Energy-Aware Collaborative Perception in HetVNets: Balancing Accuracy and Sustainability

Mengying Zhou, Shaobin Wang, Qiang Duan, Aaron Yi Ding, Xin Wang, and Yang Chen

Abstract

Environmental perception is essential for autonomous vehicles. Collaborative perception, supported by vehicular communication technologies like C-V2X, aggregates data from nearby sensors to extend sensing range and improve decision accuracy. However, it raises communication overhead and computational energy because vehicles process more data than their own sensor input. The issue is exacerbated in heterogeneous vehicular networks (HetVNets). Our measurements show that, under equal workloads, low-capability vehicles incur 10% higher computational latency and energy, resulting in a lower energy-to-accuracy gain ratio and a higher marginal cost than high-capability peers.

In reality, users prioritize driving range over marginal accuracy gains. We therefore advocate that collaborative perception should move beyond pure accuracy maximization to include user-centric sustainability goals. We present GreenFusion, an energy-aware collaborative perception scheme that embeds explicit sustainability and fairness metrics. GreenFusion adapts each vehicle's engagement and role according to information value and capability, enabling selective sharing of noteworthy data. In evaluations, GreenFusion maintains perception performance while reducing energy consumption for low-capability vehicles by 81.0% and 31.5% on average compared with fully connected and information-adaptive centralized baselines, respectively. In a typical driving scenario, these savings correspond to a 65.6% increase in driving range, demonstrating practical sustainability benefits without sacrificing perception. GreenFusion reframes collaborative perception from an accuracy-only objective to a balanced accuracy-energy strategy, fostering a more sustainable, practical vehicular networking framework that improves resilience, longevity, and user experience.

Introduction

Environmental perception is a key technology for transforming conventional vehicles into autonomous systems [1, 2], relying on sensors (e.g., LiDAR, cameras, radars) to capture real-time information about the surroundings. This information provides essential decision support, helping vehicles interpret traffic signals, understand road conditions, and predict the behavior of other traffic participants. Collaborative perception advances this capability by leveraging vehicular communication technologies to share information among vehicles or entities, extending the sensing range and handling more complex driving scenarios [3–5].

However, collaborative perception raises sustainability concerns for vehicle fleets [6, 7]. Existing schemes often broadcast either raw sensor data (e.g., LiDAR point clouds) or lightly preprocessed representations (e.g., feature maps), imposing substantial communication and computation costs on participants. In some cases, a vehicle processes 4–5× its own data load [5]. Meanwhile, much of the received content is redundant or low-utility, making the incremental post-fusion accuracy scene-dependent and relatively modest. More importantly, vehicular networks are often heterogeneous in compute and battery capacities, where lower-capability vehicles are overburdened by this additional energy consumption for relatively small performance improvements.

Unlike prior schemes that purely emphasize complementarity and accuracy gains, GreenFusion proposed in this article is a collaborative perception framework for HetVNets that embeds sustainability and fairness. It introduces information-value weighting and resource-penalty gating, enabling higher-capability vehicles to undertake more workload and thereby balance payoffs across heterogeneous participants. Using these factors, GreenFusion calculates adaptive voxel-level engagement scores to regulate each vehicle's sensing scope and transmission policy. The engagement scores are computed on a central coordinator server, with this role rotating among participants to prevent any single participant from being overburdened. Empirical results show that GreenFusion reduces energy consumption for low-capability vehicles by 81.0% versus a fully

Mengying Zhou is with the School of Computing and Artificial Intelligence, Shanghai University of Finance and Economics, China, and MOE Key Laboratory of Interdisciplinary Research of Computation and Economics, Shanghai University of Finance and Economics, China; Mengying Zhou, Shaobin Wang, Xin Wang and Yang Chen are with the College of Computer Science and Artificial Intelligence, Fudan University, China, and the Shanghai Key Lab of Intelligent Information Processing, Fudan University, China; Qiang Duan is with the Information Sciences and Technology Department, Pennsylvania State University, USA; Aaron Yi Ding is with the Department of Engineering Systems and Services, Delft University of Technology, The Netherlands.

Digital Object Identifier: 10.1109/MCOM.001.2500266

 0163-6804/26/$25.00 © 2026 IEEE

connected baseline and 31.5% versus an information-adaptive central baseline without sacrificing perception. Task latency also drops by 79.0% and 30.1%, respectively. By achieving a 65.6% driving-range increase (+216.7 km) in urban conditions, GreenFusion delivers a sustainable, resilient, and deployable vehicular networking framework for next-generation mobility.

Background and Motivation

Collaborative Perception

Collaborative perception signifies a transformative shift in vehicular sensing. Unlike traditional perception systems that rely solely on individual vehicles' onboard sensors, it utilizes communication technologies like C-V2X to exchange information among vehicles and with roadside infrastructure [6]. This facilitates a more comprehensive and accurate understanding of the environment, supporting advanced decision-making and diverse downstream applications.

A classic example of collaborative perception's application is its integration with Advanced Driver Assistance Systems (ADAS) [3, 7]. In cases where a vehicle's view is obstructed, information from nearby entities can be used to detect hidden obstacles, allowing ADAS to initiate timely evasive actions. Collaborative perception also enriches semi-autonomous driving functionalities (e.g., route planning) by providing a broader, more detailed perspective. Within intra-vehicle contexts, collaborative perception supports the development of information and entertainment applications by sharing data relevant to user-specific requirements [8].

Beyond enhancing individual vehicle functions, collaborative perception is integral to next-generation Intelligent Transportation Systems (ITS) [2, 9, 10]. By integrating communication-enabled vehicular networks with urban infrastructure such as traffic signal controllers, it supports essential tasks like traffic flow optimization, public transportation scheduling, autonomous fleet coordination, and facility management. Moreover, collaborative perception contributes to urban sustainability by addressing traffic-related challenges [6], such as minimizing fuel consumption and greenhouse gas emissions, thus facilitating an eco-friendly and efficient transportation system.

Information Exchange Strategies. Collaborative perception relies heavily on the type and volume of data exchanged among entities. Based on the stage of data processing, information fusion is typically classified into three categories [3]: early, intermediate, and late fusion, each presenting distinct trade-offs in data volume, communication overhead, and perception accuracy. Early fusion involves transmitting raw sensor data (e.g., images, point clouds) [1, 4], maximizing information completeness but incurring high communication and energy costs. Intermediate fusion shares preprocessed perceptual features (e.g., BEV features, learned embeddings) [5, 11], reducing transmission overhead while preserving relevant details. However, it requires careful selection of partition points between local processing and data sharing. Late fusion transmits high-level results (e.g., detected objects, trajectories) [9], minimizing communication and computation costs but potentially sacrificing accuracy for fine-grained tasks.

Energy						
Vehicle	**Computational**		**Communication**		**All**	
	Δ(W)	↑(%)	Δ(W)	↑(%)	Δ(W)	↑(%)
High	1.6W	14.6%	4.2W	N/A}	5.8W	51.7%
Low	2.5W	24.1%	3.9W	N/A}	6.4W	61.1%
Latency						
Vehicle	**Computational**		**Communication**		**All**	
	Δ(ms)	↑(%)	Δ(ms)	↑(%)	Δ(ms)	↑(%)
High	192ms	96.0%	120ms	N/A	312ms	156.0%
Low	220ms	119.6%	126ms	N/A	346ms	188.0%

TABLE 1. Energy and latency burden for high- and low-capability vehicles under collaboration.

Coordination Strategies. Coordination in collaborative perception typically follows either decentralized or centralized paradigms. In decentralized approaches, such as the fully connected scheme [3], all vehicles communicate directly via all-to-all broadcasts and independently perform local fusion using received information. This results in replicated computation across vehicles and requires sufficient onboard resources. In centralized approaches [12], vehicles transmit data to a central node, which conducts global fusion and broadcasts the results. The decentralized paradigm offers greater resilience in dynamic or infrastructure-limited environments but may face scalability challenges due to high communication overhead. In contrast, the centralized scheme adopts a star-shaped architecture that reduces communication load and offloads computation, at the cost of increased reliance on infrastructure.

Quantitative Analysis of Collaborative Perception Burden

Collaborative perception enhances perception accuracy for autonomous driving systems but introduces significant costs in energy consumption and latency. To quantify the additional burden of collaborative perception on vehicles with different capabilities, we conduct a case study in a semi-realistic environment derived from real-world conditions. The emulated collaborative scenario involves two vehicles performing the same perception task under identical conditions, with the only difference being their computational capabilities: one vehicle has 8 CPU cores, the other 4 cores. We use the widely adopted V2XSet LiDAR-Track dataset [11] to emulate typical traffic scenarios with multiple obstacles, dynamic elements, and detailed communication traces. Each vehicle loads raw 3D point clouds from its LiDAR sensors and exchanges them via V2X links. After data exchange, both execute the same object-detection model on identical inputs. This case study runs for two hours, during which energy consumption is monitored with pyRAPL [12], a lightweight tool based on Intel's Running Average Power Limit technology for estimating energy usage.

Disproportionate Cost vs Benefit. In collaborative perception, increased communication and computation demands lead to energy and latency costs that are disproportionate to the accuracy improvements achieved. As Table 1 shows, enabling collaboration raises fleet-wide energy consumption by 12.2 W, an uplift of 56.4%. Latency costs

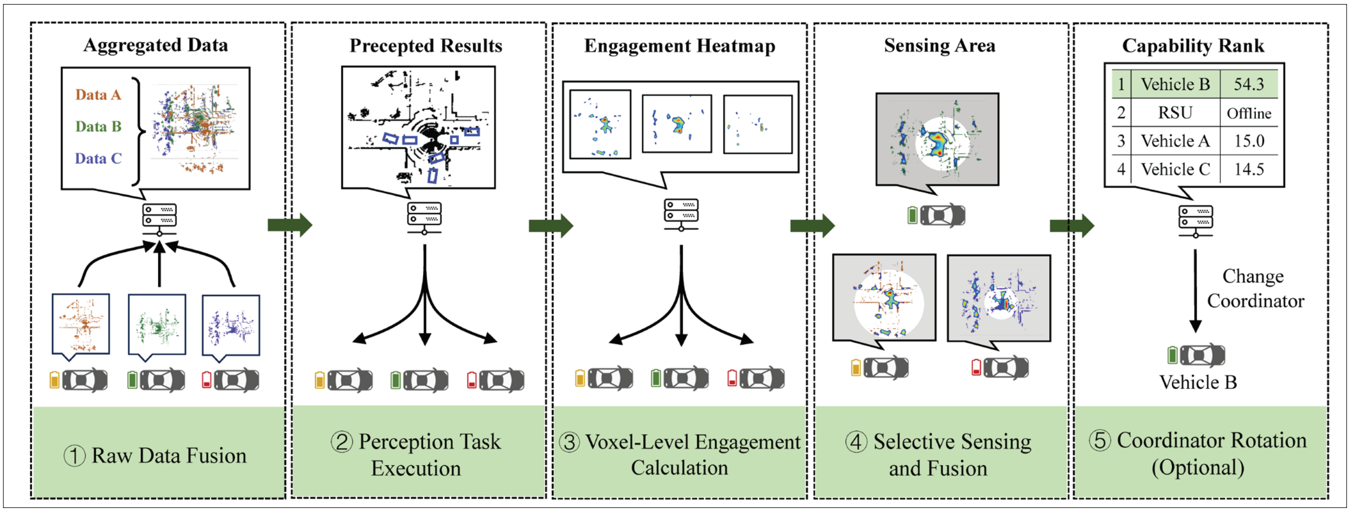

FIGURE 1. Workflow of GreenFusion in the collaborative perception task.

rise by 312 ms and 346 ms for the two vehicles, corresponding to 156.0% and 188.0% increases, respectively. Among these two overheads, most of the additional energy comes from communication, whereas the latency penalty is driven primarily by extra computation. The key benefit of collaboration is an improvement in Average Precision (AP) at IoU = 0.7 for vehicle detection from 0.625 to 0.701. This implies that each 1 percentage point increase in accuracy requires 7.4% more energy. Moreover, as baseline accuracy rises, incremental gains demand disproportionately greater resources. This is a classic diminishing-returns law widely present in real-world systems, where early improvements are resource-efficient, but later gains become increasingly costly.

Disproportionate Burden on Low-Capability Vehicles. The computational disparity between the two vehicles imposes a disproportionate energy and latency burden on the low-capability one. Table 1 shows that the vehicle with high computational capability introduces lower additional computation delay and energy during collaborative perception. In contrast, the low-capability vehicle experiences approximately 10% higher computation latency and energy for the same workload. Given equivalent perceptual improvement, this yields a lower energy-to-accuracy gain ratio and higher marginal cost for low-capability vehicles. While this computational gap does not directly affect communication efficiency, slower processing by the low-capability vehicle forces extra synchronization delays, which in turn impact the overall responsiveness of the collaborative network.

Exacerbating Range Anxiety. The above quantitative findings clearly illustrate that collaborative perception raises the problem of draining battery power and shortening operational lifespan, especially for electric vehicles with limited battery capacity. From a qualitative perspective, this heightened energy demand would further exacerbate range anxiety for drivers [13], which is a critical concern in human-vehicle interaction. If low-battery vehicles simply disable collaborative perception, they must rely solely on their onboard sensors. In low-vehicle-density environments, losing even a few vehicles will significantly decrease overall perception accuracy and safety. Consequently, drivers must carefully find a delicate balance between the benefits of improved safety and the drawbacks of shortened battery life.

Goal Transition: Balancing Accuracy and Sustainability

The quantified costs of collaborative perception expose a fundamental trade-off between maximizing detection performance and preserving vehicle range. A prior study [7] demonstrates that users in practice prioritize range retention over marginal accuracy gains. Therefore, collaborative perception should evolve beyond pure accuracy maximization to incorporate user-centric sustainability goals, including optimizing energy use while maintaining safety, balancing cost-benefit across the fleets, and even assisting low-battery vehicles in reaching nearby chargers.

However, existing solutions fall short of fully achieving these goals. One representative approach employs selective information strategies, such as sensing-area partitioning [4] or spatial-confidence filtering [5], to reduce redundancy by limiting transmissions to relevant data. Another class of solutions shifts from ad hoc broadcasts to centralized coordination [4, 9, 12] with dedicated orchestration algorithms. Nevertheless, these methods primarily optimize latency, neglect onboard energy costs, and assume participant homogeneity, failing to address real-world heterogeneity and its implications for energy fairness.

As a successor, GreenFusion advances an energy-driven paradigm via fine-grained engagement scoring and capability-aware coordinator rotation. By embedding explicit sustainability and fairness metrics, it aligns technical performance with user-centric and environmentally responsible mobility.

GreenFusion Design

Following the idea of refining the goals of collaborative perception to balance overall sustainability and perception accuracy, we introduce GreenFusion, an energy-aware framework for HetVNets.

GreenFusion Workflow

Figure 1 illustrates GreenFusion's five-step process in each round of collaborative perception: raw data fusion, perception task execution, voxel-level engagement calculation, selective sensing, and coordinator rotation. Steps ❸ to ❺, voxel-level engagement management and coordinator rotation, are the core designs to realize GreenFusion's energy awareness. GreenFusion operates in a centralized communication architecture, with a coordinator server managing vehicle interactions, fusing perceptual data, and broadcasting results to all participants.

In step ❶, each vehicle transmits its raw 3D point cloud data to the coordinator via communication modules. Each vehicle's information value varies, which is calculated at the voxel level based on factors such as vehicle capability, battery status, and perception effectiveness in the previous round. After aggregating and fusing the data, the coordinator runs perception tasks (e.g., object detection) and broadcasts the results to all participants in step ❷.

Steps ❸ to ❺ constitute the core of GreenFusion. The coordinator generates a voxel-level engagement heatmap for each vehicle based on the fused global view and individual capabilities. This heatmap is used to guide each vehicle in performing selective sensing and adjusting its data-sharing scope for the next round in step ❹. While in step ❺, to prevent any single participant from being overburdened, GreenFusion employs a coordinator rotation mechanism. If the current coordinator reaches a predefined workload threshold, the role will be transferred to a more capable participant within the network, ensuring the overall sustainability of the fleet.

Voxel-Level Engagement Management

Full participation in every perception round is neither necessary nor always beneficial. In certain cases, it may add communication overhead and noise, whereas partial engagement already yields optimal utility. Following this principle, GreenFusion generates voxel-level engagement heatmaps that assign a participation degree to each participant's voxels for the next round. These degrees determine

- The sensing scope, which may expand or shrink coverage,
- The transmission policy, which gives precedence to high-engagement voxels while downsampling or skipping low-engagement ones.

Guided by these heatmaps, vehicles concentrate their sensing and sharing on the high-value voxels, prioritizing information quality over quantity.

Engagement Heatmap Calculation. Voxel-level engagement management begins with generating an engagement heatmap for each vehicle. These heatmaps represent the appropriate engagement degree for each voxel in the next round perception (step ❸ of Fig. 1). Each voxel receives an engagement score E_v that accounts for both its informational value and the resources required for processing, concisely expressed as:

$$E_v = I_v \cdot [\Psi_v C_v]^{1-I_v^{\beta}} \quad (1)$$

The information term I_v quantifies voxel importance as the generalized mean $I_v = [\lambda R_v^{\eta} + (1 - \lambda) O_v^{\eta}]^{1/\eta}$, where R_v measures non-redundancy via the JS divergence between the voxel distribution and the global fused map, and O_v captures task relevance through the normalized mutual information between the voxel and the set of identifiable objects. A larger divergence or higher mutual information yields a higher I_v.

The formula includes a resource-gating factor, $\Psi_v C_v$, that represents energy and computation constraints. It collapses toward zero when either resource is scarce, thereby reducing the voxel's engagement score E_v. The energy-penalty term is $\Psi_v = \exp[-\alpha e_v/(\mathrm{SoC}_i + \varepsilon)]$ where e_v is the estimated energy to transmit and process the voxel and SoC_i is the vehicle's state of charge. The compute-readiness term is

$$C_v = \sqrt{1 - \mathrm{ops}(v) \,/ \big(F_i\, U_i^{\mathrm{idle}} + \varepsilon\big)},$$

where (v) is the voxel's operation count, F_i denotes maximum compute throughput, and U_i^{idle} is the CPU idle rate.

The exponent factor $1 - I_v^{\beta}$ implements a soft break-glass mechanism. When $I_v \ll 1$, the engagement score E_v is tightly constrained by resource limits, whereas for highly informative voxels ($I_v \to 1$), the factor approaches zero, effectively neutralizing the resource gate and ensuring that mission-critical data are transmitted even under severe resource constraints.

Motion Adjustment. To accommodate vehicle mobility, scene flow is applied to describe the mobility-induced spatial shifts and refine the heatmaps to align with the anticipated movement. These motion vectors predict object movement based on speed and direction inferred from the previous perception results, allowing the heatmaps to adjust for spatial shifts. This alignment ensures that voxel-level engagement accurately reflects the expected scene in the next round.

Selective Sensing and Fusion. Once the voxel-level engagement heatmaps are generated, the coordinator server distributes them to the respective vehicles, specifying the appropriate engagement degrees of voxels in the perception space. Using this information, vehicles perform selective sensing by dynamically adjusting their sensing range. Starting from a baseline sensing range, the vehicle incrementally expands its sensing coverage and assesses the engagement degrees of newly included voxels. When all newly added voxels exhibit engagement degrees below a predefined threshold T_s, this radius is taken as the sensing boundary for the current round. Step ❹ in Fig. 1 depicts the sensing range of each vehicle. This strategy directs vehicles to concentrate on high-value regions, minimizing resource expenditure in less informative areas.

After adjusting the sensing range and collecting LiDAR data, each vehicle prepares data for fusion in the next round according to the heatmap. Only voxel data exceeding a fusion threshold T_f is shared within the collaborative vehicular network. This selective fusion further ensures that only information-rich voxels are transmitted to the coordinator server for aggregation and perception tasks, maintaining perception accuracy while conserving energy and bandwidth.

Coordinator Rotation

Unlike fully connected networks where vehicles use broadcast to communicate with all others, GreenFusion establishes a coordinator-centered vehicular network to manage participant engagement and facilitate efficient information exchange. The coordinator role is strategically assigned via a rotation mechanism to enhance the transferability of coordination tasks and improve the sustainability of heavily burdened participants. This mechanism prefers to find a Roadside Unit (RSU) in scenarios with edge server support due to its

Unlike fully connected networks where vehicles use broadcast to communicate with all others, GreenFusion establishes a coordinator-centered vehicular network to manage participant engagement and facilitate efficient information exchange.

Participant	Battery (MJ)	CPU (cores)	Network Condition (Bandwidth-Latency-Loss)
Vehicle 1	200	12	400Mbps-0ms-0%, 100Mbps-10ms-0.5%, 50Mbps-25ms-1%
Vehicle 2	180	8	
Vehicle 3	150	4	
RSU	Powered	24	

TABLE 2. Machine configurations for evaluation.

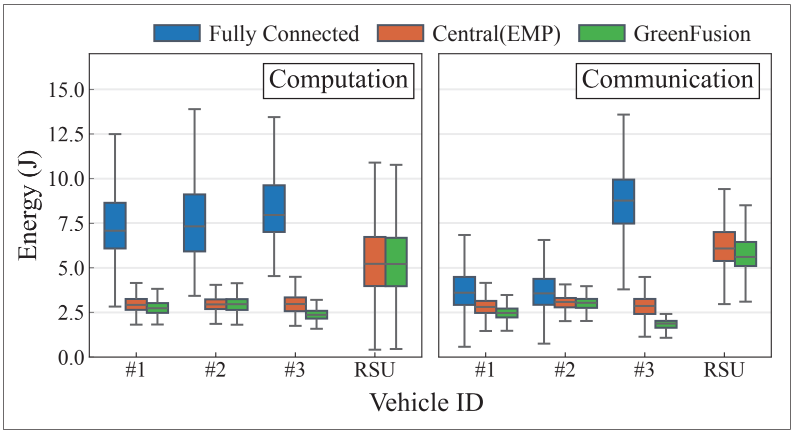

FIGURE 2. Comparison of computation and communication energy of fully connected, central, and GreenFusion schemes.

stable power supply and typically sufficient computational capability. However, when an RSU is unavailable or its resources are limited, GreenFusion dynamically selects a vehicle as the coordinator based on computational power, sensing capabilities, and battery level.

The current coordinator continuously monitors its own resources and triggers rotation when preset thresholds are approached. At each rotation check, it aggregates the voxel-level engagement scores E_v for every vehicle. The engagement degree of a participant is defined as the sum $\Sigma_v E_v$ of all its voxel scores. Using this metric, the coordinator assigns the role to the vehicle with the highest engagement degree, which usually holds the largest portion of relevant data. Granting this vehicle the coordinator role allows it to process its data locally, eliminating the need to transmit its data outward.

Step ❺ in Fig. 1 lists the engagement degree of each participant for the current round. Ranked results identify Vehicle B as the most capable node. The arrow from the RSU to Vehicle B depicts the transfer of the coordinator role to Vehicle B due to the RSU about to go offline. This policy keeps the role with the participant best positioned to handle it, and prevents any single participant from being overly burdened.

Performance Evaluation

We evaluate GreenFusion in terms of energy efficiency, perception accuracy, and extended travel range, comparing its performance with that of current collaborative architectures.

Experimental Settings

Dataset: Evaluations are conducted on the V2XSet LiDAR-Track dataset [11], which encompasses two dominant real-world driving scenarios of urban and highway across 12 scenes featuring representative roadway types (e.g., straight and curved segments, entrance ramps, intersections). Each scene includes two or three vehicles, each equipped with a LiDAR sensor capturing point clouds at 10 Hz (2 MB per frame). Vehicle speeds range from 20 to 58 km/h, and each recording lasts 5–25 s. The object detection model is trained on the V2XSet training set and remains frozen during all experiments. To ensure consistency, each scene is repeated five times for a total runtime of 11 hours.

HetVNet Setup: The HetVNet testbed consists of three vehicles and one RSU, each with distinct attributes, including battery capacity, computing power, RSU latency, and bandwidth, as summarized in Table 2. Test scenes selected from the V2XSet dataset include both two-vehicle and three-vehicle cases. Two-vehicle scenarios use the configurations of Vehicle 1 and Vehicle 2, while three-vehicle scenarios cover all vehicles. Each scenario also includes an RSU without sensing capabilities, serving as the primary coordinator for engagement management, data aggregation, and object detection. Energy consumption is measured with pyRAPL.

Communication Setup: All V2X links are emulated with the Qualcomm 9150 C-V2X chipset [14, 15], the hardware backbone of the Snapdragon Automotive Platform used by BMW, Mercedes-Benz, and Volvo. The emulation follows 5G NR PC5 sidelink parameters, supporting 0–25 ms one-way latency and 50–400 Mbps peak throughput [15].

To simulate real-world channel dynamics, two impairment strategies are introduced to reflect adverse conditions such as extreme weather, dense traffic, tunnels, and canyons [15]:

- Random perceptual block loss: each LiDAR frame is divided into a 3×3×3 grid, with three blocks randomly dropped to model spatially localized losses under adverse weather;
- Channel fluctuation: periodic packet-loss and delay bursts emulate intermittent link degradation and non-line-of-sight effects in dense traffic, tunnels, and canyons. Communication parameters are detailed in Table 2.

Parameter Settings: In our experiments, the parameters in Formula 1 are empirically set to $\lambda = 0.5$, $\varepsilon = 1$, $\alpha = 5$, and $\beta = 2$. These correspond to the trade-off between redundancy and task relevance, mean selectivity, energy penalty, and gating strictness, respectively. For different scenarios and tasks, we recommend selecting λ, ε, α, β within [0.5, 0.7], [0.3, 1], [4, 7], and [1, 2.5], where accuracy and energy savings remain robust.

Results and Discussion

This section evaluates the energy efficiency and perception accuracy of GreenFusion against two representative baselines: the fully connected scheme and the central scheme. The fully connected scheme [3] serves as a basic decentralized approach, enabling direct all-to-all communication without coordination or redundancy control. The central scheme, based on EMP [4], implements adaptive spatial partitioning, with each vehicle's zone boundaries dynamically assigned by the central coordinator based on data redundancy and network state. Lastly, we present a case study demonstrating GreenFusion's effectiveness in energy savings and driving-range extension.

Energy Efficiency. Figure 2 compares the computation and communication energy of the three schemes. Relative to the fully connected baseline, GreenFusion reduces vehicle-side computation energy by 64.0%. Of this reduction, 37% is attributable to computation offloading via the centralized architecture, as demonstrated by the performance of the central scheme. GreenFusion further achieves additional savings via selective sensing and fusion, especially for resource-constrained vehicles (e.g., Vehicle 3). Although both the central scheme and GreenFusion introduce RSUs to centralize computation, overall fleet energy consumption decreases. Notably, the RSU can operate on a commodity server supported by existing edge computing infrastructure, enabling cost-effective support for ITS facilities [10].

Meanwhile, we observe that for low-capability vehicles, limited compute capacity prolongs processing and airtime, making communication energy comparable to computation. GreenFusion also effectively mitigates this, reducing Vehicle 3's communication power from 8.76 J/s to 1.87 J/s, achieving a 78.6% reduction. Fine-grained selective transmission further lowers communication energy by 12.3% compared with the central scheme, underscoring the necessity of communication optimization as well. Furthermore, we quantify the control-plane messaging overhead at 0.025 J/s, representing a negligible 0.93% of the total communication energy in GreenFusion.

The above gains are more significant under adverse conditions (e.g., dense urban traffic, extreme weather) that cause frame loss and link degradation. Figure 3 plots energy savings of GreenFusion versus the two baselines for low-capability vehicles. Against the fully connected scheme, GreenFusion yields an average 81.0% reduction across scenarios, with larger gains as network quality worsens. Compared with the central scheme, it still achieves 31.5% average savings. Additionally, latency likewise decreases by 79.0% and 30.1% relative to the fully connected and central schemes, respectively.

Perception Accuracy. Selective data sharing may degrade perception accuracy. However, compared with lossless central fusion, GreenFusion shows only a slight accuracy decrease on the AP@0.7 metric, from 0.704 to 0.702. Despite a slight accuracy drop, GreenFusion still outperforms the non-collaborative baseline (0.702 vs. 0.632). Notably, fully connected fusion performs worst at 0.692, likely because redundant and noisy transmissions pollute useful data.

Under adverse network conditions, accuracy would decline further. GreenFusion's AP@0.7 drops modestly from 0.702 to 0.640, with fully connected and central showing comparable decreases. While at the more lenient AP@0.3 threshold, GreenFusion shows the smallest degradation (–3.87%), compared with fully connected (–4.52%) and central (–4.65%), indicating that prioritizing timely, high-gain, non-redundant information better preserves accuracy under network stress.

Energy-Saving Efficiency and Extended Range. We monitor the energy consumption of three vehicles over a period of time, as shown in Fig. 4. Evidently, energy consumption with GreenFusion (dashed line) is consistently lower than that of the central scheme (solid line), demonstrating GreenFusion's efficacy in prolonging battery life. Without GreenFusion, Vehicle 1 consumes 3.4 kJ in 8.5 s, whereas with GreenFusion activated, it consumes only 2.06 kJ in the same period, achieving a significant 39.41% energy-saving efficiency. Enabling GreenFusion can substantially extend the vehicle's range. Based on the U.S. standard electric vehicle average energy consumption of 151.4 Wh/km [13], a vehicle with a 50 kWh battery can travel 547.0 km under GreenFusion mode, compared with only 330.25 km without GreenFusion, extending the range by 216.7 km, an increase of 65.6%. Furthermore, the gentler slope of the dashed line suggests that GreenFusion slows battery depletion, which is a useful feature when driving with a low battery.

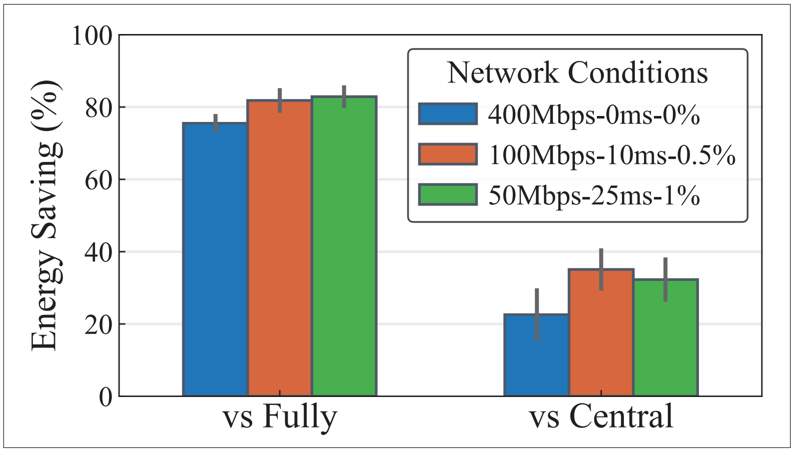

FIGURE 3. Energy savings of GreenFusion vs fully and central for the low-capability vehicle.

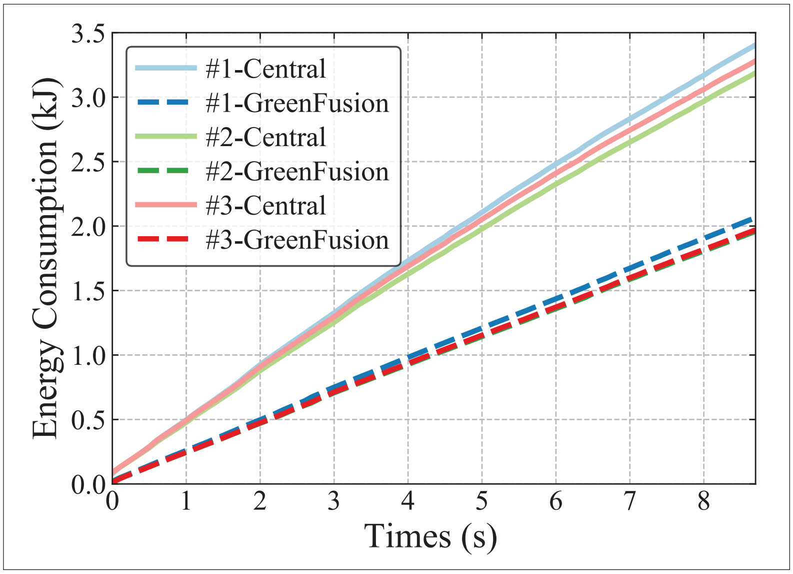

FIGURE 4. Energy consumption with GreenFusion (dashed) and central (solid) schemes.

Open Challenges and Research Directions

This section discusses open challenges and potential research directions for strengthening collaborative perception, focusing on supporting low-capability vehicles, incentivizing user participation, and co-designing energy-driven mobility.

Assistance for Vehicles without Perception Abilities

GreenFusion explores the potential of collaborative perception to assist vehicles lacking advanced sensing hardware such as LiDAR. Although these vehicles might not possess sophisticated perception modules, they typically have basic communication capabilities. In a collaborative perception context, sensor-rich vehicles can identify potential risks and promptly issue warning messages to perception-limited vehicles, which significantly enhances overall road safety by sharing essential situational awareness. Furthermore, collaborative perception offers a solution for low-battery vehicles via Teleoperated Driving (ToD) [3, 7], enabling remote intervention to extend their range to reach nearby charging stations. Nevertheless, practical deployment faces communication compatibility challenges across heterogeneous vehicles. There is an urgent need to develop universally compatible communication standards and optimized lightweight data encoding schemes to promote broader adoption of collaborative perception.

Incentives for Engaging in Collaborative Perception

Incentive mechanisms can motivate participants in collaborative networks to share onboard resources and actively contribute to perception tasks [10]. However, designing fair, effective, and scalable incentive frameworks that can handle vehicles' diverse interests and capabilities remains an open challenge. One direction is to explore reputation-based incentives, where vehicles earn reputation and higher priority through consistent contributions in collaborative networks, redeemable when assistance is needed. Additionally, dynamic pricing strategies for participants' energy and resources can also adapt to real-time demands and resource availability, facilitating cost-effective resource sharing and scheduling. Techniques such as double auctions and non-cooperative game theory may further encourage perception task offloading and enhance infrastructure utilization, fostering a more robust and cooperative vehicular communication environment.

Energy-Driven Interdisciplinary Co-Design for Mobility

GreenFusion illustrates the shift toward an energy-driven interdisciplinary paradigm that integrates technology, policy, and economics beyond isolated optimizations. This trend is reflected in policy and standards such as the U.S. Department of Energy's system-level efficiency programs for infrastructure and connected and automated vehicles (CAVs), ChargeScape's joint scheduling platform by major automakers, and the Next Generation Mobile Networks Alliance's emphasis on energy-aware design for future mobile ecosystems.

Achieving large-scale impact requires coordinated action among stakeholders. Automakers and infrastructure providers should jointly develop energy-efficient hardware and deploy green mobility platforms and device clusters to meet end users' expectations for safer, more economical, and durable mobility solutions. The key challenge is to transform a set of isolated energy-driven technologies into a cohesive interdisciplinary ecosystem that aligns communication systems, economic models, and urban planning with sustainability goals, stakeholder incentives, and deployment constraints to enable an equitable and resilient mobility future.

Conclusion

We present GreenFusion, an energy-aware collaborative framework designed to balance accuracy and energy consumption in HetVNets. By managing vehicle participation according to their capabilities, battery levels, and workloads, GreenFusion implements adaptive engagement mechanisms and coordinator rotation to optimize overall energy usage and ensure equitable benefits. GreenFusion shifts collaborative perception from a purely complementary model to a mutually assisting paradigm that balances accuracy and energy, emphasizing the resilience, sustainability, and longevity of collaborative vehicular networks. GreenFusion's practical and user-friendly design holds promise for real-world applicability, advancing next-generation sustainable autonomous driving systems. However, GreenFusion assumes honest participants. Future work will integrate lightweight malicious-behavior detection to defend against data poisoning. We also plan to conduct multi-city, at-scale road trials to validate its robustness under diverse real-world networks.

Acknowledgment

This work was supported in part by the Fundamental Research Funds for the Central Universities (No. 2025110481), National Natural Science Foundation of China (No. 62072115, No. 62472101), Shanghai Science and Technology Innovation Action Plan Project (No. 22510713600). Xin Wang and Yang Chen are the corresponding authors.

References

[1] Z. Liu *et al.*, "BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation," *Proc. ICRA*, 2023, pp. 2774–81.
[2] S. Liu *et al.*, "Communication Challenges in Infrastructure-Vehicle Cooperative Autonomous Driving: A Field Deployment Perspective," *IEEE Wireless Commun.*, vol. 29, no. 4, 2022, pp. 126–31.
[3] Y. Han *et al.*, "Collaborative Perception in Autonomous Driving: Methods, Datasets, and Challenges," *IEEE Intelligent Transportation Systems Mag.*, vol. 15, no. 6, 2023, pp. 131–51.
[4] X. Zhang *et al.*, "EMP: Edge-Assisted Multi-vehicle Perception," *Proc. MobiCom*, 2021, pp. 545–58.
[5] Y. Hu *et al.*, "Where2comm: Communication-Efficient Collaborative Perception via Spatial Confidence Maps," *Proc. NeurIPS*, 2022, pp. 4874–86.
[6] Z. Zhou *et al.*, "BEGIN: Big Data Enabled Energy-Efficient Vehicular Edge Computing," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 82–89, 2018.
[7] D. Katare *et al.*, "A Survey on Approximate Edge AI for Energy Efficient Autonomous Driving Services," *IEEE Commun. Surveys and Tutorials*, vol. 25, no. 4, 2023, pp. 2714–54.
[8] G. Qu *et al.*, "Tactical Data-Oriented Decentralized Communication Scheme in Vehicle-Assisted Networks: Challenges and Solutions," *IEEE Commun. Mag.*, vol. 62, no. 1, 2024, pp. 56–61.
[9] Y. He *et al.*, "VI-Eye: Semantic-Based 3D Point Cloud Registration for Infrastructure-assisted Autonomous Driving," *Proc. MobiCom*, 2021, pp. 573–86.
[10] X. Chen *et al.*, "Vehicle as a Service (VaaS): Leverage Vehicles to Build Service Networks and Capabilities for Smart Cities," *IEEE Commun. Surveys and Tutorials*, vol. 26, no. 3, 2024, pp. 2048–81.
[11] R. Xu *et al.*, "V2X-ViT: Vehicle-to-Everything Cooperative Perception with Vision Transformer," *Proc. ECCV*, 2022, pp. 107–24.
[12] T. Bahreini *et al.*, "VECMAN: A Framework for Energy-Aware Resource Management in Vehicular Edge Computing Systems," *IEEE Trans. Mobile Computing*, vol. 22, no. 2, 2023, pp. 1231–45.
[13] G. Oh *et al.*, "Vehicle Energy Dataset (VED), A Large-Scale Dataset for Vehicle Energy Consumption Research," *IEEE Trans. Intelligent Transportation Systems*, vol. 23, no. 4, 2022, pp. 3302–12.
[14] Qualcomm Technologies, Inc., "C-V2X Technical Performance Frequently Asked Questions," https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/c-v2x_technical_performance_faq.pdf, 2019, Accessed: 2025-08-05.
[15] T. Zugno *et al.*, "Toward Standardization of Millimeter-Wave Vehicle-to-Vehicle Networks: Open Challenges and Performance Evaluation," *IEEE Commun. Mag.*, vol. 58, no. 9, 2020, pp. 79–85.

Biographies

Mengying Zhou (zhoumengying@sufe.edu.cn) is an Assistant Professor at the School of Computing and Artificial Intelligence, Shanghai University of Finance and Economics. She received her Ph.D. in Computer Science from Fudan University. Her research interests include edge MLSys and urban computing.

Shaobin Wang (sbwang23@m.fudan.edu.cn) received his Bachelor of Engineering degree in Internet of Things Engineering from Zhengzhou University. His research interests include network measurement and distributed social systems.

Qiang Duan (qduan@psu.edu) is currently a Professor of Information Sciences and Technology at Pennsylvania State University Abington College. He received his Ph.D. in Electrical Engineering from the University of Mississippi. His research interests include computer networking and ubiquitous intelligence.

Aaron Yi Ding (Aaron.Ding@tudelft.nl) is a Senior Associate Professor with tenure at TU Delft and Adjunct Professor at the University of Helsinki. He received his Ph.D. in Computer Science from the University of Helsinki. His research interests include edge computing, mobile and distributed IoT systems.

Xin Wang (xinw@fudan.edu.cn) is a professor at College of Computer Science and Artificial Intelligence, Fudan University. He received his Ph.D. degree in Computer Science from Shizuoka University, Japan. His main research interests include quality of network services, next-generation network architecture, and network coding.

Yang Chen [SM] (chenyang@fudan.edu.cn) is a Professor with the College of Computer Science and Artificial Intelligence, Fudan University. He received his BSc and Ph.D. degrees in Electronic Engineering from Tsinghua University. His research interests include social computing, Internet architecture, and mobile computing. He is a senior member of ACM and a Fellow of the Royal Geographical Society.