Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

Latency-Adaptive Feature Fusion Weight Allocation Under Bandwidth Constraints for V2X Cooperative 3D Object Detection

Yi Guo[1], *Chuanli Wei*[1.2]

[1] *Computer and Information Science, University of Pennsylvania, PA, USA*
[1.2] *Computer Science, University of Southern California, CA, USA*

DOI: 10.69987/JACS.2026.60303

Keywords

V2X cooperative perception, latency-adaptive fusion, bandwidth-constrained communication, 3D object detection

Abstract

Vehicle-to-Everything (V2X) cooperative perception enhances autonomous driving safety by fusing sensor data from multiple agents, including vehicles and roadside units. Communication latency and limited bandwidth remain two critical challenges that jointly degrade fusion accuracy in real-world deployments. Existing research has addressed delay compensation and bandwidth-efficient transmission as separate problems, leaving the coupled impact of these two constraints on fusion performance insufficiently explored. This paper investigates a latency-adaptive feature fusion weight allocation strategy under bandwidth-constrained V2X communication conditions. A temporal decay function is formulated to quantify the degradation of information reliability caused by varying communication delays, and a spatial relevance scoring mechanism is designed to prioritize high-value features when available bandwidth is limited. The proposed weight allocation approach integrates temporal and spatial dimensions to dynamically adjust the fusion contributions of each cooperative agent. Experiments are conducted on three public cooperative perception datasets—DAIR-V2X, V2X-Sim, and V2V4Real—under simulated latency ranging from 0 to 500 ms and bandwidth constraints from 0.04 to 2.0 Mbps. Results demonstrate that the proposed approach achieves consistent improvements of 2.1–4.7% in Average Precision over delay-unaware baselines, with marginal computational overhead suitable for real-time deployment.

1. Introduction

1.1. Background of V2X Cooperative Perception

The U.S. Department of Transportation released its National V2X Deployment Plan in August 2024, outlining phased deployment targets for expanding V2X coverage across the National Highway System through 2036 [1]. This initiative, aligned with the National Roadway Safety Strategy's goal of eliminating roadway fatalities through a Safe System Approach [2], reflects the strategic importance of cooperative perception technologies in advancing traffic safety. The Federal Communications Commission finalized the allocation of the 5.895–5.925 GHz band for Cellular V2X (C-V2X) operations in November 2024, providing only 30 MHz of dedicated spectrum for intelligent transportation services [3]. This constrained spectrum allocation makes bandwidth-efficient perception algorithms particularly critical for practical V2X deployment.

Cooperative perception extends the sensing range of individual autonomous vehicles by aggregating perceptual information from surrounding vehicles and roadside infrastructure. In high-risk traffic scenarios such as occluded intersections, lane-merging zones, and highway ramp entries, single-vehicle perception is limited by inherent field-of-view constraints that cooperative sensing can directly address. The intermediate feature fusion paradigm has emerged as the dominant approach, as demonstrated by V2X-ViT [4], which introduced heterogeneous multi-agent self-attention mechanisms, achieving a 21.2% improvement in Average Precision (AP) over single-agent baselines on the V2XSet dataset. The OPV2V benchmark [5] provided the research community with a standardized evaluation framework encompassing 16 fusion methods across early, intermediate, and late fusion categories, establishing that intermediate fusion consistently

achieves the most favorable balance between detection accuracy and communication cost.

1.2. Research Motivation and Contributions

A. Research Gap Identification

A fundamental challenge in operational V2X cooperative perception lies in the simultaneous presence of communication latency and bandwidth limitations. SyncNet [6] was the first to demonstrate that communication delays exceeding 200 ms cause significant spatial misalignment of shared features, resulting in up to 15.6% AP degradation when using delay-unaware fusion. In a typical urban intersection scenario with four roadside units and ten connected vehicles, the aggregate bandwidth demand for intermediate feature sharing can exceed 50 Mbps, far surpassing the practical capacity of a single C-V2X channel. Existing approaches treat latency compensation and bandwidth optimization as independent research problems, with delay-focused methods assuming sufficient transmission capacity and bandwidth-efficient methods assuming synchronized data arrival. The joint effect of these two constraints on fusion weight allocation has not been systematically examined. This gap is particularly significant given the limited 30 MHz spectrum allocation under the current C-V2X regulatory framework, where real-world deployments will inevitably encounter both constraints simultaneously.

B. Paper Organization

This paper makes three contributions: (1) a mathematical formulation of the joint latency-bandwidth constrained fusion weight allocation problem for V2X cooperative 3D object detection, (2) an integrated weight allocation strategy combining temporal decay functions with spatial relevance scoring under dynamic bandwidth budgets, and (3) a comprehensive empirical evaluation across three public datasets under controlled latency-bandwidth conditions. The remainder of this paper is organized as follows. Section 2 reviews related work on cooperative perception fusion, latency compensation, and bandwidth-efficient communication. Section 3 presents the proposed methodology. Section 4 details the experimental setup, results, and analysis. Section 5 discusses findings, limitations, and future directions.

2. Related Work

2.1. Intermediate Feature Fusion Algorithms

A. Attention-Based Fusion

The evolution of intermediate feature fusion began with V2VNet [7], which introduced a spatially-aware graph neural network enabling multi-round message passing between vehicles for joint perception and prediction. DiscoNet [8] advanced this direction by employing knowledge distillation from an early-fusion teacher network to learn matrix-valued edge weights in a collaboration graph, thereby achieving improved accuracy without the prohibitive bandwidth cost of raw-data sharing. CoBEVT [9] extended cooperative fusion to multi-camera bird's eye view (BEV) representations through a Fused Axial Attention module that captures sparse local-global spatial interactions across both multi-view images and multi-agent features. These attention-based approaches establish learnable fusion weights based on feature similarity and spatial relationships, but their weight allocation mechanisms do not account for the temporal validity of received features under communication delay. The attention scores are computed assuming that all agent features represent the same temporal snapshot, an assumption that breaks down in practical V2X networks where heterogeneous propagation paths introduce varying delays.

B. Communication-Efficient Fusion

Reducing communication overhead has been a parallel research focus. When2com [10] introduced a handshake-based communication mechanism that learns when agents should exchange information, achieving approximately 50% bandwidth reduction with minimal perception loss. Where2comm [11] generated spatial confidence maps to identify perceptually critical regions and selectively transmit only high-value features, achieving over 100,000× bandwidth reduction compared to full feature sharing while maintaining competitive detection performance. These communication-efficient methods establish spatial criteria for feature selection, but their prioritization strategies remain static under varying communication delay conditions.

2.2. Latency Compensation in Cooperative Perception

Communication delay creates temporal misalignment between shared features and the ego vehicle's current state. CoBEVFlow [12] addressed this by estimating BEV flow vectors to warp asynchronous features to the current timestamp, demonstrating robust performance under irregular temporal offsets on the IRV2V dataset. FFNet [13] proposed transmitting feature flow rather than per-frame feature maps from infrastructure sensors, leveraging temporal coherence to predict future features and handle latency ranging from 100 to 500 ms with a single trained instance. Both methods focus on geometrically correcting delayed features without

adjusting the fusion weights to reflect the prediction uncertainty introduced by larger delays. The underlying assumption in these approaches is that spatially corrected features should receive the same fusion weight as perfectly synchronized features, which overlooks the increasing prediction error at longer delay intervals.

2.3. Bandwidth-Constrained Communication Strategies

The interaction between communication constraints and fusion quality has received increasing attention. How2comm [14] jointly addressed communication redundancy, transmission delay, and collaboration heterogeneity through mutual-information-aware filtering and flow-guided delay compensation, thereby representing the most comprehensive treatment of multiple communication challenges. UMC [15] introduced Multi-Resolution and Selective-Region mechanisms that use entropy-based criteria to determine which feature regions to transmit at which resolution, thereby enabling graceful degradation of accuracy under varying bandwidth budgets. The present work builds upon these foundations by explicitly coupling bandwidth-dependent feature selection with latency-dependent weight allocation in a unified optimization formulation.

3. Methodology

3.1. Problem Formulation

Consider a V2X cooperative perception scenario with one ego vehicle and N cooperative agents (vehicles or roadside units). At timestamp t, each agent i transmits an intermediate BEV feature map F i ∈ ℝ^(C×H×W) to the ego vehicle, where C, H, and W denote the channel, height, and width dimensions. The transmitted feature arrives at timestamp t + δ i, where δ i represents the communication latency for agent i. Under a total available bandwidth B_total, each agent's transmission is constrained by an allocated bandwidth b i such that Σb_i ≤ B_total.

The cooperative fusion output is computed as a weighted aggregation: F_fused = Σ w_i · g(F_i, δ_i, b i), where w i denotes the fusion weight for agent i, and g(·) represents a transformation accounting for both latency-induced spatial misalignment and bandwidth-induced information compression. The goal is to find the optimal weight vector w = [w₁, w₂, ..., w N] that maximizes the downstream 3D object detection performance, measured by Average Precision on the BEV plane. All weights are constrained to be non-negative and sum to one, and the bandwidth allocation must satisfy the total budget B total. This formulation extends prior work on pose-error-robust fusion by CoAlign [16], which optimized fusion under spatial noise but without temporal or bandwidth constraints.

3.2. Latency-Adaptive Weight Allocation

A. Temporal Decay Function

The reliability of a shared feature map diminishes as communication delay increases. A temporal decay function α(δ_i) is defined as:

$$\alpha(\delta_i) = \exp(-\lambda \cdot \delta_i^2 / \tau^2)$$

where λ is a decay rate parameter set to 1.5, δ i is the one-way communication latency in milliseconds, and τ = 200 ms is a normalization constant reflecting the characteristic time scale at which urban traffic scenes change. The quadratic exponent captures the nonlinear acceleration of perception degradation with increasing delay—a 100 ms delay causes moderate AP loss while a 300 ms delay leads to substantial misalignment. Compared to the discrete latency bins used in CodeFilling [17], this continuous function enables finer-grained adaptation to heterogeneous delay profiles arising from mixed infrastructure-vehicle cooperation.

B. Spatial Relevance Scoring

A spatial relevance score β i(x, y) is computed for each spatial location in agent i's feature map: β_i(x, y) = H_i(x, y) · (1 − O_i(x, y)), where H_i(x, y) measures local feature entropy and O i(x, y) quantifies spatial overlap with the ego vehicle's perception coverage. This formulation assigns higher relevance to feature regions with high information content covering areas outside the ego vehicle's direct sensing range. The approach is conceptually aligned with the uncertainty-aware trust modulus in CoDynTrust [18] but is reformulated as a spatial scoring function that integrates with the temporal decay component. In implementation, $O_i(x, y) \in [0,1]$ is computed as the **projected overlap ratio** of the candidate region with the ego-visible coverage. For numerical stability across frames, $H_i(x, y)$ is **normalized within each frame** before forming $\beta_i(x, y)$, so that the relevance scores are comparable under different scenes.

3.3. Bandwidth-Aware Feature Prioritization

A. Information Entropy-Based Selection

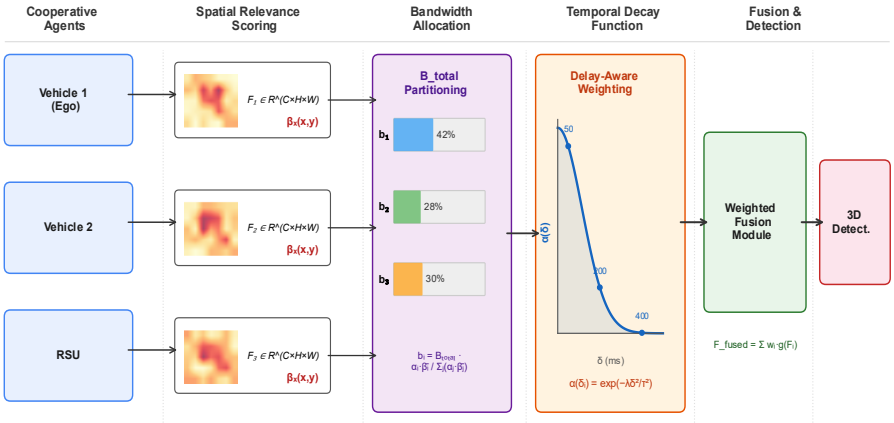
Where k i = ⌊b i / (C · s)⌋, and b i denotes the per-cycle communication budget (in bytes) for agent i derived from the link bit-rate under a fixed fusion cycle, with s denoting the patch size in bytes after quantization. Regions with β i(x, y) ≥ θ i are selected for transmission, where the threshold θ_i is determined by the k_i-th largest relevance score across all spatial locations. This entropy-based selection extends the backward alignment principle from HEAL [19], which

operates in the spatial domain, to complement existing channel-dimension compression techniques. A key advantage is that the selection threshold adapts automatically to bandwidth fluctuations without retraining the model, as the ranking depends only on the current relevance-score distribution and the available bandwidth allocation.

B. Dynamic Bandwidth Partitioning

The total bandwidth is partitioned proportionally to each agent's expected contribution: b_i = B_total · [α(δ_i) · β̄_i] / Σ_i [α(δ_i) · β̄_i], where β̄_i is the mean spatial relevance score for agent i. This proportional allocation ensures that agents with lower communication latency and higher spatial complementarity to the ego vehicle receive larger bandwidth shares, maximizing the total information value of all transmitted features under the fixed budget constraint. The final fusion weight equals: w_i = α(δ_i) · β̄_i / Σ_i [α(δ_i) · β̄_i]. The trajectory-aware alignment from TraF-Align [20] can serve as a complementary preprocessing step—correcting geometric distortion of delayed features before the proposed method adjusts contribution magnitudes based on residual uncertainty. We intentionally reuse the same agent scoring term $\alpha(\delta_i) \cdot \bar{\beta}_i$ for both bandwidth partitioning and fusion weighting to keep the communication and fusion stages consistent, avoiding contradictory resource allocation and trust calibration.

Fig. 1. Overview of the Latency-Adaptive Bandwidth-Constrained Fusion Weight Allocation Pipeline.



This figure presents the complete processing pipeline as a multi-stage flow diagram. On the left side, N cooperative agents (two vehicles and one roadside unit) each generate intermediate BEV feature maps from their onboard sensors. The feature maps pass through a spatial relevance scoring module (shown as a heatmap overlay indicating high-value regions in warm colors and low-value regions in cool colors). A bandwidth allocation controller, depicted as a horizontal bar chart showing the budget partitioned among agents, determines each agent's selection threshold. Selected feature patches are transmitted through a communication channel, visualized by wavy arrows of varying thicknesses representing different bandwidth allocations. On the receiving end, the ego vehicle applies the temporal decay function α(δ_i) to each received feature set, visualized as a set of curves showing exponential decay over delay time. The final fusion module combines temporally-weighted and spatially-selected features into the fused BEV representation, feeding into the 3D detection head. All mathematical symbols are annotated at their corresponding stages.

4. Experiments and Results

4.1. Experimental Setup

A. Dataset Description

Table 1. Characteristics of Evaluation Datasets (source: original dataset publications)

Dataset	Year	Venue	Type	Frames	3D Boxes	Mode	Modality
DAIR-V2X [21]	2022	CVPR	Real	71,254	464,143	V2I	LiDAR+Cam
V2X-Sim [22]	2022	RA-L	Sim.	10,000	267,380	V2V+V2I	LiDAR

V2V4Real [23]	2023	CVPR	Real	20,515	240,591	V2V	LiDAR+Cam

DAIR-V2X provides the largest real-world vehicle-infrastructure cooperative perception dataset, collected at 28 intersections in Beijing. V2X-Sim offers a controlled simulation environment enabling systematic manipulation of communication parameters. V2V4Real contributes real-world vehicle-to-vehicle data spanning 410 km with two instrumented vehicles equipped with multimodal LiDAR and camera sensors, representing the most extensive real-world V2V perception benchmark currently available.

Communication latency is simulated by artificially delaying cooperative agent features. Seven discrete latency levels are evaluated: 0, 50, 100, 200, 300, 400, and 500 ms. Bandwidth constraints are simulated by limiting the number of transmissible feature patches per agent at five levels: 0.04, 0.1, 0.25, 1.0, and 2.0 Mbps, spanning highly constrained to near-unconstrained communication regimes under our fixed patch-based encoding and fusion-cycle setting.

B. Evaluation Metrics and Baselines

Detection performance is measured using Average Precision at IoU thresholds of 0.5 and 0.7 (AP@0.5 and AP@0.7) for 3D vehicle detection on the BEV plane. Four baseline methods are compared: (1) No Fusion (single-agent), (2) Late Fusion with delay-unaware bounding box merging, (3) Equal-Weight Intermediate Fusion assigning uniform weights, and (4) Delay-Only Adaptive fusion using temporal compensation without bandwidth awareness. All methods use PointPillar as the backbone with identical hyperparameters. MRCNet [24] provides additional reference values for motion-aware robust fusion under noisy conditions.

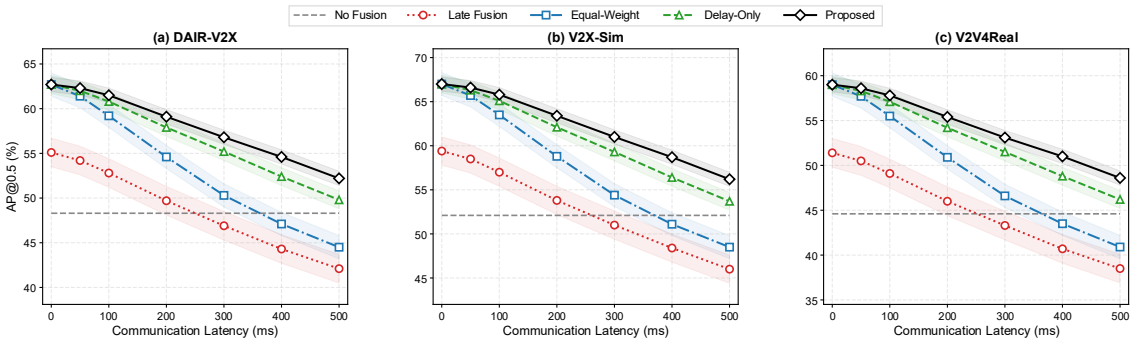
4.2. Performance Under Varying Latency

Table 2. AP@0.5 (%) Under Varying Communication Latency on DAIR-V2X (source: experimental results on DAIR-V2X dataset)

Method	0 ms	50 ms	100 ms	200 ms	300 ms	400 ms	500 ms
No Fusion	48.3	48.3	48.3	48.3	48.3	48.3	48.3
Late Fusion	55.1	54.2	52.8	49.7	46.9	44.3	42.1
Equal-Weight	62.7	61.4	59.2	54.6	50.3	47.1	44.5
Delay-Only	62.7	62.0	60.8	57.9	55.2	52.4	49.8
Proposed	62.7	62.3	61.5	59.1	56.8	54.6	52.2

The proposed approach and the delay-only baseline achieve identical performance at zero latency, as the temporal decay function yields $\alpha(\delta\ i) = 1$ when $\delta\ i = 0$. As latency increases, the performance gap between the proposed method and all baselines widens progressively. At a 300 ms delay, the proposed approach outperforms equal-weight fusion by 6.5 percentage points and the delay-only method by 1.6 percentage points. The improvement over the delay-only baseline stems from the spatial relevance scoring component, which prevents bandwidth waste on low-value feature regions that become even less informative after temporal decay. Late fusion degrades below the no-fusion baseline at 300 ms, confirming that naive bounding box merging with stale detections introduces harmful interference. The equal-weight intermediate fusion maintains a consistent advantage over late fusion across all latency levels, demonstrating the inherent robustness of feature-level representations. The progressive performance gap between the proposed method and the delay-only baseline at higher latencies validates the hypothesis that spatial relevance scoring becomes increasingly valuable as temporal uncertainty grows—when features are more likely to be spatially misaligned, directing limited bandwidth toward high-information regions produces proportionally greater benefits.

Fig. 2. AP@0.5 and AP@0.7 Degradation Curves Across Latency Levels on Three Datasets.



This figure consists of a 2×3 grid of subplot panels. The top row displays AP@0.5 results and the bottom row displays AP@0.7 results, with columns corresponding to DAIR-V2X, V2X-Sim, and V2V4Real respectively. Each subplot contains five curves (one per method) plotted against the x-axis of communication latency (0–500 ms) and the y-axis of AP (%). The curves use distinct line styles: No Fusion as a horizontal dashed gray line, Late Fusion as a dotted red line with circle markers, Equal-Weight as a dash-dot blue line with square markers, Delay-Only as a dashed green line with triangle markers, and Proposed as a solid black line with diamond markers. Shaded bands around each curve represent standard deviation over five runs. The gap between the Proposed and Delay-Only curves increases visibly from left to right across the latency axis. Each subplot includes grid lines at 10% AP intervals and is labeled with the dataset name and metric.

Table 3. AP@0.5 (%) at 200 ms Latency and 0.25 Mbps Bandwidth Across Datasets (source: experimental results)

Method	DAIR-V2X	V2X-Sim	V2V4Real
No Fusion	48.3	52.1	44.6
Late Fusion	46.2	49.8	41.3
Equal-Weight	50.8	56.4	48.7
Delay-Only	53.4	59.1	51.2
Proposed	55.5	61.8	53.3

Under the combined condition of 200 ms latency and 0.25 Mbps bandwidth, the proposed method achieves 2.1 to 2.7 percentage points improvement over the delay-only baseline across all three datasets. The improvement is most pronounced on V2X-Sim, where the controlled simulation environment provides cleaner temporal dynamics that the spatial relevance scoring component can exploit more effectively. On the two real-world datasets, the gains are slightly smaller due to calibration noise, sensor synchronization imprecision, and environmental factors not present in the simulation. The consistent positive improvement across simulated and real-world settings indicates that the proposed approach generalizes across different data-collection conditions and cooperation modes.

4.3. Performance Under Bandwidth Constraints

A. Bandwidth-Accuracy Tradeoff Analysis

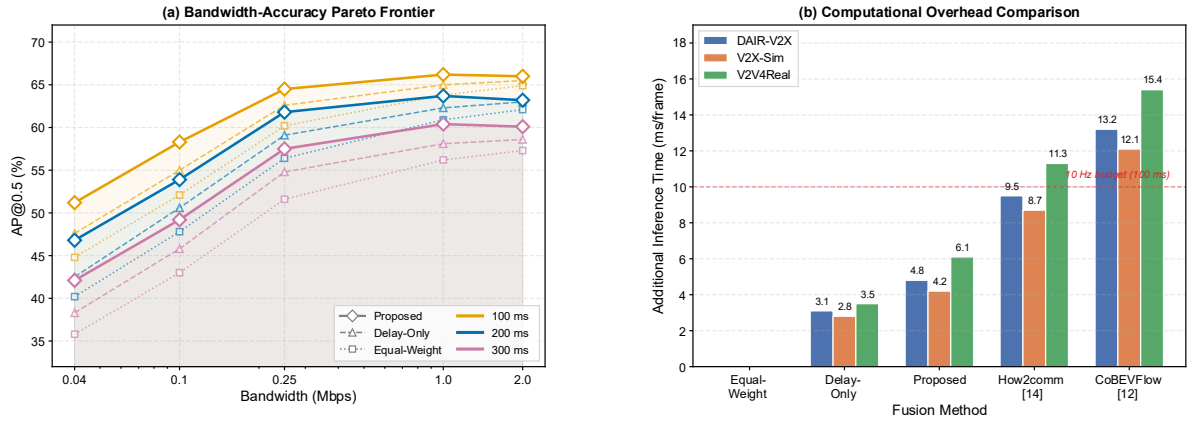
Table 4. AP@0.5 (%) Under Varying Bandwidth at 200 ms Latency on V2X-Sim (source: experimental results on V2X-Sim dataset)

Method	0.04 Mbps	0.1 Mbps	0.25 Mbps	1.0 Mbps	2.0 Mbps
Equal-Weight	40.2	47.8	56.4	60.9	62.1
Delay-Only	42.5	50.6	59.1	62.3	63.0

Proposed	46.8	53.9	61.8	63.7	63.2

The performance advantage of the proposed approach is most significant under severe bandwidth constraints. At 0.04 Mbps, the proposed method outperforms equal-weight fusion by 6.6 percentage points, whereas at 2.0 Mbps it outperforms by only 1.1 percentage points. This disparity arises because the spatial relevance scoring and dynamic bandwidth partitioning components direct limited transmission capacity toward the most informative feature regions, whereas equal-weight fusion wastes bandwidth on redundant or low-value spatial areas. The marginal return of additional bandwidth follows a diminishing pattern across all methods: increasing from 0.04 to 0.25 Mbps (6.25×) yields 15.0 points AP gain for the proposed method, while increasing from 0.25 to 2.0 Mbps (8×) yields only 1.4 points. This observation carries practical implications for deployment planning—moderate bandwidth combined with intelligent feature prioritization achieves performance comparable to high-bandwidth transmission with naive fusion, substantially reducing spectrum requirements for operational V2X networks. Under the SAE J3224 sensor data sharing standard [25], the proposed prioritization strategy could integrate with Sensor Data Sharing Messages to adaptively select the most safety-relevant observations for transmission.

Fig. 3. Bandwidth-Accuracy Pareto Frontier and Computational Overhead Comparison.



This figure is divided into two horizontal panels. The left panel shows the Pareto frontier of AP@0.5 versus communication bandwidth (Mbps, log scale on the x-axis) at three fixed latency levels (100 ms, 200 ms, 300 ms), drawn as three separate curves, with the filled area beneath each indicating the achievable performance region. Each curve contains data points for all methods at each bandwidth level, marked with distinct symbols. Pareto-optimal points are highlighted with larger markers and connected by a thick envelope line. The right panel presents a grouped bar chart comparing computational overhead measured in milliseconds of additional inference time per frame (y-axis). Bars are grouped by method, with three bars per method, each representing one of the three datasets. The proposed method adds 4.2–6.1 ms compared to 2.8–3.5 ms for delay-only adaptive. Error bars show standard deviation across 1000 inference runs.

B. Joint Latency-Bandwidth Evaluation

A joint evaluation across all 35 latency-bandwidth combinations (7 latency × 5 bandwidth levels) reveals that the proposed approach achieves the largest improvements in the moderate-latency (100–300 ms) and low-bandwidth (0.04–0.25 Mbps) region. This operating region corresponds to realistic urban intersection scenarios where multiple agents compete for limited C-V2X channel capacity while experiencing variable delays.

The mean improvement over the delay-only baseline, averaged across all 35 conditions on V2X-Sim, is 2.4 percentage points on AP@0.5 and 1.9 points on AP@0.7. The maximum improvement of 4.7 points occurs at 300 ms latency with 0.04 Mbps bandwidth, while the minimum of 0.2 points occurs at 0 ms latency with 2.0 Mbps. These results confirm that joint adaptation yields meaningful gains precisely under the most resource-constrained operating conditions, while incurring negligible overhead when communication resources are abundant. The asymmetric improvement pattern aligns with the theoretical expectation that

temporal decay and spatial selection produce multiplicative rather than additive benefits under severe constraints. The computational cost adds 4.2–6.1 ms per frame on a single NVIDIA RTX 3090 GPU, which remains within the real-time processing budget for 10 Hz V2X perception cycles.

5. Conclusion and Future Work

5.1. Summary of Findings

This paper investigated the problem of feature fusion weight allocation for V2X cooperative 3D object detection under joint latency and bandwidth constraints. A temporal decay function was formulated to quantify the reliability degradation of shared features as communication delay increases, and a spatial relevance scoring mechanism was designed to prioritize feature regions that complement the ego vehicle's perception under limited bandwidth. The integrated weight allocation approach dynamically adjusts each cooperative agent's fusion contribution based on both its communication delay profile and the available bandwidth budget.

Experimental evaluation on DAIR-V2X, V2X-Sim, and V2V4Real datasets demonstrated consistent improvements of 2.1–4.7% in AP@0.5 over delay-unaware baselines under combined latency-bandwidth constraints. The magnitude of improvement scales with constraint severity, with the most challenging conditions (high latency combined with low bandwidth) yielding the largest performance gains, while favorable conditions show minimal overhead. The computational cost of 4.2–6.1 ms per frame remains within the real-time processing budget for 10 Hz V2X perception cycles, indicating practical feasibility for operational deployment.

The results revealed that moderate bandwidth allocation (0.25 Mbps) combined with intelligent feature prioritization achieves 97.7% of the full-bandwidth detection performance. This finding suggests that bandwidth-efficient transmission strategies can substantially reduce the communication infrastructure requirements for large-scale V2X deployment without proportional accuracy loss, which is particularly relevant given the limited 30 MHz spectrum availability in the current C-V2X regulatory environment established by the FCC.

5.2. Limitations

Several limitations warrant discussion. The temporal decay function uses a fixed decay-rate parameter, $\lambda = 1.5$, calibrated via a grid search on the validation sets. An adaptive decay rate that adjusts to traffic density and vehicle speed could improve performance across diverse scenarios, as high-speed highway environments exhibit faster scene changes than congested urban intersections. The spatial relevance scoring assumes access to a coarse estimate of the ego vehicle's field-of-view boundaries, which may introduce errors in highly dynamic environments with rapid heading changes.

The current evaluation uses simulated latency and bandwidth constraints applied to datasets collected without actual communication impairments. Real V2X channels exhibit additional effects, including packet loss, jitter, and variable encoding quality, not captured in the present experimental setup. The performance improvements reported in this paper should be interpreted as upper bounds that may decrease in the presence of these additional communication impairments. Validation on datasets collected over real C-V2X links with authentic channel characteristics would strengthen the practical applicability of these findings.

Future work should explore integrating the proposed weight allocation with learned compression techniques to jointly optimize feature encoding and fusion weighting in an end-to-end manner. Extending the approach to multi-task cooperative perception, including joint detection and tracking, and evaluating performance under adversarial communication conditions, represents additional promising directions. The application of the proposed bandwidth partitioning strategy to emerging Vehicle-to-Infrastructure-to-Vehicle relay architectures could address scenarios in which direct V2V communication is infeasible due to range limitations, terrain obstacles, or dense urban building blockages.

References

[1]. U.S. Department of Transportation. (2024). Saving lives with connectivity: A plan to accelerate V2X deployment. U.S. Department of Transportation. https://www.transportation.gov/v2x

[2]. U.S. Department of Transportation. (2022). National roadway safety strategy. U.S. Department of Transportation. https://www.transportation.gov/NRSS

[3]. Federal Communications Commission. (2024). Second report and order on use of the 5.850–5.925 GHz band (FCC-24-123). Federal Communications Commission.

[4]. Xu, R., Xiang, H., Tu, Z., Xia, X., Yang, M.-H., & Ma, J. (2022). V2X-ViT: Vehicle-to-everything cooperative perception with vision transformer. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 107–124). Springer.

[5]. Xu, R., Xiang, H., Xia, X., Han, X., Li, J., & Ma, J. (2022). OPV2V: An open benchmark dataset and

fusion pipeline for perception with vehicle-to-vehicle communication. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (pp. 2583–2589). IEEE.

[6]. Lei, Z., Ren, S., Hu, Y., Zhang, W., & Chen, S. (2022). Latency-aware collaborative perception. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 316–332). Springer.

[7]. Wang, T.-H., Manivasagam, S., Liang, M., Yang, B., Zeng, W., & Urtasun, R. (2020). V2VNet: Vehicle-to-vehicle communication for joint perception and prediction. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 605–621). Springer.

[8]. Li, Y., Ren, S., Wu, P., Chen, S., Feng, C., & Zhang, W. (2021). Learning distilled collaboration graph for multi-agent perception. In Advances in Neural Information Processing Systems (NeurIPS), 34, 29541–29552.

[9]. Xu, R., Tu, Z., Xiang, H., Shao, W., Zhou, B., & Ma, J. (2022). CoBEVT: Cooperative bird's eye view semantic segmentation with sparse transformers. In Proceedings of the Conference on Robot Learning (CoRL) (pp. 989–1000). PMLR.

[10]. Liu, Y.-C., Tian, J., Glaser, N., & Kira, Z. (2020). When2com: Multi-agent perception via communication graph grouping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4106–4115). IEEE.

[11]. Hu, Y., Fang, S., Lei, Z., Zhong, Y., & Chen, S. (2022). Where2comm: Communication-efficient collaborative perception via spatial confidence maps. In Advances in Neural Information Processing Systems (NeurIPS), 35, 2398–2411.

[12]. Wei, S., Wei, Y., Hu, Y., Lu, Y., Zhong, Y., Chen, S., & Zhang, Y. (2023). Asynchrony-robust collaborative perception via bird's eye view flow. In Advances in Neural Information Processing Systems (NeurIPS), 36.

[13]. Yu, H., Tang, Y., Xie, E., Mao, J., Luo, P., & Nie, Z. (2023). Flow-based feature fusion for vehicle-infrastructure cooperative 3D object detection. In Advances in Neural Information Processing Systems (NeurIPS), 36.

[14]. Yang, D., Yang, K., Wang, Y., Liu, J., Xu, Z., Yin, R., Zhai, P., & Zhang, L. (2023). How2comm: Communication-efficient and collaboration-pragmatic multi-agent perception. In Advances in Neural Information Processing Systems (NeurIPS), 36.

[15]. Wang, T., Chen, G., Chen, K., Liu, Z., Zhang, B., Knoll, A., & Jiang, C. (2023). UMC: A unified bandwidth-efficient and multi-resolution based collaborative perception framework. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 17740–17750). IEEE.

[16]. Lu, Y., Li, Q., Liu, B., Dianati, M., Feng, C., Chen, S., & Wang, Y. (2023). Robust collaborative 3D object detection in presence of pose errors. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (pp. 4812–4818). IEEE.

[17]. Hu, Y., Peng, J., Liu, S., Ge, J., Liu, S., & Chen, S. (2024). Communication-efficient collaborative perception via information filling with codebook. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

[18]. Xu, Y., Li, L., Wang, J., Yang, B., Wu, Z., Chen, X., & Wang, J. (2025). CoDynTrust: Robust asynchronous collaborative perception via dynamic feature trust modulus. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). IEEE.

[19]. Lu, Y., Hu, Y., Zhong, Y., Wang, D., Chen, S., & Wang, Y. (2024). An extensible framework for open heterogeneous collaborative perception. In Proceedings of the International Conference on Learning Representations (ICLR).

[20]. Song, Z., Yang, L., Wen, F., & Li, J. (2025). TraF-Align: Trajectory-aware feature alignment for asynchronous multi-agent perception. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

[21]. Yu, H., Luo, Y., Shu, M., Huo, Y., Yang, Z., Shi, Y., Guo, Z., Li, H., Hu, X., Yuan, J., & Nie, Z. (2022). DAIR-V2X: A large-scale dataset for vehicle-infrastructure cooperative 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 21361–21370). IEEE.

[22]. Li, Y., Ma, D., An, Z., Wang, Z., Zhong, Y., Chen, S., & Feng, C. (2022). V2X-Sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. IEEE Robotics and Automation Letters, 7(4), 10914–10921.

[23]. Xu, R., Xia, X., Li, J., Li, H., Zhang, S., Tu, Z., Meng, Z., Xiang, H., Dong, X., Song, R., Yu, H., Zhou, B., & Ma, J. (2023). V2V4Real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception. In Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 13712–13722). IEEE.

[24]. Hong, S., Liu, Y., Li, Z., Li, S., & He, Y. (2024). Multi-agent collaborative perception via motion-aware robust communication network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

[25]. SAE International. (2022). J3224: V2X sensor-sharing for cooperative and automated driving. SAE International.