



基于V2X通信的协同感知

周盛 清华大学电子系

2025.11.03 @RISTA前沿大讲堂

- 目录 -

CONTENT

1 研究背景

2 研究内容

① 传输：传输什么？

② 组网：与谁协同？

3 数据集与实验平台

4 思考与展望



- 目录 -

CONTENT

1 研究背景

2 研究内容

① 传输：传输什么？

② 组网：与谁协同？

3 数据集与实验平台

4 思考与展望

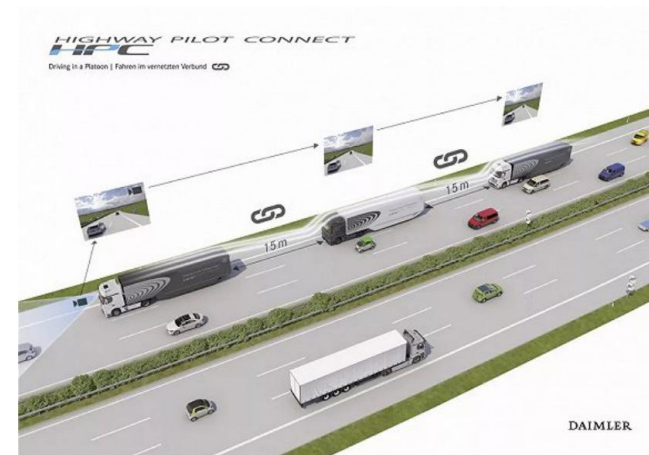
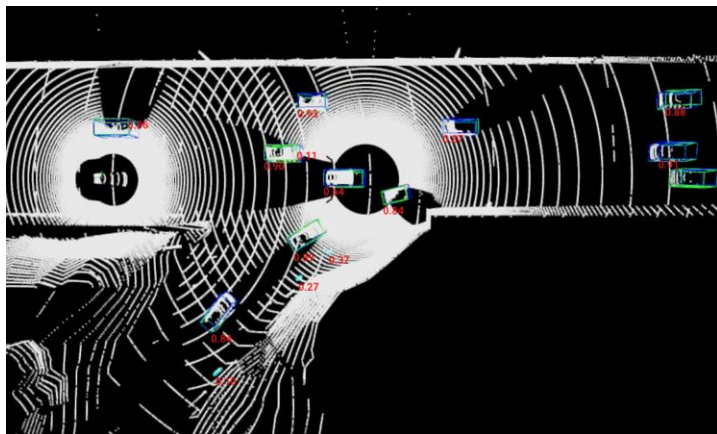


自动驾驶系统的构成

环境感知

预测规划

控制决策



- 周围有什么物体？
- 这些物体未来如何运动？
- 本车应该怎么做？

当今主流自动驾驶技术基于单车智能



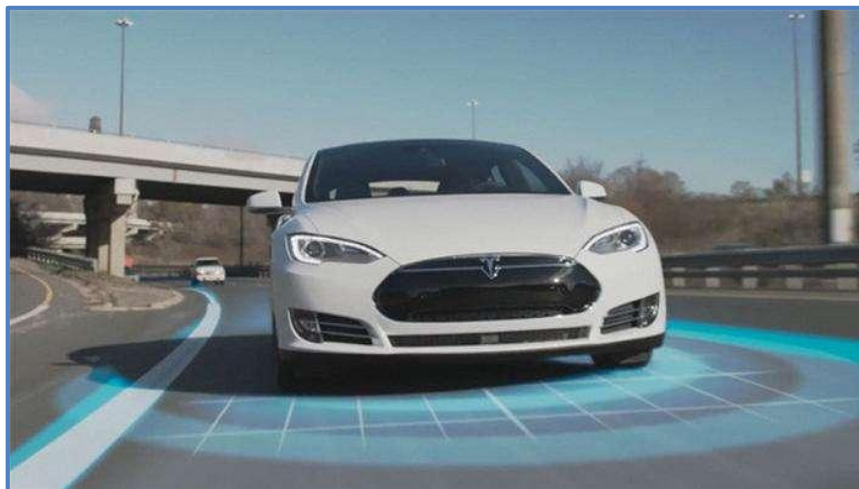
Waymo
Autonomous
Driving

#2 148 vehicles, 1,454,173 miles tested, 110 disengagements



Baidu
Apollo

#1 4 vehicles, 108,300 miles tested, 6 disengagements



Tesla
FSD

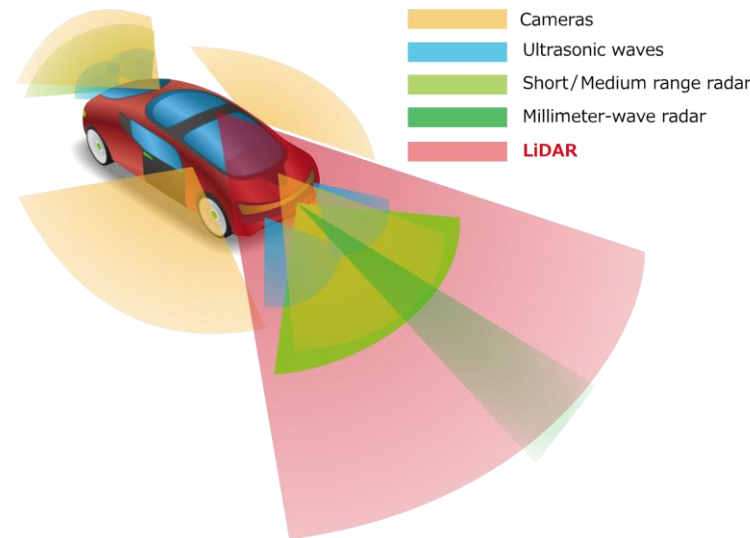


Huawei
ADS

单车智能的局限性

准确可靠的**环境感知**是安全自动驾驶的前提

- 采用冗余互补的**多模态**传感器：摄像头、毫米波雷达、激光雷达等
- 感知困难场景呈现**长尾分布**，需要大量数据训练模型



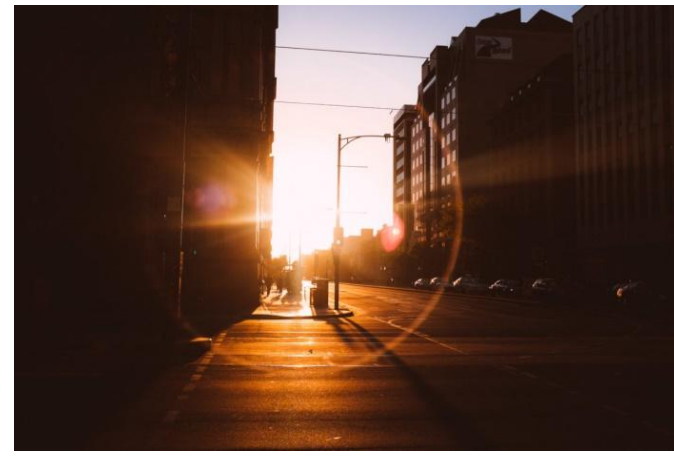
单车传感器的限制：



遮挡



雨、雪、雾



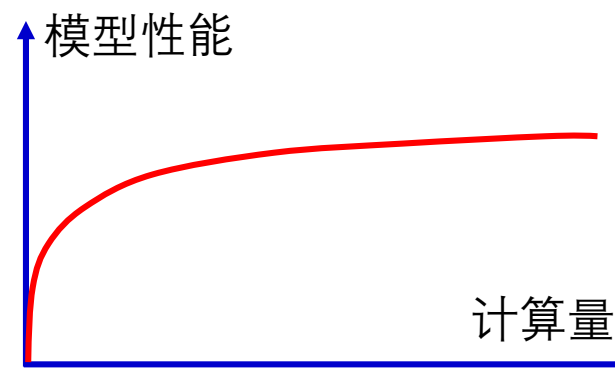
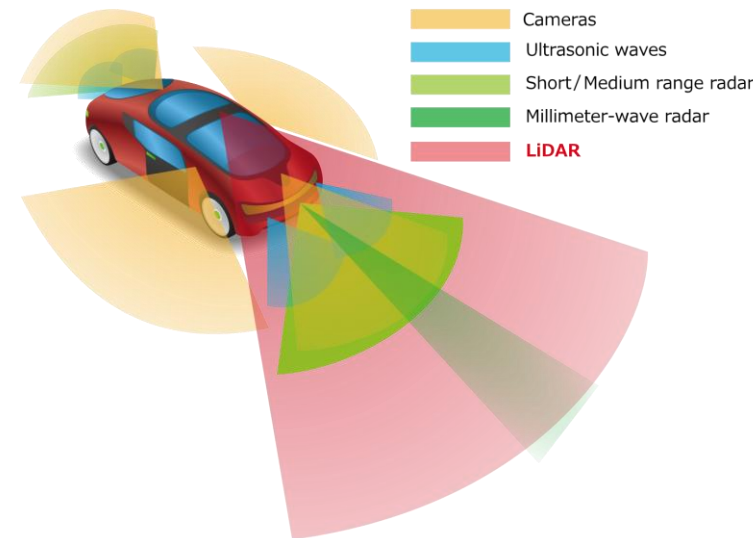
太阳直射

单车智能的局限性

准确可靠的环境感知是安全自动驾驶的前提

- 采用冗余互补的**多模态**传感器：摄像头、毫米波雷达、激光雷达等
- 感知困难场景呈现**长尾分布**，需要大量数据训练模型

特斯拉的教训（“长尾”的目标视角）



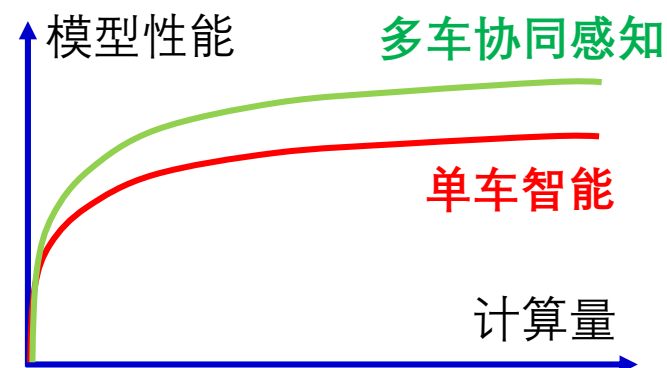
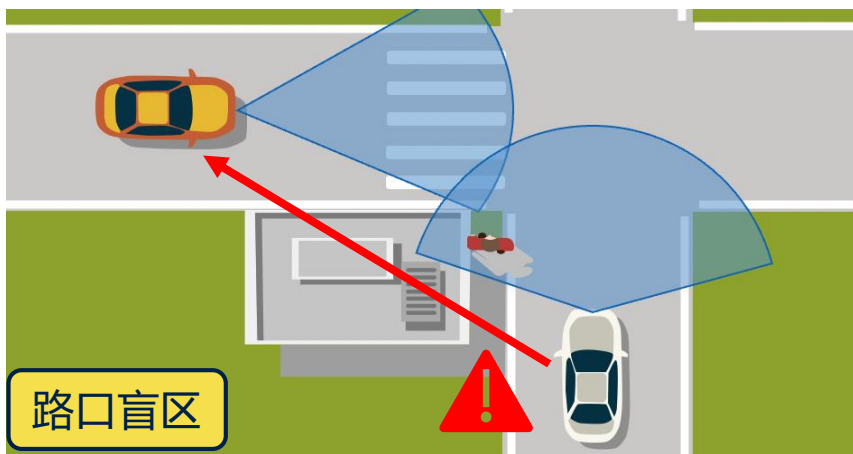
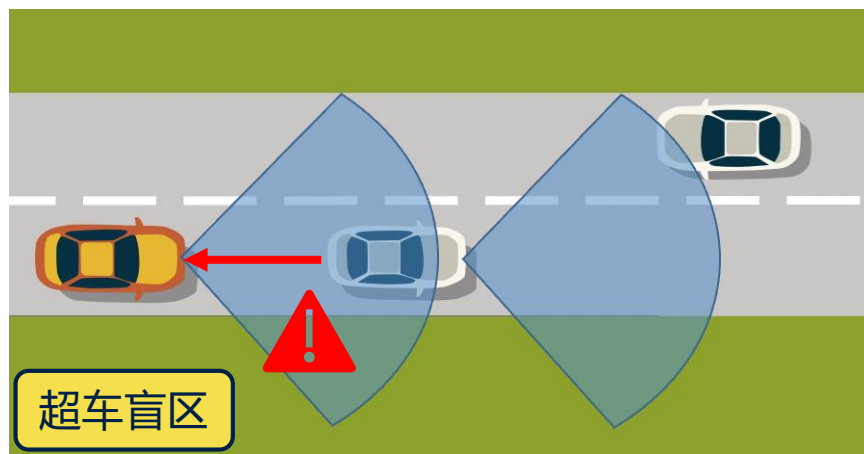
对于基于深度学习的视觉任务，
感知性能随计算量/数据增加的边际收益递减

中国实现自动驾驶的独特挑战

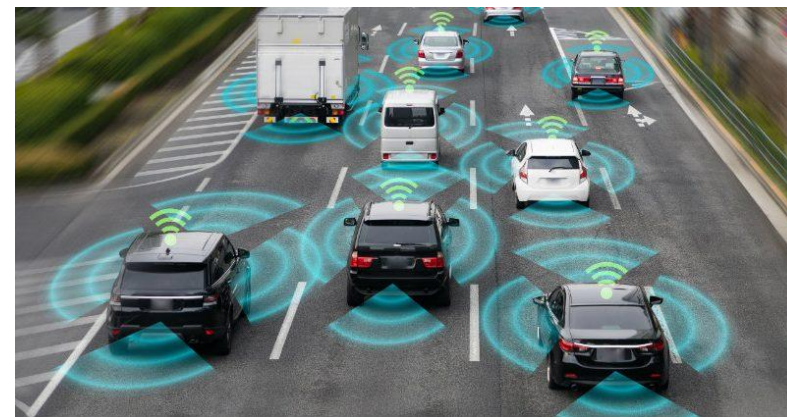


从单车智能到多车协同感知

单车的感知能力限制在视线所及(Field-of-View)，因遮挡会产生**视野盲区**



利用车联网 (Vehicle-to-Everything, V2X)，分享传感器数据，扩展网联车辆的感知范围



以至于多车协同控制

自动驾驶的一种终极形态



协同感知固有问题

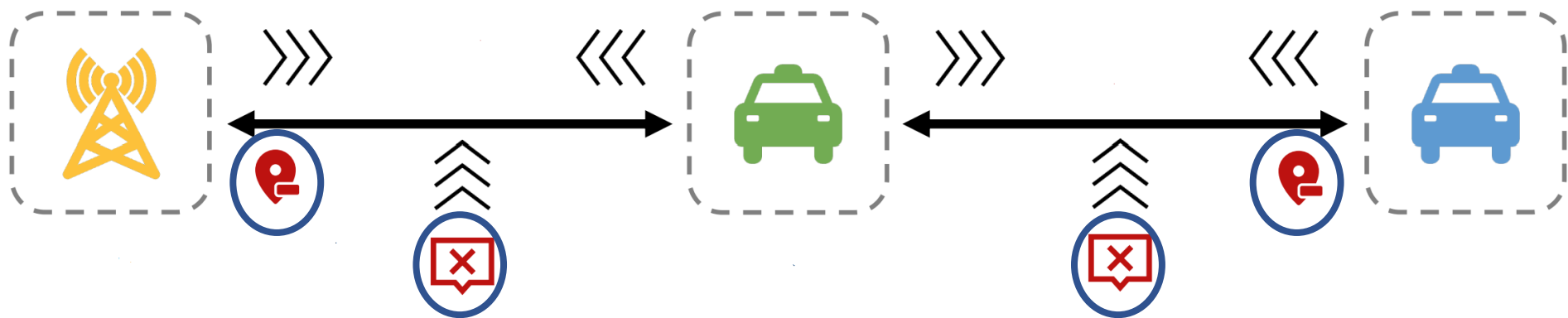
- 不同视角信息的融合依赖准确**数据对齐**
- 感知精度与数据量**非强相关**：数据并不是“越多越好”

无线通信引入的问题

- 有限带宽：必须进行信息**压缩**，传输更有价值的信息
- 随机信道：工作频段**高**、天线高度**低**、终端移动性**高**、环境影响**大**
- 通信延时：信息**时效性**下降对信息效用有负面影响

任务导向的信息压缩

非理想条件的融合算法



- 目录 -

CONTENT

1 研究背景

2 研究内容

① 传输：传输什么？

② 组网：与谁协同？

3 数据集与实验平台

4 思考与展望

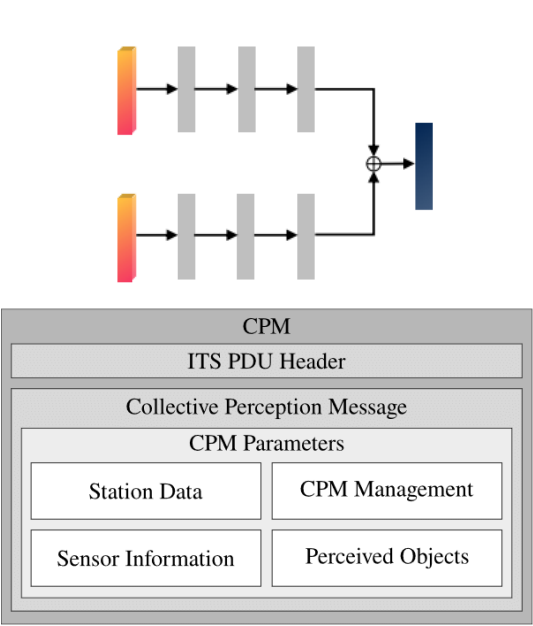


协同感知中的多源信息融合方法

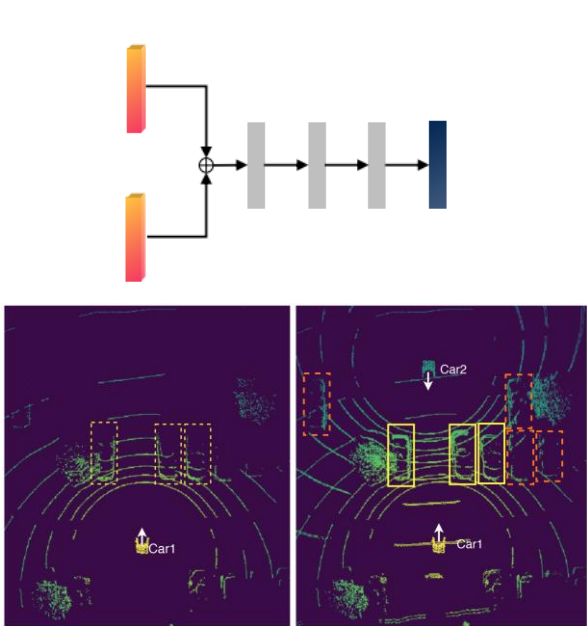
任务导向的信息压缩：传什么&传多少？

融合模式	目标级融合	原始级融合	特征级融合
通信数据量	B-KB级	MB级	KB-MB级
同质性要求	无	无	相同的模型

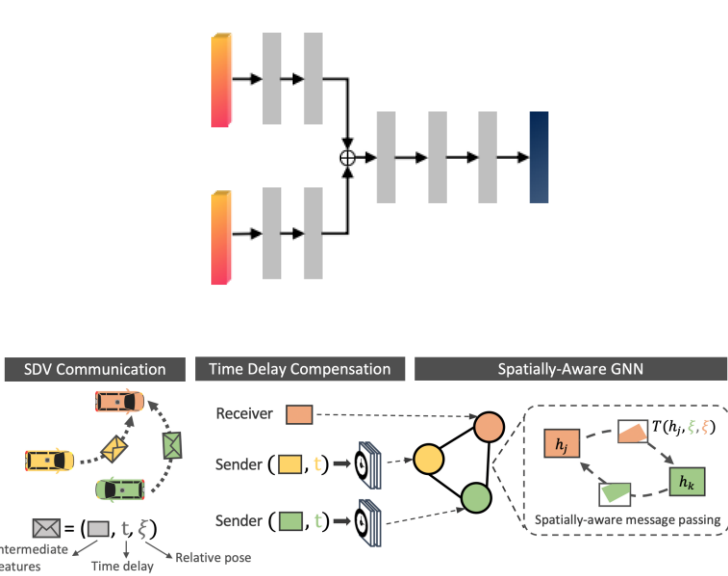
核心问题：
任务导向的
信源压缩和
跨层优化！



CPM^[1]: **目标级融合**



Cooper^[2]: **原始级点云融合**



V2VNet^[3]: **基于GNN的特征融合**

[1] "TR 103 562 - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Analysis of the Collective Perception Service (CPS)", TC ITS Tech. Rep. V2.1.1
[2] Chen Q, Tang S, Yang Q, et al. Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds[C], 2019 IEEE 39th ICDCS. IEEE, 2019: 514-524.
[3] Wang T H, Manivasagam S, Liang M, et al. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction[C], ECCV. Springer, Cham, 2020: 605-621.

基于特征级融合的BEV (Bird-Eye-View) 感知

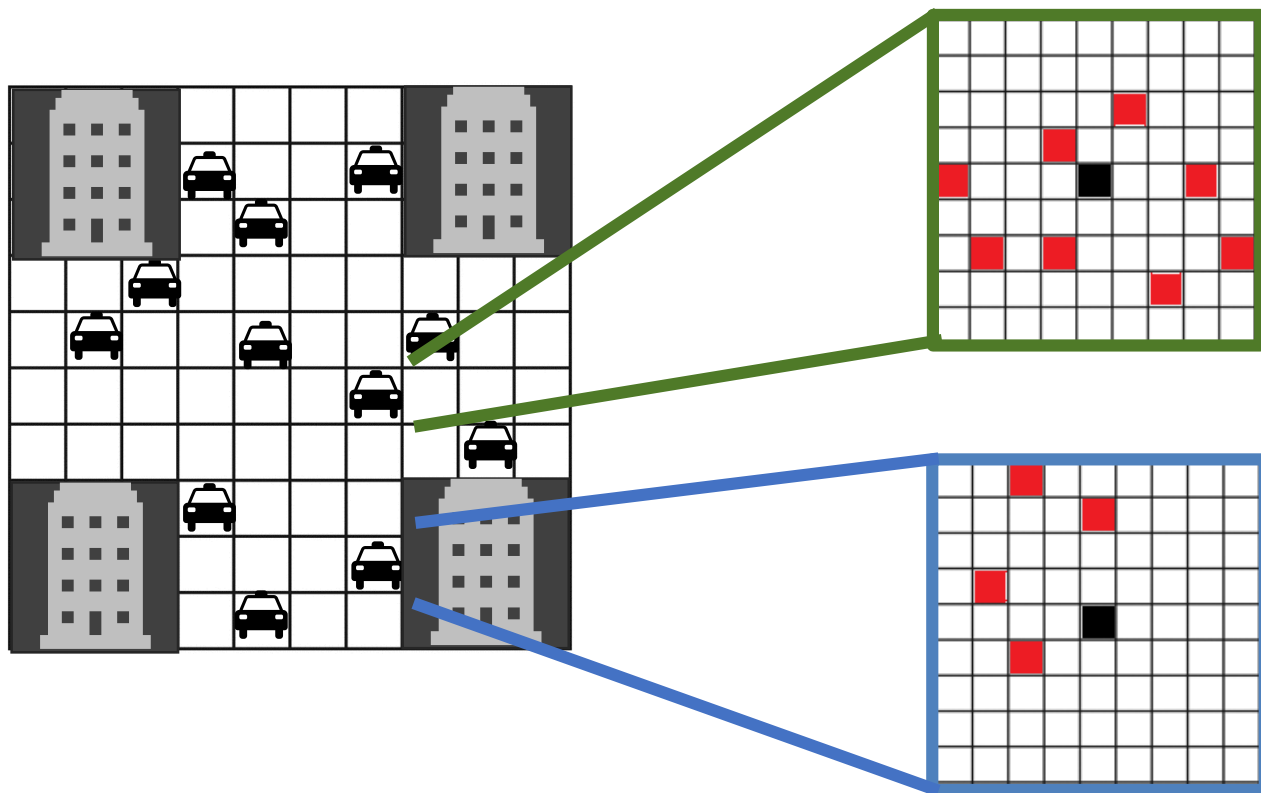
- 基于BEV特征表达，构建多视角多模态共享的特征空间
- 蕴含统一时空坐标，利于融合

对特征图进行进一步压缩

- 基于视角重叠关系的选取
- 基于检测置信度的选取
- 基于特征匹配度的选取

融合算法如何对抗非理想因素？

- 位置与位姿误差、通信丢包和时延



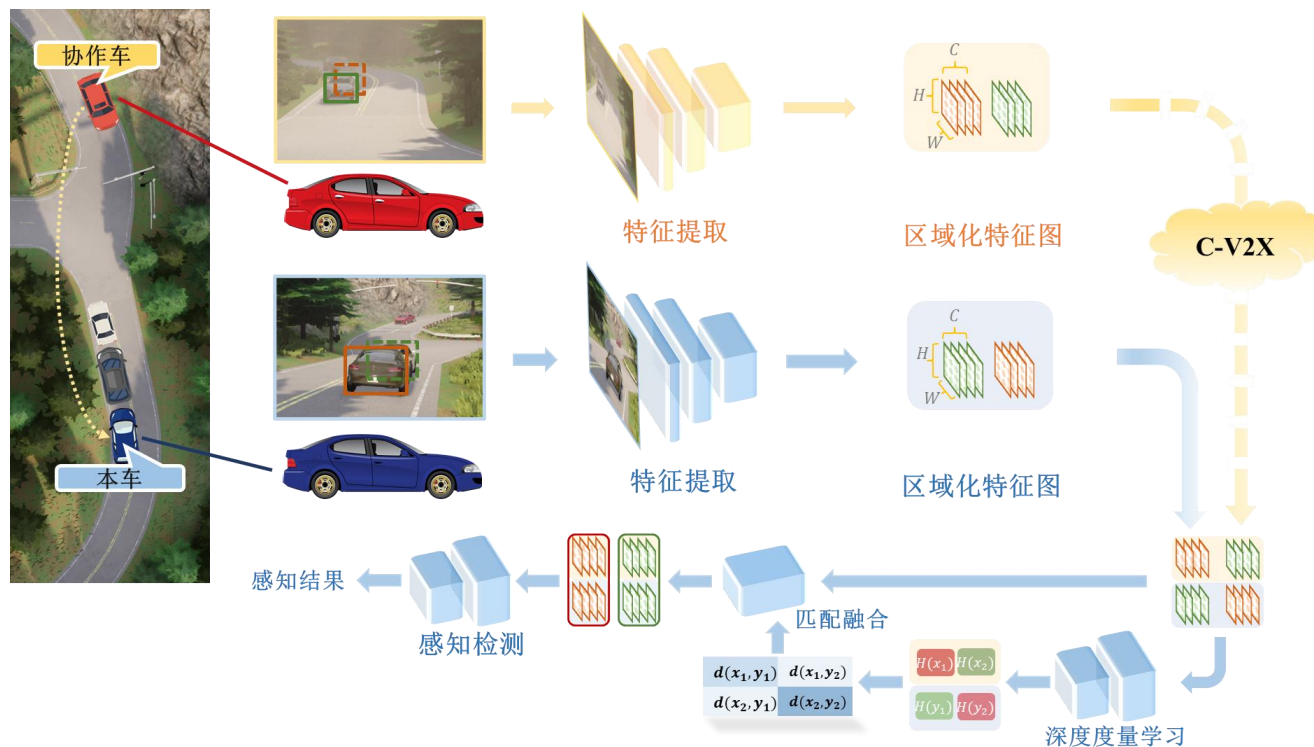
[4] Li, Zhiqi, et al. "Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers." European conference on computer vision. Cham: Springer Nature Switzerland, 2022.

[5] Hu, Yue, et al. "Where2comm: Communication-efficient collaborative perception via spatial confidence maps." Advances in neural information processing systems 35 (2022): 4874-4886.

[6] Wang, Tianhang, et al. "UMC: A Unified Bandwidth-efficient and Multi-resolution based Collaborative Perception Framework." arXiv preprint arXiv:2303.12400 (2023).

将深度度量学习 (Deep Metric Learning) 方法引入传统两阶段目标检测范式

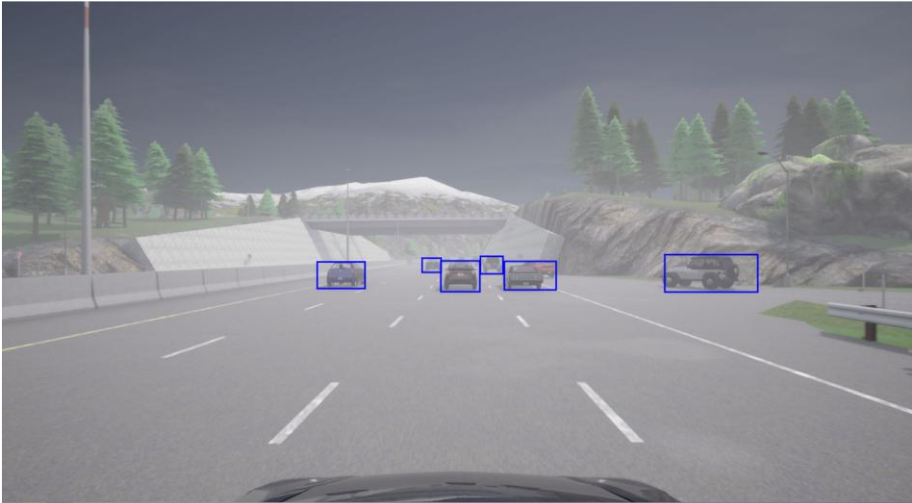
- 从特征级进行融合，传输筛选后的 Regional Feature Map，相对较小的通信数据量
- 结合了双目视觉算法和车辆重识别算法的优势：无需准确的位置和姿态信息！



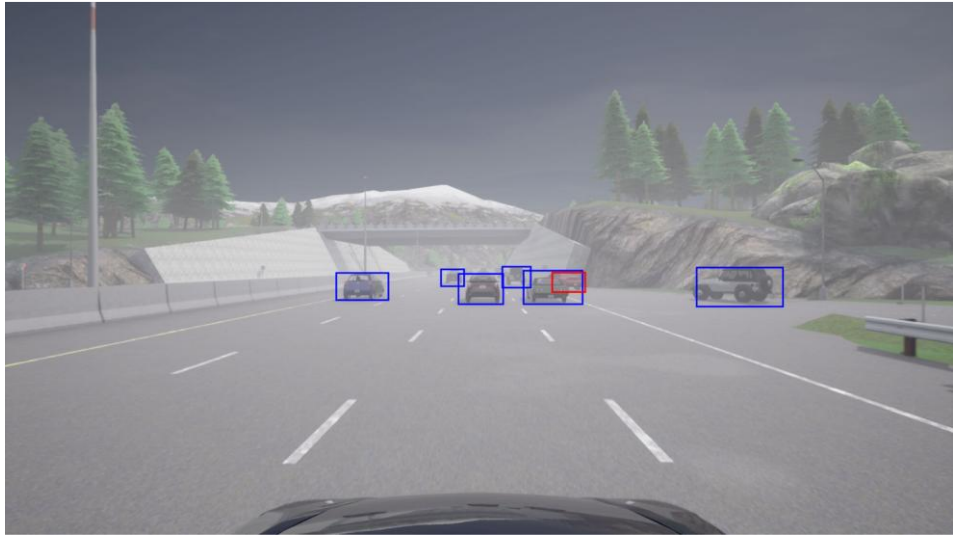
R. Mao, J. Guo, Y. Jia, J. Dong, Y. Sun, S. Zhou, and Z. Niu "MoRFF: Multi-View Object Detection for Connected Autonomous Driving under Communication and Localization Limitations," *VTC2023-Fall*, Oct. 2023.

大视差场景下的多源图像信息融合，提升遮挡目标的检测精度

本车
视角



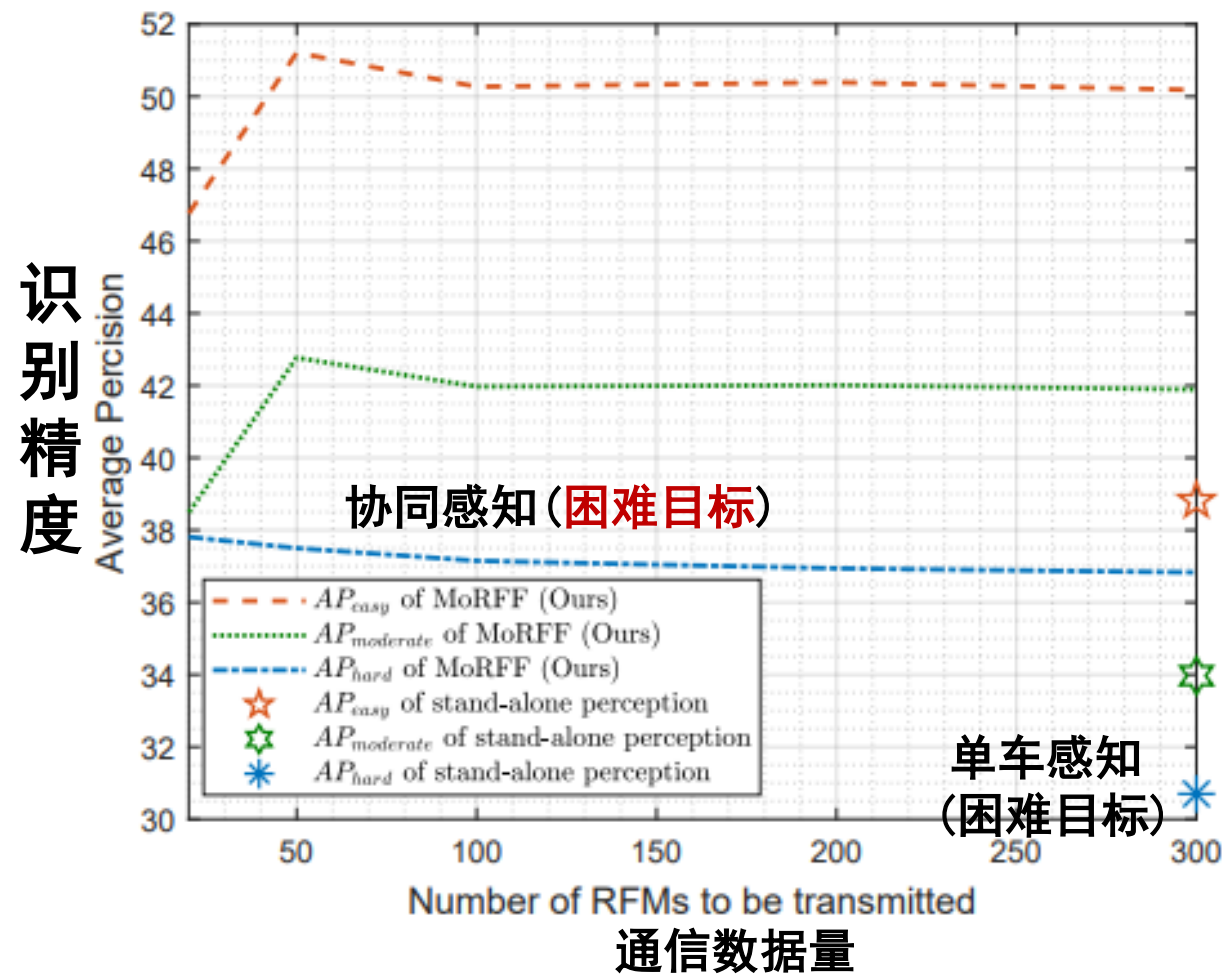
路侧
单元
视角



Methods	AP (easy)	AP (hard)	Inference Time (s)
Mono	0.78	0.60	0.10
Fusion w/o matching	0.49	0.35	0.12
Ours	0.77	0.69	0.18

对于部分遮挡的困难物体取得近9%的精度提升

算法可支持调整区域特征图 RFM 数量，实现检测性能提升与计算/传输开销的Trade-off



	Scenario 1 (Intersection)			Scenario 2 (T-junction)			Scenario 3 (Steep ramp)		
	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
Mono	39.50	31.82	31.22	55.04	53.79	52.58	35.56	29.25	25.43
Ours	45.43	37.22	33.28	57.38	56.19	55.51	39.60	32.63	27.56
w/o matching	39.72	31.97	31.20	55.20	52.56	51.92	31.83	25.96	23.11

	Scenario 4 (Highway on-ramp)			Scenario 5 (Intersection)			Scenario 6 (Mountain road)		
	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
Mono	31.43	26.25	22.93	32.42	31.63	31.13	60.76	52.36	39.70
Ours	50.29	36.85	36.35	52.05	43.25	42.80	62.25	54.50	39.64
w/o matching	34.86	27.36	26.26	34.74	31.35	29.38	51.08	44.23	33.26

Latency	Easy	Moderate	Hard
Sync (0 s)	50.184	41.891	36.833
1 frame (0.5 s)	50.182	41.889	36.832
2 frames (1.0 s)	50.180	41.888	36.831
3 frames (1.5 s)	50.175	41.889	36.831
4 frames (2.0 s)	50.175	41.889	36.831

数据量的极致压缩：利用条件扩散模型？

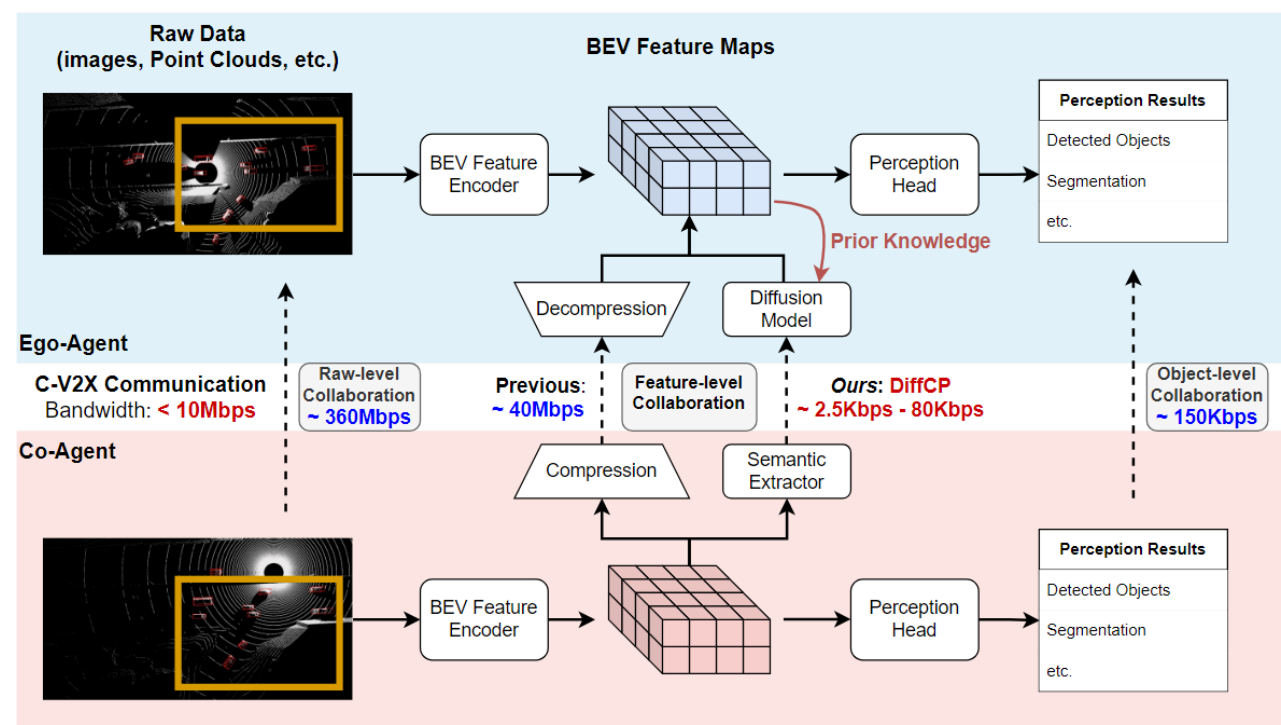


To guess the unseen: 先验知识

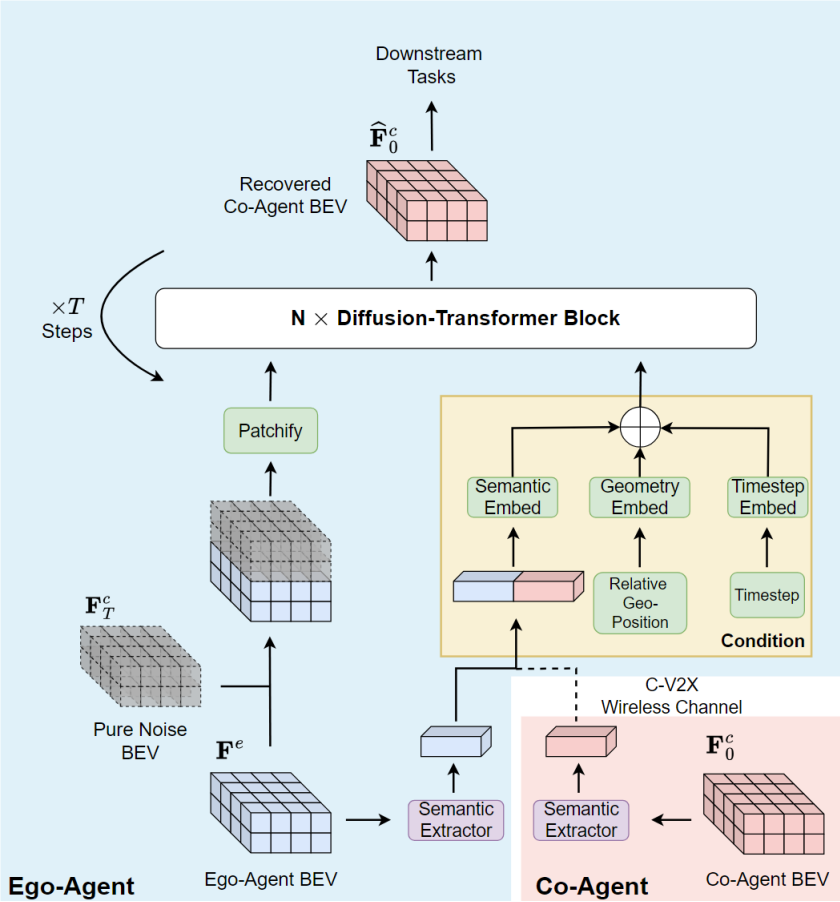
基于条件扩散模型的协同编码算法DiffCP

利用生成式模型，捕捉视角间的几何先验与语义差异

- 各终端的空间位置
 - 协作终端感知信息的语义表征
 - 利用条件扩散模型在本地终端重建协作感知信息
- 所需传输数据 (~ Kbps)



感知流程



训练流程

DiffCP算法对协作感知特征的重建效果

- 以本地感知特征和协作感知语义向量为条件
- 从纯噪声张量逐渐重建协作感知特征

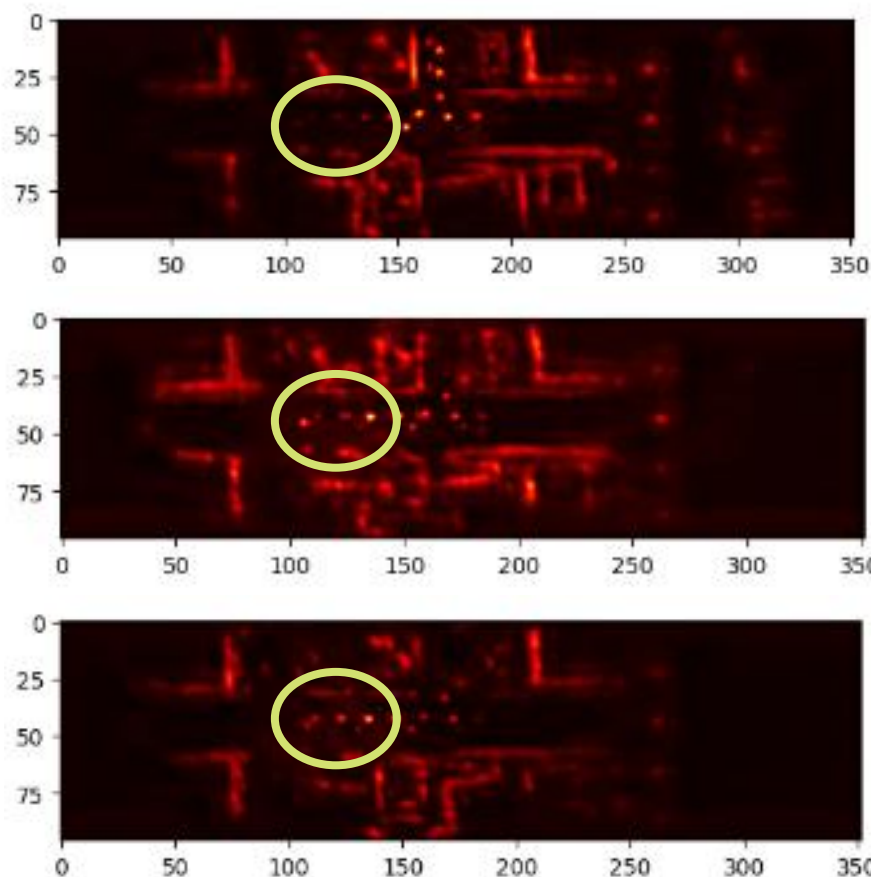
本地感知BEV特征

与协作感知BEV特征的MSE: **0.550**

重建的协作感知BEV特征

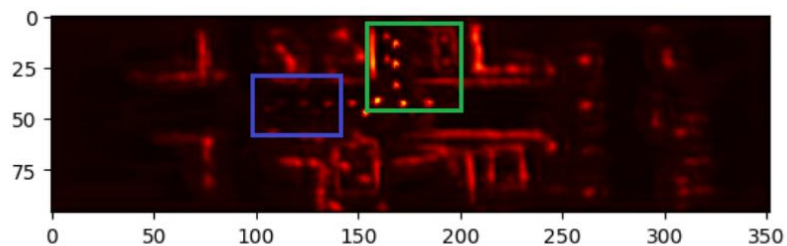
与协作感知BEV特征的MSE: **0.271**

协作感知BEV特征 (真值)

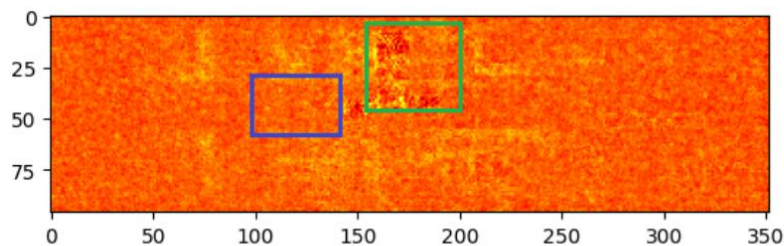


DiffCP算法在通信、计算与重建精度上的折中关系

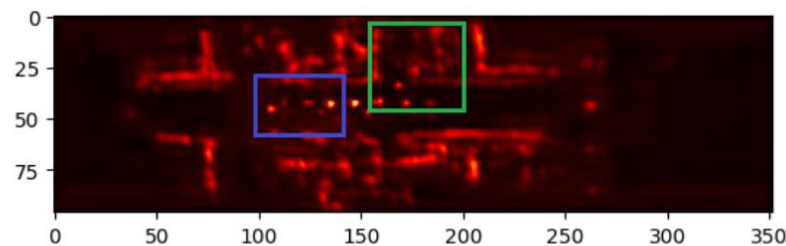
- 通信数据量越大，重建精度越高
- 通信数据量越大，达到最优重建精度所需的迭代次数越低
- 过多的扩散迭代步数反而会降低重建精度（扩散漂移）



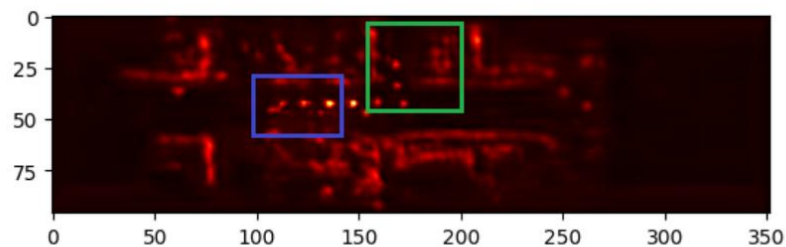
(a) Ego-agent BEV, MSE=0.550



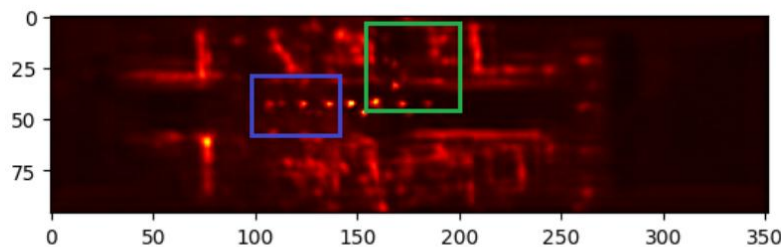
(b) Step=2, MSE=5.507



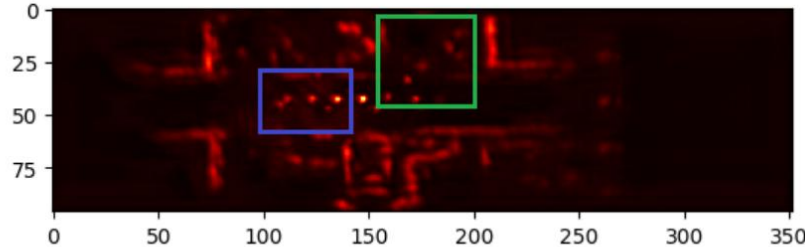
(c) Step=5, MSE=0.271



(d) Step=10, MSE=0.285



(e) Step=15, MSE=0.294

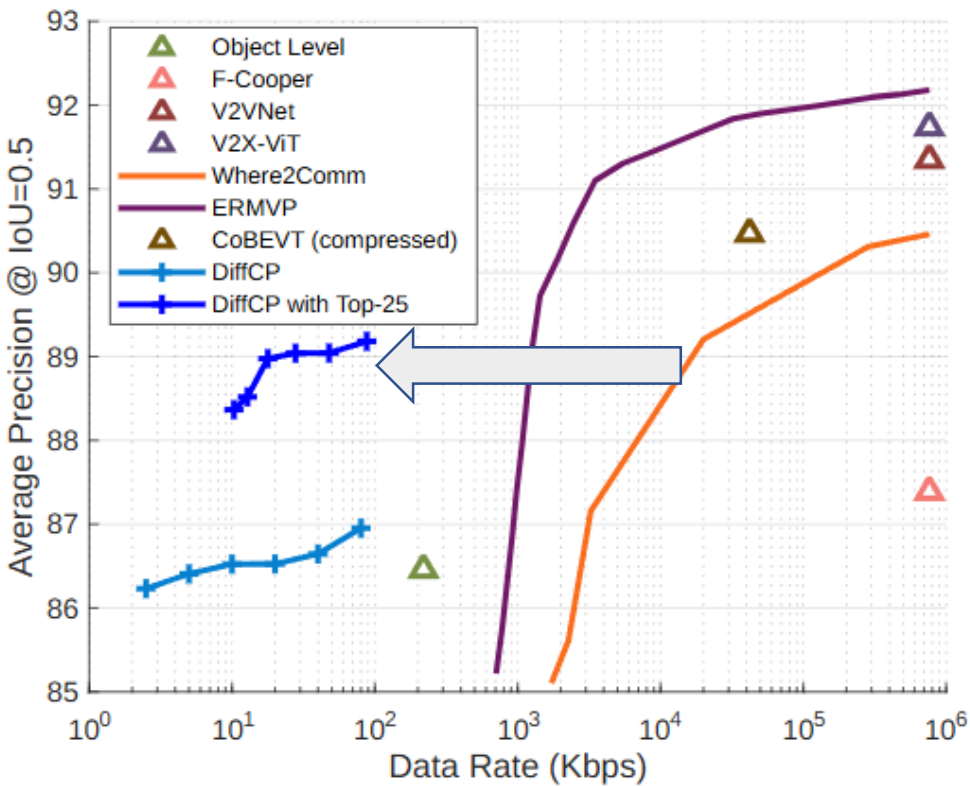


(f) Co-agent BEV, Target

DiffCP算法对协同感知任务的效用

- 仅需**5%**的数据率即可达到目标级协同增益
- 在SOTA特征级协同算法上实现**14.5倍**压缩
- 在**极低带宽**通信系统中表现出优异性能与鲁棒性

Method/Metric	Data Rate	AP@IoU=0.5	AP@IoU=0.7
No Collaboration	0	73.25	58.22
Object-level	219 Kbps	86.46	79.21
F-Cooper [7]	745 Mbps	87.39	79.39
V2VNet [8]	745 Mbps	91.35	82.43
V2X-ViT [9]	745 Mbps	91.74	83.31
Where2Comm [15]	745 Mbps	90.46	84.22
ERMVP [13]	745 Mbps	92.18	85.59
CoBEVT (compressed) [10]	41 Mbps	90.47	84.76
+ DiffCP	80 Kbps	86.95	74.50
+ DiffCP + Top-25	87.8 Kbps	89.18	78.75
+ Top-25 only	7.8 Kbps	80.73	62.14



R. Mao, H. Wu, Y. Jia, Z. Nan, Y. Sun, S. Zhou, D. Gunduz and Z. Niu, “DiffCP: Ultra-Low Bit Collaborative Perception via Diffusion Model,” 2025 IEEE Conference on Robotics and Automation (ICRA), May 2025.

新一代自动驾驶范式——端到端自动驾驶

自动驾驶系统构成

- 感知：周围有什么物体？
- **预测：这些物体未来如何运动？**
- 规控：我（本车）应该怎么做？

数据、模型、算力增长

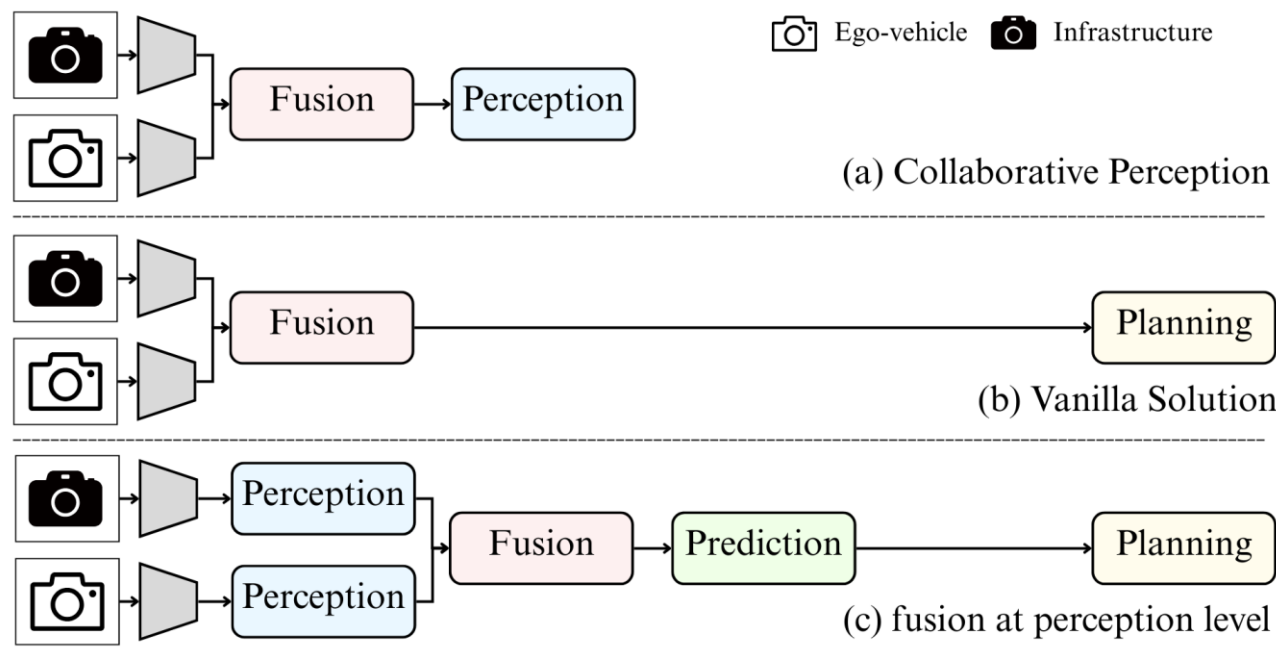


端到端自动驾驶

- 感知+预测+规控**全流程梯度可传**
- 与基于规则的模型相比，能更好处理长尾场景
- 减少由于分模块设计的累积误差对整体性能的影响
- 具备数据驱动的持续学习潜力

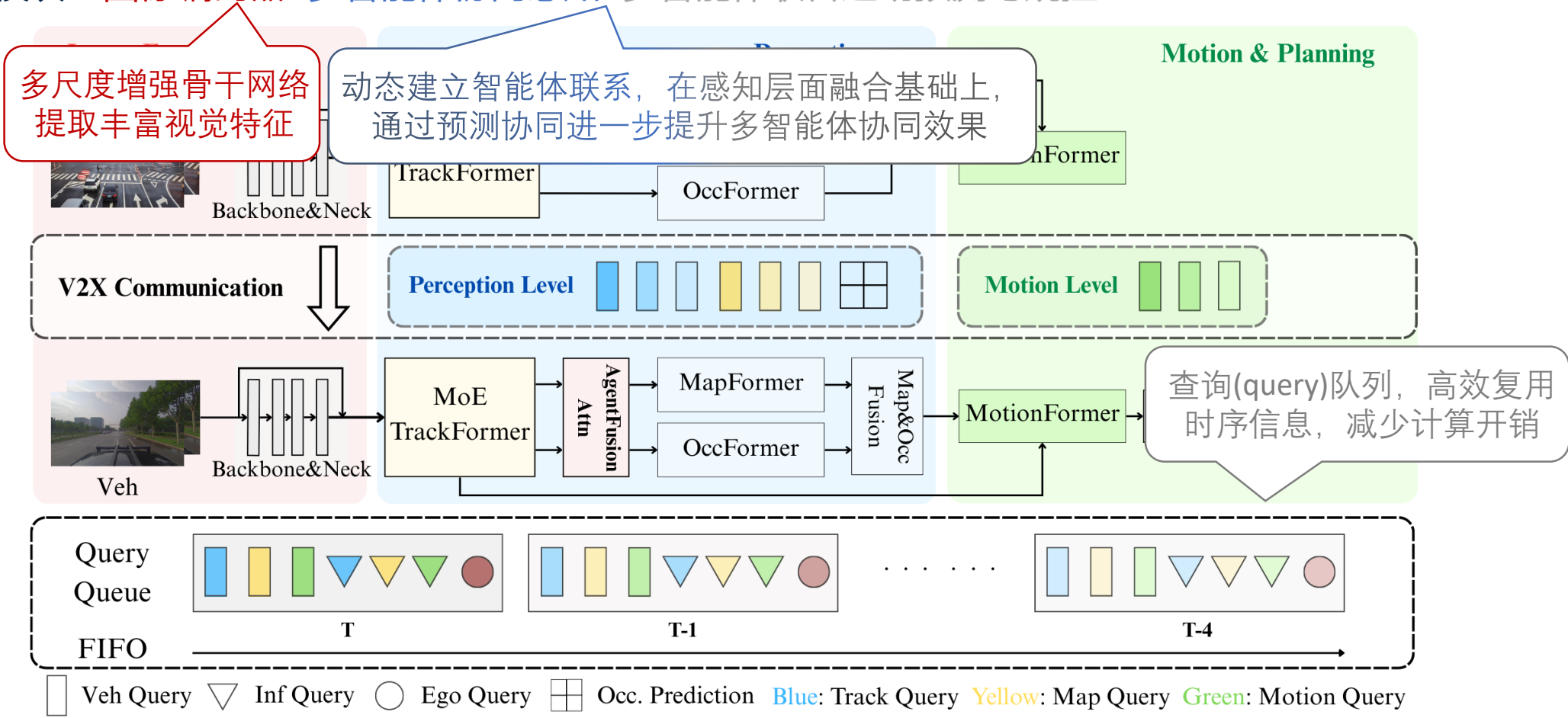
端到端自动驾驶与车联网技术结合？

- 联合感知输出并非直接服务于本车控制目标，不能满足自动驾驶的最终规控需求
- **挑战：**端到端框架下多智能体如何进行信息交互和协同？



感知、预测与规控多层次协同端到端自动驾驶

- 三大模块：图像编码器+多智能体协同感知+多智能体联合运动预测与规控



多层次协同效果

- 同SOTA端到端多智能体协同算法相比
 - 检测性能提升**39.7%**，轨迹追踪性能提升**77.2%**，建图性能提升23.4%
 - 预测误差降低**10.2%**，规控L2误差降低**25.6%**，碰撞率降低20.0%

Method	Detection		Tracking			Online Mapping	
	mAP↑	mATE↓	AMOTA↑	AMOTP↓	Recall↑	Lane (%)	Crossing (%)
UniAD* (Hu et al. 2023)	0.181	0.795	0.197	1.685	0.299	13.3	8.7
SparseDrive* (Sun et al. 2024)	0.324	0.743	0.130	1.680	0.386	-	23.9
UniV2X [†] (Yu et al. 2025)	0.302	0.786	0.241	1.559	0.375	17.7	19.7
Ours	0.422	0.649	0.427	1.280	0.535	19.4	24.3

Method	L2 Error (m)↓				Collision Rate (%)↓				Off-Road Rate (%)↓			
	1s	2s	3s	Avg.	1s	2s	3s	Avg.	1s	2s	3s	Avg.
UniAD* (Hu et al. 2023)	1.26	2.22	3.06	2.18	0.88	1.18	1.32	1.13	0.00	0.15	1.18	0.44
SparseDrive* (Sun et al. 2024)	1.02	1.69	2.37	1.69	4.83	5.51	5.16	5.17	-	-	-	-
CooperNaut (Cui et al. 2022)	3.83	5.26	6.69	5.26	0.59	1.92	1.63	1.38	-	-	-	-
UniV2X (Yu et al. 2025)	1.45	2.19	3.04	2.23	0.15	0.15	0.44	0.25	0.29	0.00	0.44	0.24
Ours	0.88	1.57	2.54	1.66	0.15	0.29	0.15	0.20	0.29	0.59	0.15	0.34

端到端自动驾驶多智能体多层级协同设计

在实测数据集DAIRV2X上的最终规控效果



Z. Song, C. Xia, C. Wang, H. Yu, S. Zhou and Z. Niu, “UniMM-V2X: MoE-Enhanced Multi-Level Fusion for End-to-End Cooperative Autonomous Driving,” submitted to AAAI’26.

- 目录 -

CONTENT

1 研究背景

2 研究内容

① 传输：传输什么？

② 组网：与谁协同？

3 数据集与实验平台

4 思考与展望



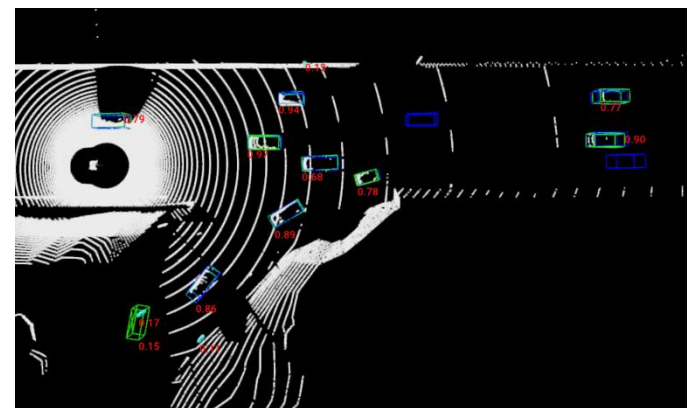
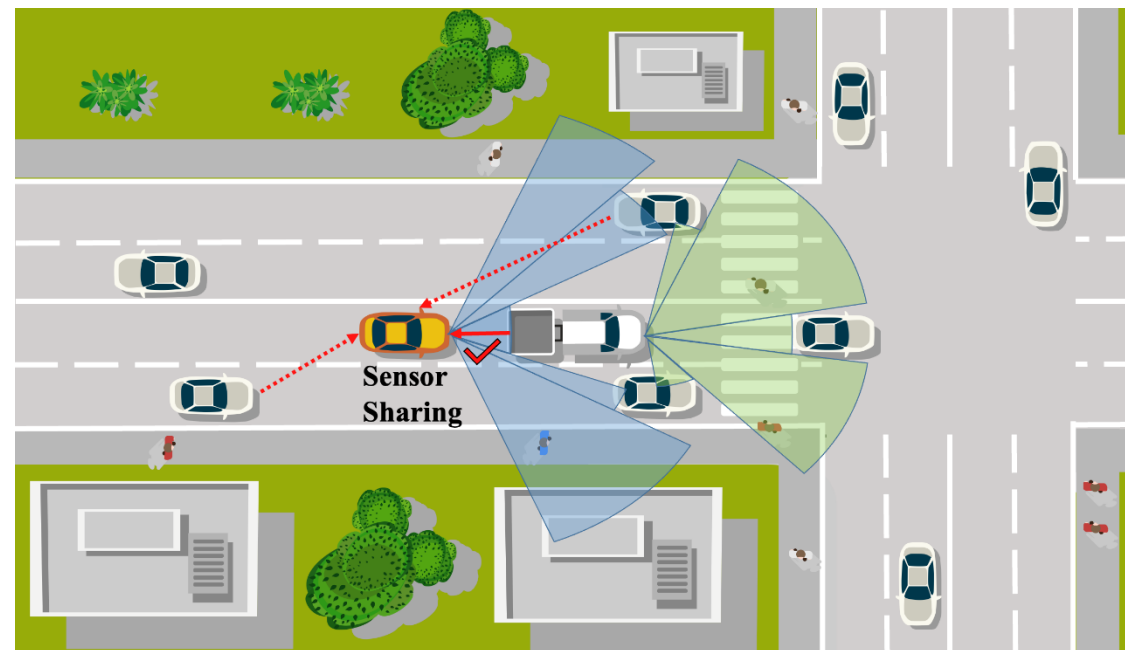
基于多臂赌博机的传感信息调度算法

- 在 t 时刻，用户车选择某辆服务车进行协同
 - 引入环境的感知难度 ω_t
 - 通信传输速率要求 $B \log(1 + \text{SNR}) \geq R$
 - 检测模型采用动态神经网络以达到感知精度要求
- 应该选谁进行协同感知？协同增益如何度量？**

$$g_i(t) = c_0(t) - c_i(t)$$
$$= \sum_{j \in \mathcal{O}_t} w_j^{(t)} \left(1 - \Phi \left(\mathcal{X}_{0,j}^{(t)} \right) \right) \Phi \left(\mathcal{X}_{0,j}^{(t)} \cup \tilde{\mathcal{X}}_{i,j}^{(t)} \right)$$

Undetected by
onboard sensors

Detected with
additional data



基于多臂赌博机的传感信息调度算法

优化目标：实现最优平均协同**增益**

$$\begin{aligned} \max_{a_1, \dots, a_T} \quad & \frac{1}{T} \sum_{t=1}^T g_{a_t}(t), \\ \text{s.t.} \quad & a_t \in \mathcal{V}_t, \end{aligned}$$

一个Online Restless Multi-Armed Bandit (RMAB)问题

- 传统UCB算法:

$$\tilde{g}_i(t) = \hat{g}_i + \beta \sqrt{2 \ln t / n_i}$$

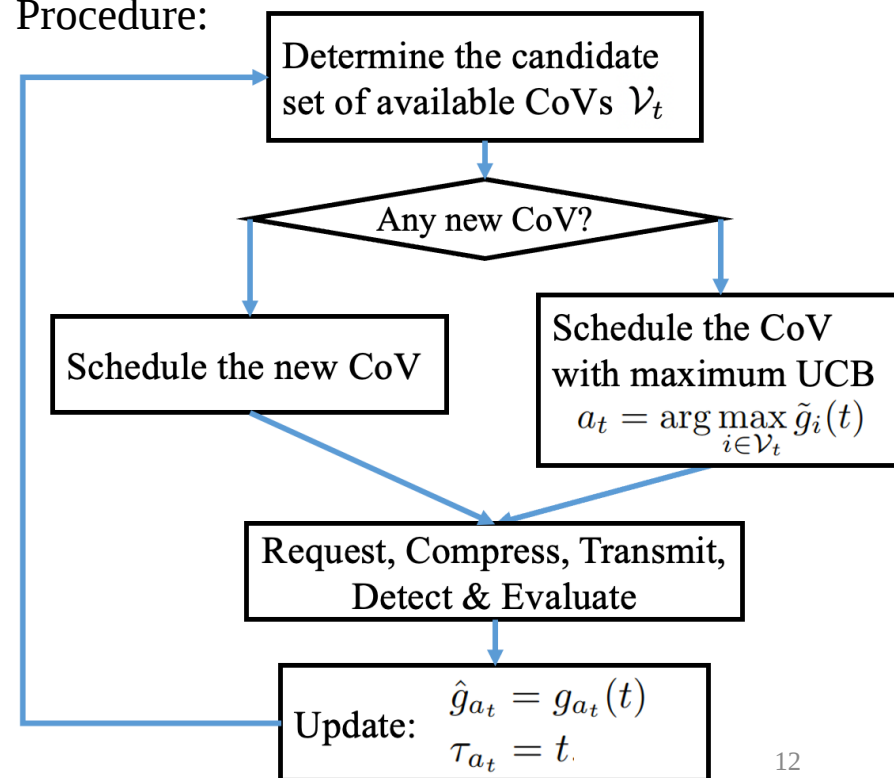
- 所提算法: $\tilde{g}_i(t) = \boxed{\hat{g}_i} + \boxed{\beta \sqrt{t - \tau_i}}$

Last-seen gain

Confidence Bound
increases over time



Procedure:



12

性能保证

Theorem 2 (Dynamic Candidates). *Let $\beta = 15\sigma \log \sigma^{-1}$. For a sufficient long trip $T \geq \Omega(\sigma^{-2})$ with a dynamic candidate CoV set satisfying $|\mathcal{V}_t| \leq 2$, the expected average learning regret of MASS is bounded by*

$$\bar{R}_{MASS} \leq O(\sigma^2 \log^3(1/\sigma)) + 2\lambda,$$

Lower bound for
any online algorithm

$$O(\sigma^2)$$

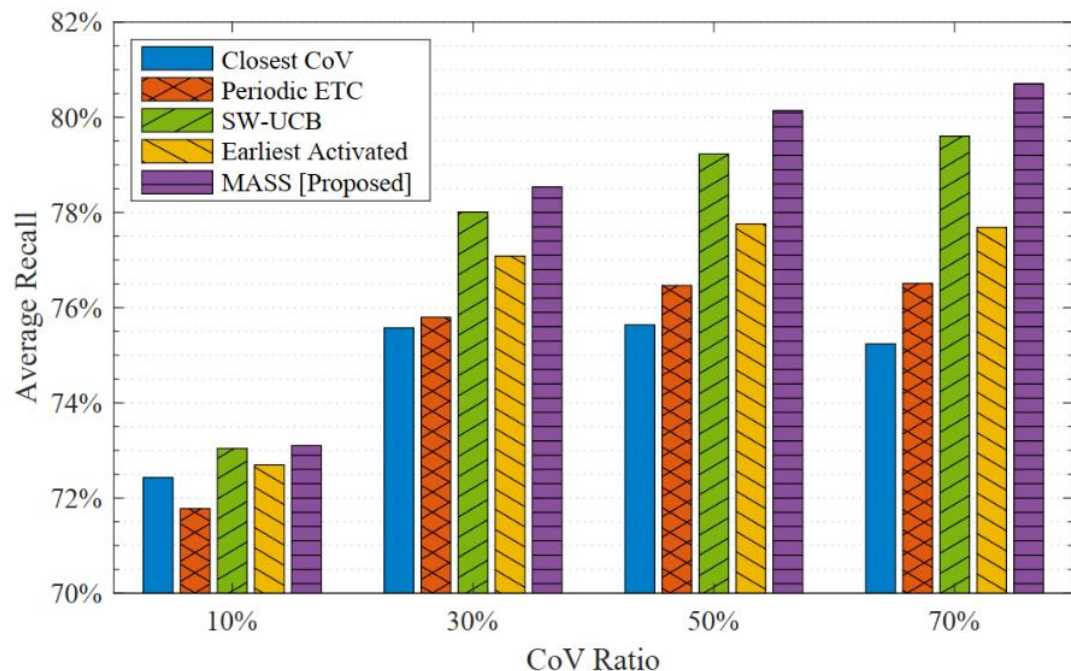
Conjecture 1. *Let $\beta = 15\sigma \log \sigma^{-1}$. For a sufficient long trip $T \geq \Omega(\sigma^{-2})$ with a dynamic candidate CoV set satisfying $|\mathcal{V}_t| \leq V_{\max}$, the expected average learning regret of MASS is bounded by*

$$\bar{R}_{MASS} \leq O(V_{\max} \sigma^2 \log^3(1/\sigma)) + 2\lambda.$$

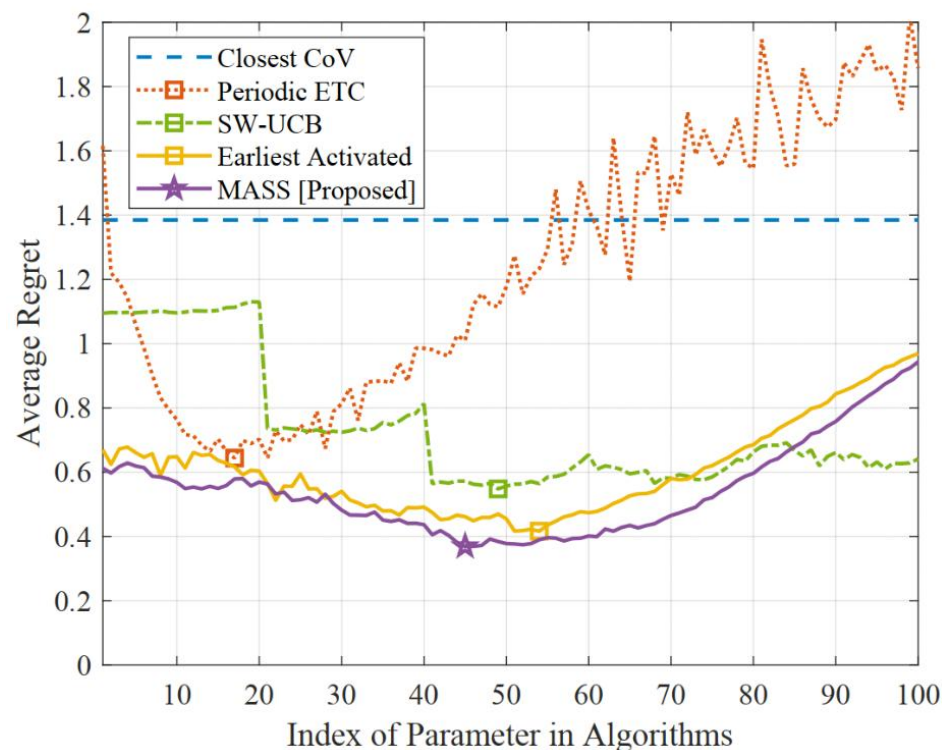
where λ is the arrival rate of candidate CoVs. When $\lambda \leq O(V_{\max} \sigma^2 \log^3(1/\sigma))$, the expected average learning regret is bounded by

$$\bar{R}_{MASS} \leq O(V_{\max} \sigma^2 \log^3(1/\sigma)).$$

基于多臂赌博机的传感信息调度算法



所提算法相较于基线算法能显著提升
Average Recall



所提算法相较于基线算法能显著降低
Learning Regret

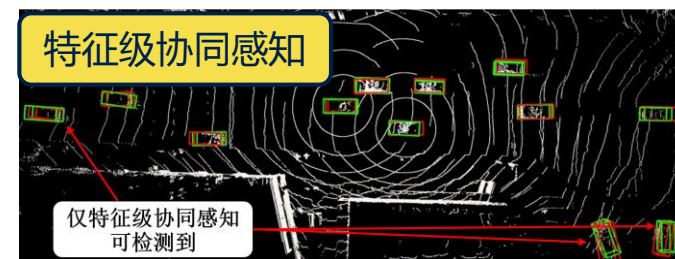
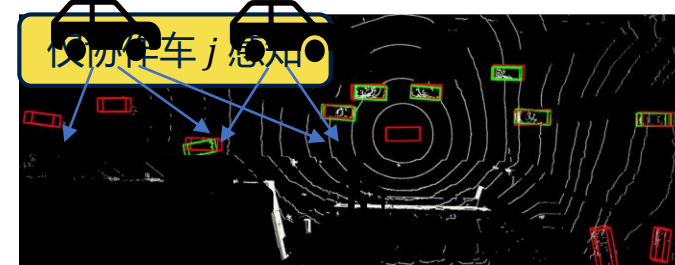
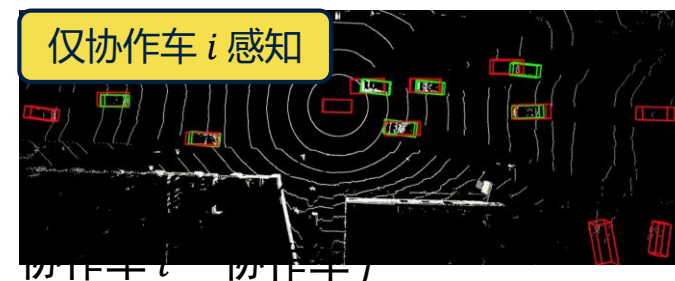
Y. Jia, R. Mao, Y. Sun, **S. Zhou**, and Z. Niu, "[MASS: Mobility-Aware Sensor Scheduling of Cooperative Perception for Connected Automated Driving](#)," *IEEE Trans. Veh. Technol.*, 2023, 72(11):14962-14977.

问题描述

- 决策：调度**协作车辆** $a_t \in \mathcal{V}_t \rightarrow$ 调度**协作车辆集合** $\mathcal{A}_t \subseteq \mathcal{V}_t$
- 目标：最大化加权检测召回率
- 约束：总通信带宽 W_t

问题难点

- 多车协作效用**深度耦合（非线性）**
 - 情况1：多传感源冗余
 - $g_t(\{i, j\}) < g_t(\{i\}) + g_t(\{j\})$
 - 情况2：存在仅协同感知才可检测的目标
 - $g_t(\{i, j\}) > g_t(\{i\}) + g_t(\{j\})$



特征级协同感知的经验研究

- 数据集：仿真数据集OPV2V^[21]、V2XSet^[22]
- 统计发现：在不少于3个协作车辆的场景中，绝大多数（99.94%、99.81%）目标的正确检测源于其中**某一个协作车辆**或**某两个协作车辆**的感知数据

感知目标集合 $\mathcal{P}_t(\mathcal{S})$

- **定义**：调度协作车辆组合 $\mathcal{S} \subseteq \mathcal{V}_t$ 时，可检测的目标集合

$$\mathcal{P}_t(\mathcal{S}) = \left\{ n \in \mathcal{O}_t \mid \Phi \left(\bigcup_{i \in \mathcal{S}} \mathcal{X}_{i,n}^{(t)} \right) = 1 \right\}$$

- 感知目标集合的**二阶近似**

一阶感知目标集合

二阶感知目标集合

$$\mathcal{P}_t(\mathcal{S}) = \left(\bigcup_{i \in \mathcal{S}} \mathcal{P}_t^{(1)}(i) \right) \cup \left(\bigcup_{i,j \in \mathcal{S}, i \neq j} \mathcal{P}_t^{(2)}(i,j) \right)$$

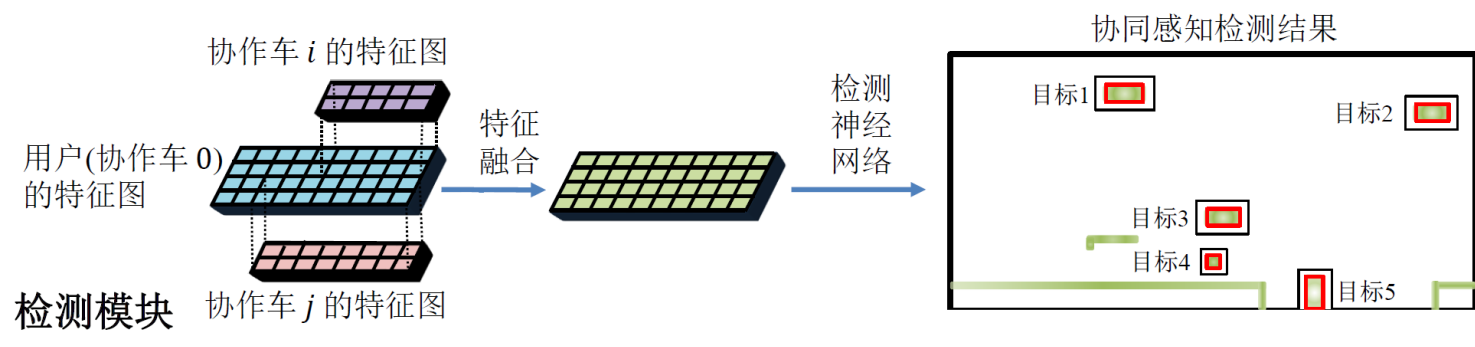
$\Phi(\cdot)$ ：检测神经网络

$\mathcal{X}_{i,n}^{(t)}$ ：协作车传感数据与目标 n 关联的部分

- 一阶感知目标集合： $\mathcal{P}_t^{(1)}(i) = \mathcal{P}_t(\{i\})$
- 二阶感知目标集合： $\mathcal{P}_t^{(2)}(i,j) = \mathcal{P}_t(\{i,j\}) \setminus \mathcal{P}_t^{(1)}(i) \setminus \mathcal{P}_t^{(1)}(j)$

拓展到多协作车的调度：检测回放机制

- 针对每个检测出目标的邻域，取出每个协作车辆的特征图、每两个车辆融合的特征图的对应位置，拼接成一张特征图，再次输入检测头检测
- 维护经验感知目标集合 $\hat{\mathcal{P}}^{(1)}(i) \leftarrow \mathcal{P}_t(\{i\})$, $\hat{\mathcal{P}}^{(2)}(i, j) \leftarrow \mathcal{P}_t(\{i, j\}) \setminus \mathcal{P}_t(\{i\}) \setminus \mathcal{P}_t(\{j\})$



拓展到多协作车的调度：优化问题

问题表述与求解

- 协作效用
- 优化问题表述为

$$g_t(\mathcal{A}_t) = \sum_{n \in \mathcal{P}_t(\mathcal{A}_t)} w_n^{(t)}$$

$\mathcal{P}_t(\mathcal{A}_t)$: 协作车集合 $\mathcal{A}_t$ 的感知目标集合

$w_n^{(t)}$: 目标重要性权重

$$\begin{aligned} \max_{\mathcal{A}_t \subseteq \mathcal{V}_t} \quad & \sum_{t=1}^T g_t(\mathcal{A}_t), \\ \text{s. t.} \quad & \sum_{i \in \mathcal{A}_t} B_{t,i} \leq W_t, \forall t = 1, 2, \dots, T. \end{aligned}$$

- 预算约束下最大覆盖问题(budgeted maximum coverage)的推广形式, NP-难问题

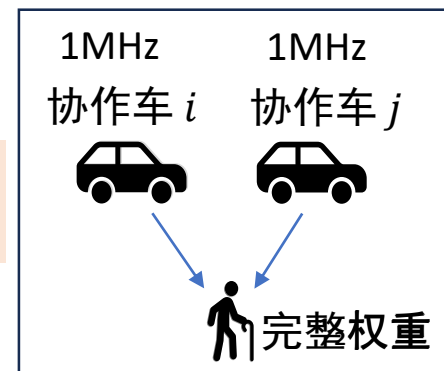
- 求解思路: 对于二阶感知目标集合, 引入“部分检测”的效用

$$g_t^+(\mathcal{S}) = \boxed{g_t(\mathcal{S})} + \lambda \sum_{n \in \mathcal{P}_t^C(\mathcal{S})} w_n(t) \boxed{\max_{\substack{i \in \mathcal{S} \\ j \notin \mathcal{S}}} \mathbb{I}\{n \in \mathcal{P}_t^{(2)}(i,j)\} \frac{B_{t,i}}{B_{t,i} + B_{t,j}}}$$

协作效用

调度车辆的最大带宽比例

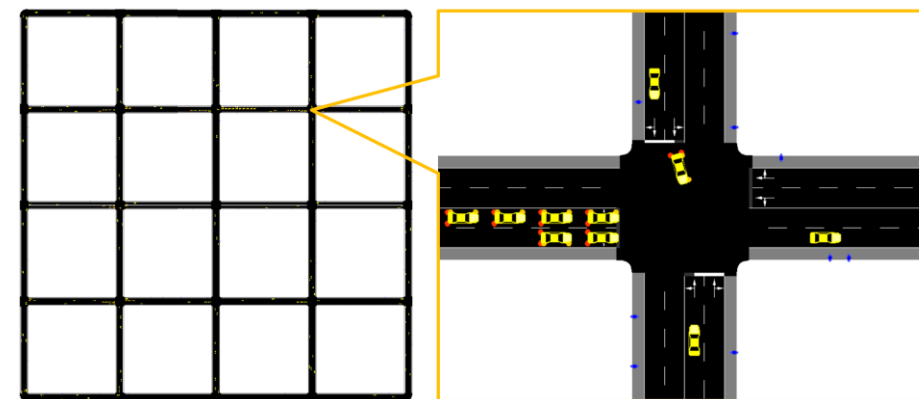
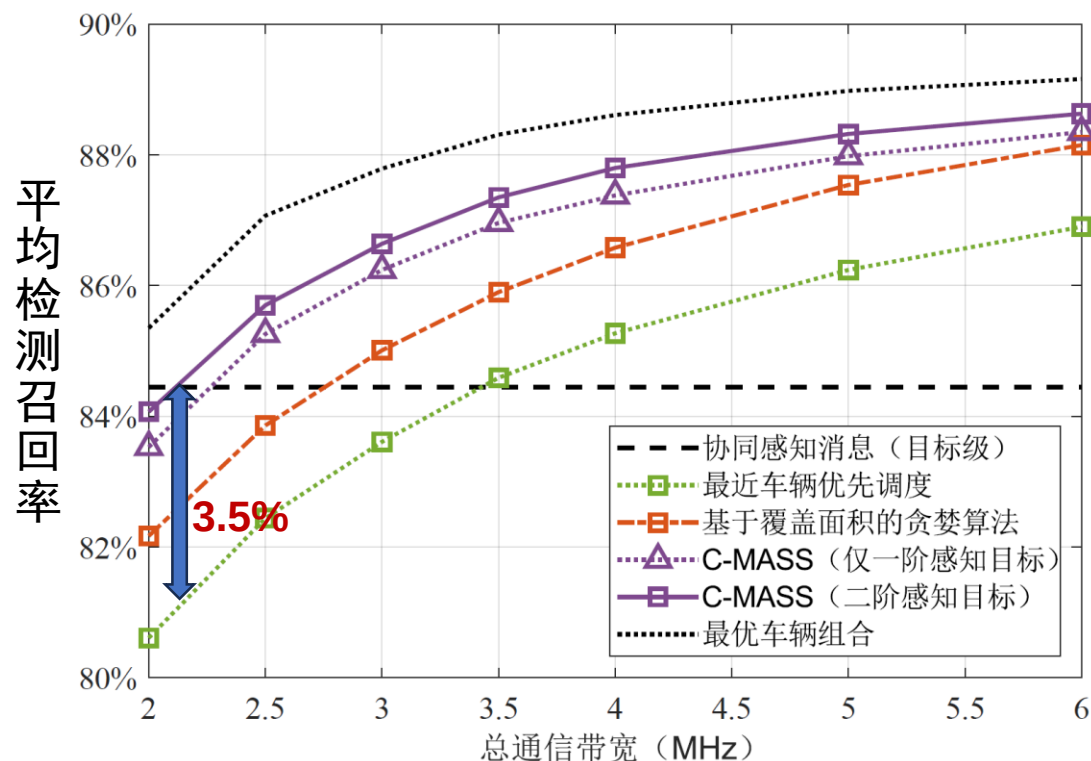
$\mathcal{P}_t^{(2)}(i,j)$: 二阶感知目标集合
 $B_i(t)$: 协作车带宽代价



- 参数 $\lambda \in [0,1]$ 平衡因为二阶感知目标而调度车辆的倾向

拓展到多协作车的调度：仿真结果

- 时隙长度0.1秒，连续10,000个时隙的驾驶轨迹
- 协作车辆比例：50%，可用通信带宽：2~6MHz



所提算法对比基线算法可最高提升召回率达 **3.5%**
(80.6%→84.1%)

利用二阶感知目标集合相较于仅利用一阶感知目标集合，提升召回率 **0.5%**

Y. Jia, R. Mao, Y. Sun, Z. Nan, S. Zhou, and Z. Niu, "C-MASS: Combinatorial Mobility-Aware Sensor Scheduling for Collaborative Perception with Second-Order Topology Approximation," IEEE Trans. Vehicular Technology, accepted.

- 目录 -

CONTENT

1 研究背景

2 研究内容

① 传输：传输什么？

② 组网：与谁协同？

3 数据集与实验平台

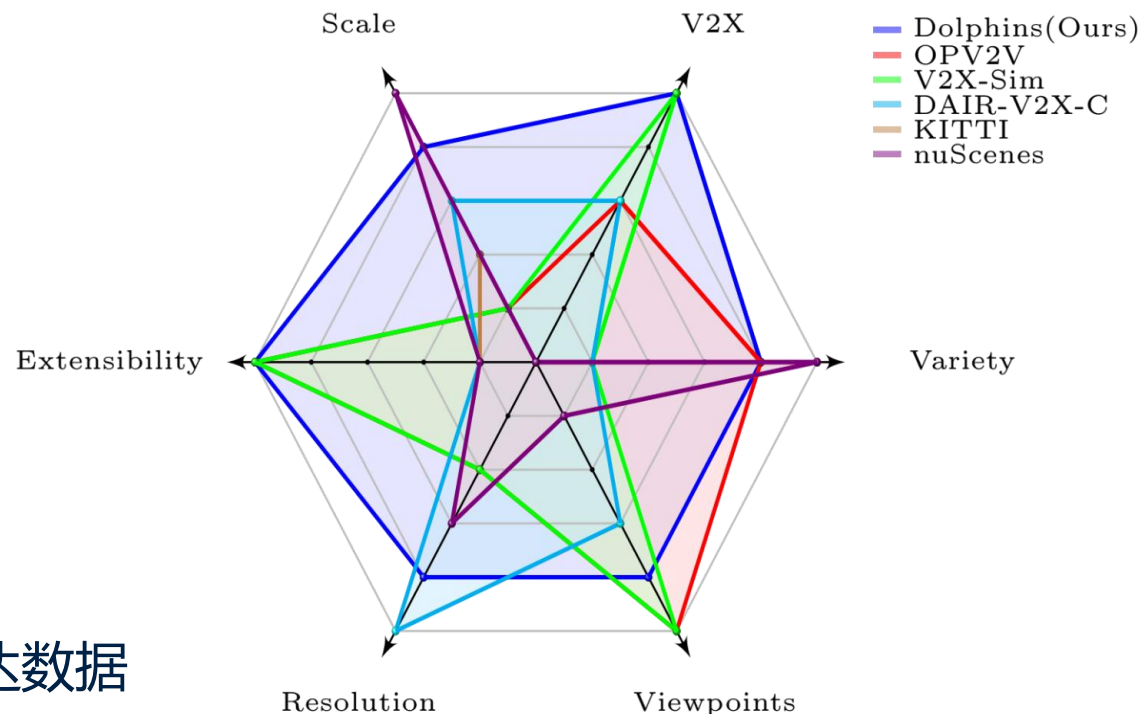
4 思考与展望



大规模协同感知数据集——DOLPHINS

DOLPHINS: Dataset for cOLlaborative Perception enabled Harmonious and INterconnected Self-driving

- 基于开源交通模拟器 CARLA 构建生成
- 对比其他自动驾驶数据集具有六大特点：
 - **大规模**
 - 共 42376 帧数据
 - 共 292549 个3D标注物体
 - 精准的地理位置和校准矩阵
 - **多场景**: 6个典型自动驾驶场景
 - **多视角 (V2X)** : 目前支持三个视角
 - **多模态高分辨率**: 可生成Full-HD图像和64线雷达数据
 - **可扩展性**: 提供完整的API可根据需要生成数据



www.dolphins-dataset.net

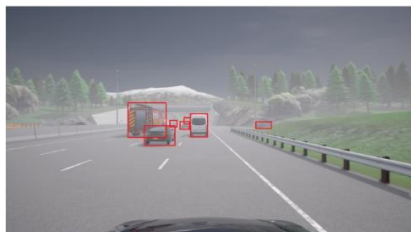
R. Mao, J. Guo, Y. Jia, Y. Sun, **S. Zhou**, and Z. Niu "DOLPHINS: Dataset for collaborative perception enabled harmonious and interconnected self-driving," *Asian Conf. Comput. Vision (ACCV)*, Dec. 2022.

大规模协同感知数据集——DOLPHINS

DOLPHINS: Dataset for cOLlaborative Perception enabled Harmonious and INterconnected Self-driving

■ 多视角+多模态:

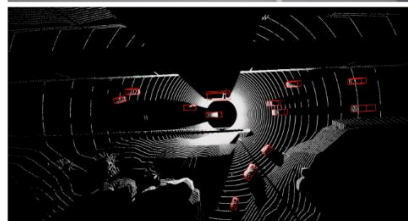
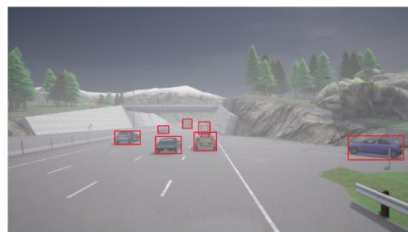
- 包括本车、辅助车(Aux Vehicle)及路侧单元(RSU)
- 每帧数据由时间对齐的多视角的点云和图像组成



(a) Ego vehicle



(b) RSU



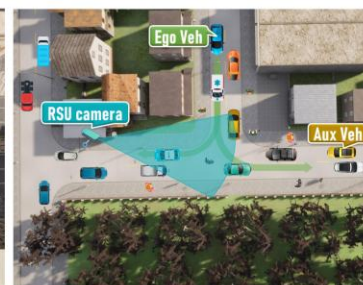
(c) Aux vehicle

■ 多场景:

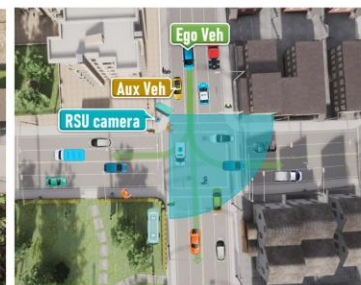
- 包含6个典型自动驾驶场景: 城市十字路口、丁字路口、匝道、山路等



(a) Scenario 1 (7046 frames)



(b) Scenario 2 (7020 frames)



(c) Scenario 3 (7043 frames)



(d) Scenario 4 (7057 frames)



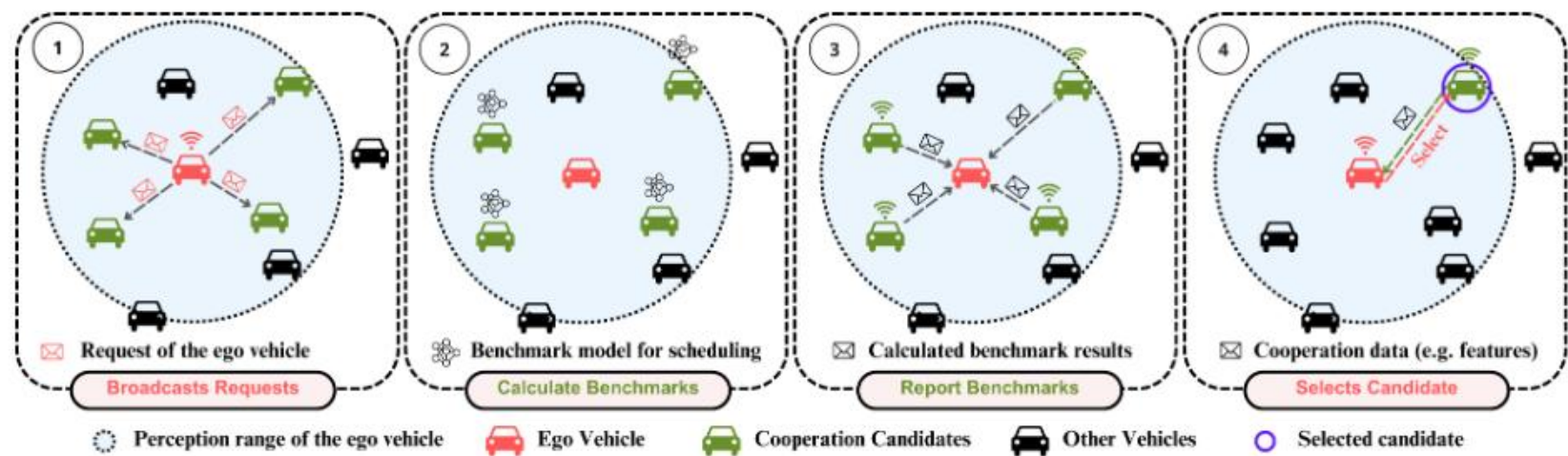
(e) Scenario 5 (7011 frames)



(f) Scenario 6 (7199 frames)

大规模协同感知数据集 —— WHALES

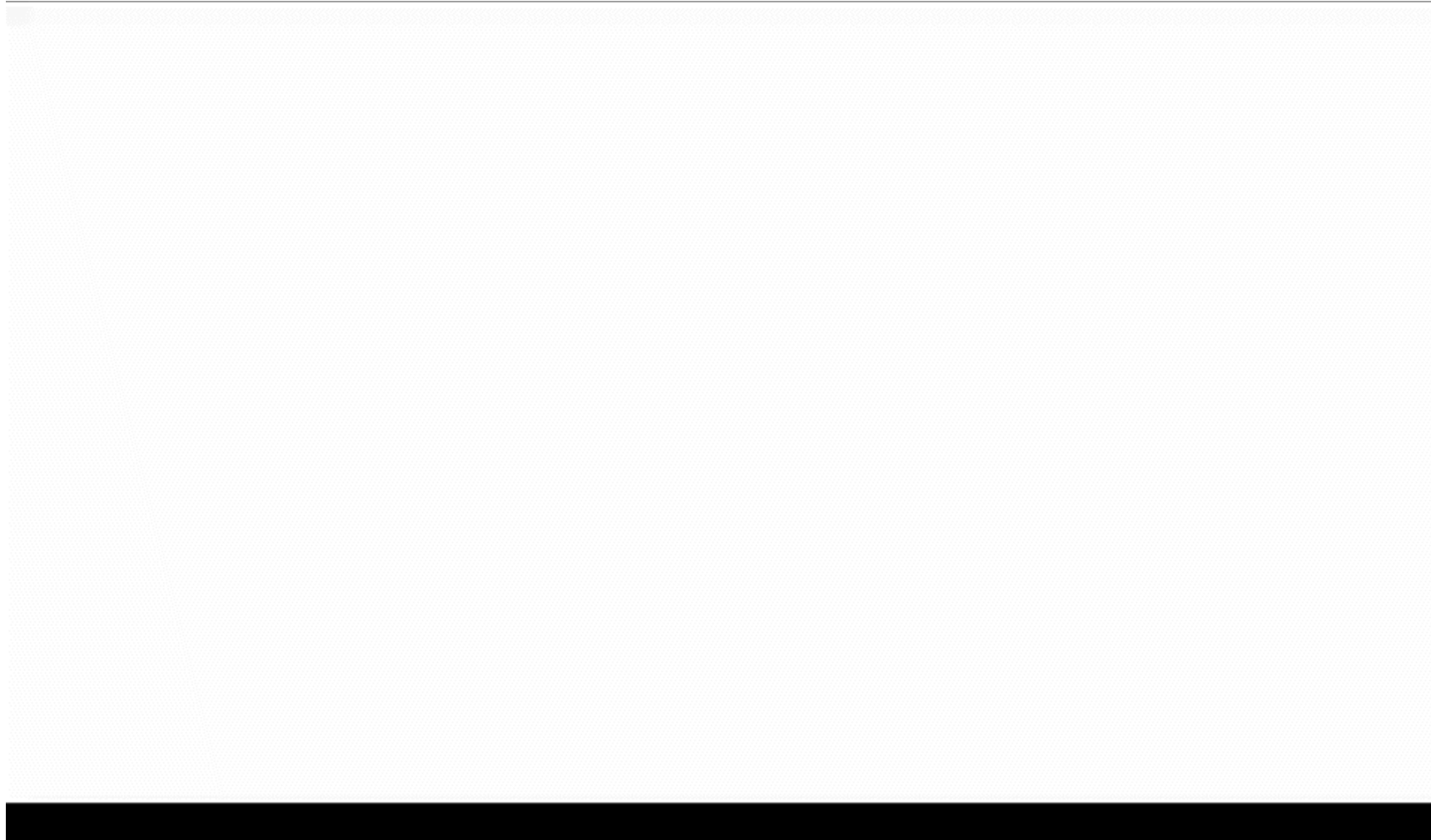
- **Advantages:** Large volume, Numerous agents, More cooperative tasks, High label density, More label information
- **Benchmark** methods in 3D stand-alone and cooperative object detection, and agent scheduling.
- **Implemented** with MMdetection3D, easy to reproduce & modify
- Code and dataset available at <https://github.com/chensiweiTHU/WHALES>



mAP Scores on 3D Object Detection using Different Scheduling Policies (50m/100m)

Inference \ Training	No Fusion	Closest Agent	Single Random	Multiple Random	Full Communication
No Fusion	50.9/26.2	50.9/23.3	51.3/25.3	50.3/22.9	45.6/18.8
Closest Agent	39.9/20.3	58.4/30.2	58.3/32.6	57.7/30.5	55.4/10.8
Single Random	43.3/22.8	57.9/31.0	58.4/33.3	57.7/31.4	55.0/14.6
MASS	55.5/11.0	58.8/33.7	58.9/34.0	57.3/32.3	54.1/27.4
Historical Best	54.8/29.6	58.6/31.7	58.9/34.0	58.3/32.6	54.1/27.4
Multiple Random	34.5/16.9	60.7/35.1	61.2/37.1	61.4/36.4	58.8/12.9
Full Communication	29.1/10.5	63.7/38.4	64.0/39.9	64.7/41.3	65.1/39.2

Large scale dataset for CP —— WHALES



<https://github.com/chensiweiTHU/WHALES>

R. Wang, S. Chen, Z. Song, S. Zhou*, “**WHALES: A Multi-agent Scheduling Dataset for Enhanced Cooperation in Autonomous Driving**,” 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Oct. 2025.

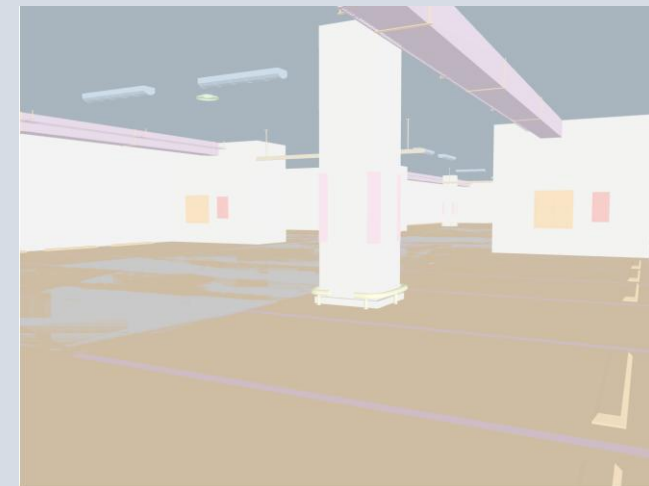
停车场环境下的网联自动驾驶仿真平台

逼近真实场景的仿真环境

- 高解析度3D场景（基于CARLA）
- 模拟光照条件的变化
- 支持动态和交互场景
 - 可进行闭环测试



assign
radio
materials



- 基于Ray-Tracing的信道生成



全局视角

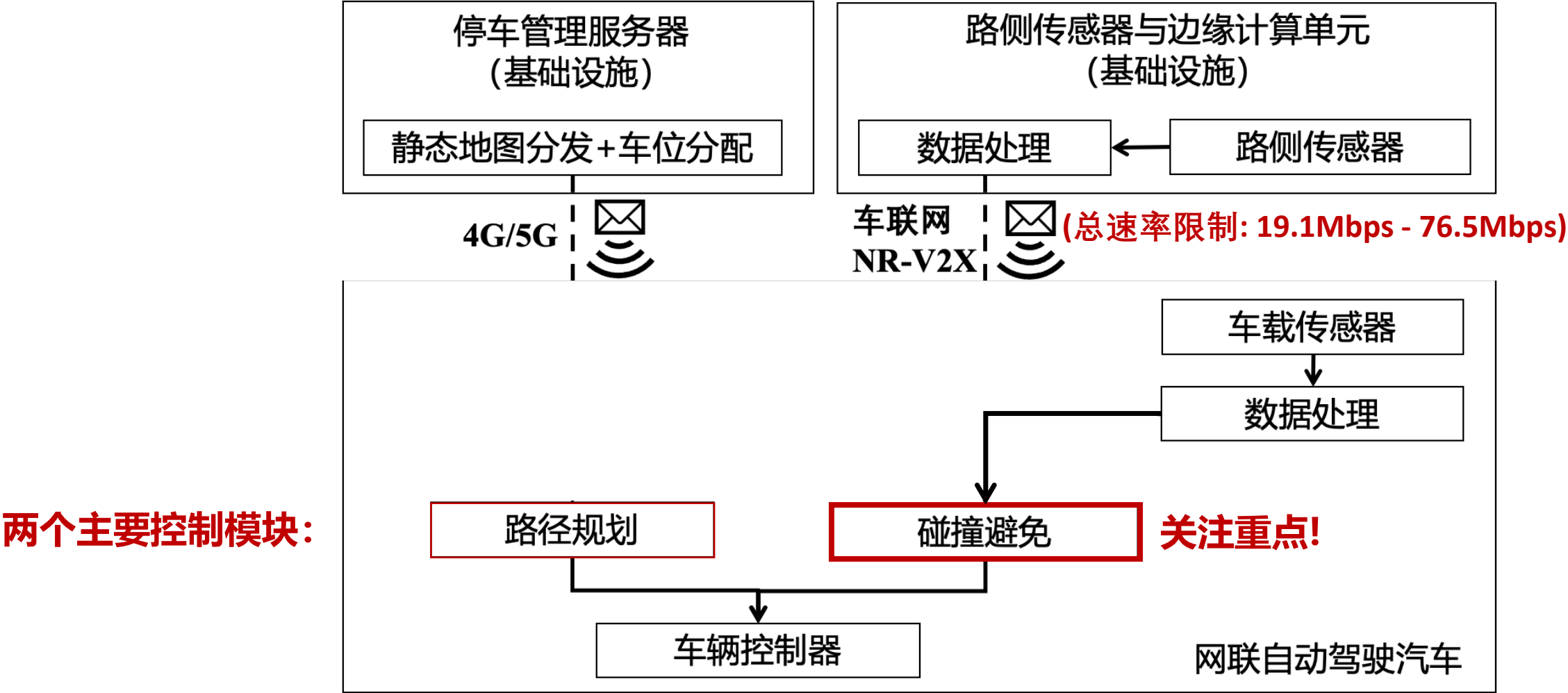


主车视角



路侧单元视角

网联自动代客泊车系统：系统框架

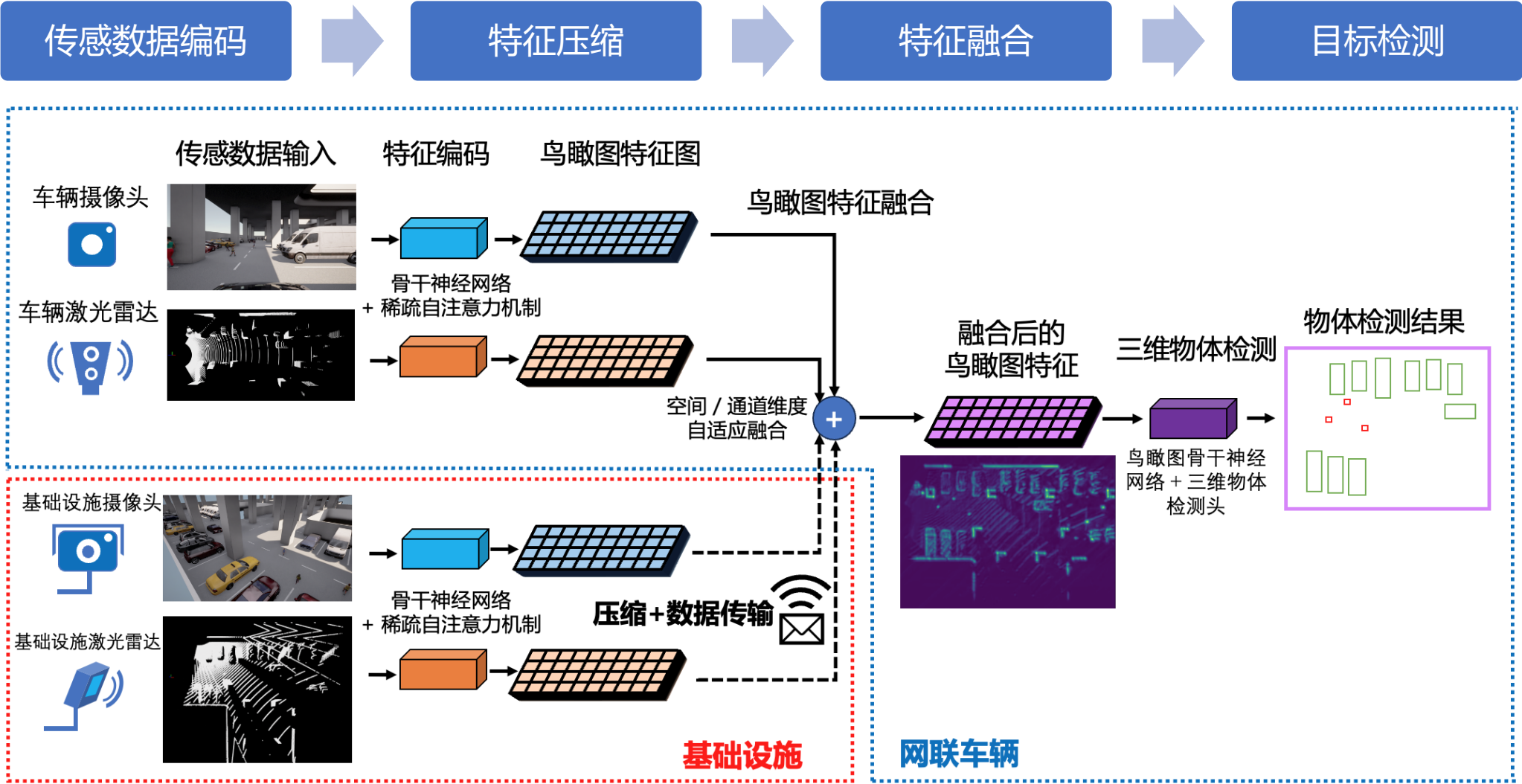


Y. Jia, J. Zhang, S. Lu, B. Fan, R. Mao, S. Zhou, and Z. Niu, "Infrastructure-Assisted Collaborative Perception in Automated Valet Parking: A Safety Perspective," *IEEE VTC-Spring'24*, May 2024.

网联自动代客泊车系统: 基于鸟瞰图特征的协同感知



- 基于鸟瞰图特征的多源多模态传感信息融合框架



网联自动代客泊车系统：感知-控制联合仿真

- 场景 1：行人从停泊车辆之间穿越内部道路

- 蓝色框: 检测到的目标
- 红色框: 未检测到的目标

- 单车感知



- 基础设施辅助的协同感知



- 场景：行人从停泊车辆之间穿越内部道路
 - 在给定传感器配置和巡航速度下，统计20次随机实验中的事故数量
 - 激光雷达相比于摄像头可靠性更高，因为更擅长提前检测半遮挡的物体

传感器配置		车辆自动驾驶巡航速度							
自动代客泊车车辆	基础设施	3.0m/s	4.0m/s	5.0m/s	6.0m/s	7.0m/s	8.0m/s	9.0m/s	10.0m/s
摄像头	无	0/20	1/20	3/20	3/20				
	摄像头		0/20	0/20	1/20	2/20	2/20		
	摄像头+激光雷达			0/20	0/20	1/20	2/20	1/20	
摄像头+激光雷达	无			0/20	0/20	1/20	1/20	1/20	
	摄像头				0/20	0/20	2/20	1/20	1/20
	摄像头+激光雷达						0/20	0/20	1/20

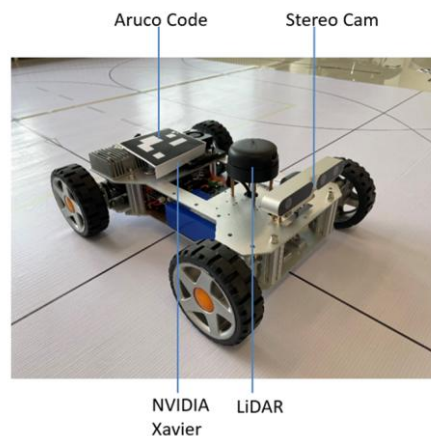
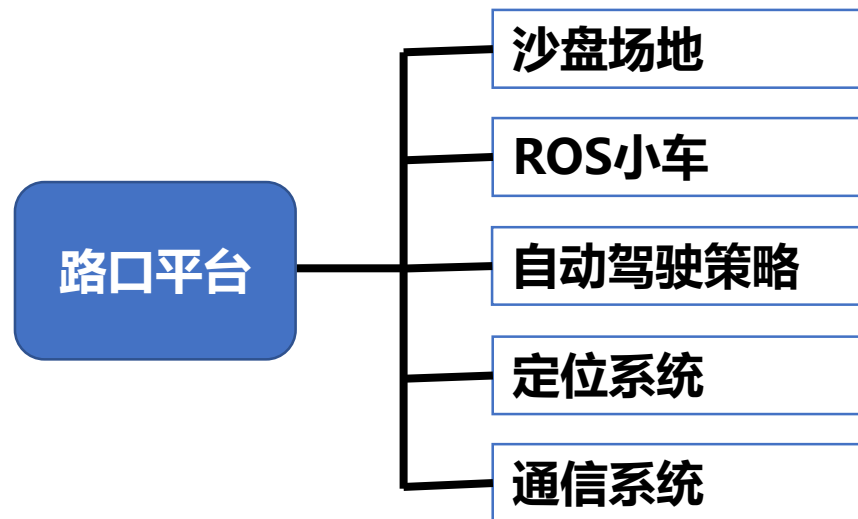
利用协同感知，自动代客泊车的安全巡航速度可提升7~11km/h



停车时间和离场时间显著降低50% (60s -> 30s)

智能路口沙盘平台：系统设计

平台设计：双向双车道（左转/直行）路口，小车装载**激光雷达**和**双目摄像头**



智能路口沙盘平台演示：协同自动驾驶

智能路口沙盘平台设计：演示无红绿灯自动驾驶路口通行



- 目录 -

CONTENT

1 研究背景

2 研究内容

① 传输：传输什么？

② 组网：与谁协同？

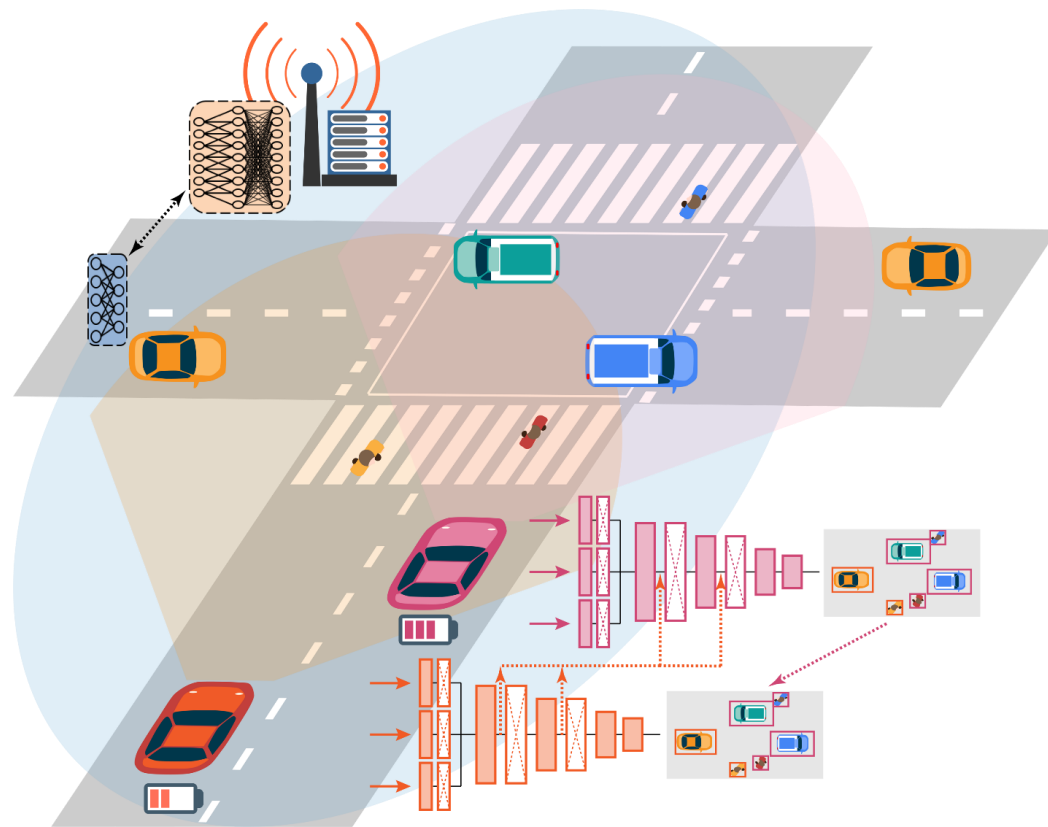
3 数据集与实验平台

4 思考与展望



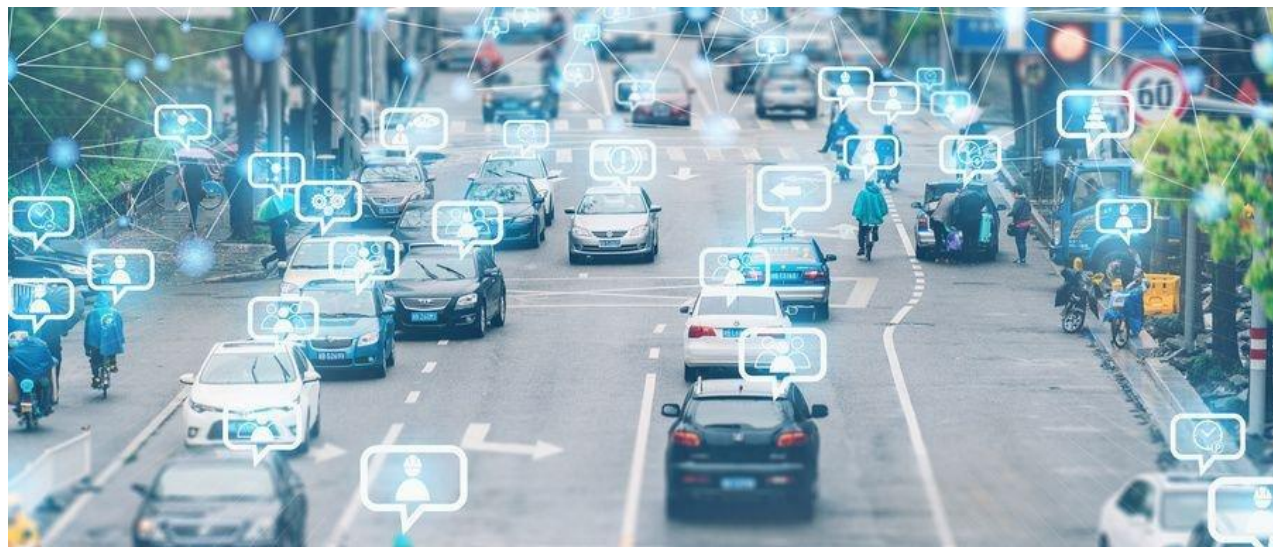
无人系统协同感知中的若干开放科学问题：

- 高效的多源、多模态传感数据融合算法设计
- 非精准位姿（外参）条件下的鲁棒协同
- 能耗受限、通信资源受限下的信源信道联合编码
- 新的VLM模型如何实现多体协同
- 无线网络的协同感知“容量”



汽车：代步工具 → 安全可靠的座舱、移动的私人助理

车路协同、万物智联：
实现无盲区的安全驾驶



多媒体座舱：
沉浸式智慧体验



- **未来车辆**具有强大的**传感、通信、处理及自供电**能力，可以充当**移动云或移动基站**，有助于解决城市中边缘云和小基站密集部署的瓶颈
- 载有物质（人、货）、能量（充放电）、信息（移动云、移动基站）智能车辆为城市中各种终端提供**机会服务**，并把**智能**传播到城市各个角落



谢谢大家！
请批评指正

报告人：周盛