

SRA-CP: Spontaneous Risk-Aware Selective Cooperative Perception

Jiaxi Liu^a, Chengyuan Ma^{*a}, Hang Zhou^a, Weizhe Tang^a, Shixiao Liang^a, Haoyang Ding^b, Xiaopeng Li^a and Bin Ran^a

^aDepartment of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, 53706, Wisconsin, United States

^bSchool of Computer, Data and Information Sciences, University of Wisconsin-Madison, Madison, 53706, Wisconsin, United States

ARTICLE INFO

Keywords:

Cooperative perception
Vehicle-to-Everything
Object detection
Blind spot analysis
Spontaneous Risk-Aware Selective Cooperative Perception (SRA-CP)

ABSTRACT

Cooperative perception (CP) offers significant potential to overcome the limitations of single-vehicle sensing by enabling information sharing among connected vehicles (CVs). However, existing generic CP approaches need to transmit large volumes of perception data that are irrelevant to the driving safety, exceeding available communication bandwidth. Moreover, most CP frameworks rely on pre-defined communication partners, making them unsuitable for dynamic traffic environments. This paper proposes a *Spontaneous Risk-Aware Selective Cooperative Perception (SRA-CP)* framework to address these challenges. SRA-CP introduces a decentralized protocol where connected agents continuously broadcast lightweight perception coverage summaries and initiate targeted cooperation only when risk-relevant blind zones are detected. A perceptual risk identification module enables each CV to locally assess the impact of occlusions on its driving task and determine whether cooperation is necessary. When CP is triggered, the ego vehicle selects appropriate peers based on shared perception coverage and engages in selective information exchange through a fusion module that prioritizes safety-critical content and adapts to bandwidth constraints. We evaluate SRA-CP on a public dataset against several representative baselines. Results show that SRA-CP achieves less than 1% average precision (AP) loss for safety-critical objects compared to generic CP, while using only 20% of the communication bandwidth. Moreover, it improves the perception performance by 15% over existing selective CP methods that do not incorporate risk awareness.

1. Introduction

Multi-agent cooperative perception (CP) has emerged as a promising paradigm to overcome the limitations of single-agent sensing by enabling agents to share information with each other. In road traffic environments, a single vehicle's sensing capability is often obstructed by occlusions, resulting in blind zones that lead to hesitation in decision-making and increased collision risk with surrounding participants. These issues are particularly pronounced in scenarios such as unprotected left turns or pedestrians suddenly appearing from behind parked vehicles. With the rapid advancement of connected and automated vehicle (CAV) technologies, ensuring safe and complete perception becomes even more critical, especially for autonomous driving in complex environments [Zha et al., 2025]. In such contexts, CP offers valuable potential to enhance safety.

However, most existing CP studies remain limited to simulations or small-scale experimental setups conducted under ideal and controlled conditions. Achieving large-scale deployment of CP in real-world traffic still faces two major challenges. The first challenge lies in the gap between the massive volume of sensing data generated by CAVs and the limited bandwidth of vehicular communication networks [Hu et al., 2022]. For example, intermediate features extracted from onboard sensors at a rate of 5–20 Hz can produce up to 2MB per frame, translating to a potential transmission rate of 300 Mbps. Such data loads are far beyond what even advanced wireless systems (e.g., 5G) can support in dense environments, especially when vehicles attempt to transmit full perception data simultaneously, as the **Generic CP** shown in Figure 1 (a). Moreover, the problem is exacerbated in real-world traffic, where the number of dynamic agents is large and constantly changing. If every pair of agents were required to maintain real-time communication, the bandwidth burden would grow quadratically with the number of agents. In reality, most of the information being shared is unnecessary—an individual vehicle's local perception is often sufficient for safe driving in the majority of

*Corresponding author

ORCID(s): 0009-0001-2749-6435 (J. Liu); 0000-0002-6337-0450 (C. Ma*); 0000-0003-3286-341X (H. Zhou); 0000-0002-5264-3775 (X. Li)

SRA-CP: Spontaneous Risk-Aware Selective Cooperative Perception

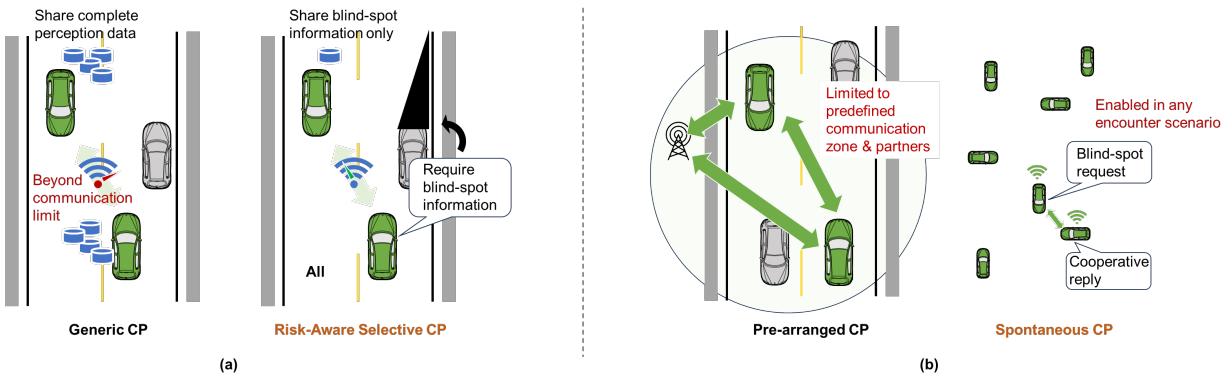


Figure 1: Comparison between (a) **Generic CP** with full-time information exchange VS the proposed **Risk-aware selective CP** activated by risky blind-spot events; and (b) **Pre-arranged CP** constrained by predefined communication partners VS **Spontaneous CP** enabling dynamic ad-hoc cooperation in arbitrary encounter situations.

situations. Even in CP-required cases, not all detected elements need to be transmitted. A more efficient strategy is to share only the information that is both unseen by the receiving vehicle and potentially hazardous to its driving decisions. This observation motivates the concept of **Risk-aware selective CP**, where each vehicle evaluates the risk level of objects it perceives and only transmits those that satisfy two conditions simultaneously: (i) the object lies within another vehicle’s blind zone, and (ii) the object poses a potential safety risk. For instance, in Figure 1, a parked roadside vehicle obstructs a portion of the scene, and the oncoming vehicle provides supplementary information to complete perception. While recent studies have explored selective CP strategies that only share blind-zone content—resulting in significant communication reduction compared to generic CP [Qiu et al., 2025]—they still do not account for the traffic risk relevance of shared content. Our previous work validated that only a small fraction (0.1%) of driving scenarios actually require CP [Ma et al., 2025]. Thus, the insight behind risk-aware selective CP serves as the foundation for the communication-efficient strategy proposed in this study.

The second challenge concerns how to construct communication pathways in a scalable and dynamic traffic system. To the best of our knowledge, most existing CP studies are conducted within **Pre-arranged** communication zones and among predefined partner vehicles, as illustrated in Figure 1(b). While pre-arranged communication enables stable point-to-point connections under idealized network assumptions—and even allows for global optimization of communication topologies and grouping strategies [Dong et al., 2022]—these setups are typically limited to experimental testbeds with a fixed number of specified agents. They fail to generalize to open-world traffic environments, where a small number of unfamiliar connected vehicles may encounter each other spontaneously across large spatial areas and at unpredictable times. This limitation highlights the urgent need for a decentralized and self-organized communication mechanism. To address this, we propose a **Spontaneous CP** framework that builds upon the selective CP principle. Each connected vehicle operates independently, broadcasting minimal information about its own perception coverage. Only when a risk-relevant blind zone is detected does it initiate a CP request. Neighboring vehicles, upon receiving the request, respond cooperatively if they are capable of contributing, as illustrated in Figure 1(b). This mechanism leverages a key principle: connected vehicles can identify whether they can be assisted in blind-zone completion by evaluating other vehicles’ relative positions and their shared perception coverage. As a result, the handshake process is realized through lightweight broadcasting during regular operation and precise, selective information exchange triggered only when necessary.

To bridge the above two gaps, we propose a **Spontaneous Risk-Aware Selective Cooperative Perception (SRA-CP)** framework. It introduces a spontaneous collaboration mechanism composed of two modes: a routine mode, where vehicles continuously broadcast only their perception coverage maps; and a triggered mode, where a vehicle initiates CP only when it detects a risk-relevant blind zone, and neighboring connected agents are capable of assisting. Built upon this mechanism, we develop a **risk-aware hierarchical perception fusion model** that ensures efficient CP by adaptively prioritizing critical information within any available communication bandwidth. The model consists of four key components: a shared feature encoder, a risk-aware communication module, a dual-attention fusion network, and a multi-task decoder. This architecture enables vehicles to selectively fuse the most important perceptual features based on spatial occlusion and safety relevance under any given communication constraints. We evaluate the proposed

framework on a public dataset by comparing it against three baselines: generic CP, a state-of-the-art selective CP method without risk awareness, and a no-CP setup, in terms of communication cost and Average Precision (AP) for key-object detection. Results demonstrate that SRA-CP achieves comparable or superior performance with significantly reduced communication overhead. Specifically, compared to generic CP, SRA-CP reduces the transmission volume to 80% while incurring only a 0.1 drop in AP. Compared to the selective CP baseline without risk modeling, SRA-CP improves AP for critical objects by 15% under the same communication budget, showcasing its potential for scalable deployment in large-scale, dynamic traffic environments.

The main contributions of this paper are as follows:

- We address the bandwidth bottleneck in multi-agent CP by proposing a **risk-aware selective CP** strategy, which prioritizes the transmission of perceptual elements based on their impact on driving safety. This approach enables efficient use of limited communication resources under varying bandwidth constraints.
- We propose the **Spontaneous Risk-Aware Selective Cooperative Perception (SRA-CP)** framework, which supports dynamic, on-demand handshakes between agents without predefined regions or communication partners. This design enables scalable and low-cost CP in large-scale, real-world traffic environments with self-organizing connected agents.
- We validate the proposed framework on a public dataset and show that **SRA-CP** achieves comparable performance to generic CP while using only 20% of the communication volume, with less than 1% drop in AP. Compared to a cutting-edge selective CP baseline that does not consider driving risk, SRA-CP improves critical object detection accuracy by 15%.

2. Related work

2.1. Cooperative Perception (CP)

CP allows multiple agents to share perceptual information to achieve a more complete understanding of their surroundings. This paradigm addresses key limitations of single-agent perception such as occlusion and limited sensing range [Chen et al., 2019a, Liu et al., 2020]. There could be different downstream perception tasks to be fulfilled with CP, such as 3D object detection [Xu et al., 2023, Li and Pei, 2024, Xiang et al., 2024, Yu et al., 2023], lane detection [Jahn et al., 2024, El Boukili et al., 2025], object tracking [Chiu et al., 2024, Zimmer et al., 2024, Zhong et al., 2025]. CP has been implemented through various fusion schemes, including early [Chen et al., 2019b, Yang et al., 2025], intermediate [Liu et al., 2020, Wang et al., 2020, Xu et al., 2022a], and late fusion [Liu et al., 2024, Sarlak et al., 2025]. Early fusion directly shares raw sensor data (e.g., LiDAR point clouds or images), aligns them in a common coordinate frame, and jointly processes the fused measurements through a single perception network. Intermediate fusion exchanges intermediate feature maps extracted by each agent's backbone. These features are spatially aligned and combined before the detection head. Late fusion transmits only high-level perception outputs such as object boxes or tracks, which are then associated and merged at the cooperative layer, offering low communication cost at the expense of reduced ability to recover missed local detections.

Although these schemes achieve a certain trade-off between communication cost and perception fusion performance, the widespread transmission of perceptual information in multi-agent scenarios remains a significant challenge.

2.2. Selective Information Sharing

A major challenge in CP is reducing the communication burden while maintaining perception quality. Recent work such as Where2comm [Hu et al., 2022] has proposed to use spatial confidence maps to identify perceptually important regions and selectively transmit only the features from those areas. This strategy improves the perception accuracy under limited bandwidth by avoiding indiscriminate sharing of all information. There are also other CP works that share the same thoughts [Yang et al., 2023a, Liu et al., 2020, Yang et al., 2023b]. However, perceptual accuracy alone is not a sufficient criterion in driving scenarios. In practice, much of the perceptual improvement may not contribute to driving decisions or safety [Ma et al., 2025, Van Brummelen et al., 2018, Pan et al., 2024, Gao et al., 2024]. For example, perceiving a distant vehicle with higher precision may not change the ego vehicle's behavior. Thus, a key limitation of current selective strategies is the lack of risk-awareness. They fail to distinguish between information that is perceptually useful and information that is safety-critical. A risk-aware selection mechanism is needed to ensure that perception elements that are both unseen and pose potential safety risks are shared.

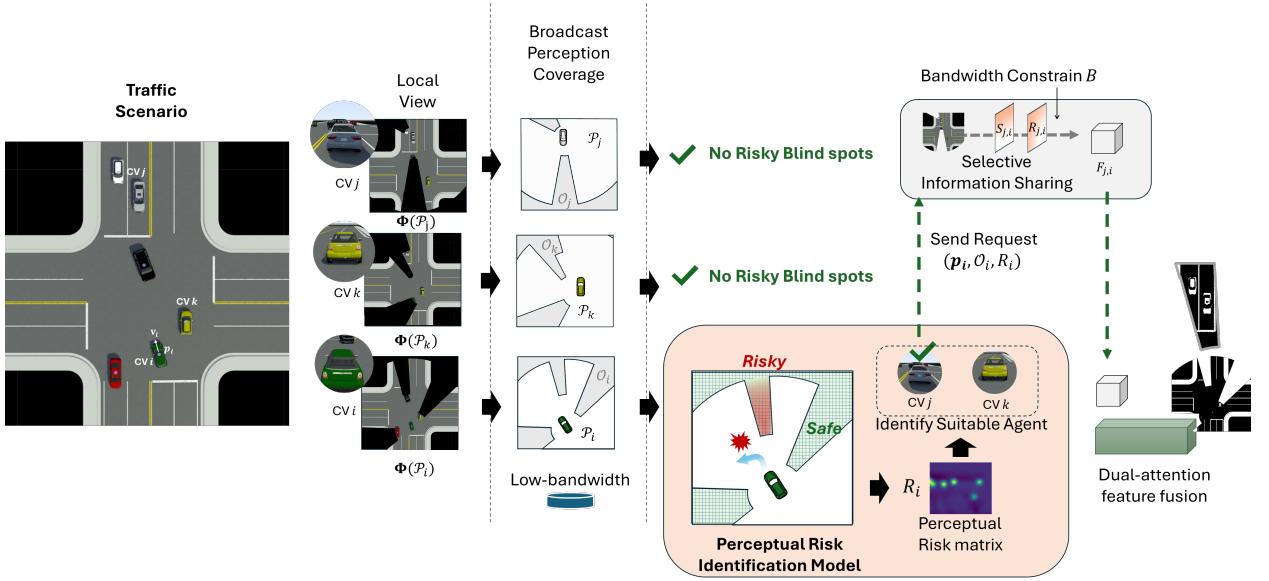


Figure 2: Spontaneous Risk-Aware Selective Cooperative Perception (SRA-CP)

2.3. Communication Paradigms for CP

The communication paradigm for CP is also an important topic in CP. In most cases, prior studies assume a limited and fixed set of collaborators operating within a bounded area. These assumptions simplify the design of interaction protocols and enable direct coordination [Feng et al., 2018, Yu et al., 2019]. Most existing CP frameworks also adopt such settings [Chen et al., 2019a, Liu et al., 2020]. However, these conditions are difficult to satisfy in real-world traffic environments, where agents are numerous, highly dynamic, and distributed across large spatial regions. In addition, only a small fraction of encounters truly require cooperation, and participating agents are often unfamiliar with one another. These limitations highlight the need for a scalable, deployable, and self-organized CP communication paradigm. Recent work in agentic AI has begun to explore the notion of *spontaneous cooperation* [Wu et al., 2024, Godhwani et al., 2025, Mirsky et al., 2022], where collaboration emerges dynamically based on local context and shared objectives. Building on this insight, our proposed SRA-CP framework leverages the property that connected vehicles can share their perception coverage, allowing other agents to identify opportunities to fulfill blind-zone completion needs. This enables the spontaneous formation of cooperation links without prior coordination or global knowledge.

3. Problem Description

As illustrated in Figure 2, we consider a dynamic road network with multiple Connected Vehicle (CV) agents indexed by e, j, k , in which e denotes the ego vehicle. At a certain time t (all variables hereafter are defined at time t unless otherwise specified), each vehicle has a physical state represented by its position $\mathbf{p}_e = (x_e, y_e)$ and velocity $\mathbf{v}_e = (v_e^x, v_e^y)$, taking ego vehicle as an example. For a given connected agent e , its perception range from a bird's-eye view (e.g., the spatial coverage of LiDAR) is denoted by \mathcal{P}_e , with the corresponding sensed information $\Phi(\mathcal{P}_e)$. The blind zone—areas not observable by the agent—is denoted as \mathcal{O}_e . In most cases, such blind zones have a negligible impact on driving safety, as illustrated by \mathcal{O}_j and \mathcal{O}_k in Figure 2. However, in some cases, such as a vehicle approaching from the opposite direction within \mathcal{O}_e that affects a left-turn decision, the blind zone can pose a significant risk. This study focuses on identifying such risky blind zones and selectively completing them via CP with limited communication bandwidth.

We design the SRA-CP framework in which the ego connected vehicle e continuously broadcasts its own position \mathbf{p}_e , velocity \mathbf{v}_e , and perception coverage \mathcal{P}_e (requiring only low bandwidth). It then receives broadcasted data from nearby vehicles within its communication zone \mathcal{Z}_e —a circular area of radius l_c —which includes the set of neighboring agents \mathcal{W}_e . Based on this shared information, the agent determines whether it has a risky blind zone that can be supplemented by any surrounding vehicle, and if so, initiates a spontaneous CP handshake and performs cooperative

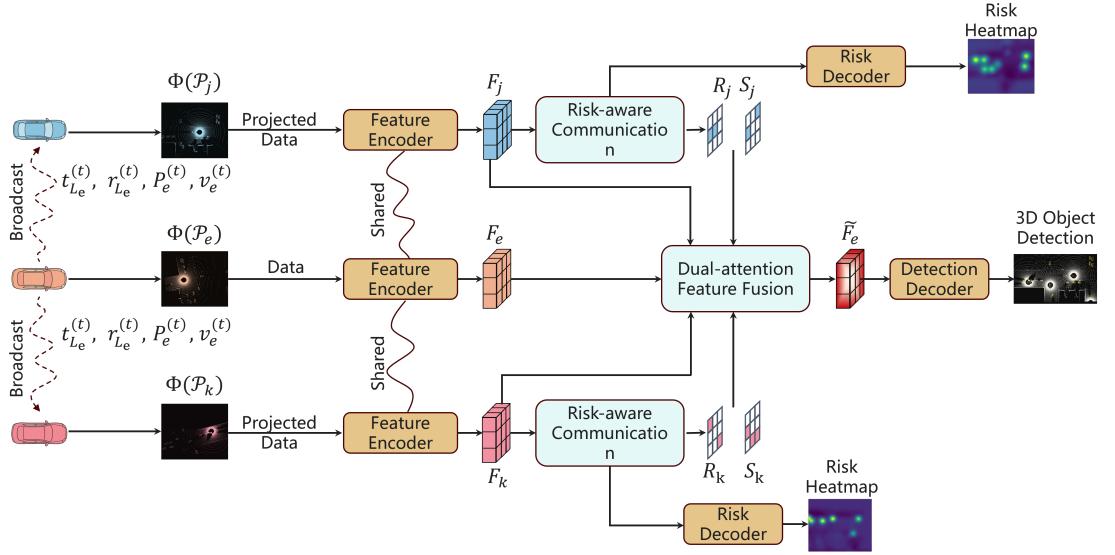


Figure 3: End-to-end architecture of Selective Information Sharing and Fusion. Each co-operative vehicle $i \in \{e, k, j\}$ projects its raw point cloud $\Phi(\mathcal{P}_i)$ to the ego Bird's-Eye-View (BEV) frame and encodes it through a shared feature-encoder, yielding F_i . **Risk-aware communication** (Sec. 4.3.2) attaches two light-weight masks—the spatial mask S_i and the risk mask R_i —to the feature map and broadcasts only these three tensors, avoiding transmission of raw point clouds. The ego car receives the partner streams and performs **dual-attention feature fusion** (Sec. 4.3.3): a safety-focused selector prunes partner features with (S_i, R_i) and a location-wise multi-head attention block aligns the surviving cells with the ego map F_e , producing \tilde{F}_e . Finally, two heads operate on \tilde{F}_e : (i) a Risk Decoder refines a dense risk heat-map, and (ii) a Detection Decoder outputs 3-D bounding boxes.

fusion. Specifically, the ego connected vehicle e first evaluates the risk level of its perception blind zones using the proposed **perceptual risk identification** model, resulting in a perceptual risk matrix \mathcal{R}_e . If no risky blind zones are detected (e.g., as in the case of other CVs j and k), the process at t terminates. If risky blind zones are identified, the vehicle proceeds to select an appropriate target connected agent. Based on the shared perception coverage from neighboring agents (e.g., \mathcal{P}_j and \mathcal{P}_k), the vehicle determines whether any connected agent can compensate for its occluded regions (e.g., agent j in the illustrated case). Note that if no suitable connected agents are available to provide blind-zone compensation, the ego vehicle relies solely on its own onboard perception for decision-making—e.g., by stopping to continue observation. Once a candidate is selected, ego vehicle e sends a CP request to agent j , including its current position \mathbf{p}_e , blind zone \mathcal{O}_e , and the computed risk matrix \mathcal{R}_e . Upon receiving the request, agent j invokes the proposed **selective information sharing** model, which, under the given bandwidth constraint B_{bytes} , selects and transmits the most informative features $F_{j,e}$ to supplement agent e 's perception. Finally, agent e performs cooperative fusion via a **dual-attention feature fusion** model to integrate the received features and complete its understanding of the occluded region.

In this study, we focus on CP using LiDAR data, which provides accurate geometric structure, consistent performance under varying illumination, and reliable spatial measurements for dynamic traffic environments. These properties make LiDAR particularly suitable for blind-zone estimation and risk-aware perception. It is worth noting that the proposed framework is modality-agnostic. Although our implementation uses LiDAR as the primary sensing modality, the methodology can be extended to other perception inputs, such as video data.

3.1. Notation and Symbols

Table 1: Notation used throughout the paper (unified).

Symbol	Type	Meaning / Unit
<i>Sets, indices, regions</i>		
t	scalar	Time index.
i, j, k	index	Generic agent indices.
e	index	Ego agent.
$\mathbf{p}_i = (x_i, y_i)$	vector	2D position of agent i in BEV/ego frame (at time t , unless stated otherwise).
$\mathbf{v}_i = (v_i^x, v_i^y)$	vector	2D velocity of agent i .
\mathcal{P}_i	set	Perception coverage (field of view, FoV) of agent i .
$\Phi(\cdot)$	map	Region → perceived sensory information; e.g., $\Phi(\mathcal{P}_i)$ is the perception information of agent i .
\mathcal{O}_i	set	Blind zone of agent i .
\mathcal{Z}_i	set	Local communication region of agent i (disk of radius l_c).
l_c	scalar	Communication radius defining \mathcal{Z}_i (m).
\mathcal{W}_i	set	Neighboring connected agents within \mathcal{Z}_i .
<i>SRA-CP protocol</i>		
$\rho_{i,j}$	scalar	Pairwise collision risk score between i and j .
τ_r	scalar	Threshold on ρ to trigger risk-aware sharing.
\mathcal{R}	matrix	Risk matrix collecting $\rho_{i,j}$.
$R_i, R_i^{(d)}, R_i^{(s)}, R_i^n$	scalar	Object i 's total risk and distance/speed/intersection components.
\hat{R}_i	scalar	Clipped/normalized risk of object i in [0, 1].
R_{gt}	map	Ground-truth risk heatmap.
D_e	set	Potentially dangerous agents for ego e , selected from \mathcal{W}_e using \mathcal{R} .
<i>Perceptual risk identification model</i>		
$\mathbf{u} = (x, y)$	vector	2D BEV grid cell center in ego coordinates.
$\mathcal{G}_{\mathbf{u}}$	grid	BEV grid cell centered at $\mathbf{u} = (x, y)$.
$o(\mathbf{u}) \in [0, 1]$	field	Occupancy at BEV cell \mathbf{u} .
Π_{BEV}	op	3D→BEV projection operator.
$\kappa, \sigma(\cdot)$	func	Smoothing kernel; squashing function for occupancy.
ϑ	angle	Ray azimuth (rad).
r	scalar	Range of BEV grid cell (m).
$\mathbf{r}(s; \vartheta)$	curve	Ray parameterization along azimuth ϑ .
$T(\mathbf{u})$	scalar	Line-of-sight transmittance to \mathbf{u} .
$\lambda, \Delta s, K$	scalars	Beer-Lambert attenuation; step; number of samples along a ray.
$\chi_{\text{fov}}(\mathbf{u})$	gate	FoV gate {0, 1}.
$P_{\text{occ}}(\mathbf{u})$	prob	Occlusion probability at \mathbf{u} .
$\tau_{\text{occ}}, K_t, \tau_t$	scalars	Occlusion threshold; number of temporal frames; temporal consensus threshold.
$\mathcal{O}_e(\mathbf{u}), \tilde{\mathcal{O}}_e$	mask	Instantaneous and stabilized blind-zone masks.
$\mathbf{T}_{e \leftarrow w}$	matrix	Rigid transform from world to ego frame.
z	scalar	Vertical coordinate (height) in the ego frame.
<i>Selective information sharing and fusion</i>		
$F_i \in \mathbb{R}^{C \times H \times W}$	tensor	BEV feature map of agent i ; C channels, $H \times W$ the height/width of grid.
$C_{s,j}, C_{r,j}$	map	Spatial- and risk-confidence maps on partner j .
$S_j, R_j \in \{0, 1\}^{H \times W}$	mask	Spatial / risk masks (binary).
\tilde{F}_i	tensor	Masked feature patch to transmit from partner i .
$F_{j,e}$	tensor	Partner j 's selected feature patch transmitted to ego e .
$f_{\text{enc}}, f_{\text{dec}}$	net	Shared encoder; multi-task decoder.
$(\hat{C}, \hat{B}, \hat{R})$	out	Class scores, 3D boxes, refined risk heatmap.
K_{sel}	scalar	Top- K selected cells for transmission.
$g_{\text{sp}}(\mathbf{u}), g_{\text{risk}}(\mathbf{u})$	score	Spatial/risk gains used for selection.
$g(\mathbf{u}), \alpha$	score	Combined gain and its mixing weight $\alpha \in [0, 1]$.

Continued on next page

Table 1 (continued).

Symbol	Type	Meaning / Unit
P_e	path	Planned trajectory used by the risk head of ego agent e .
<i>Training objective and evaluation</i>		
$\mathcal{L}_{\text{total}}$	loss	Total training loss.
\mathcal{L}_{det}	loss	Detection loss.
$\lambda_{\text{risk}}, \lambda_{\text{comm}}$	weight	Weights for risk regression and communication penalty.
$\phi(\text{usage}; \text{target})$	penalty	Hinge-style penalty on over-usage of bytes.
B_{bytes}	bytes	Target per-link byte budget.
h_{hdr}	bytes	Header/metadata overhead per message.
$b_{\text{idx}}, b_{\text{feat}}, b_{\text{cell}}$	bytes	Bytes per cell index / per feature value / per cell ($b_{\text{cell}} = b_{\text{idx}} + C b_{\text{feat}}$).
U	bytes	Actual bytes used in a batch.
B_{batch}, L_b	count	Batch size; number of agents in sample b .
$M_{l,i,j}^{(b)} \in \{0, 1\}$	mask	Binary selection mask for sample b .
$\text{AP}, \text{3DAP}(\theta)$	metric	AP and 3D AP at IoU threshold θ .
$TP(\theta), FP(\theta), FN(\theta)$	count	True/false positives and false negatives at θ .
$\mathcal{I}_{\text{risk}}(\tau)$	set	Subset filtered by risk threshold τ for Risk-AP.
$\Delta\text{Risk-AP}/\text{KB}$	metric	$\Delta\text{Risk-AP}$ per KB ($\Delta\text{Risk-AP}/\text{KB}$).
<i>Risk label generation</i>		
$\alpha_d, \alpha_s, \alpha_n$	scalar	Weights for distance-, speed-, and intersection-based risk components in the overall risk score R .
λ_d, λ_n	scalar	Decay rates for distance-based and intersection-based risk terms.
m	index	index of intersections.
\mathbf{q}_m	vector	Center location of the m -th intersection.
Q	set	Set of all intersection center locations.
v_i	scalar	Speed magnitude of object i .
ϵ	scalar	Small positive constant to avoid division by zero in the speed-based risk normalization.

4. Methodology

As mentioned in the previous section, the proposed SRA-CP framework is designed to operate in two phases: during normal operation, each vehicle broadcasts basic perception coverage information with minimal bandwidth; when a risk-relevant blind zone is detected, it initiates a targeted CP link and transmits only the most critical information within the available communication bandwidth. The framework relies on a **perceptual risk identification model** to assess the risk level of blind zones. Upon identifying a suitable cooperative agent, the responder employs a **selective information sharing model** to determine which features to transmit under bandwidth constraints. The receiving agent then performs cooperative fusion using a **dual-attention feature fusion model** to produce an enhanced perception result. The following subsections detail each of these four key components.

4.1. SRA-CP Protocol

The core idea of the proposed SRA-CP protocol is as follows: at each time t , the ego vehicle e periodically broadcasts a compact coverage map of \mathcal{P}_e to all nearby agents \mathcal{W}_e within \mathcal{Z}_e . This map summarizes which BEV cells are currently visible and which are likely occluded (Sec. 4.2), without exposing raw sensor data. Each neighbor does the same, enabling every agent to infer who can potentially compensate for its blind zones. When a risky blind area is detected, the ego triggers an on-demand handshake with one suitable partner and proceeds with selective sharing and fusion under the current byte budget. In practice, a risk threshold τ_r determines whether the detected blind-zone risk warrants initiating the cooperative handshake. Then, based on the received coverage maps, each agent constructs an inter-object risk matrix $\mathcal{R} = [\rho(e, i) \mid i \in \mathcal{W}_e]$ by evaluating pairwise risks. From this vector, the potentially dangerous set \mathcal{D}_e is identified. If an agent $i \in \mathcal{D}_e$ also lies in \mathcal{O}_e , the partner transmits only the features covering that region to assist perception completion.

For example, as illustrated in the intersection scenario in Figure 2, there are six vehicles, among which i, j , and k are connected agents. In the first step, each connected agent broadcasts its local perception coverage $\mathcal{P}_i, \mathcal{P}_j$, and

\mathcal{P}_k to others. Since only coverage maps are shared—without detailed perception content—this step incurs negligible communication overhead.

Next, each agent performs an inter-object risk estimation over the observed objects in the scene and generates a risk vector \mathcal{R} to estimate whether they need external information from other agents to help with their perception. In this scenario, agent k and agent j find no risky blind spot, so they do not need further external information from other agents. However, agent i finds it can not see the potentially risky objects in the blind zone of the black vehicle, which is risky to its driving intention, and agent j finds it can help the detection of agent i . Therefore, agent j sends the information of the potentially risky zone to agent i to complete its perception.

4.2. Perceptual risk identification model

The **Perceptual Risk Identification Model** takes the individual perception $\Phi(\mathcal{P}_i)$ as input and produces a risk matrix \mathcal{R}_i over the blind zone \mathcal{O}_i , indicating the safety-critical importance of each location with respect to the ego vehicle's driving decisions.

SRA-CP requires a light-weight, geometry-based estimate of the ego vehicle's blind zones to prioritize compensation from partners. We adopt a BEV visibility model that is fast, rule-based, and admits a continuous formulation for analysis. Let $\mathcal{G}_{\mathbf{u}}$ denote a BEV grid with cell centers $\mathbf{u} = (x, y)$ in ego coordinates, z is the vertical coordinate of \mathbf{u} in the ego frame used for 2.5D occupancy computation, ego pose $\mathbf{T}_{e \leftarrow w}$ (world \rightarrow ego), and a 2.5D occupancy field $o(\mathbf{u}) \in [0, 1]$ obtained from the LiDAR sweep $\Phi(\mathcal{P}_i)$ by height-thresholding and kernel density aggregation:

$$o(\mathbf{u}) = \sigma \left(\max_{z \in [z_{\min}, z_{\max}]} \kappa * \sum_{\mathbf{p} \in \mathcal{L}} \delta(\Pi_{\text{BEV}}(\mathbf{T}_{e \leftarrow w} \mathbf{p}) - (\mathbf{u}, z)) \right), \quad (1)$$

where Π_{BEV} projects 3D points to the BEV cell, κ is a spatial smoothing kernel, and σ is a squashing function ensuring $o \in [0, 1]$ (e.g., $\sigma(a) = 1 - e^{-a}$). For a BEV direction $\vartheta = \text{atan}2(y, x)$ and range $r = \|\mathbf{u}\|_2$, define the ray parameterization $\mathbf{r}(s; \vartheta) = s [\cos \vartheta, \sin \vartheta]^T$, $s \in [0, r]$. The line-of-sight transmittance to \mathbf{u} is modeled by a Beer–Lambert integral over occupancy:

$$T(\mathbf{u}) = \exp \left(- \int_0^r \lambda o(\mathbf{r}(s; \vartheta)) ds \right), \quad \lambda > 0, \quad (2)$$

with discrete approximation on grid steps Δs :

$$T(\mathbf{u}) \approx \exp \left(- \lambda \Delta s \sum_{k=0}^K o(\mathbf{r}(k \Delta s; \vartheta)) \right), \quad K \Delta s \approx r. \quad (3)$$

Cells outside the sensor field-of-view (FOV) or range are treated as fully occluded by an FOV gate $\chi_{\text{fov}}(\mathbf{u}) \in \{0, 1\}$; we define the occlusion probability and the binary blind-zone mask as

$$P_{\text{occ}}(\mathbf{u}) = 1 - \chi_{\text{fov}}(\mathbf{u}) T(\mathbf{u}), \quad \mathcal{O}_{\text{e}}(\mathbf{u}) = \mathbb{I}[P_{\text{occ}}(\mathbf{u}) > \tau_{\text{occ}}], \quad (4)$$

with threshold $\tau_{\text{occ}} \in (0, 1)$. To reduce flicker, we temporally stabilize the mask by warping the last K_t frames into the current ego frame using odometry and taking a robust union:

$$\bar{\mathcal{O}}_{\text{e}}(\mathbf{u}) = \mathbb{I} \left[\frac{1}{K_t} \sum_{t'=t-K_t+1}^t \mathcal{O}_{\text{e}}^{(t')}(\mathbf{T}_{e \leftarrow e(t')}(\mathbf{u})) > \tau_t \right]. \quad (5)$$

The mask $\bar{\mathcal{O}}_{\text{e}}$ is used as a compressed coverage summary and to increase selection gains in risky blind zones. Specifically, let $g_{\text{sp}}(\mathbf{u})$ and $g_{\text{risk}}(\mathbf{u})$ are spatial/risk scores (Sec. 4.3.2), the budgeted gain can be

$$g(\mathbf{u}) = \alpha g_{\text{sp}}(\mathbf{u}) g_{\text{risk}}(\mathbf{u}) + (1 - \alpha) \bar{\mathcal{O}}_{\text{e}}(\mathbf{u}) g_{\text{risk}}(\mathbf{u}), \quad \alpha \in [0, 1], \quad (6)$$

which prioritizes risky and occluded cells under a rate/byte budget.

4.3. Selective Information Sharing and Fusion

The **Selective Information Sharing and Fusion Model** describes the full pipeline from selecting the target agent for cooperation, to determining which features to share, and finally to integrating the received features on the ego vehicle. The overall framework is illustrated below.

Selective Information Sharing and Fusion is the model layer that operationalizes the SRA-CP contract: given the neighbors and a per-link budget, it learns what to communicate and how to fuse. Concretely, Selective Information Sharing and Fusion produces lightweight spatial and risk masks, sparsifies partner features under a given communication budget, and performs risk-aware fusion on the ego agent using the fused ego map (Figure 3). This converts communication bandwidth into safety-relevant detections by prioritizing risky \times occluded regions. The pipeline has four building blocks:

1. **Shared feature encoder** $f_{\text{enc}}(\cdot)$ that transforms each LiDAR sweep $\Phi(\mathcal{P}_i)$ into a BEV feature tensor $F_i \in \mathbb{R}^{C \times H \times W}$;
2. **Risk-aware communication module** (Figure 4) that derives a spatial mask S_i and a risk mask R_i from F_i and these masks will be used as a reference in the Dual-attention feature fusion process to decide which features should be shared;
3. **Dual-attention feature fusion module** (Figure 5) that selects the features $\{\tilde{F}_j\}_{j \neq e}$ to be shared to the ego vehicle based on the spatial mask S_j and the risk mask R_j and transmits $\{\tilde{F}_j\}_{j \neq e}$ to the ego vehicle and merges them with the ego feature map F_e and outputs the fused representation \tilde{F}_e in the ego vehicle's coordinate system;
4. **Multi-task decoder** that predicts both 3D bounding boxes and a dense risk heat-map.

4.3.1. Feature Encoding

During the training process each CV i encodes its LiDAR sweep $\Phi(\mathcal{P}_i)$ with a shared PointPillar BEV encoder Lang et al. [2019] in the same structure, yielding $F_i \in \mathbb{R}^{C \times H \times W}$. Features are expressed in a common ego BEV frame using the known pairwise poses which is transmitted with the coverage map. The backbone within the same structure feeds two light heads to derive a spatial confidence map and a risk map. The spatial confidence map stores the confidence score of the features from the spatial perspective, which means which feature is spatially important for perception. And the risk confidence map stores the confidence score of the features from the traffic risk perspective, which means which feature is essential in terms of traffic importance. Both of these two confidence maps will guide the communication process to choose which features to communicate and the later fusion.

4.3.2. Risk-Aware Communication

The aim of this module is to reduce the communication bandwidth while preserving the balance of safety relevance and spatial relevance. Each partner summarizes where its features are informative (spatial saliency) and where they are safety-critical for the ego (risk), then the CVs will combine the scores together to select which features are more important for the ego vehicle and they will send only the most important parts under a given budget.

On each partner j , two lightweight heads process F_j (Figure 4):

- Spatial-confidence head outputs $C_{s,j}$.
- Risk-confidence head outputs $C_{r,j}$.

Under a given communication budget, adaptive sampling will perform Top-K selection over non-ego grid cells based on their spatial and risk scores separately in the scene and then it will produce binary masks $S_j, R_j \in \{0, 1\}^{H \times W}$. Under a given per-link budget, SRA-CP combines the two cues (union) and serializes only cells within the mask as in the safety-focus feature selection part in the Figure 5. In practice this masks the feature map:

$$\tilde{F}_j = F_j \odot (S_j \vee R_j),$$

so only areas that are spatially salient and safety-critical are transmitted. This concentrates communication on occluded or risky regions that matter for decision-making, keeps privacy by avoiding raw points, and gracefully adapts to tighter budgets by shrinking the selected area.

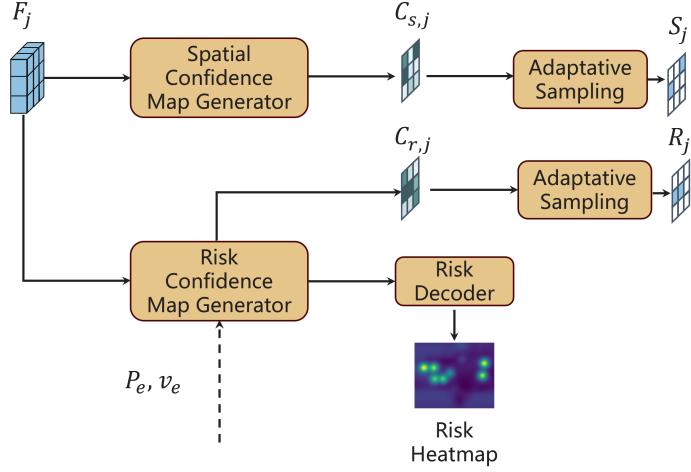


Figure 4: Risk-aware communication pipeline executed on each partner vehicle j . The shared feature map F_j is processed by two light-weight heads: (i) *Spatial-confidence map generator* produces a spatial confidence map $C_{s,j}$ that highlights semantically important cells; an adaptive sampling module is used to select a sparse binary spatial mask based on scenario S_j for transmission. (ii) *Risk-confidence map generator* uses F_j together with the ego planned trajectory P_e and speed v_e to compute a risk map $C_{r,j}$. Adaptive sampling converts it into a binary risk mask R_j . Both masks (S_j, R_j) are sent to the ego vehicle, while a miniature Risk Decoder can optionally convert $C_{r,j}$ into a dense risk heat-map for supervision training.

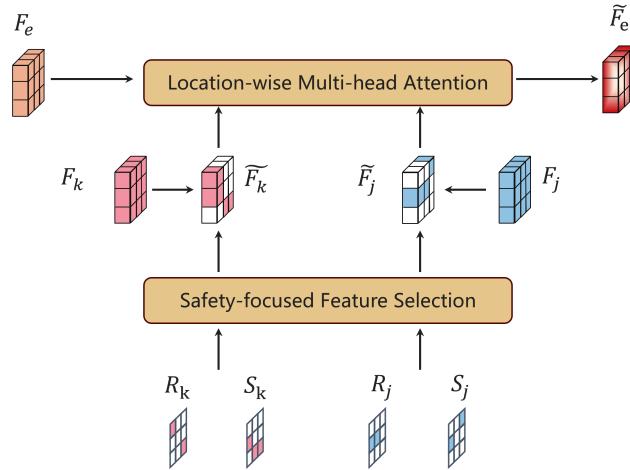


Figure 5: Dual-attention feature fusion. Remote feature tensors F_k and F_j are first filtered by a Safety-focused Feature Selection block that combines each partner's spatial mask S_i and risk mask R_i , yielding sparsified maps \tilde{F}_k and \tilde{F}_j . The ego map F_e and the sparsified partner maps are then fused by a location-wise multi-head attention module that performs per-cell key-query interactions, producing an enriched representation \tilde{F}_e . This two-stage design discards bandwidth-hungry, low-value regions before attention, so both communication and computation focus on areas that are simultaneously safety-critical and semantically informative. During this process, only three low-bandwidth tensors (\tilde{F}_j, S_j, R_j) leave the vehicle, preserving privacy and saving channel capacity.

4.3.3. Dual-Attention Feature Fusion

At the ego agent, the masked partner maps $\{\tilde{F}_j\}$ and the local map F_e are fused in two stages (Figure 5). First, a safety-focused selector re-applies (S_j, R_j) to suppress any residual clutter and enforce budget consistency. Second, We fuse ego and partner features in a location-wise manner, for each BEV cell \mathbf{u} , the ego feature provides the query, while only partners that selected this cell contribute keys and values. This yields an attention distribution over the

relevant collaborators, ensuring that information is aggregated only where communication actually provided features. A residual update then produces the fused representation \tilde{F}_e .

To handle small spatial misalignment, the module can optionally attend within a local window around \mathbf{u} , but still restricts computation to cells indicated by partner selection. This keeps the complexity proportional to the number of communicated cells, making the fusion efficient under sparse CP.

This kind of design limits computation to a small set of safety-relevant cells, improves alignment under occlusion, and avoids flooding the decoder with low-value regions. When no partner data arrives, the module naturally falls back to the ego features without architectural changes.

4.3.4. Budgeted Selection and Training Objective

The fused tensor is decoded as $(\hat{C}, \hat{B}, \hat{R}) = f_{\text{dec}}(\tilde{F}_e)$, where \hat{C} are class scores, \hat{B} are 3-D boxes, and \hat{R} is the refined risk heat-map.

Budgeted selection. Given a budget per link, Selective Information Sharing and Fusion Model ranks non-ego BEV cells by the gain $g(\mathbf{u})$ (Sec. 4.2) and selects the top K_{sel} cells subject to the budget. Let the per-cell byte cost be $b_{\text{cell}} = b_{\text{idx}} + C \cdot b_{\text{feat}}$ and header overhead h_{hdr} . For a byte budget B_{bytes} , the capacity in cells is

$$K_{\text{sel}} = \max\left(0, \left\lfloor \frac{B_{\text{bytes}} - h_{\text{hdr}}}{b_{\text{cell}}} \right\rfloor\right), \quad (7)$$

where, h_{hdr} is a fixed header cost (bytes).

Budget-aware training. To make the bandwidth–accuracy trade-off controllable at training time, we add a communication regularizer that penalizes over-usage relative to a target budget; this does not change runtime budget, but shapes the model’s selection behavior. The total loss denoted by $\mathcal{L}_{\text{total}}$ is as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \lambda_{\text{risk}} \|\hat{R} - R_{\text{gt}}\|_2^2 + \lambda_{\text{comm}} \phi(U; B_{\text{bytes}}). \quad (8)$$

where:

- **Detection loss.** $\mathcal{L}_{\text{det}} = \mathcal{L}_{\text{conf}} + \mathcal{L}_{\text{reg}}$ is the standard detection loss. The classification term $\mathcal{L}_{\text{conf}}$ is a focal loss with $\alpha = 0.25$, $\gamma = 2.0$, computed on BEV anchors and normalized by the number of positives. The regression term \mathcal{L}_{reg} is a weighted Smooth-L1 loss over 7 box codes per anchor.

- **Risk regression.** The risk loss is a mean-squared error between predicted and ground-truth BEV risk maps:

$$\|\hat{R} - R_{\text{gt}}\|_2^2.$$

- **Communication over-usage penalty.** The term $\phi(U; B_{\text{bytes}}) = \max(0, U/B_{\text{bytes}} - 1)$. penalizes only communication *above* the target budget, aligning learned masks with the desired budget without changing the runtime protocol.

The usage definitions can be calculated from:

$$U = B_{\text{batch}} \cdot h_{\text{hdr}} + \left(\sum_{b=1}^{B_{\text{batch}}} \sum_{l=2}^{L_b} \sum_{i,j} M_{l,i,j}^{(b)} \right) \cdot (b_{\text{idx}} + C \cdot b_{\text{feat}}), \quad (9)$$

where B_{batch} is the batch size of this training, $M_{l,i,j}^{(b)} \in \{0, 1\}$ is the non-ego mask (for all the masks of ego is $l=1$), h_{hdr} is a fixed header bytes cost, b_{idx} is per-cell index bytes cost, C is the channel dimension, and b_{feat} is bytes per feature value.

5. Experimental Setup

5.1. Datasets

We use the OPV2V dataset [Xu et al., 2022b] as the base dataset. OPV2V is a synthetic multi-vehicle CP benchmark generated by the OpenCDA co-simulation of SUMO [Krajzewicz et al., 2012] and CARLA [Dosovitskiy et al., 2017].

OPV2V contains 73 scenarios (average ~ 25 s) across multiple CARLA towns, where 2 \sim 7 connected vehicles record 64-channel LiDAR from their own viewpoints. We follow the standard frame-level counts of 6765/1981/2170 for train/val/test, respectively. Importantly, OPV2V natively covers a diverse set of driving situations without any additional sampling from our side. The included situations comprise:

- **Overtaking / Lane Change:** fast lateral maneuvers with transient occlusions.
- **Left-turn and Right-turn Intersections:** cross-traffic under partial observability (pedestrians/cyclists may emerge from blind zones).
- **On-ramp Merging:** gap selection and speed adjustment with strong temporal risk.
- **Unprotected Crossroads:** multiple agents with conflicting trajectories.
- **Head-on Encounters:** close-range opposing traffic forming highly critical regions.
- **Straight Driving (Low-risk Baseline):** low-complexity scenes for calibration.
- **Multi-agent Cooperation:** ≥ 3 vehicles jointly negotiating maneuvers.

To illustrate why risk-aware cooperation is meaningful, we provide illustrative exemplars from OPV2V for the above situations in Figure 6. These thumbnails are for visualization only and do not change the dataset composition.

To strengthen generalization and avoid leakage, we keep the natural scenario composition of OPV2V, but ensure that train/val/test have comparable proportions of each situation (e.g., intersections, merging, head-on). The unit of assignment is the entire scenario (all its frames stay in one split), preventing temporal leakage while reducing distributional drift between splits.

We control the number of agents per frame (2–7) by matching their histograms across splits within $\pm 5\%$. The qualitative exemplars in Figure 6 shows the different scenarios that are inherently covered by the dataset organized by us.

5.2. Risk label generation

To facilitate risk-aware CP using the OPV2V dataset [Xu et al., 2022b], we generate risk annotations based on spatial, kinematic, and traffic-contextual information, further refined by expert domain knowledge—particularly in complex environments such as intersections. The final risk score for each object is computed as a weighted combination of three sub-components:

$$R_i = \alpha_d R_i^{(d)} + \alpha_s R_i^{(s)} + \alpha_m R_i^{(n)}, \quad (10)$$

where R_i denotes the overall risk score for object i , and $R_i^{(d)}$, $R_i^{(s)}$, and $R_i^{(n)}$ correspond to distance-based, speed-based, and intersection-based risk scores, respectively. The weights $\alpha_d = 0.5$, $\alpha_s = 0.3$, and $\alpha_n = 0.2$ were selected based on empirical tuning and expert input.

- **Distance-Based Risk:** Objects located closer to the ego vehicle are more likely to pose an immediate threat. We quantify this via an exponential decay function of the Euclidean distance:

$$R_i^{(d)} = \exp(-\lambda_d \cdot \|\mathbf{p}_i - \mathbf{p}_e\|_2), \quad (11)$$

where \mathbf{p}_i and \mathbf{p}_e denote the positions of object i and the ego vehicle, respectively. The parameter λ_d controls the decay rate of risk with distance.

- **Speed-Based Risk:** Rapidly approaching vehicles or those with high relative speed introduce dynamic hazards. We model this component as:

$$R_i^{(s)} = \frac{|v_i - v_e|}{\max_j |v_j - v_e| + \epsilon}, \quad (12)$$

where v_i is the velocity of object i , v_e is the ego vehicle's speed, and ϵ is a small constant to avoid division by zero. This formulation emphasizes relative speed normalized across the scene.

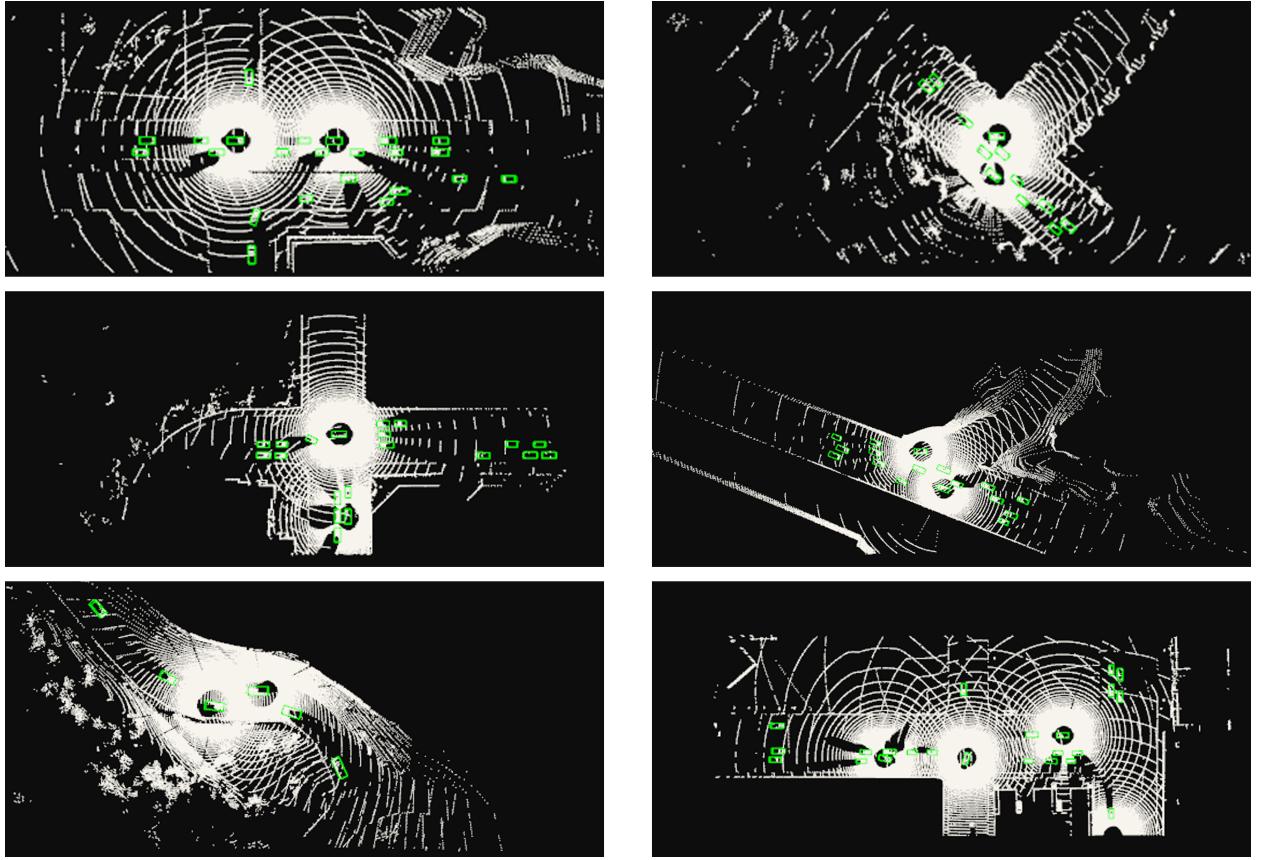


Figure 6: Representative exemplars from OPV2V illustrating scenarios that are inherently covered by the dataset and organized by us.

- **Intersection-Based Risk:** Intersections are inherently high-risk regions due to complex traffic flows, occlusion, signal compliance issues, and the presence of vulnerable road users. We begin by measuring proximity to intersections:

$$R_i^{(n)} = \exp\left(-\lambda_t \cdot \min_{\mathbf{q}_m \in Q} \|\mathbf{p}_i - \mathbf{q}_m\|_2\right), \quad (13)$$

where $Q = \{\mathbf{q}_m\}$ denotes known intersection coordinates and λ_m adjusts the decay with distance to intersections.

Normalization: Finally, we clip the combined risk score to the range [0, 1] for stable learning:

$$\hat{R}_i = \min(1, \max(0, R_i)). \quad (14)$$

5.3. Baselines

To contextualize the standard AP results, we compare the following baselines under the same backbone, BEV grid, IoU thresholds, synchronization window, and quantization:

- **Where2Comm (Spatial-only baseline) [Hu et al., 2022].** A representative spatial-communication method that learns where to communicate based solely on spatial saliency without explicit risk or task-aware weighting. Each agent predicts a binary mask indicating informative BEV cells, and only those regions are transmitted for feature fusion. This baseline captures the benefit of geometry-aware but task-agnostic cooperation.
- **Upper Bound (fully connected).** Fully connected communication with no budget, transmitting all partner features for fusion; serves as a performance ceiling.

- **Lower Bound (single-agent).** No cooperative communication. It measures the capability of the ego-only detector.
- **Fixed-Neighbor (equal-budget).** The total communication budget is equally divided among all non-ego neighbors. Within each neighbor, features (e.g., grid cells or point clusters) are uniformly sampled at random. This baseline isolates the effect of adaptive link-wise budget allocation from uniform distribution.
- **Random-Cell.** A global uniform sampler randomly selects exactly K feature cells from all non-ego agents, regardless of their spatial location or risk relevance. This baseline evaluates the effectiveness of our selective content transmission compared to random feature selection under the same bandwidth constraint.

5.4. Implementation details

Model and feature encoding. We adopt a PointPillar BEV backbone. The PillarVFE uses 64 channels. The BEV backbone has 3/5/8 blocks with filters 64/128/256 and deconv of 128 channels; a shrink header downsamples to 256 channels for heads. We attach three lightweight heads: classification (per cell 2 anchors), regression (7 parameters per anchor), and a risk head (1 per anchor) to produce dense risk heatmaps.

Voxel/grid and anchors. Voxel size is $0.4 \times 0.4 \times 4$ m with LiDAR range $[-140.8, -38.4, -3, 140.8, 38.4, 1]$ m. The BEV grid is $H=192$, $W=704$ (feature stride 4). Anchors follow $(l, w, h)=(3.9, 1.6, 1.56)$ with yaw $\{0^\circ, 90^\circ\}$; $NMS = 0.15$, and the positive, negative thresholds are 0.6 and 0.45 separately.

Training setup. Optimizer: We select Adam with learning rate= 2×10^{-4} as the optimizer and the selection of weight_decay is 0.01 and eps=1e-10. For the learning rate schedule, we set cosine annealing for 50 epochs with 10-epoch warmup (with the warmup learning rate= 2×10^{-5} , and the minimal learning rate= 5×10^{-6}). During the training of all the models, we set the batch size as 8. In terms of connecting agent numbers, we cut the number of agents up to 5 CAVs. Data augmentation includes x-axis flip, random rotation ($\pm 45^\circ$), and scaling (0.95–1.05). Voxelization caps are 32 points/voxel, with train/test voxel maxima as 32k/70k separately.

Inference and post-processing. We decode detection and risk heatmaps after fusion. Evaluation uses $\text{IoU} \in \{0.3, 0.5, 0.7\}$; risk-aware AP uses $\tau \in \{0.2, 0.3, 0.4\}$. We log per-frame communication rate and bytes for the report.

5.5. Evaluation protocols and metrics.

We use 3D Average Precision (3DAP) to assess object detection performance. Given a detection is considered correct if the Intersection over Union (IoU) between the predicted and ground-truth 3D bounding box exceeds a threshold θ , the AP is computed based on the precision-recall curve.

We report 3DAP under three IoU thresholds:

$$\theta \in \{0.3, 0.5, 0.7\},$$

corresponding to different levels of localization strictness.

Let $TP(\theta)$, $FP(\theta)$, and $FN(\theta)$ be the number of true positives, false positives, and false negatives under threshold θ , respectively. Precision and recall are defined as:

$$\text{Precision}(\theta) = \frac{TP(\theta)}{TP(\theta) + FP(\theta)}, \quad \text{Recall}(\theta) = \frac{TP(\theta)}{TP(\theta) + FN(\theta)}. \quad (15)$$

3DAP is then computed as:

$$3\text{DAP}(\theta) = \int_0^1 \text{Precision}(\theta, r) dr, \quad (16)$$

where $\text{Precision}(\theta, r)$ is interpolated at recall level r .

To assess the influence of risk understanding on perception, we compute 3DAP selectively over high-risk regions determined by thresholding the predicted risk map.

Let $\mathcal{I}_{risk}(\tau) = \{i \mid \hat{R}_i > \tau\}$ be the set of objects or regions identified as risky with a risk threshold τ . We evaluate detection performance on this subset, denoted as $3\text{DAP}_{risk}(\theta, \tau)$:

$$3\text{DAP}_{risk}(\theta, \tau) = 3\text{DAP} \text{ evaluated on } \mathcal{I}_{risk}(\tau), \text{ with IoU threshold } \theta. \quad (17)$$

We report results for:

$$\theta \in \{0.3, 0.5, 0.7\}, \quad \tau \in \{0.2, 0.3, 0.4\}.$$

This metric captures how well the model perceives objects in scenarios that are potentially dangerous or require immediate attention, reflecting the synergy between risk assessment and spatial awareness.

We evaluate under two protocols:

- **P1: Fixed-bandwidth.** Per-link budget $B_{\text{bytes}} \in \{0.5, 0.7, 1, 2, 3, 5, 10\}$ KB/frame. Each method is tuned per B_{bytes} ; we report Risk-AP, Δ Risk-AP/KB, and data transmission volume. This stresses efficiency at scarce bandwidth.
- **P2: Fixed-performance.** Given a target Risk-AP (e.g., $\geq X$), we report minimal bytes and latency to reach it. This answers how much bandwidth is necessary for a safety line.

6. Results and Discussion

6.1. Main Results (Standard AP)

We report standard detection AP at IoU 0.3/0.5/0.7 denoted as AP30, AP50 and AP70 separately, across baselines and our method. See Table 2 for overall comparison. It should be noted that the communication budget of Where2comm, Fixed-Neighbor, Random-Cell and ours are 20% of the fully connected situation like the settings of the method Upper Bound. And there is no communication in Lower Bound method. As shown in the table, our model achieves consistently competitive performance across all IoU thresholds, with only marginal differences compared to the strongest baselines, while remaining close to the upper bound. This indicates that both our approach and the baseline methods are able to effectively leverage the advantages of CP.

6.2. Risk-Aware Evaluation

We further evaluate Risk-Aware AP by filtering ground-truths above risk thresholds $\tau \in \{0.2, 0.3, 0.4\}$. Results are summarized in Tables 3. Compared with the overall AP results, where our model and the baselines perform similarly in Table 2, the risk-aware evaluation reveals a clearer distinction. As shown in Table 3, our method consistently outperforms the baselines across all IoU thresholds, especially under higher risk conditions ($\tau = 0.3, 0.4$). The performance of our model remains close to the upper bound while the spatial-only baseline drops significantly as risk increases. This demonstrates that our design better preserves detection robustness when encountering high-risk or safety-critical objects, validating the effectiveness of the communication protocol and SRA-CP coordination. In other words, although both methods achieve comparable aggregate perception accuracy, our framework exhibits stronger risk sensitivity and resilience, which are essential for safety-oriented CP. We additionally visualize risk-aware example heatmaps (Sec. 6.6).

6.3. P1: Pareto efficiency under fixed bandwidth

To further examine model performance under resource-constrained conditions, we plot AP30/AP50/AP70 vs. communication cost (KB/frame) in Figure 7 and Risk-AP30/AP50/AP70 vs. communication cost (KB/frame) in Figure 8. These plots provide a quantitative view of how perception accuracy scales with bandwidth usage.

Across the 0.5–10 KB/frame regime, our proposed SRA-CP configuration consistently dominates the Pareto frontier, achieving higher safety-aware gains per byte compared to baseline methods. For example, at 5 KB/frame, our approach yields approximately +4.7% Risk-AP50 improvement over the baseline while maintaining comparable communication overhead. This demonstrates that the method’s communication sparsification and the inference fusion jointly enable efficient and safety-preserving cooperation.

6.4. P2: Minimal cost to reach a safety line

Figure 9 reports the minimum bandwidth (KB/frame) required to achieve specific Risk-AP30/AP50/AP70 targets under different risk thresholds. Across all nine subplots in Figure 9, our method consistently requires fewer bytes per frame to reach the same Risk-AP target compared to the baseline, demonstrating superior efficiency across all thresholds $\tau \in \{0.2, 0.3, 0.4\}$.

It is worth noting that our target values for Risk-AP were determined in a principled way: for each risk threshold τ , we set the target AP values (for AP30, AP50, and AP70) to 0.9 \times , 0.8 \times , and 0.7 \times of the upper bound performance,

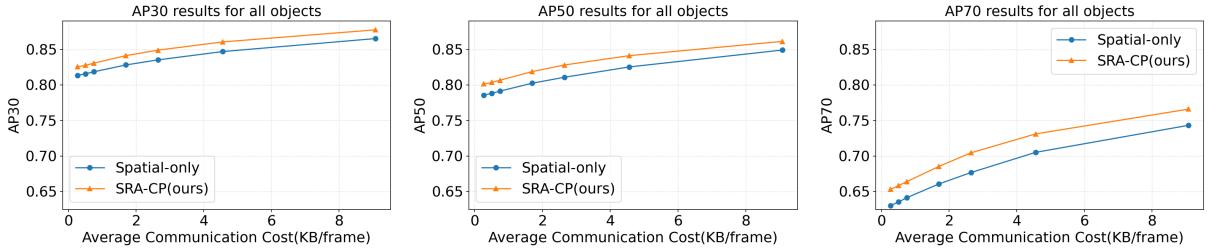


Figure 7: Comparison of perception accuracy (AP30, AP50, AP70) under varying communication costs (KB/frame) across all objects.

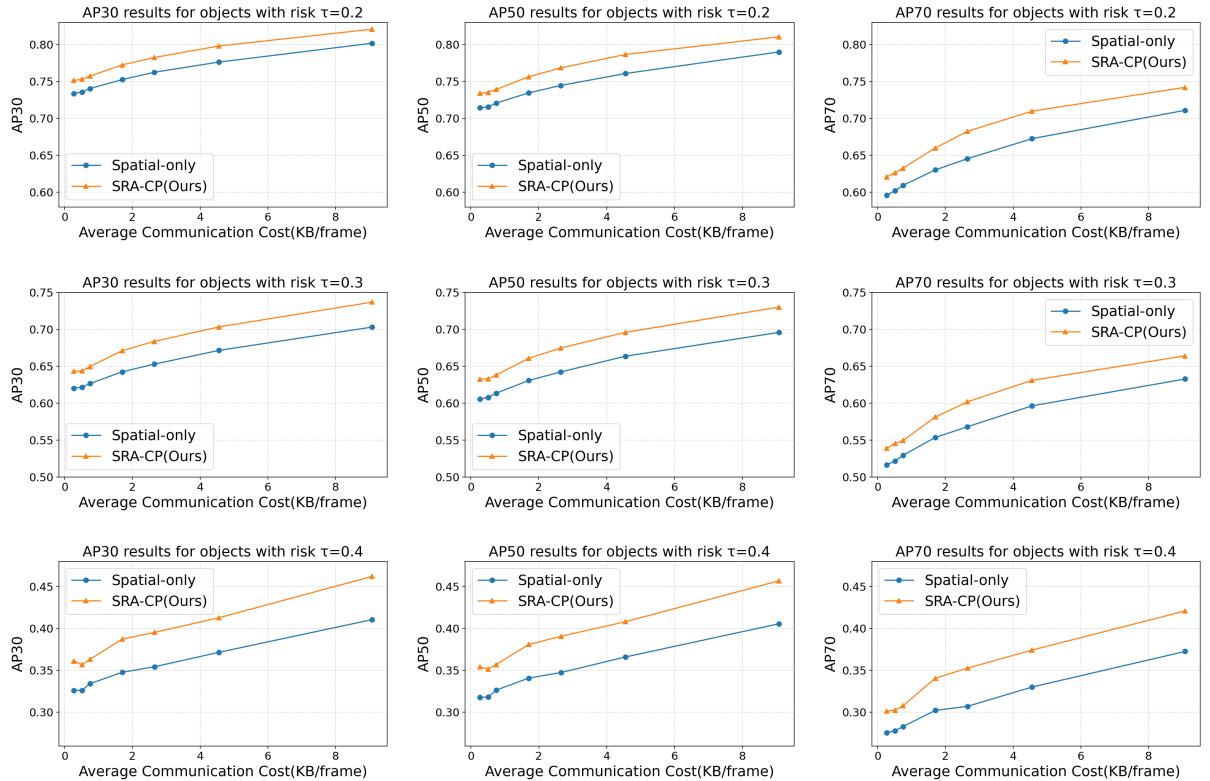


Figure 8: Comparison of perception accuracy (AP@30, AP@50, AP@70) under varying communication costs (KB/frame) for objects with different risk levels, defined by risk thresholds $\tau \in \{0.2, 0.3, 0.4\}$.

respectively. This provides a reasonable and balanced target scale—stringent enough to challenge the communication strategy, yet attainable for well-designed cooperative frameworks.

For example, at $\tau = 0.2$ and AP50=0.75, our method reaches the target Risk-AP using only 1.3 KB/frame, compared to the baseline's 3.3 KB/frame. The advantage becomes even more pronounced under higher risk thresholds: at $\tau = 0.4$ and AP70=0.42/0.38, the baseline fails to achieve the required AP target across all IoU levels (Figure 9(c, f)), while our method maintains strong performance with 5.1 and 9 KB/frame. This indicates that when perception becomes safety-critical, the baseline communication policy saturates its bandwidth without sufficient accuracy gain, whereas our SRA-CP-driven policy continues to deliver usable, risk-aware perception outputs.

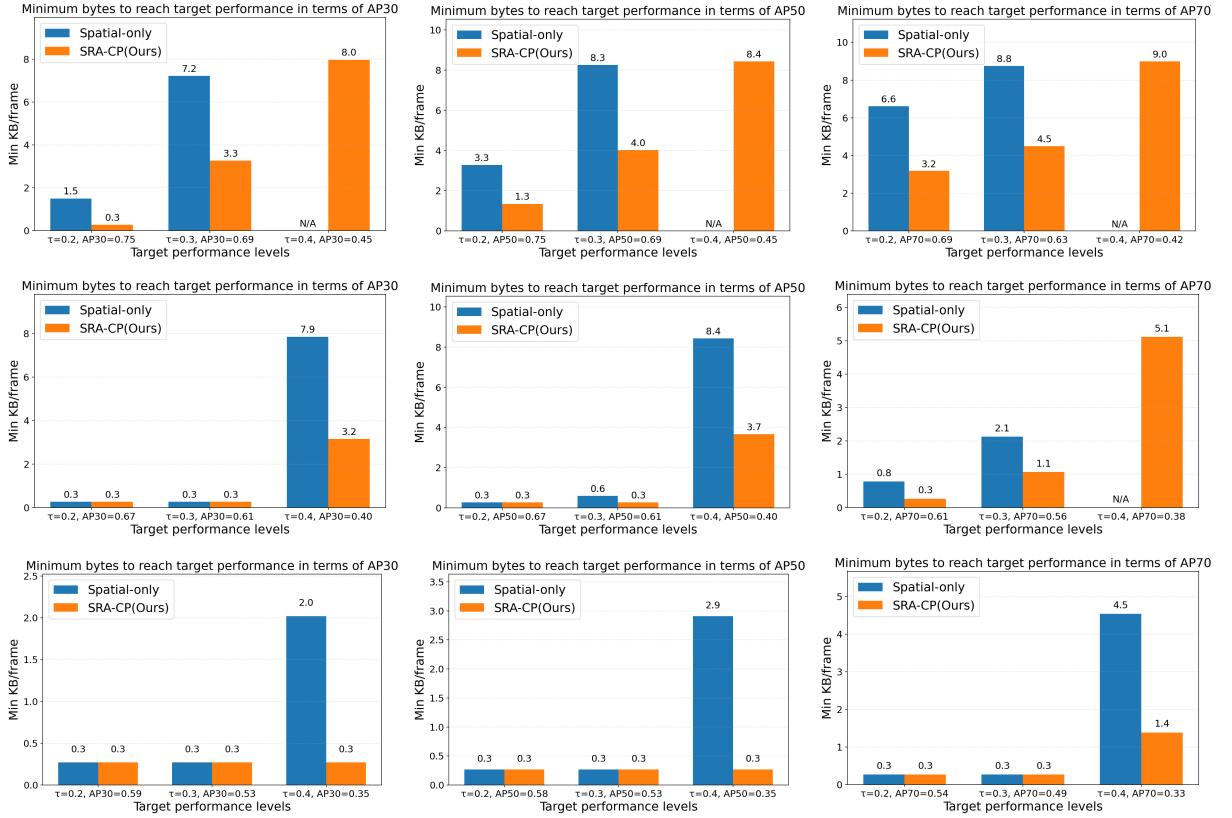


Figure 9: Comparison of different CP methods in terms of the minimum communication bandwidth (KB/frame) required to achieve specific perception accuracy levels (AP30, AP50, AP70) for objects with varying risk levels, categorized by risk thresholds $\tau \in \{0.2, 0.3, 0.4\}$.

Table 2

Detection performance comparison (the communication transmission volume of all the baselines are 20% of the Upper Bound) (*AP score higher is better*).

Method	AP30	AP50	AP70
Upper Bound	0.9057	0.8955	0.7996
Ours	0.8920	0.8731	0.7979
Where2comm (spatial-only)	0.8902	0.8791	0.7928
Fixed-Neighbor (equal-budget, ours-union)	0.8341	0.8159	0.6857
Random-Cell (ours-union)	0.8337	0.8156	0.6861
Lower Bound	0.8190	0.7908	0.6263

6.5. Ablation Studies

We ablate key communication choices like gate type (S-only, R-only, Union), blind-zone estimation (on/off) to see whether the modules of our method are actually working.

Gate Mode Analysis We compare three gate configurations under the same 5 kB/frame bandwidth: spatial-only (S-only), risk-only (R-only), and our hybrid Union gate that integrates both spatial and risk cues as shown in Table 4. The results in Table 4 report Risk-Aware AP at IoU=0.3/0.5/0.7 across $\tau \in \{0.2, 0.3, 0.4\}$.

Across all thresholds and IoU levels, the proposed Union gate consistently outperforms both S-only and R-only variants. At $\tau=0.3$, for instance, Union improves AP50 from 0.6636 (S-only) and 0.6722 (R-only) to 0.6959, while at $\tau=0.4$ the gap widens to over +4.2% compared with S-only. Similarly, the AP70 metric rises from 0.3302 (S-only)

Table 3Risk-aware detection performance across risk thresholds ($\tau=0.2/0.3/0.4$). Higher is better.

Method	Risk τ	AP30	AP50	AP70
Upper Bound	0.2	0.8461	0.8411	0.7745
	0.3	0.7659	0.7632	0.6962
	0.4	0.5003	0.4994	0.4704
Ours	0.2	0.8365	0.8315	0.7667
	0.3	0.7642	0.7622	0.6998
	0.4	0.4963	0.4955	0.4702
Where2comm (spatial-only)	0.2	0.8203	0.8136	0.7512
	0.3	0.7412	0.7354	0.6807
	0.4	0.4701	0.4553	0.4177
Fixed-Neighbor (equal-budget)	0.2	0.7644	0.7519	0.6519
	0.3	0.6640	0.6553	0.5705
	0.4	0.3610	0.3565	0.3171
Random-Cell	0.2	0.7641	0.7511	0.6505
	0.3	0.6670	0.6578	0.5723
	0.4	0.3737	0.3685	0.3238
Lower Bound	0.2	0.7531	0.7357	0.6191
	0.3	0.6483	0.6374	0.5381
	0.4	0.3631	0.3581	0.3111

Table 4Risk-aware AP at IoU=0.3/0.5/0.7 for different gate modes (5k budget) across risk thresholds τ .

Gate	Metric	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$
S-only	AP30	0.7763	0.6714	0.3716
	AP50	0.7608	0.6636	0.3661
	AP70	0.6725	0.5963	0.3302
R-only	AP30	0.7861	0.6811	0.3959
	AP50	0.7731	0.6722	0.3894
	AP70	0.6795	0.6002	0.3544
Union (ours)	AP30	0.7981	0.7032	0.4128
	AP50	0.7866	0.6959	0.4082
	AP70	0.7097	0.6308	0.3742

and 0.3544 (R-only) to 0.3742. These gains demonstrate that combining spatial coverage with risk awareness yields complementary benefits—risk-only gating favors safety-critical regions but may miss peripheral context, whereas spatial-only gating ensures broader coverage but wastes bandwidth on low-risk areas.

By unifying both criteria, the Union gate adaptively allocates transmission priority based on spatial relevance and estimated collision risk, effectively balancing perception completeness and communication efficiency. This hybrid gating thus provides a more stable and risk-sensitive communication policy, enabling the system to maintain higher detection performance even as τ increases.

Blind-Zone Estimation To examine whether the model benefits from explicitly prioritizing safety-critical blind areas, we conduct an ablation study on the Union gating scheme with and without blind-zone weighting under a fixed 5 kB/frame communication budget. The results in Table 5 report Risk-Aware AP at IoU=0.3/0.5/0.7 across risk thresholds $\tau \in \{0.2, 0.3, 0.4\}$.

Table 5Risk-aware AP at IoU=0.3/0.5/0.7 for Union gate with/without blind-zone weighting (5k budget) across τ .

Setting	Metric	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$
Union (no blind)	AP30	0.7912	0.6877	0.3973
	AP50	0.7803	0.6778	0.3912
	AP70	0.6948	0.6104	0.3542
Union (blind on, ours)	AP30	0.7981	0.7032	0.4128
	AP50	0.7866	0.6959	0.4082
	AP70	0.7097	0.6308	0.3742

Across all IoU and risk thresholds, enabling blind-zone weighting consistently improves detection performance. Compared to the vanilla Union gate, our method achieves an average gain of +0.7%, +1.2%, and +1.6% for AP30, AP50 and AP70, respectively. The improvement becomes more pronounced as the risk threshold increases. For instance, at $\tau=0.4$, the Risk-AP70 rises from 0.3542 to 0.3742, representing a relative gain of +5.6%. This pattern suggests that the proposed weighting mechanism effectively allocates communication bandwidth toward regions with higher occlusion and potential collision risk.

Qualitatively, this mechanism acts as a “safety amplifier”: when cooperative views overlap poorly or when agents observe asymmetric blind spots, the weighting function adaptively increases the transmission priority of uncertain spatial zones. As a result, even under the same bandwidth constraint, more informative features are propagated to neighboring vehicles, enhancing risk-aware perception robustness in safety-critical scenarios.

6.6. Visualization & Case Study

Figure 10 illustrates a challenging unprotected left-turn scenario with dense cross-traffic. The ego vehicle intends to turn left, yet its LiDAR alone cannot observe the incoming traffic hidden behind other vehicles’ occlusions. These blind-zone regions coincide with locations where high-risk background vehicles are approaching, making the timely restoration of occluded agents crucial for safe maneuver planning.

We compare four communication strategies: a random-cell baseline, spatial-only, risk-only, and our Union (SRA-CP) method. The spatial-only, risk-only, and Union methods all operate under the same fixed communication budget, whereas the random baseline uses a significantly higher budget, illustrating how communication volume alone does not guarantee performance.

Ours vs. Spatial-only and Risk-only. Despite using the same byte budget, the three strategies prioritize cells differently:

Spatial-only focuses solely on geometric visibility difficulty. It successfully identifies cells that are hard to perceive but often fails to emphasize high-risk agents located in traffic-conflict regions. As a result, it may transmit cells that are geometrically interesting yet irrelevant for imminent collision risk, while missing the truly dangerous ones.

Risk-only allocates nearly all bandwidth to the high-risk region. This improves awareness of hazardous agents but ignores spatial fusion quality, often leading to incomplete or noisy reconstructions because difficult-to-fuse regions receive insufficient coverage.

Union (Ours) balances both spatial fusion difficulty and collision risk. It means SRA-CP suppresses low-value regions and forms a dense transmission corridor aligned with the ego–background conflict path, precisely where the occluded vehicle lies. As shown in the detection overlays, Union restores the hidden vehicle more reliably and aligns closer with the ground truth than either single-objective method.

Ours vs. Random-cell Communication. Even with a much larger number of transmitted cells, the random-cell method performs poorly. Because cells are sampled uniformly at random, it often allocates bandwidth to irrelevant free-space areas while failing to cover the critical blind-zone region at the correct moment. Consequently, the recovered detection remains incomplete or inconsistent despite the inflated budget.

In contrast, Union (SRA-CP) pinpoints and transmits only the essential cells—those that influence collision risk or improve multi-agent fusion quality—and thus reconstructs the critical occluded vehicle with dramatically fewer bytes.

What Gets Transmitted (Per-cell Transmission Maps). The transmission maps further confirm each method’s behavior:

Spatial-only spreads bytes broadly across many cells—high coverage but low efficiency. Risk-only over-concentrates in a compact region—high focus but weak contextual support. Random shows noisy, unstructured coverage even with a large budget—no semantic prioritization. Union (Ours) exhibits an intelligent, elongated high-density band that tracks the potential collision trajectory while maintaining minimal peripheral context.

This pattern matches the ablation findings in Table 4: combining spatial difficulty and risk factors yields the most efficient allocation strategy.

In summary, our experiments show that SRA-CP consistently dominates existing cooperative-perception baselines in the communication–safety trade-off. Under the same communication budget, SRA-CP matches or exceeds the cutting-edge spatial-only selective method in perception accuracy, while delivering notably higher perception accuracy for safety-critical objects. When sweeping the per-link budget, our method traces the Pareto frontier: for any given bandwidth it attains the best risky-object detection, and for any target perception accuracy it requires substantially fewer transmitted bytes than competing schemes. Qualitative case studies at unprotected intersections further illustrate that SRA-CP automatically concentrates messages on risky blind-zone cells, allowing the ego vehicle to recover occluded, dangerous agents earlier and more reliably during driving.

7. Conclusion

This paper presents a novel Spontaneous Risk-Aware Selective Cooperative Perception (SRA-CP) framework to address the scalability and bandwidth challenges of multi-agent cooperative perception in dynamic traffic environments. We first design a protocol in which connected agents continuously broadcast their perception coverage with very low communication cost and initiate on-demand handshakes when risk-relevant blind zones are detected. For a certain connected agent, we propose a perceptual risk identification model to detect and quantify risk-critical occlusions, a selective information sharing model to determine which features to transmit under bandwidth constraints, and a dual-attention feature fusion model to integrate received features into the ego agent’s perception output.

Extensive evaluations on a public dataset were conducted against five baseline methods, each targeting a different aspect of the problem. These include a cutting-edge selective CP method, a fully connected CP setting as an upper bound, a no-CP setup as a lower bound, and another 2 methods: fixed neighbor allocation and random feature Sampling to evaluate the effects of communication target selection and content-level feature prioritization, respectively. Experimental results show that SRA-CP achieves less than 1% loss for safety-critical objects compared to generic CP, while using only 20% of the communication bandwidth. Moreover, compared to the cutting-edge selective CP method, SRA-CP improves the AP for critical objects by 15% under the same bandwidth budget, demonstrating its communication efficiency and risk-awareness advantage.

As future work, we are collecting real-world driving data using our lab’s connected vehicles. We plan to further evaluate the framework on this in-house dataset and conduct field tests to assess its real-world applicability and robustness.

References

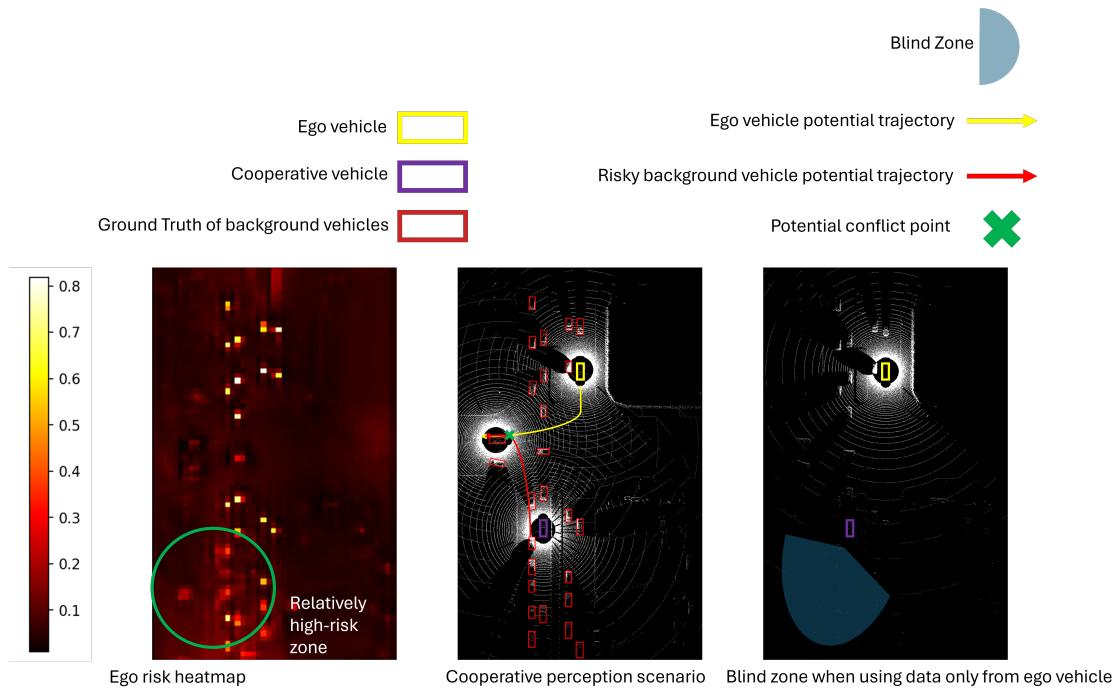
- Qi Chen, Xu Ma, Sihai Tang, Jingda Guo, Qing Yang, and Song Fu. F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds. In *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing (SEC)*, pages 88–100. ACM/IEEE, 2019a.
- Qi Chen, Sihai Tang, Qing Yang, and Song Fu. Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds. In *Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 514–524. IEEE, 2019b. doi: 10.1109/ICDCS.2019.00058.
- Hsu-kuang Chiu, Chien-Yi Wang, Min-Hung Chen, and Stephen F. Smith. Probabilistic 3d multi-object cooperative tracking for autonomous driving via differentiable multi-sensor kalman filter. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 18458–18464. IEEE, 2024. doi: 10.1109/ICRA57147.2024.10610487.
- Liang Dong, Zheng Yang, Xinjun Cai, Yi Zhao, Qiang Ma, and Xin Miao. Wave: Edge-device cooperated real-time object detection for open-air applications. *IEEE Transactions on Mobile Computing*, 22(7):4347–4357, 2022.
- Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017.
- Brahim El Boukili, Mohammed-Hicham Zaggaf, and Lhoussain Bahatti. Cooperative lane keeping assist: Design and evaluation of a v2v lane perception sharing approach. *Journal of Robotics and Control*, 6(5):2239–2248, 2025. doi: 10.18196/jrc.v6i5.26784.

- Yiheng Feng, Chunhui Yu, and Henry X. Liu. Spatiotemporal intersection control in a connected and automated vehicle environment. *Transportation Research Part C: Emerging Technologies*, 89:364–383, 2018. doi: 10.1016/j.trc.2018.02.001.
- Bolin Gao, Jiaxi Liu, Hengduo Zou, Jiaxing Chen, Lei He, and Keqiang Li. Vehicle-road-cloud collaborative perception framework and key technologies: A review. *IEEE Transactions on Intelligent Transportation Systems*, 2024.
- Kirin Godhwani, Adam S. R. Parker, Matthew E. Taylor, William Yeoh, and Reuth Mirsky. Towards spontaneous cooperation in multi-agent reinforcement learning using explicit goal recognition. In *RLC 2025 Workshop on Cooperative and Competitive Multi-Agent Reinforcement Learning (CoCoMARL)*, 2025. Poster paper.
- Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. In *Advances in Neural Information Processing Systems*, volume 35, pages 4874–4886, 2022.
- Lennart Lorenz Freimuth Jahn, Seongjeong Park, Yongseob Lim, Jinung An, and Gyeungho Choi. Enhancing lane detection with a lightweight collaborative late fusion model. *Robotics and Autonomous Systems*, 175:104680, 2024.
- Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, Laura Bieker, et al. Recent development and applications of sumo-simulation of urban mobility. *International journal on advances in systems and measurements*, 5(3&4):128–138, 2012.
- Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12697–12705, 2019.
- Rongsong Li and Xin Pei. Multi-V2X: A large scale multi-modal multi-penetration-rate dataset for cooperative perception, 2024.
- Jiaxi Liu, Bolin Gao, Wei Zhong, Yanbo Lu, and Shuo Han. Adaptive optimization strategy and evaluation of vehicle-road collaborative perception algorithm in real-time settings. *Computers and Electrical Engineering*, 120:109785, 2024.
- Yen-Cheng Liu, Junjiao Tian, Nathaniel Glaser, and Zsolt Kira. When2com: Multi-agent perception via communication graph grouping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4106–4115, 2020.
- Chengyuan Ma, Hangyu Li, Keke Long, Hang Zhou, Zhaohui Liang, Pei Li, Hongkai Yu, and Xiaopeng Li. Real-time identification of cooperative perception necessity in road traffic scenarios. Available at SSRN 4973353, 2025.
- Reuth Mirsky, Ignacio Carlucho, Arrasy Rahman, Elliot Fosong, William Macke, Mohan Sridharan, Peter Stone, and Stefano V. Albrecht. A survey of ad hoc teamwork: Definitions, methods, and open problems. *arXiv preprint arXiv:2202.10450*, 2022.
- Fenglian Pan, Yinwei Zhang, Jian Liu, Larry Head, Maria Elli, and Ignacio Alvarez. Reliability modeling for perception systems in autonomous vehicles: A recursive event-triggering point process approach. *Transportation Research Part C: Emerging Technologies*, 169:104868, 2024. doi: 10.1016/j.trc.2024.104868.
- Huan Qiu, Jian Zhou, Bijun Li, Qin Zou, Youchen Tang, and Man Luo. Map4comm: A map-aware collaborative perception framework with efficient-bandwidth information fusion. *Information Fusion*, page 103567, 2025.
- Ahmad Sarlak, Rahul Amin, and Abolfazl Razi. Extended visibility of autonomous vehicles via optimized cooperative perception under imperfect communication. *Transportation Research Part C: Emerging Technologies*, 180:105350, 2025. doi: 10.1016/j.trc.2025.105350.
- Jessica Van Brummelen, Marie O'Brien, Dominique Gruyer, and Homayoun Najjaran. Autonomous vehicle perception: The technology of today and tomorrow. *Transportation Research Part C: Emerging Technologies*, 89:384–406, 2018. doi: 10.1016/j.trc.2018.02.012.
- Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Bin Yang, Wenyuan Zeng, and Raquel Urtasun. V2VNet: Vehicle-to-vehicle communication for joint perception and prediction. In *Computer Vision – ECCV 2020*, volume 12347 of *Lecture Notes in Computer Science*, pages 605–621. Springer, 2020.
- Zengqing Wu, Run Peng, Shuyuan Zheng, Qianying Liu, Xu Han, Brian I. Kwon, Makoto Onizuka, Shaojie Tang, and Chuan Xiao. Shall we team up: Exploring spontaneous cooperation of competing LLM agents. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 5163–5186, Miami, Florida, USA, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-emnlp.297. URL <https://aclanthology.org/2024.findings-emnlp.297/>.
- Hao Xiang, Zhao Liang Zheng, Xin Xia, Runsheng Xu, Letian Gao, Zewei Zhou, Xu Han, Xinkai Ji, Mingxi Li, Zonglin Meng, Li Jin, Mingyue Lei, Zhaoyang Ma, Zihang He, Haoxuan Ma, Yunshuang Yuan, Yingqian Zhao, and Jiaqi Ma. V2X-Real: A large-scale dataset for vehicle-to-everything cooperative perception. In *Computer Vision – ECCV 2024*, 2024.
- Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. V2X-ViT: Vehicle-to-everything cooperative perception with vision transformer. In *Computer Vision – ECCV 2022*, volume 13699 of *Lecture Notes in Computer Science*, pages 107–124. Springer, 2022a.
- Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2583–2589. IEEE, 2022b.
- Runsheng Xu, Xin Xia, Jinlong Li, Hanzhao Li, Shuo Zhang, Zhengzhong Tu, Zonglin Meng, Hao Xiang, Xiaoyu Dong, Rui Song, Hongkai Yu, Bolei Zhou, and Jiaqi Ma. 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)*, pages 13712–13722, 2023.
- Dingkang Yang, Kun Yang, Yuzheng Wang, Jing Liu, Zhi Xu, Rongbin Yin, Peng Zhai, and Lihua Zhang. How2comm: Communication-efficient and collaboration-pragmatic multi-agent perception. *Advances in Neural Information Processing Systems*, 36:25151–25164, 2023a.
- Kun Yang, Dingkang Yang, Jingyu Zhang, Hanqi Wang, Peng Sun, and Liang Song. What2comm: Towards communication-efficient collaborative perception via feature decoupling. In *Proceedings of the 31st ACM international conference on multimedia*, pages 7686–7695, 2023b.
- Wenbin Yang, Hang Yu, Xiangfeng Luo, and Shaorong Xie. Density-aware early fusion for vehicle collaborative perception. *IEEE Intelligent Transportation Systems Magazine*, 17(2):33–47, 2025.
- Chunhui Yu, Yiheng Feng, Henry X. Liu, Wanjing Ma, and Xiaoguang Yang. Corridor level cooperative trajectory optimization with connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, 105:405–421, 2019. doi: 10.1016/j.trc.2019.06.002.
- Haibao Yu, Wenxian Yang, Hongzhi Ruan, Zhenwei Yang, Yingjuan Tang, Xu Gao, Xin Hao, Yifeng Shi, Yifeng Pan, Ning Sun, Juan Song, Jirui Yuan, Ping Luo, and Zaiqing Nie. V2X-Seq: A large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

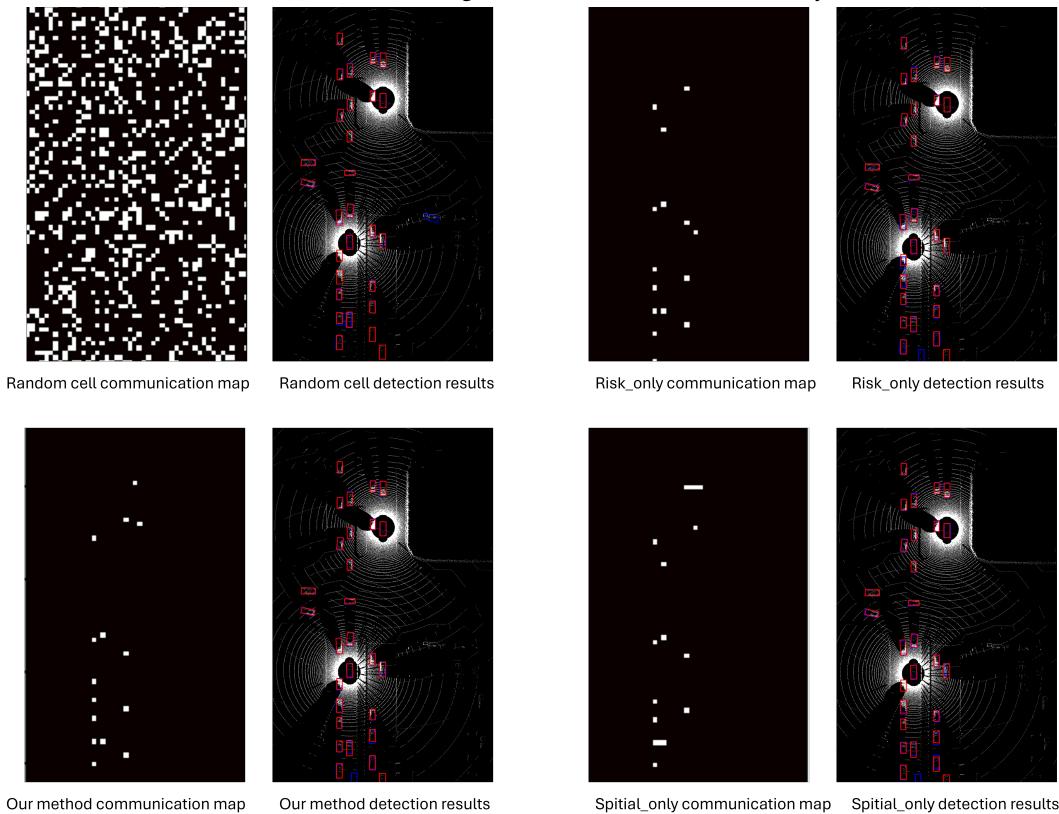
Yuanyuan Zha, Wei Shangguan, Junjie Chen, Linguo Chai, Weizhi Qiu, and Antonio M López. Heterogeneous multiscale cooperative perception for connected autonomous vehicles via v2x interaction. *IEEE Internet of Things Journal*, 2025.

Jiaru Zhong, Jiahao Wang, Jiahui Xu, Xiaofan Li, Zaiqing Nie, and Haibao Yu. Cooptrack: Exploring end-to-end learning for efficient cooperative sequential perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. doi: 10.48550/arXiv.2507.19239. Highlight paper.

Walter Zimmer, Gerhard Arya Wardana, Suren Sritharan, Xingcheng Zhou, Rui Song, and Alois C. Knoll. Tumtraf V2X cooperative perception dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.



(a) Predictions vs. ground truth in BEV with risk overlay.



(b) Risk heatmap and per-cell transmission (Union vs. baselines).

Figure 10: Qualitative example at an unprotected intersection. Our method prioritizes risky blind-zone cells, recovering occluded targets with fewer transmission bytes.