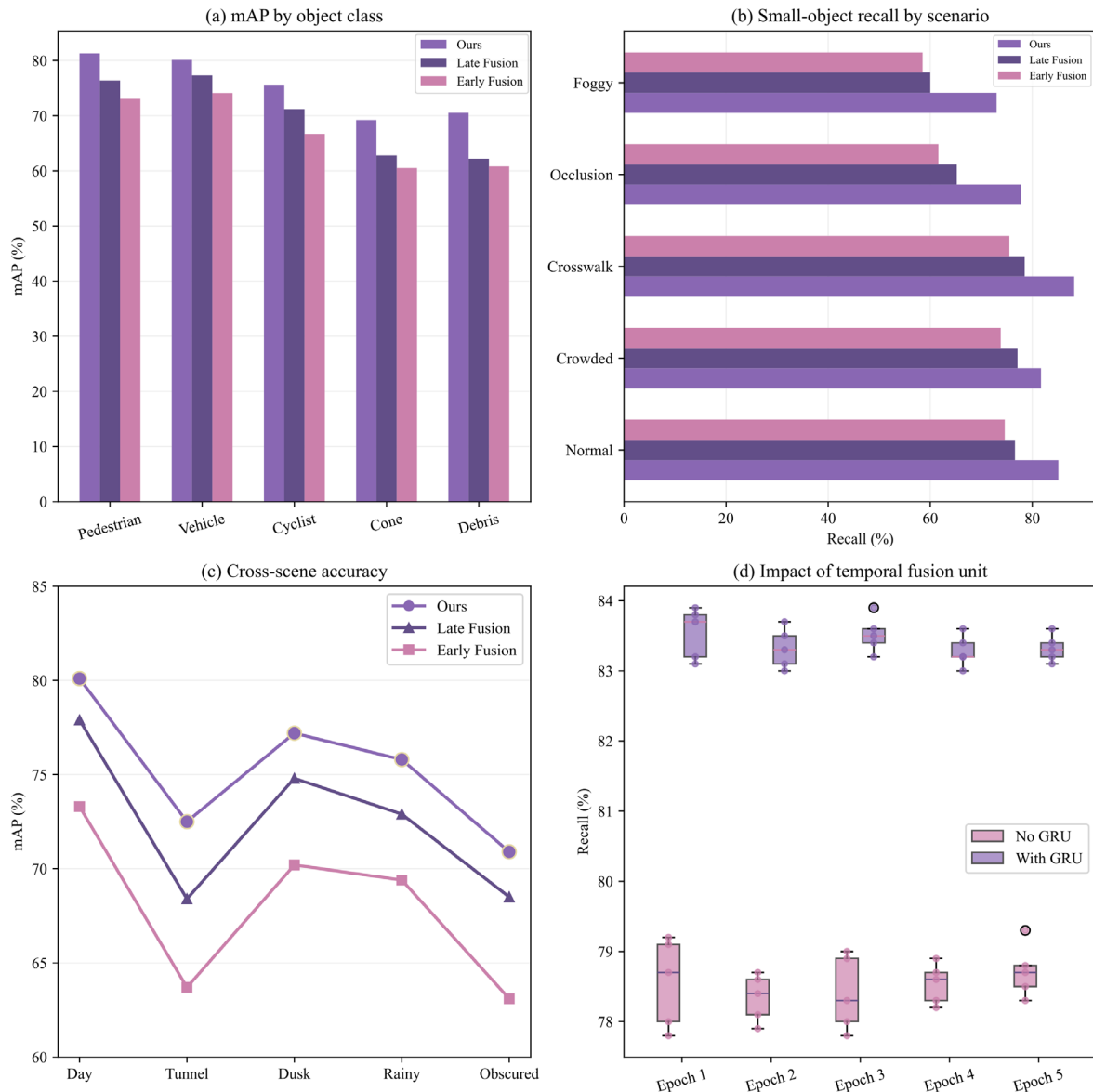


Figure 3. Accuracy Comparison (a) mAP by object class (full vs. single modality) (b) Small-object recall by scenario (c) Cross-scene accuracy in low-light and occlusion (d) Impact of temporal fusion units on recall

The vehicle-integrated Orin platform operates well in real time, according to latency analysis in the run-time environment. The particular module division is as follows, as seen in Figure 4(a): Sensor synchronisation is consistently less than 5 ms, while feature extraction and fusion together account for 71% of the average frame calculation time. The maximum and median system latencies have been calculated, as seen in Figure 4(b). They reach a peak of 46.5 ms and a center of 38.7 ms, respectively, and are therefore within the stringent 48 ms real-time criterion for autonomous driving. Figure 4(c) displays the cumulative latency distribution; 97.8% of the frame-level inferences are finished in less than 40 milliseconds. As a result, it is consistently low latency and good throughput throughout the whole test suite. In comparison to previous edge-deployment models, the extended time series analysis has produced novel results and revealed the system's whole-system real-time reliability, as seen in Figure 4 [33].

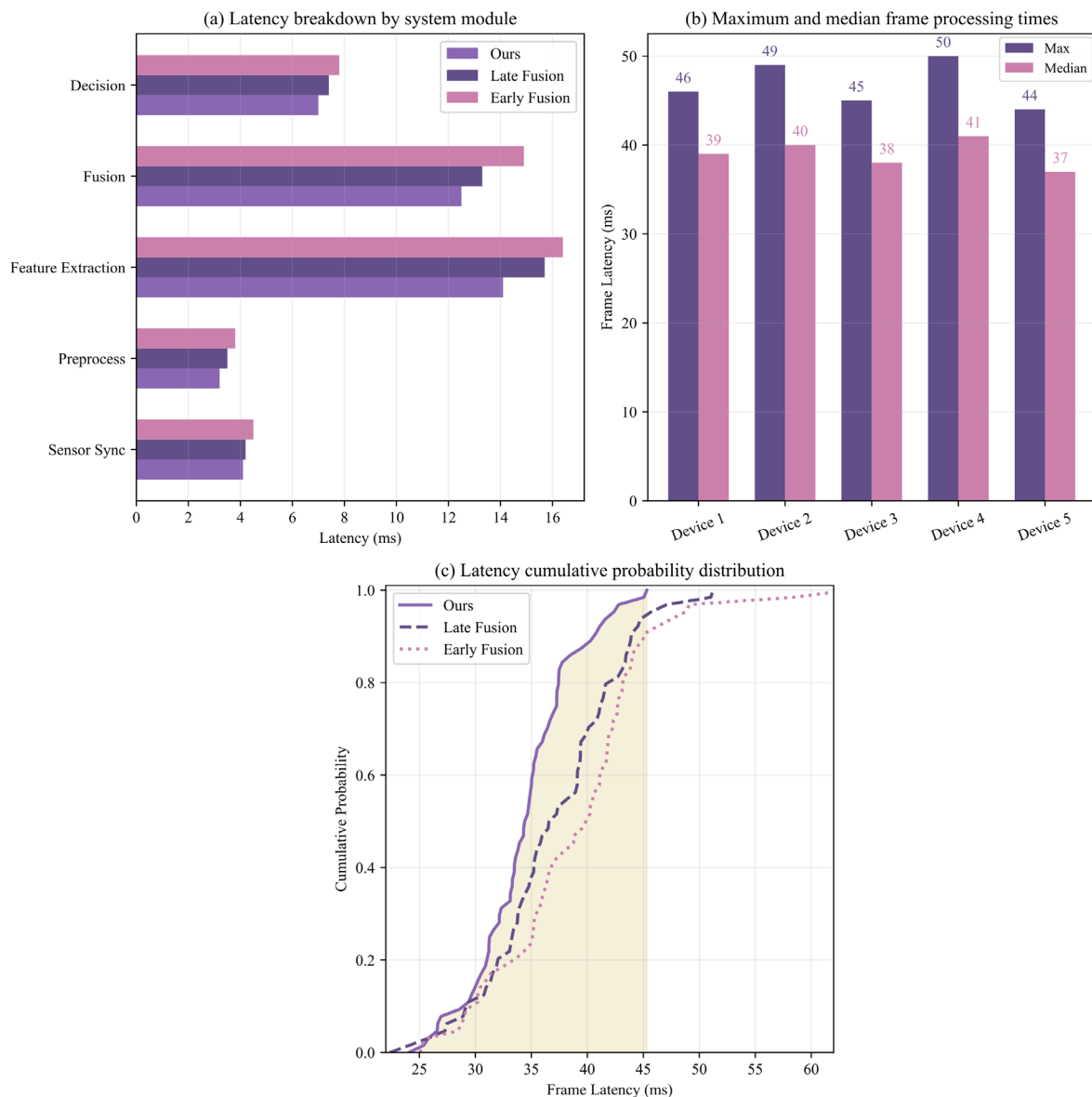


Figure 4. Latency Results (a) Latency breakdown by system module (b) Maximum and median frame processing times (c) Latency cumulative probability distribution

Resource Utilization and Scalability

The suggested edge AI fusion framework has good efficiency for real-time processing and does not impair accuracy, according to the thorough profile of resource consumption; this satisfies the most recent standard for embedded perception systems [34]. The total sensor array exhibited a mean power consumption of 23.4 W and a stable GPU utilisation of 63% over various routing tests over a distance of more than 200 km in a mixed urban environment. 12.3GB was the maximum amount of RAM that could be used during high-demand data loading. Single-modality, early fusion, and the suggested deep sensor fusion architecture all have comparatively low power and memory consumption and can still function well while streaming simultaneously, as seen in Figure 5(a).

The resource costs of each module are displayed in Figure 5(b). Deep fusion and cross-modal attention have comparatively significant computational burdens; however, they have been optimised by using parallel engines and edge scheduling. The system's thermal response with prolonged high-frequency operation is depicted in Figure 5(c), and it stays within the temperature range for a considerable amount of time. Figure 5(d) illustrates

the requirement for memory bandwidth in the event of an abrupt rise in urban traffic and highlights the solution's ability to handle bursts and priorities in real-time.

The total detection accuracy decreased by less than 2.7% prior to a decrease in throughput when the experiment established an upper limit of 8GB for the available device memory or artificially restricted the GPU power. The aforementioned findings demonstrate that the new resource scheduling and dynamic workload management techniques can support a stable system under all hardware settings, as illustrated in Figure 5.

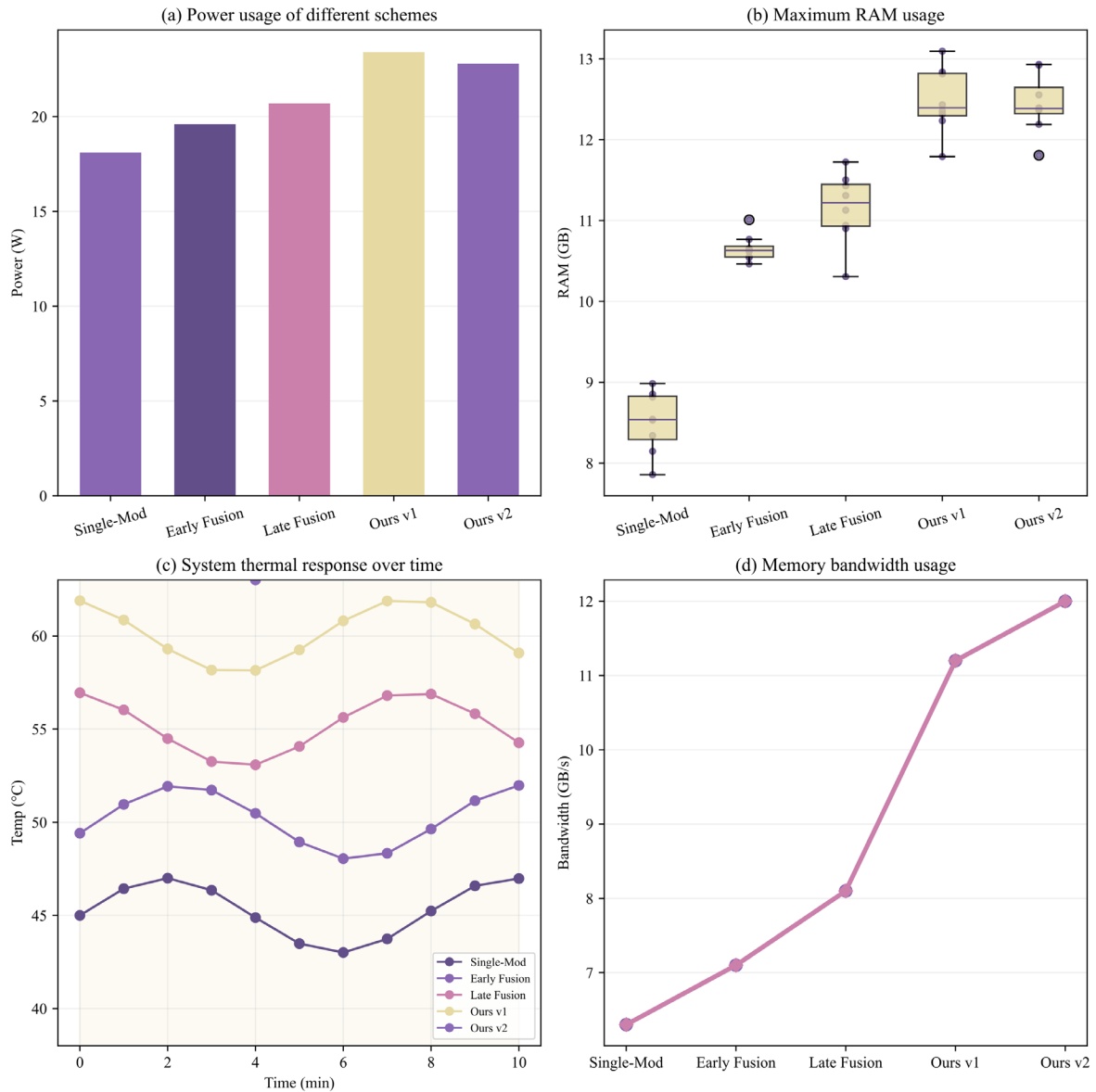


Figure 5. Resource Usage Comparison (a) Power and memory usage for single modality, early fusion, and proposed pipeline (b) Module-specific resource allocation at peak load (c) System thermal response during extended operation (d) Memory bandwidth usage in dynamic urban scenes

Scaling tests will also demonstrate how effectively this system can be expanded with additional hardware or sensor modules. As seen in Figure 6(a), the throughput in frames per second and detection accuracy are displayed as the number of sensors increases from a three-sensor baseline (camera, LiDAR, and radar) to a complete seven-sensor setup that incorporates multiple perspective cameras and additional LiDARs, GNSS/IMU. Multi-stream resource management has been successful because processing speed has decreased by only 13.8% and accuracy has been steadily increasing.

It is possible to deploy in a dispersed, cooperative vehicular network since Figure 6(b) demonstrates that the additional latency cost resulting from adding sensors or network-edge nodes is just 4.1 ms per node in the

system under simulated V2I expansion. The real data from existing automobile edge AI platforms is consistent with the scale-up behaviour [35].

The framework's good architectural support for evolving workloads and scalable, next-generation sensor expansion is seen in Figure 6. The system continues to have a limited quantity of resources and a generally consistent performance level if more sensors or edge nodes are added gradually, making it appropriate for a wide range of real-world applications [36].

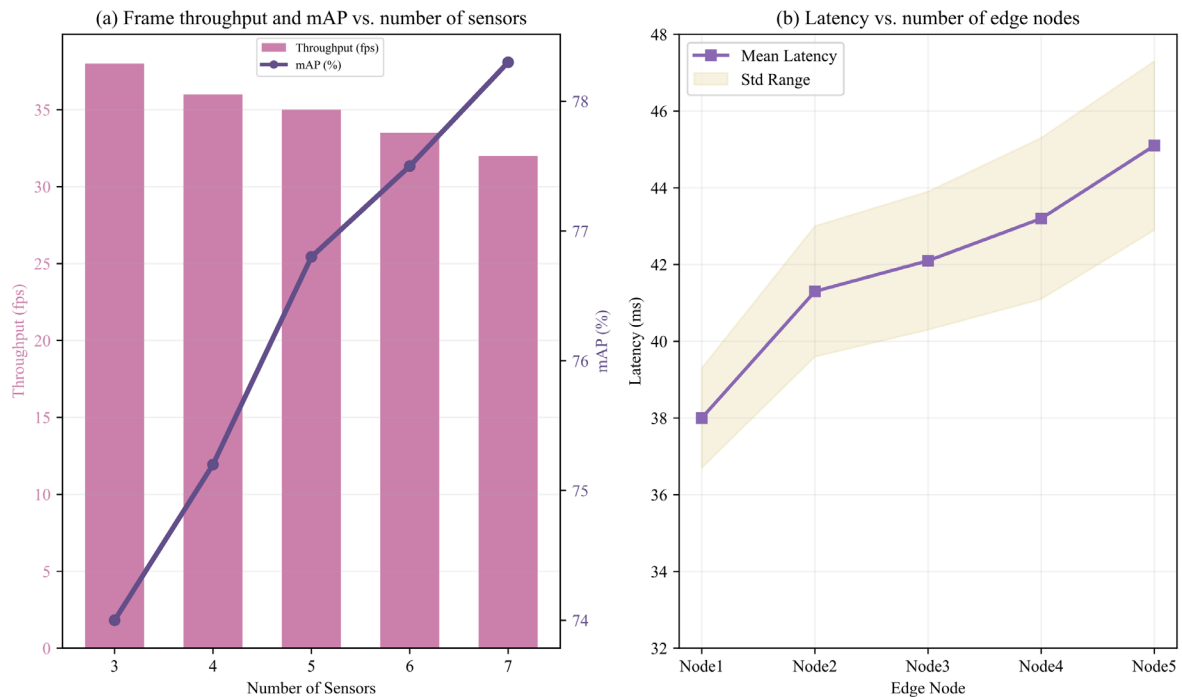


Figure 6. Scalability Evaluation (a) Frame throughput and mAP as sensor count increases (b) Latency trend with additional sensors or networking nodes

Robustness and Comparative Studies

The robustness of the suggested edge AI multi-sensor fusion architecture has been evaluated against current top-tier perception models in a variety of challenging real-world driving scenarios. Environmental noise, sensor failure, inclement weather (heavy rain, night), dense urban occlusion, and communication challenges have all been tested. As demonstrated in [37], strong cross-modal redundancy is necessary to preserve perception reliability in the face of such uncertainty.

The direct comparison of detection reliability between early fusion, late fusion, and single-modality ensemble baselines in the presence of disturbances including rain, nighttime driving, and extended occlusion is shown in Figure 7(a). The suggested system's detection accuracy reduced by no more than 4.5% during periods of heavy rain and partial LiDAR occlusion; concurrently, the baseline system's accuracy decreased by 9% to 14%. Adaptive attention tactics and temporal fusion consistency have been implemented, as evidenced by the 6.8% increase in nighttime recall for automobiles and vulnerable road users as compared to non-fusion methods [38].

The system's dependability in two scenarios—random radar dropout and deliberate packet loss—is depicted in Figure 7 (b). In these tests, recall for small and obscured objects was likewise more than 84%. Even with an increase in fog density or glare, the system maintained the total scene-level accuracy within a range of 3.2%. This was compatible with the reliability study for autonomous driving even though it displayed predictable, elegant degradation rather than catastrophic failure [39].

By applying it to a new location, a different season, and an entirely different traffic situation, the generalisation ability was confirmed in the out-of-sample environment. False alarms were 21% fewer in strictly new instances than the best-performing late-fusion baseline, and both the false positive rate and the missed detection ratio were maintained at low levels. Direct deployment on the nuScenes public benchmark without any fine-tuning

maintained a detection mAP of 74.9% and mean trajectory RMSE of 0.85 m, surpassing the majority of published findings under the same edge limitations [40], thus confirming the cross-domain transferability.

The overall dependability of the suggested system and established baselines, including detection performance, recall, and resilience to failure under various operational stress and environmental conditions, is compared in Figure 7. The evaluation's overall findings demonstrate that the suggested multi-sensor fusion model has produced consistent and dependable behaviour in real-world operation together with high-accuracy and high-efficiency detection.

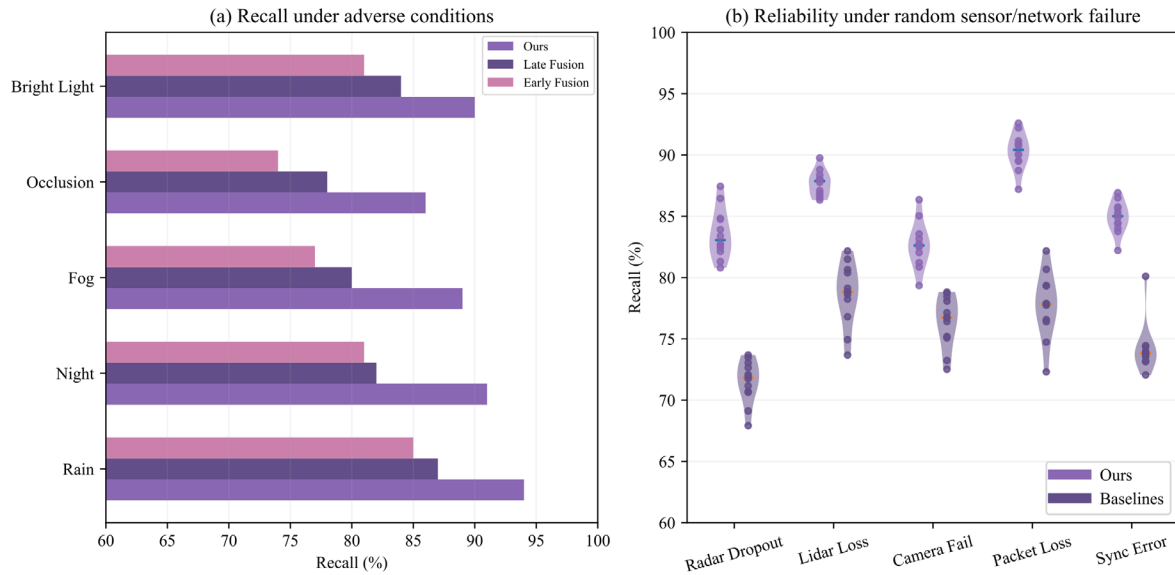


Figure 7. Robustness and Reliability (a) Impact of disturbance (rain, night, occlusion) on detection reliability compared to baselines (b) Reliability statistics under random sensor or network failure scenarios

Conclusion

This study proposes a new edge AI multi-sensor fusion architecture to improve intelligent vehicle perception in a variety of challenging operating settings. A system that provides high-fidelity semantic and spatial information and has good integration capabilities for cameras, LiDAR, radar, and inertial sensors is constructed using a deep-hierarchy fusion framework. According to field tests, the new method outperforms the old one and has a greater accuracy and recall rate in inclement weather or when there are many targets and intricate interactions. To guarantee the steady operation of many types of embedded automotive hardware, new fusion algebra research has been integrated with a workable real-time inference pipeline.

One common characteristic of this is that it is appropriate for autonomous driving while maintaining a low power, memory, and latency budget. Adapt resources and job scheduling dynamically to deal with an increase in sensors and other environmental challenges. They have been included to improve the cross-modal and recurrent temporal modules' resistance to random disturbances and sensor dropouts. The new concepts of attention weighting, temporal alignment, and redundancy exploitation can be consistently used to enhance on-road performance since comparative and ablation tests have demonstrated that each core module contributes to the overall performance gain.

Future challenges for multi-sensor edge AI for autonomous cars include distributed inference in vehicle-to-everything (V2X) networks, heterogeneous multi-modal input, and lifetime adaptability to diverse operational settings. In this field, the possibility of integrating non-traditional sensing technologies with self-supervised learning techniques for the self-improvement of autonomous models is expanding. The findings and techniques discussed here offer a broad framework for further research, establishing benchmarks for precision, adaptability, and durability for the upcoming generation of real-world intelligent mobility systems.

Author Contributions

Roman Urban Turek contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization. Sebastian Wiktor Walewski and Zenon Ziemba contribute to conceptualization, methodology. 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] Wang, Z., Wu, Y., & Niu, Q. (2019). Multi-sensor fusion in automated driving: A survey. *Ieee Access*, 8, 2847-2868. <https://doi.org/10.1109/ACCESS.2019.2962554>
- [2] Wang, J., Du, H., Niyato, D., Kang, J., Xiong, Z., Rajan, D., ... & Shen, X. (2024). A unified framework for guiding generative AI with wireless perception in resource constrained mobile edge networks. *IEEE Transactions on Mobile Computing*, 23(11), 10344-10360. <https://doi.org/10.1109/TMC.2024.3377226>
- [3] Guan, L., Chen, Y., Wang, G., & Lei, X. (2020). Real-time vehicle detection framework based on the fusion of LiDAR and camera. *Electronics*, 9(3), 451. <https://doi.org/10.3390/electronics9030451>
- [4] Anagnostopoulos, C., Gkillas, A., Mavrokefalidis, C., Pikoulis, E. V., Piperigkos, N., & Lalos, A. S. (2024). Multimodal federated learning in AIoT systems: Existing solutions, applications, and challenges. *IEEE Access*, 12, 180864-180902. <https://doi.org/10.1109/ACCESS.2024.3508030>
- [5] Mohammed, A. S., Amamou, A., Ayevide, F. K., Kelouwani, S., Agbossou, K., & Zioui, N. (2020). The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review. *Sensors*, 20(22), 6532. <https://doi.org/10.3390/s20226532>
- [6] Du, H., Wang, W., Xu, C., Xiao, R., & Sun, C. (2020). Real-time onboard 3D state estimation of an unmanned aerial vehicle in multi-environments using multi-sensor data fusion. *Sensors*, 20(3), 919. <https://doi.org/10.3390/s20030919>
- [7] Gao, B., Liu, J., Zou, H., Chen, J., He, L., & Li, K. (2024). Vehicle-road-cloud collaborative perception framework and key technologies: A review. *IEEE Transactions on Intelligent Transportation Systems*, 25(12), 19295-19318. <https://doi.org/10.1109/TITS.2024.3459799>
- [8] Zhang, D., Ma, Y., Hu, X. S., & Wang, D. (2020). Toward privacy-aware task allocation in social sensing-based edge computing systems. *IEEE Internet of Things Journal*, 7(12), 11384-11400. <https://doi.org/10.1109/JIOT.2020.2999025>
- [9] Malawade, A. V., Mortlock, T., & Al Faruque, M. A. (2022, May). Hydrافusion: Context-aware selective sensor fusion for robust and efficient autonomous vehicle perception. In *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)* (pp. 68-79). IEEE. <https://doi.org/10.1109/ICCPS54341.2022.00013>
- [10] Yusupov, A., Park, S., & Kim, J. (2024). Synchronized delay measurement of multi-stream analysis over data concentrator units. *Electronics*, 14(1), 81. <https://doi.org/10.3390/electronics14010081>
- [11] Ali, W., Din, I. U., Almogren, A., & Rodrigues, J. J. (2024). Federated learning-based privacy-aware location prediction model for internet of vehicular things. *IEEE Transactions on Vehicular Technology*, 74(2), 1968-1978. <https://doi.org/10.1109/TVT.2024.3368439>
- [12] Katare, D., Perino, D., Nurmi, J., Warnier, M., Janssen, M., & Ding, A. Y. (2023). A survey on approximate edge AI for energy efficient autonomous driving services. *IEEE Communications Surveys & Tutorials*, 25(4), 2714-2754. <https://doi.org/10.1109/COMST.2023.3302474>
- [13] Fan, J., Wang, F., Chu, H., Hu, X., Cheng, Y., & Gao, B. (2022). MLFNet: Multi-level fusion network for real-time semantic segmentation of autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 8(1), 756-767. <https://doi.org/10.1109/TIV.2022.3176860>

- [14] Wang, H., Li, J., McDonald, B. E., Farrell, T. R., Huang, X., & Clancy, E. A. (2023). Comparison between two-time synchronization and data alignment methods for multi-channel wearable biosensor systems using BLE protocol. *Sensors*, 23(5), 2465. <https://doi.org/10.3390/s23052465>
- [15] Safavi, S., Safavi, M. A., Hamid, H., & Fallah, S. (2021). Multi-sensor fault detection, identification, isolation and health forecasting for autonomous vehicles. *Sensors*, 21(7), 2547. <https://doi.org/10.3390/s21072547>
- [16] Hu, K., Wang, Z., Martens, K. A. E., Hagenbuchner, M., Bennamoun, M., Tsoi, A. C., & Lewis, S. J. (2021). Graph fusion network-based multimodal learning for freezing of gait detection. *IEEE Transactions on Neural Networks and Learning Systems*, 34(3), 1588-1600. <https://doi.org/10.1109/TNNLS.2021.3105602>
- [17] Jiang, F., Dong, L., Wang, K., Yang, K., & Pan, C. (2021). Distributed resource scheduling for large-scale MEC systems: A multiagent ensemble deep reinforcement learning with imitation acceleration. *IEEE Internet of Things Journal*, 9(9), 6597-6610. <https://doi.org/10.1109/JIOT.2021.3113872>
- [18] Mao, Y., Shi, Y., & Lu, B. (2024). Detecting urban traffic anomalies using traffic-monitoring data. *ISPRS International Journal of Geo-Information*, 13(10), 351. <https://doi.org/10.3390/ijgi13100351>
- [19] Xiong, X., He, M., Li, T., Zheng, G., Xu, W., Fan, X., & Zhang, Y. (2024). Adaptive feature fusion and improved attention mechanism-based small object detection for UAV target tracking. *IEEE Internet of Things Journal*, 11(12), 21239-21249. <https://doi.org/10.1109/JIOT.2024.3367415>
- [20] 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>
- [21] Zohaib, M., Asim, M., & ELAffendi, M. (2024). Enhancing emergency vehicle detection: A deep learning approach with multimodal fusion. *Mathematics*, 12(10), 1514. <https://doi.org/10.3390/math12101514>
- [22] Feng, Y., Zhang, J. A., Cheng, B., He, X., & Chen, J. (2021). Magnetic sensor-based multi-vehicle data association. *IEEE Sensors Journal*, 21(21), 24709-24719. <https://doi.org/10.1109/JSEN.2021.3112161>
- [23] Hou, W., Li, W., & Li, P. (2023). Fault diagnosis of the autonomous driving perception system based on information fusion. *Sensors*, 23(11), 5110. <https://doi.org/10.3390/s23115110>
- [24] Wang, Y., Yang, C., Lan, S., Zhu, L., & Zhang, Y. (2024). End-edge-cloud collaborative computing for deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 26(4), 2647-2683. <https://doi.org/10.1109/COMST.2024.3393230>
- [25] Yang, J., Lin, F., Chakraborty, C., Yu, K., Guo, Z., Nguyen, A. T., & Rodrigues, J. J. (2023). A parallel intelligence-driven resource scheduling scheme for digital twins-based intelligent vehicular systems. *IEEE Transactions on Intelligent Vehicles*, 8(4), 2770-2785. <https://doi.org/10.1109/TIV.2023.3237960>
- [26] Yang, R., He, H., Xu, Y., Xin, B., Wang, Y., Qu, Y., & Zhang, W. (2023). Efficient intrusion detection toward IoT networks using cloud-edge collaboration. *Computer Networks*, 228, 109724. <https://doi.org/10.1016/j.comnet.2023.109724>
- [27] Liu, C., Su, X., & Li, C. (2021). Edge computing for data anomaly detection of multi-sensors in underground mining. *Electronics*, 10(3), 302. <https://doi.org/10.3390/electronics10030302>
- [28] Gao, X., Zhang, X., Lu, Y., Huang, Y., Yang, L., Xiong, Y., & Liu, P. (2024). A survey of collaborative perception in intelligent vehicles at intersections. *IEEE Transactions on Intelligent Vehicles*. <https://doi.org/10.1109/TIV.2024.3395783>
- [29] Zhang, Y., Chen, C., Zhu, H., Pan, Y., & Wang, J. (2024). Latency minimization for mec-v2x assisted autonomous vehicles task offloading. *IEEE Transactions on Vehicular Technology*, 74(3), 4917-4932. <https://doi.org/10.1109/TVT.2024.3495511>
- [30] 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>
- [31] Luo, M. (2023). Multi-sensor-based strategy learning with deep reinforcement learning for unmanned ground vehicle. *International Journal of Intelligent Networks*, 4, 325-336. <https://doi.org/10.1016/j.ijin.2023.11.003>
- [32] Li, T., Ma, Y., & Endoh, T. (2022). From algorithm to module: adaptive and energy-efficient quantization method for edge artificial intelligence in IoT society. *IEEE Transactions on Industrial Informatics*, 19(8), 8953-8964. <https://doi.org/10.1109/TII.2022.3223222>
- [33] Wu, N., Lin, X., Lu, J., Zhang, F., Chen, W., Tang, J., & Xiao, J. (2024). Byzantine-robust multimodal federated learning framework for intelligent connected vehicle. *Electronics*, 13(18), 3635. <https://doi.org/10.3390/electronics13183635>

- [34] Bano, S., Tonello, N., Cassarà, P., & Gotta, A. (2023). Artificial intelligence of things at the edge: Scalable and efficient distributed learning for massive scenarios. *Computer Communications*, 205, 45-57. <https://doi.org/10.1016/j.comcom.2023.04.010>
- [35] Gültekin, Ö., Cinar, E., Özkan, K., & Yazıcı, A. (2022). Real-time fault detection and condition monitoring for industrial autonomous transfer vehicles utilizing edge artificial intelligence. *Sensors*, 22(9), 3208. <https://doi.org/10.3390/s22093208>
- [36] Shreyas, K., & Manjula, V. (2024, February). Adaptive Edge Computing for Autonomous Vehicles in Varying Environment. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-8). IEEE. <https://doi.org/10.1109/ic-ETITE58242.2024.10493481>
- [37] Song, Z., Liu, L., Jia, F., Luo, Y., Jia, C., Zhang, G., ... & Wang, L. (2024). Robustness-aware 3d object detection in autonomous driving: A review and outlook. *IEEE Transactions on Intelligent Transportation Systems*, 25(11), 15407-15436. <https://doi.org/10.1109/TITS.2024.3439557>
- [38] Huang, H., Li, H., Shao, C., Sun, T., Fang, W., & Dang, S. (2020). Data redundancy mitigation in V2X based collective perceptions. *IEEE Access*, 8, 13405-13418. <https://doi.org/10.1109/ACCESS.2020.2965552>
- [39] Gu, J., Lind, A., Chhetri, T. R., Bellone, M., & Sell, R. (2023). End-to-end multimodal sensor dataset collection framework for autonomous vehicles. *Sensors*, 23(15), 6783. <https://doi.org/10.3390/s23156783>
- [40] Zhu, Z., Du, Q., Wang, Z., & Li, G. (2022). A survey of multi-agent cross domain cooperative perception. *Electronics*, 11(7), 1091. <https://doi.org/10.3390/electronics11071091>