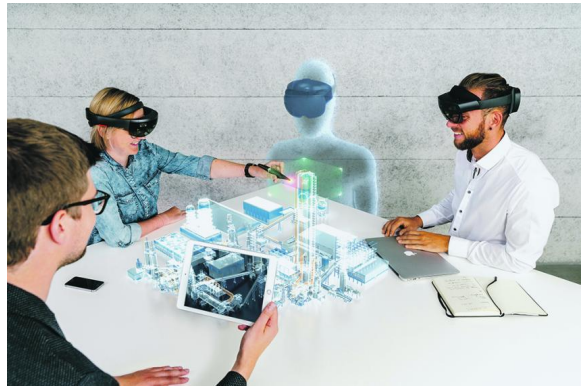
多智能体协同感知研究进展

王永才
中国人民大学 信息学院 计算机系
ycw@ruc.edu.cn
2024年12月4日

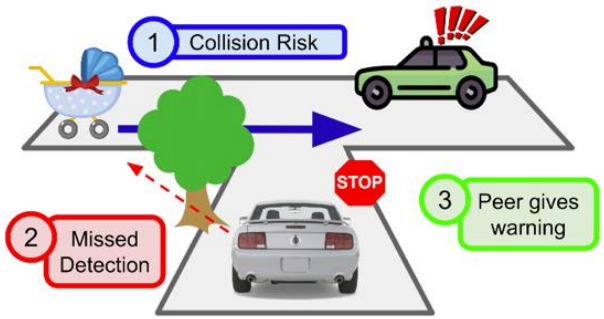
多智能体协同感知主要应用场景

AR，VR场景中的多手机协同感知



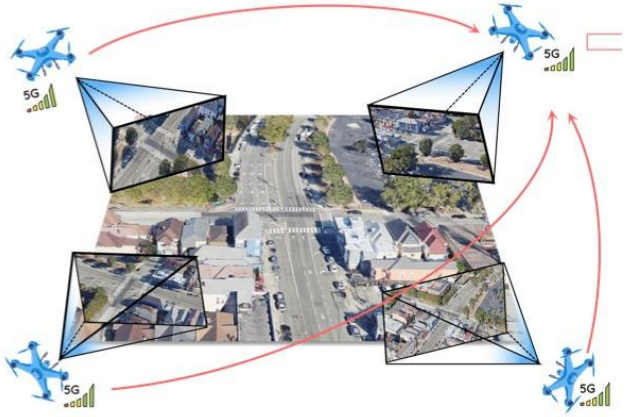
AR，VR场景中的**协同感知、协同定位、建图、交互式多人游戏**等

自动驾驶场景中的多车协同感知



多车协同感知、提高自动驾驶的安全性、主要问题是**降低通讯成本、提升检测性能**。

军民应用中多无人机的协同感知



多无人机多视角的协同感知，**提高多目标检测、追踪的准确性**。

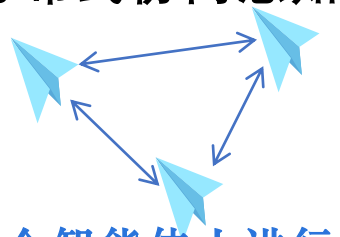
协同感知（Collaborative Perception）方法分类

协同方式上分类

- 集中式协同感知

智能体将信息汇聚到云端，在云端进行信息融合，提高云端感知能力

- 分布式协同感知



每个智能体上进行分布式融合，提升每个智能体感知能力

融合方法上分类

- 早期协同（Early Collaboration）
  - 传输和融合原始感知数据
  - 通信成本高
- 后期协同（Late Collaboration）
  - 传输和融合智能体本地的感知结果
  - 通信成本低、但无法发现各智能体都看不到的目标
- 中期协同（Intermediate Collaboration）
  - 传输和融合压缩的特征数据
  - 通信成本居中，且能融合发现大家都无法发现的目标。

融合模态上分类

- 单模态融合
  - 基于视觉的协同感知
  - 基于雷达的协同感知
    - where2com, V2VNet
- 多模态融合
- 融合视觉和雷达的协同感知
  - 基于BEV的多模态融合
    - BEV-Fusion
    - CoBEVT
  - 基于Query的多模态融合
    - Futr3D, Transfusion

目录

- CoISLAM: 多手机协同SLAM系统 （ACM MM2023）
- RoCo：多车协同感知系统 （ACM MM2024）
- SAHNet: 多车协同感知系统 （ADVEI 2024）
- GSLAMOT: 无人车同步定位建图与多目标追踪 （ACM MM2024）
- DroneMOT: 无人机多目标追踪系统（ICRA 2024）

ColSLAM: A Versatile Collaborative SLAM System for Mobile Phones Using Point-Line Features and Map Caching

Wanting Li[1], **Yongcai Wang***[1], Yongyu Guo[1], Shuo Wang[1], Yu Shao[1], Xuewei Bai[1], Xudong Cai[1], Qiang Ye[2], Deying Li[1]

1. School of Information, Renmin University of China, Beijing, China
2. Faculty of Computer Science, Dalhousie University, Halifax, Canada

发表于 ACM MM2023，CCF A
李婉婷、王永才等
中国人民大学信息学院

CoISLAM应用场景：多手机协作完成同步定位和建图（SLAM）

CoI：**Col**laborative

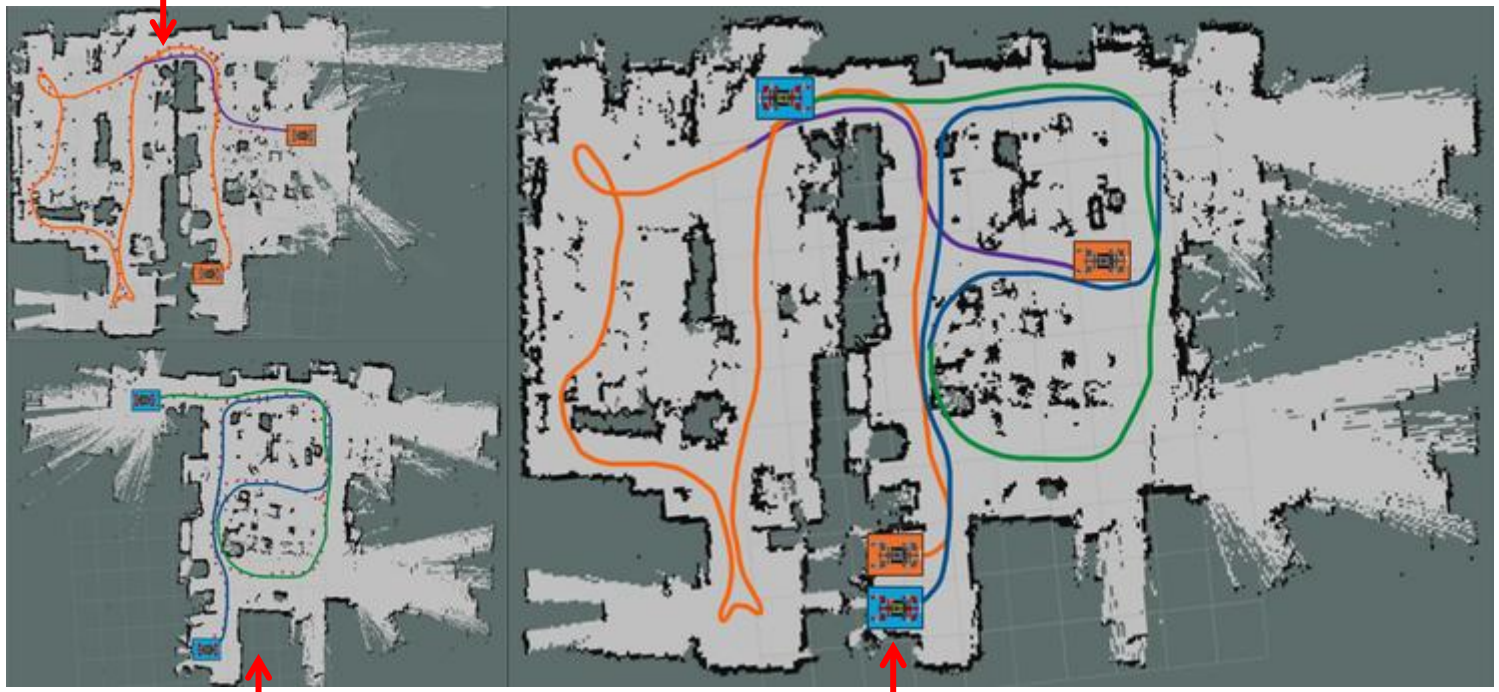
S：**S**imutanously

L：**L**ocating

A：**A**nd

M：**M**apping

Agent1的局部定位建图结果



Agent2的局部定位建图结果

云端融合之后的全局定位建图结果

关键问题：1.多智能体SLAM时，云端如何快速响应移动端地图融合的要求？

2.如何充分利用云端融合的结果，纠正移动端定位建图的偏差？

ColSLAM: A Versatile Collaborative SLAM System for Mobile Phones Using Point-Line Features and Map Caching

Demo Video

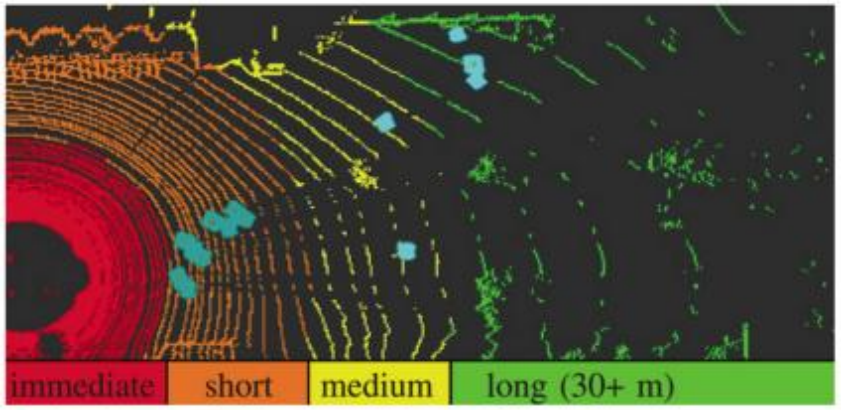
自动驾驶中的协同感知

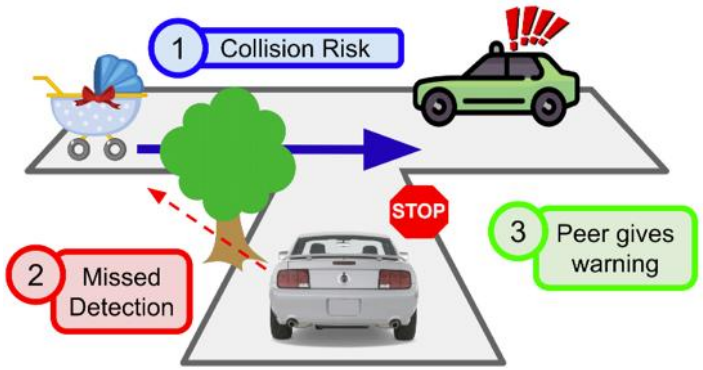
ACM MM2024，ADEVI 2024

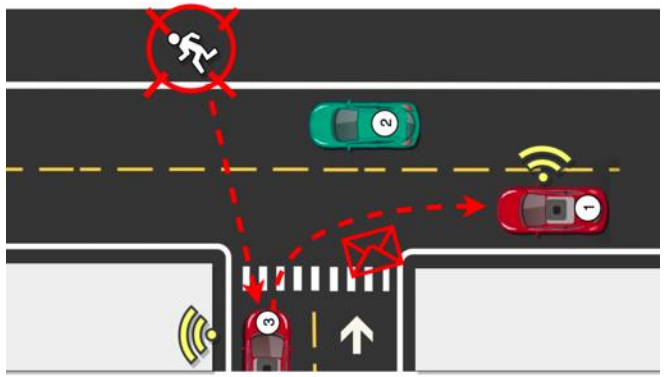
黄哲、王永才等
中国人民大学信息学院

自动驾驶中的单车感知的不足

- 单车自动驾驶已经被广泛研究[1]
- 由于雷达传感器的稀疏测量，**远距离目标感知**是个挑战[2]
- 单车感知也**受遮挡影响**[2]





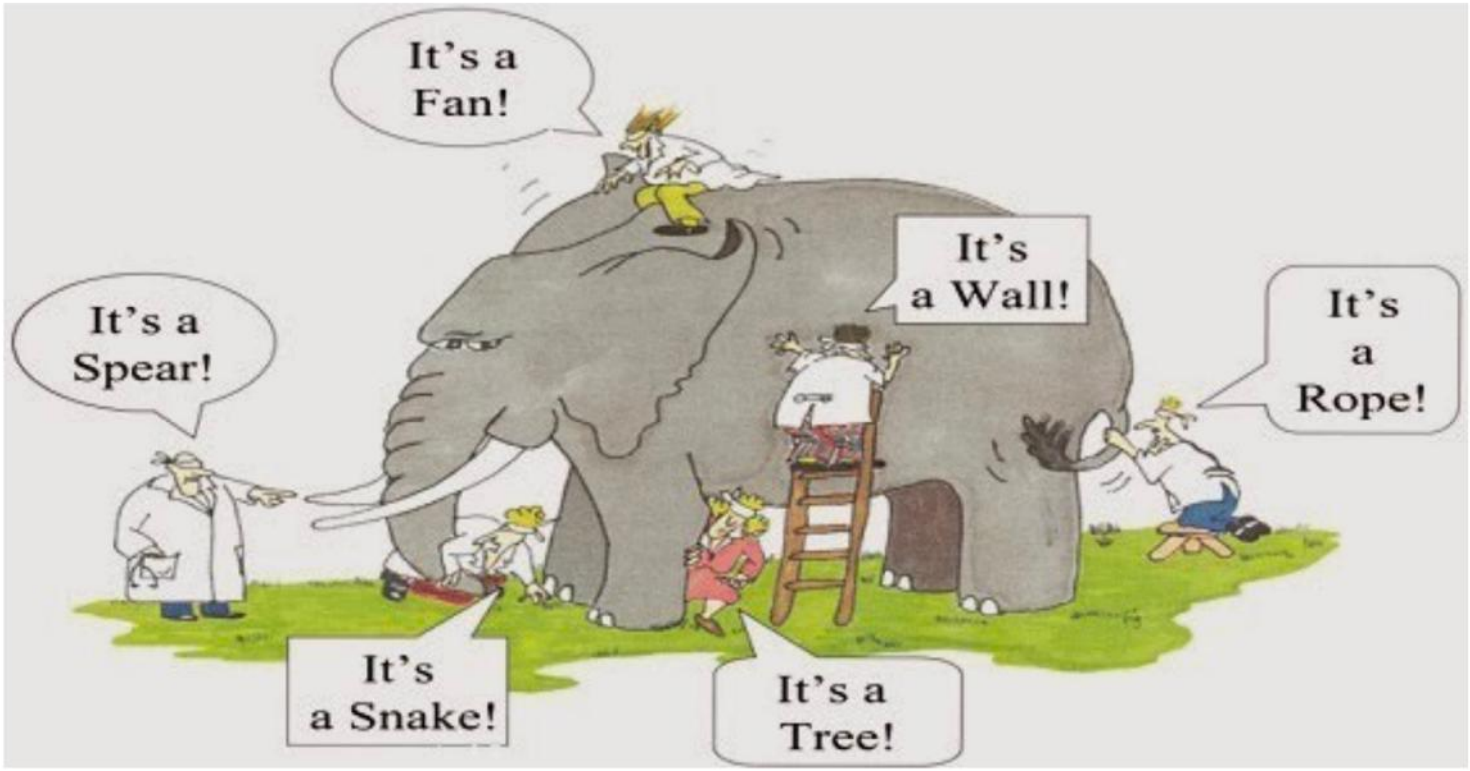


[1] Li Y, Ibanez-Guzman J. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems[J]. IEEE Signal Processing Magazine, 2020, 37(4): 50-61.

[2] Wang T H, Manivasagam S, Liang M, et al. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020: 605-621.

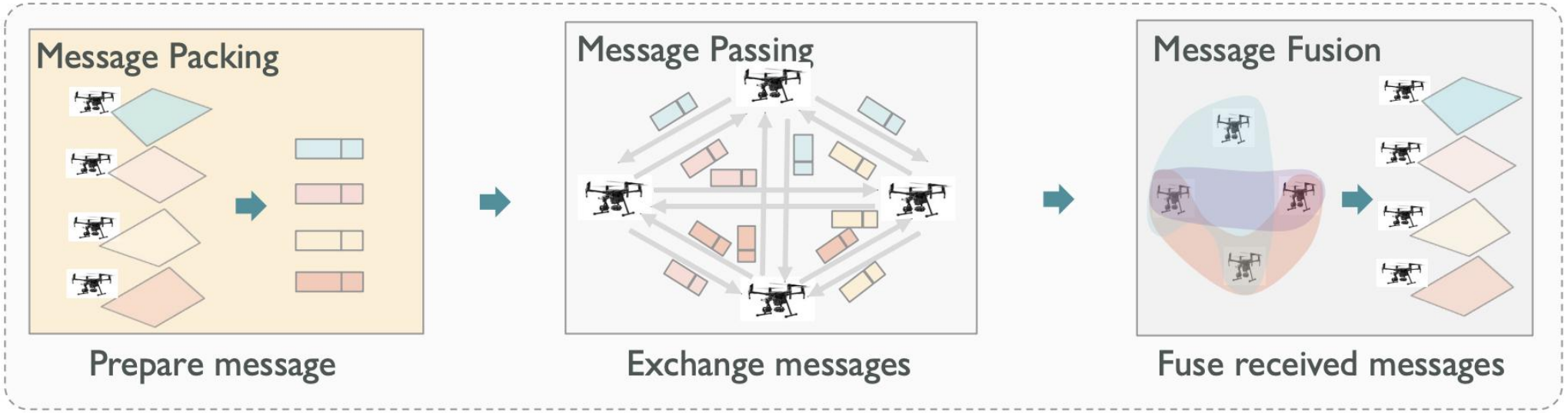
协同感知 See further. See better (Moreevidence). See through occlusion

What you see is what you get ?



Collaboration! Holistic view!

Background



V2VNet	ECCV 2020	Where2Comm	NeurIPS 2022	CoHFF	CVPR 2024
DiscoNet	NeurIPS2021	CoBEVFlow	NeurIPS 2023	CoopDet3D	CVPR 2024
V2X-ViT	ECCV 2022	CoCa3D	NeurIPS 2023		
CoBEVT	CoRL 2022	V2X-Graph	NeurIPS 2024		

RoCo: Robust Cooperative Perception By Iterative Object Matching and Pose Adjustment

Zhe Huang[1†], Shuo Wang[1†], Yongcai Wang[1*], Wanting Li[1], Deying Li[1], Lei Wang[2]

[1] Renmin University of China, School of Information

[2] University of Wollongong, School of Computing and Information Technology

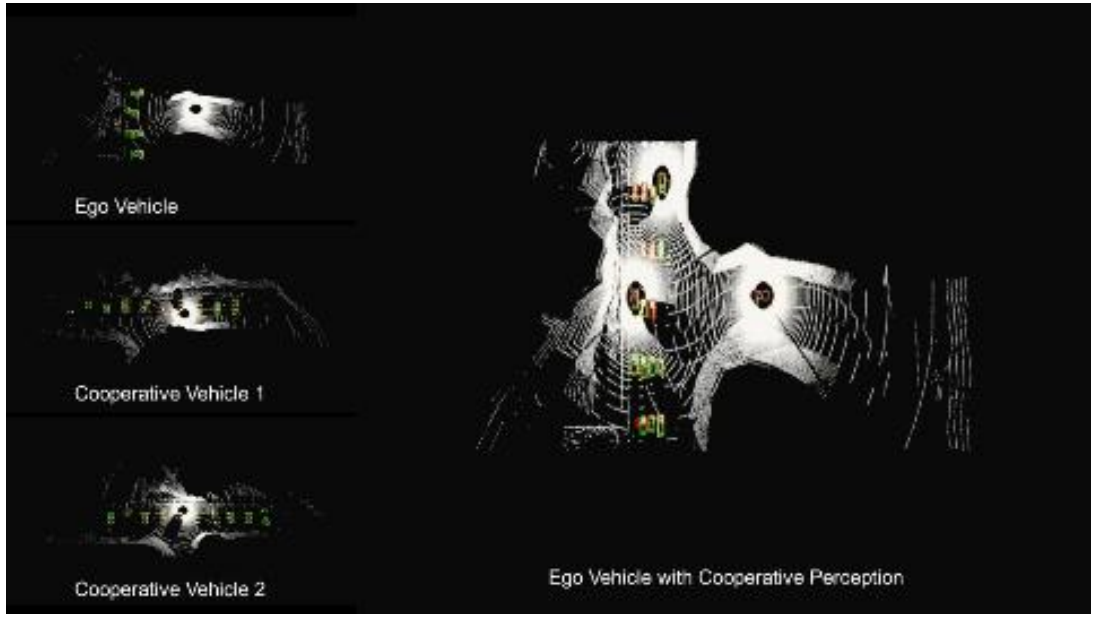
Project Page & Codes: https://github.com/HuangZhe885/RoCo

Multimedia 2024 CCF A Oral 3.97%

Background

Collaborative perception relies on **accurate poses** of agents

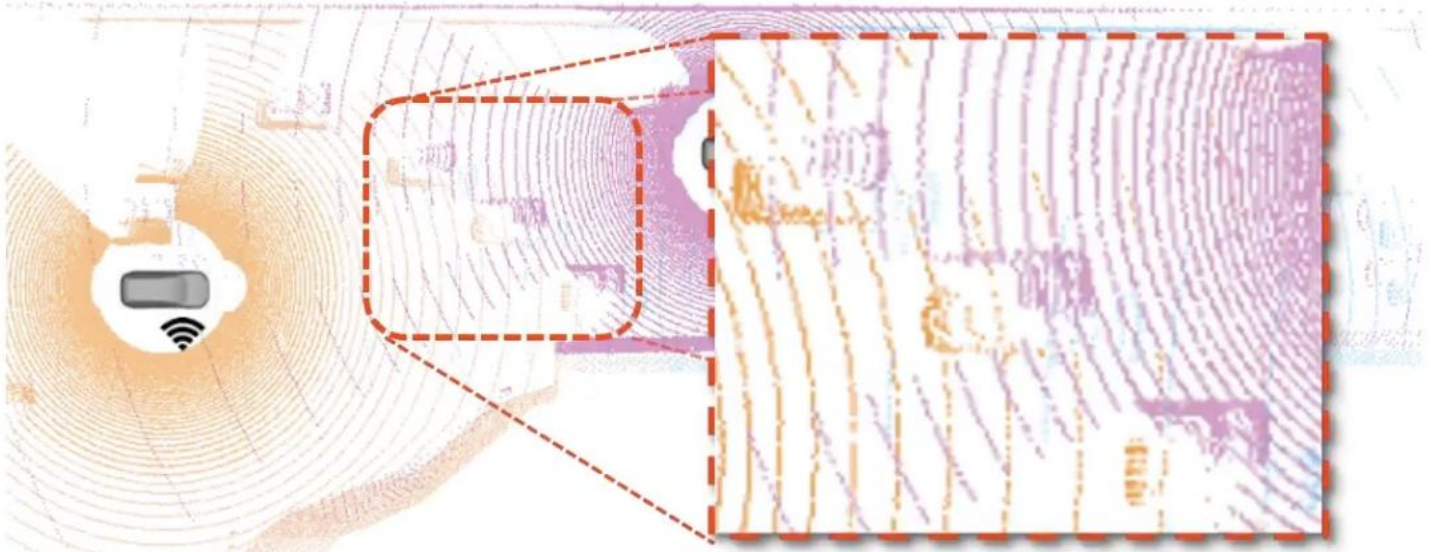
However, poses measured by localization module are **always noisy**



Visualization of collaborative agents

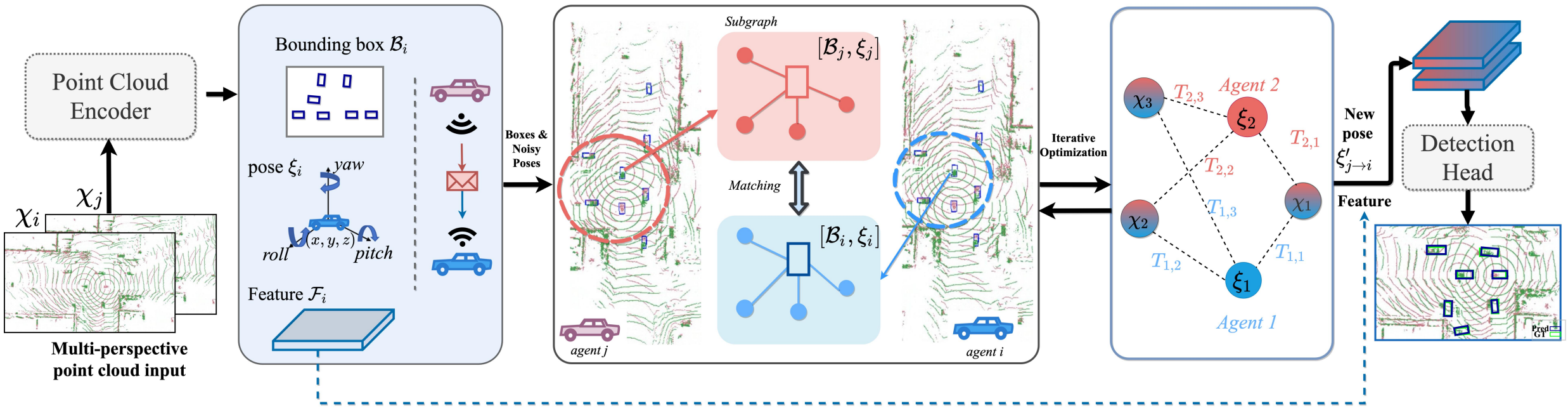
Background

- Due to inaccurate poses, data misalignment often occurs during fusion
- It leads to feature misalignment and significantly reduces collaborative performance.



Pose errors lead to point cloud misalignment

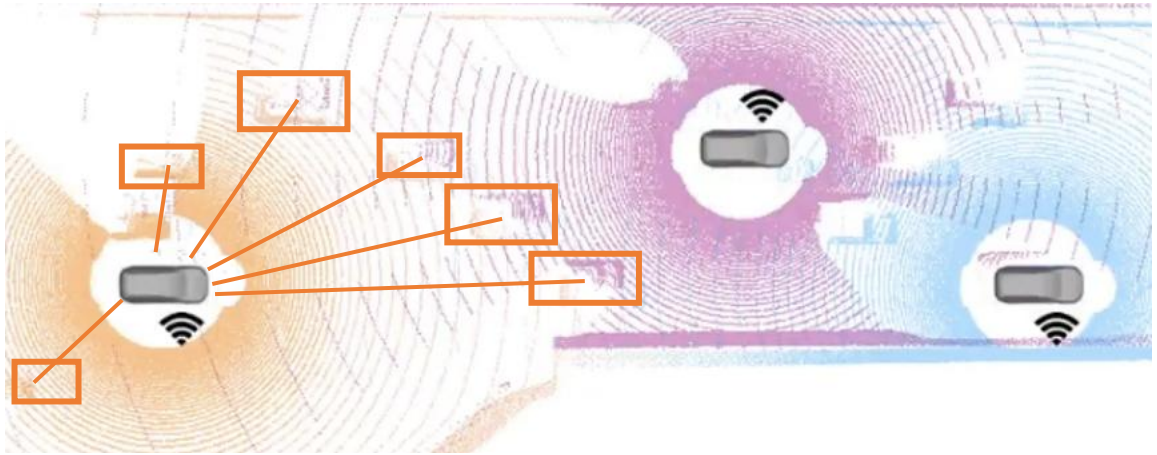
RoCo : Overview



A hybrid perception framework that **combines the strengths of late fusion and intermediate fusion**

RoCo : Overview

Single agent bounding box detection



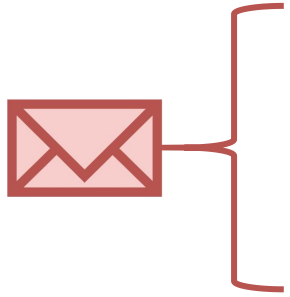
height, width, length, yaw

Bounidng box: $[x, y, z, \mathrm{h}, w, l, \theta, \sigma_x^2, \sigma_y^2, \sigma_\theta^2]$

3d center

uncertainty of x, y, yaw

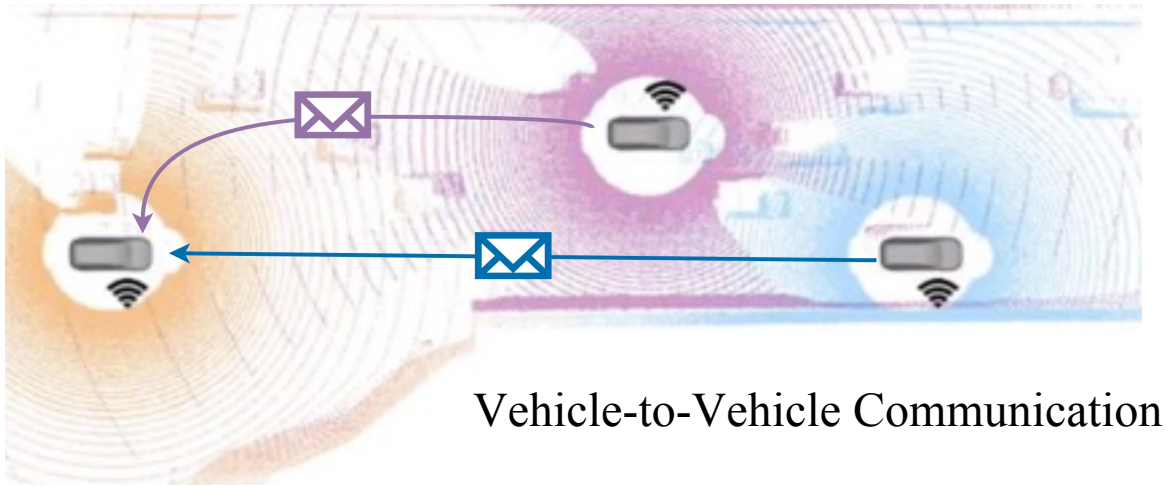
Information Sharing



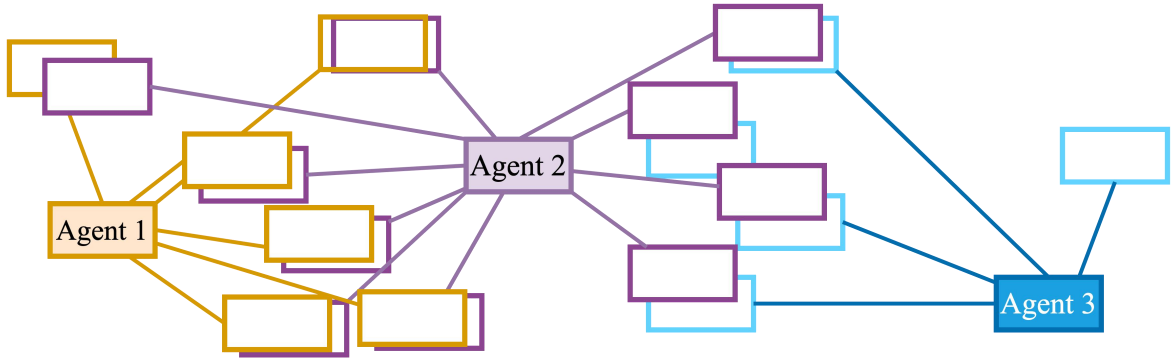
1. **Box detections**
2. **Feature map**
3. **Noisy pose**

Broadcast the message to other agents

RoCo : Overview



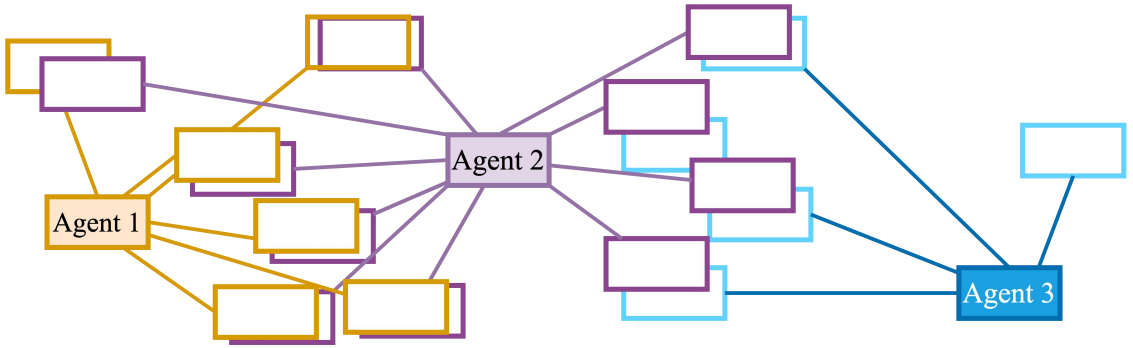
Vehicle-to-Vehicle Communication



The Ego agent aggregate all bounding boxes locally

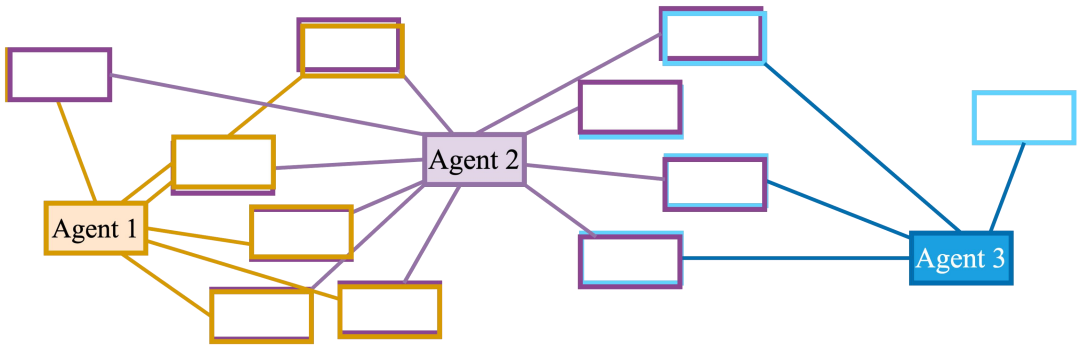
RoCo : Overview

Noisy Poses



Object matching and pose graph optimization

Corrected relative poses



RoCo : Results

Dataset	DAIR-V2X					V2XSet				
Method/Metric	AP@0.5 ↑									
Noise Level (σ_t/σ_r $(m/°)$)	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.8/0.8	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.8/0.8
F-Cooper[9]	73.4	72.3	70.5	69.2	67.1	78.3	76.3	71.2	65.9	62.0
FPV-RCNN[51]	65.5	63.1	58.0	58.1	57.5	86.5	85.3	68.7	62.1	49.5
V2VNet[40]	66.0	65.5	64.6	63.6	61.7	87.1	86.0	83.2	79.7	75.0
Self-Att[47]	70.5	70.3	69.5	68.5	67.8	87.6	86.8	85.4	83.7	82.1
V2X-ViT[45]	70.4	70.0	68.9	67.8	66.0	91.0	90.1	86.9	84.0	81.8
CoAlign[28]	74.6	73.8	72.0	70.0	69.2	91.9	90.9	88.1	85.5	82.7
CoBEVFlow[41]	73.8	73.2	70.3	-	-	-	-	-	-	-
Ours (RoCo)	76.3	74.8	73.3	71.9	71.5	91.9	91.0	90.0	85.9	84.1

Method/Metric	AP@0.7 ↑									
Noise Level (σ_t/σ_r $(m/°)$)	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.8/0.8	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.8/0.8
F-Cooper[9]	55.9	55.2	54.2	53.8	51.6	48.6	46.0	43.4	41.0	39.5
FPV-RCNN[51]	50.5	45.9	42.0	41.0	38.9	56.3	51.2	37.4	31.8	27.0
V2VNet[40]	48.6	48.3	47.8	47.5	38.0	64.6	62.0	56.2	50.7	44.9
Self-Att[47]	52.2	52.0	51.7	51.4	51.1	67.6	66.2	65.1	63.9	63.0
V2X-ViT[45]	53.1	52.9	52.5	52.2	51.3	80.3	76.8	71.8	69.0	66.6
CoAlign[28]	60.4	58.8	57.9	57.0	56.9	80.5	77.3	73.0	70.1	67.3
CoBEVFlow[41]	59.9	57.9	56.0	-	-	-	-	-	-	-
Ours (RoCo)	62.0	59.4	58.4	58.2	57.8	80.5	77.4	77.3	71.0	68.9

Experiments show that RoCo **holds the best resistance** to pose errors

RoCo: Robust Collaborative Perception By Iterative Object Matching and Pose Adjustment

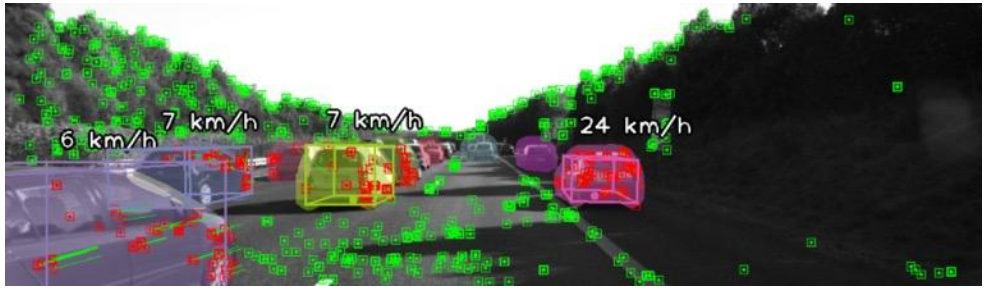
自动驾驶中的同步定位建图与目标跟踪

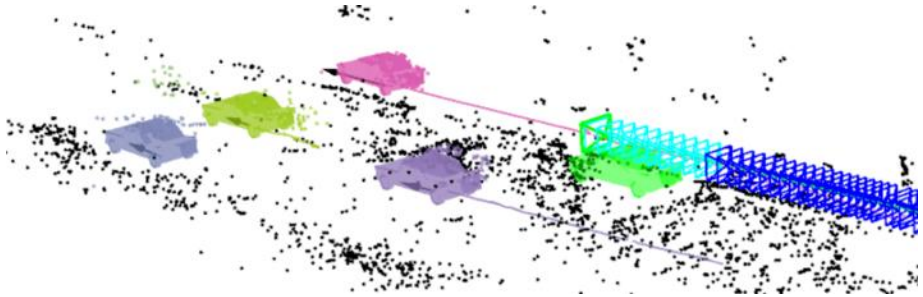
ACM MM2024
王硕、王永才等
中国人民大学信息学院

实现无人车的同步定位建图与多目标追踪系统

Background

- SLAM同步定位和建图，准确性受环境中移动目标影响很大





我们的目标：

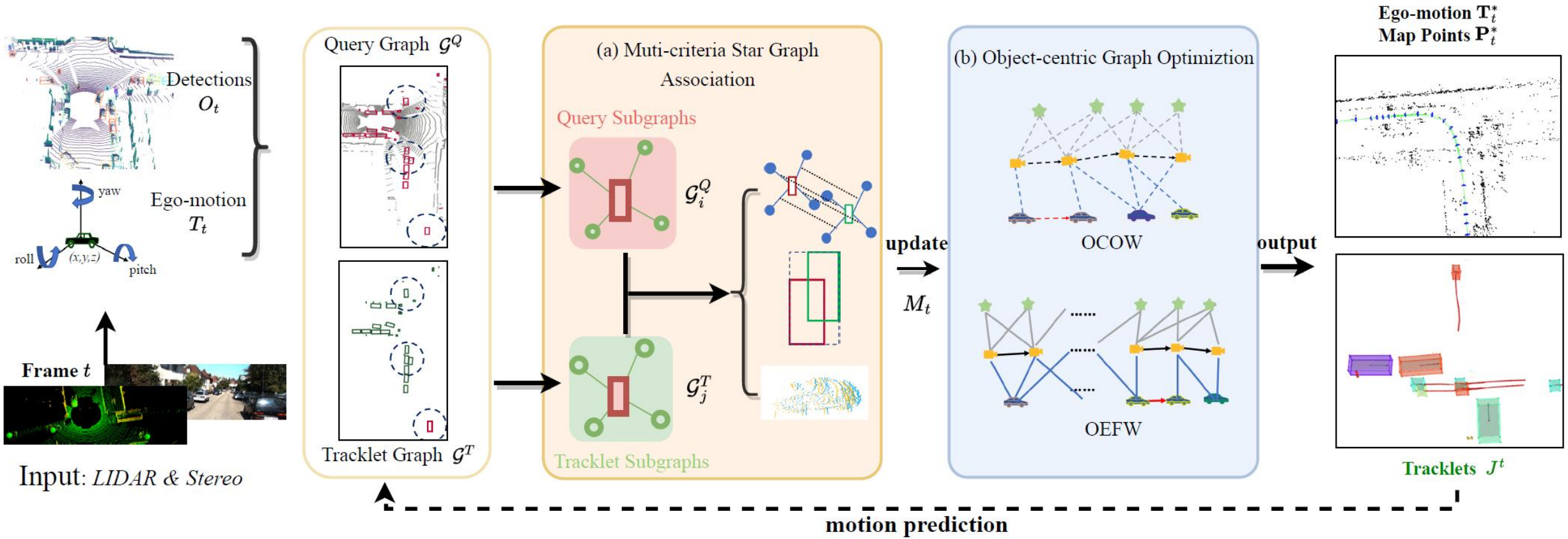
- **系统输入**
  - 无人车连续采集雷达与双目视觉相机数据
- **系统输出**
  - 无人车自身的位姿轨迹
  - 建立静态语义环境地图
  - 检测并追踪周边的移动车辆

Locating + Mapping + Tracking

GSLAMOT: A Tracklet and Query Graph-based Simultaneous Locating, Mapping, and Multiple Object Tracking System

ACM MM2024，人工智能领域顶会
王硕、王永才等
中国人民大学信息学院

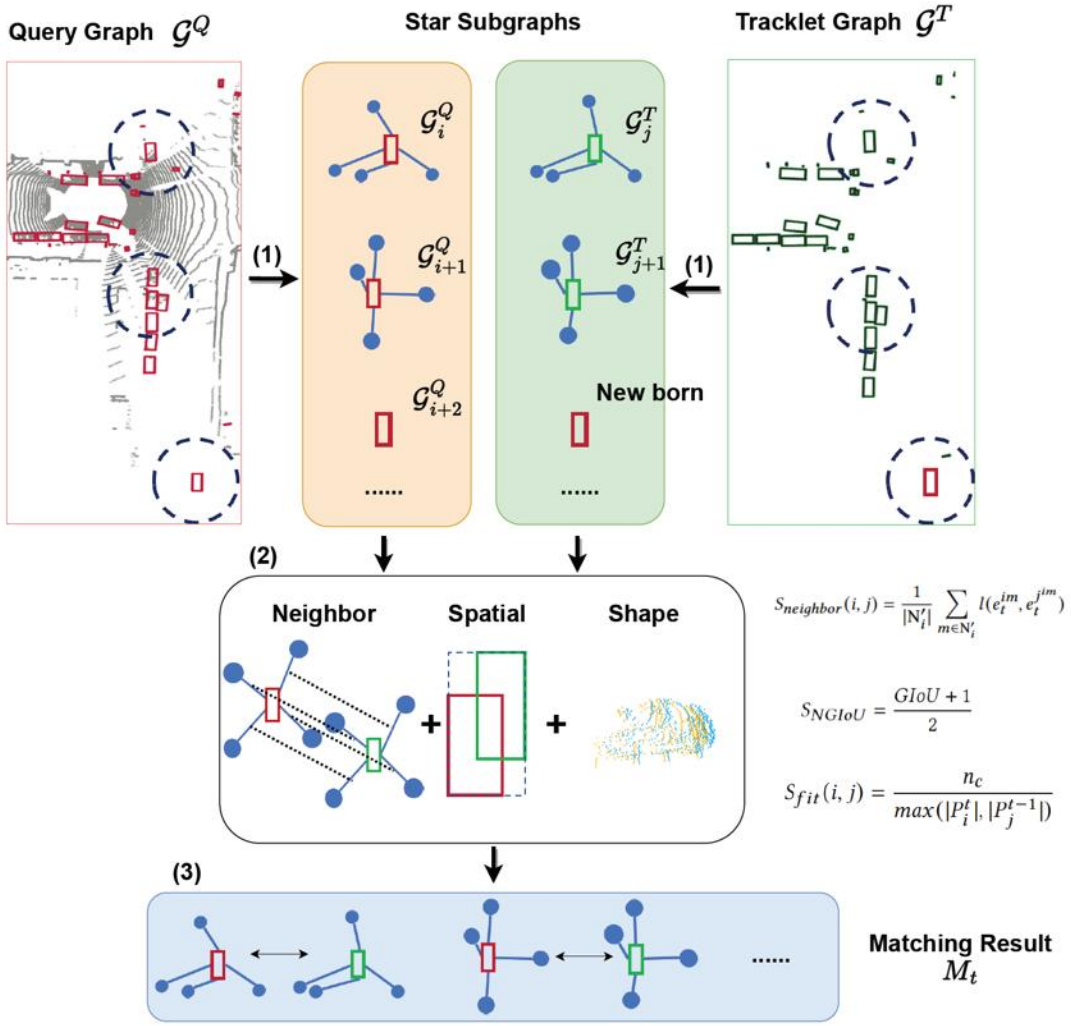
系统架构：同步定位建图与多目标追踪系统



输入　　目标检测　　当前帧查询历史目标轨迹图　　自身轨迹、建图、多目标跟踪结果

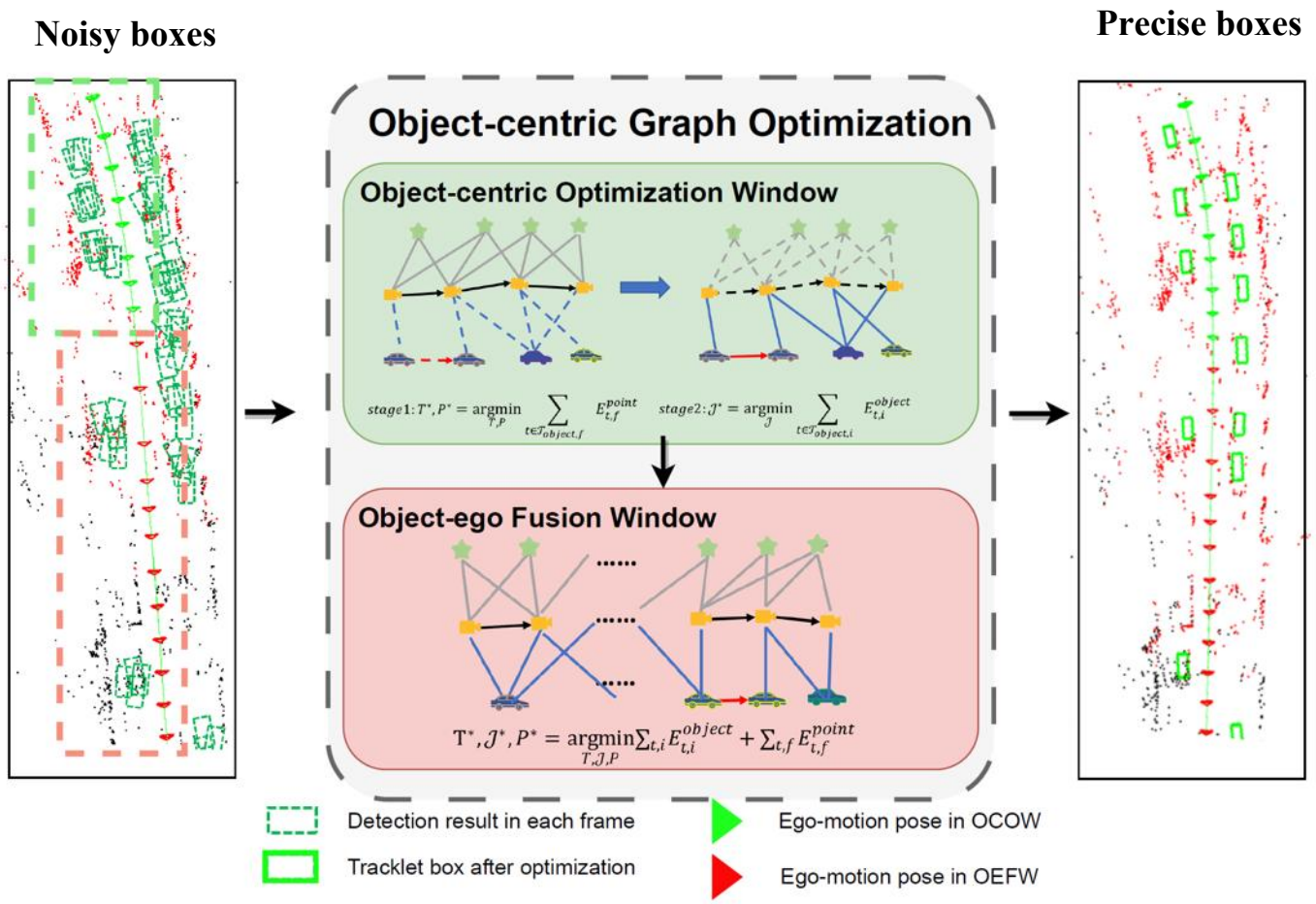
多种相似性的查询图(Query Graph)与历史轨迹图(Tracklet Graph)匹配

- 我们为当前帧的检测创建查询图(Query Graph)，并为地图中的轨迹创建轨迹图(Tracklet Graph)。
- 每个检测和轨迹都分别被分配一个星形子图。
- 我们通过评估它们的星形子图的**邻居、空间和形状一致性**来**匹配检测和轨迹。**



以目标为中心的图优化方法

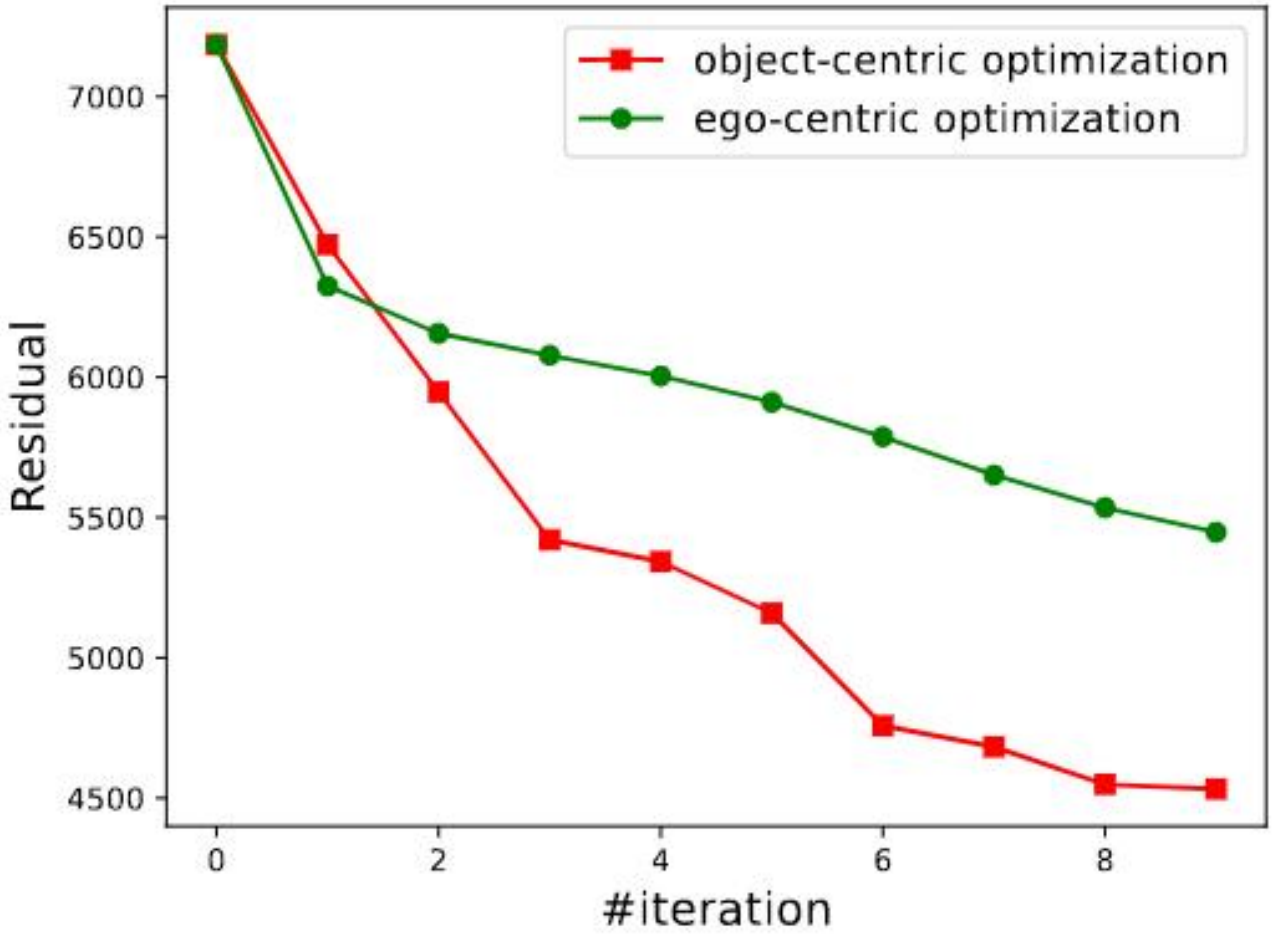
- 以目标为中心的图优化窗口（OCOW）
  - 阶段1：利用静态环境地标**估计车辆的自身运动**。
  - 阶段2：固定车辆自身运动，来**优化移动目标位姿**。
- 目标-自我融合窗口（OEFW）：
  - 一个紧密耦合的优化策略，**联合优化自我运动姿势、地图点和轨迹姿势**。



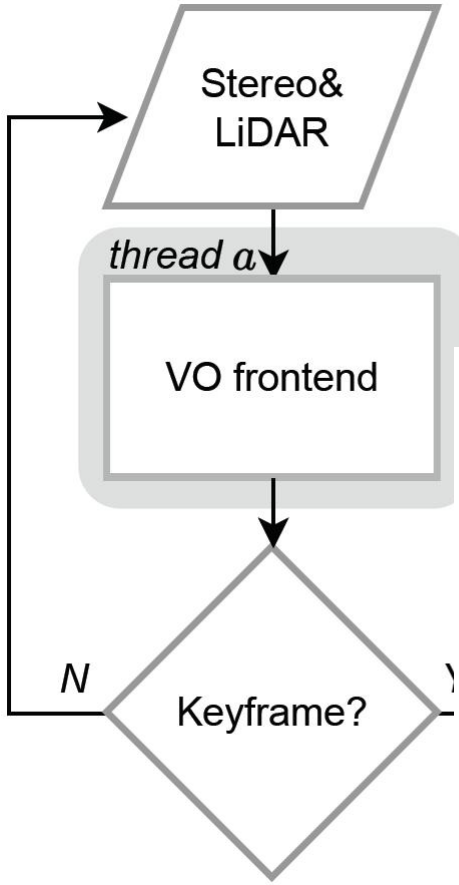
以目标为中心的图优化方法

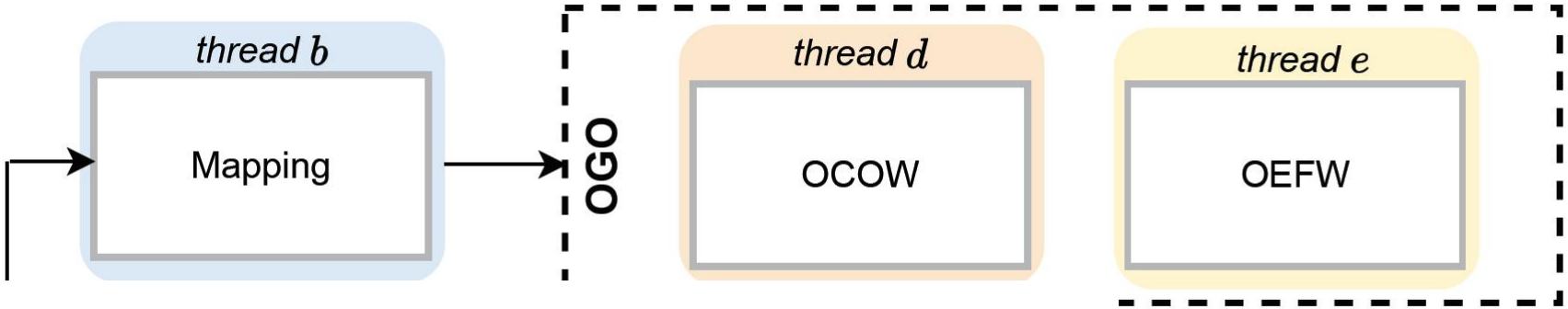
我们提的以目标为中心的优化和经典的优化方法的收敛残差对比。

误差更小、收敛更快。



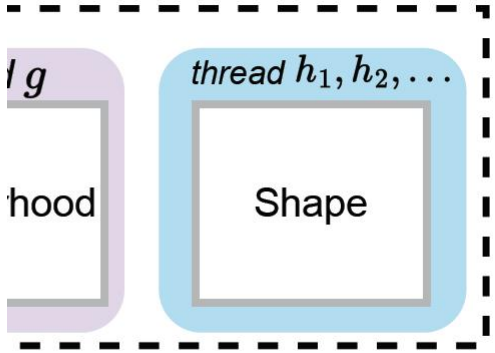
多线程的系统实现



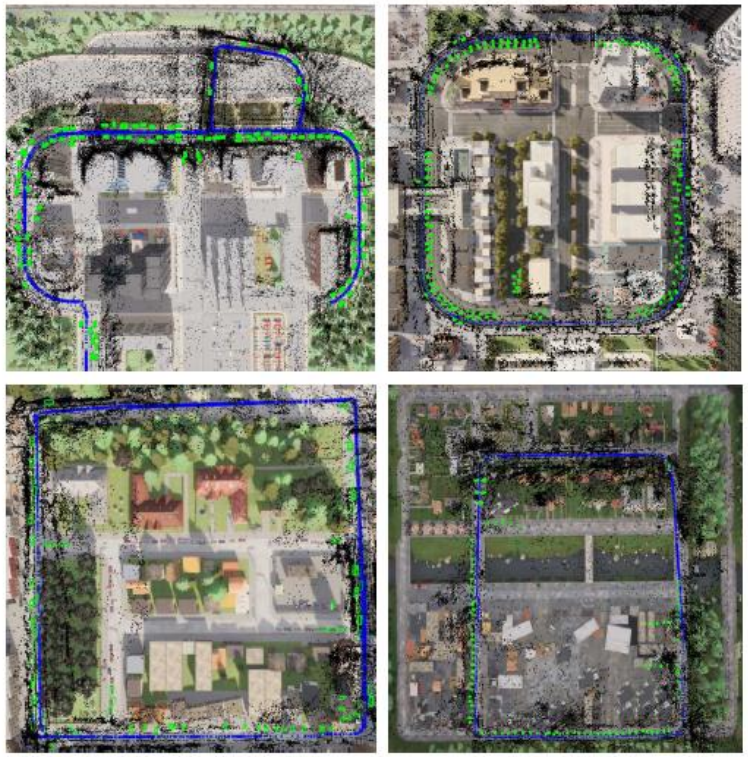


Module	*Time Per Frame(s)*↓
Ego-motion Tracking	0.031
Mapping	0.027
3D Detection†	0.022
MSSA†	0.024
OGO†	0.037
Total	0.082

†: only for keyframes.



实验数据集



我们自己构建的高密度交通流数据集

实验结果：定位与多目标追踪效果

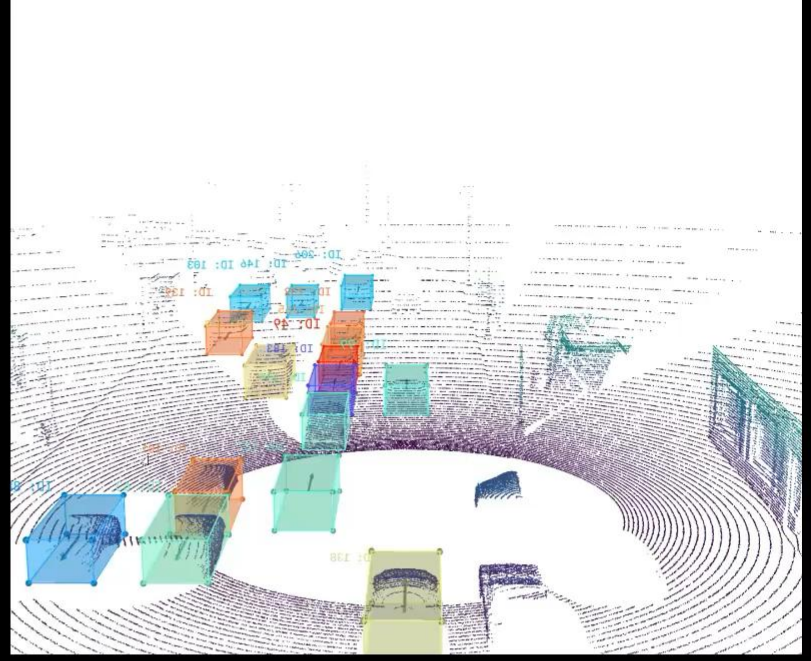
在KITTI数据集上结果显著好于所有现有方法

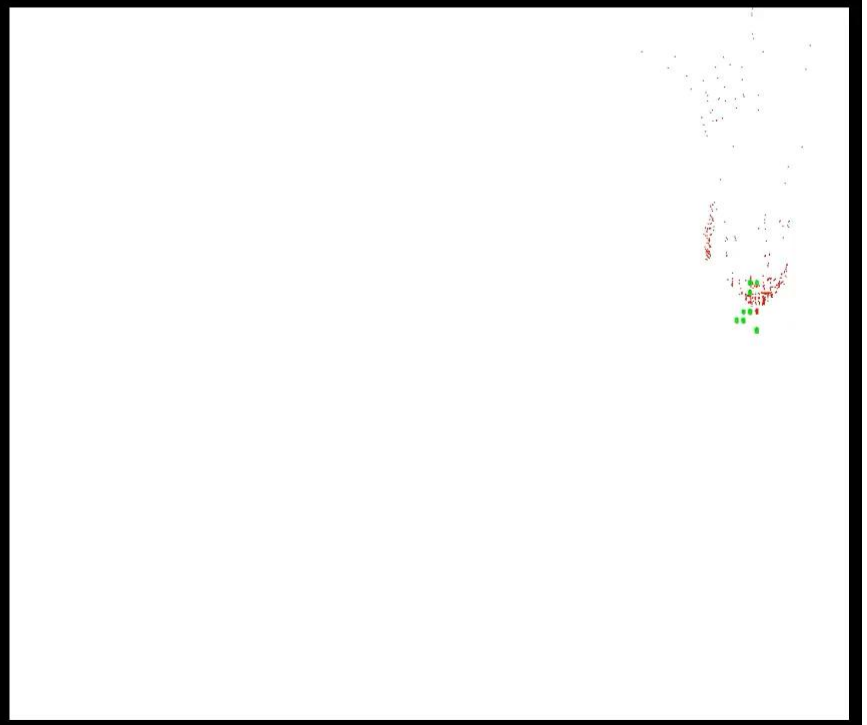
KITTI Seq.	00		01		02		03		04		05		06		07		08		Average	
Metrics(m)	RPE	APE	RPE	APE	RPE	APE	RPE	APE	RPE	APE	RPE	APE	RPE	APE	RPE	APE	RPE	APE	RPE	APE
ORBSLAM3[5]	2.09	1.46	7.52	12.70	2.3	3.5	0.84	1.44	0.6	0.25	0.91	0.93	0.92	0.99	0.49	0.49	3.06	**3.06**	2.08	2.76
DSP-SLAM[33]	1.09	1.10	3.87	**12.06**	0.94	**0.89**	1.28	**0.47**	0.64	0.73	**0.53**	0.46	0.81	**0.42**	0.5	0.48	3.17	11.99	1.40	3.18
DynaSLAM[3]	1.05	1.28	3.75	21.13	1.1	0.91	**0.68**	1.43	0.73	0.82	0.64	1.52	0.8	1.35	0.51	0.78	3.06	10.41	1.36	4.40
VDO-SLAM[42]	1.02	1.44	3.80	13.79	0.98	0.99	0.79	0.83	0.61	0.25	0.59	0.49	0.75	0.63	0.49	0.52	3.34	9.76	1.50	3.19
GSLAMOT(Ours)	**1.01**	**1.02**	**3.69**	13.1	**0.92**	0.91	**0.68**	0.57	**0.56**	**0.23**	**0.53**	**0.41**	**0.70**	0.44	**0.48**	**0.43**	**3.05**	3.16	**1.29**	**2.25**

在Waymo数据集上3D目标检测效果优于现有方法.

Method	MOTA(L1)↑	MOTA(L2)↑	Mismatch↓	MOTA(L2)↑		
				vehicle	pedestrian	cyclist
AB3DMOT[34]	-	-	-	40.1	33.7	50.39
ProbTrack[7]	48.26	45.25	1.05	54.06	48.10	22.98
CenterPoint[37]	58.35	55.81	0.74	59.38	56.64	60.0
SimpleTrack[25]	59.44	56.92	0.36	56.12	57.76	56.88
BOTT[46]	59.67	57.14	0.35	59.49	58.82	60.41
TrajectoryF[6]	-	-	-	59.7	**61.0**	**60.6**
GSLAMOT	59.69	57.10	**0.33**	60.45	60.02	60.33
GSLAMOT*	**59.75**	**57.20**	**0.33**	**60.47**	60.23	60.45

GSLAMOT: The ego-motion poses are estimated by odometry front-end.
GSLAMOT*: The groundtruth ego-motion poses are given as other MOT algorithms.





录像展示

Speed X 10

多目标追踪效果，
不同颜色代表追踪
的不同目标.

- Green box: tracklet
- Red point: local map point
- Black point: map point

总结

- 协同感知可以克服单个智能体感知能力的限制，实现更为全面、准确、可靠的感知。
- 在多手机、多车协同感知方面的成果已经比较丰富。
- 多无人机协同感知是在应用中非常必要的，也是更难的3D空间的感知问题。
- 但是现在多无人机协同感知的研究工作很少，有较大研究空间。

谢谢，Q&A

ycw@ruc.edu.cn

主页：http://www.yongcaiwang.cn

http://yongcaiwang.github.io