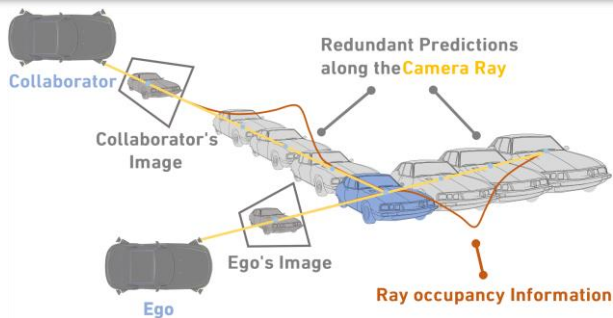# RayFusion: Ray Fusion Enhanced Collaborative Visual Perception

Shaohong Wang, Lu Bin, Xinyu Xiao, Hanzhi Zhong, Bowen Pang, Tong Wang, Zhiyu Xiang, Hangguan Shan, Eryun Liu

Motivation

- ➤ Due to depth ambiguity, each agent may predict multiple targets along a camera ray.
- ➤ Cross-view observations enable agents to record ray occupancy and validate the true 3D position.

Contributions

- ➤ A camera-only collaborative visual perception framework leveraging ray occupancy fusion to enhance localization accuracy and distinguish hard negatives.
- ➤ A unified coordinate system with motion modeling for robust spatial-temporal alignment under communication delays.
- ➤ An explicit ray-based 3D encoding that captures scene structure and reduces redundant or false-positive predictions.
- ➤ A multi-agent feature aggregation module that enhances collaboration, achieving state-of-the-art perception results on DAIR-V2X, V2XSet, and OPV2V datasets..
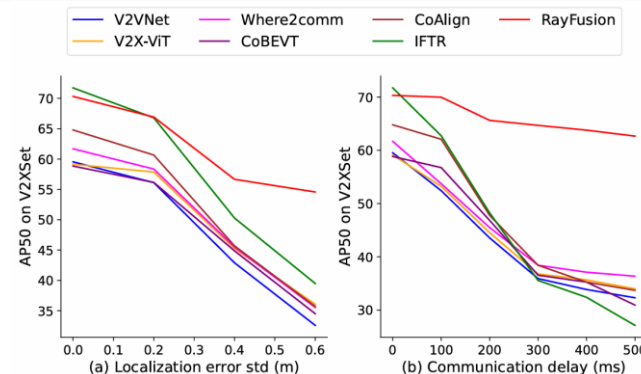
RayFusion
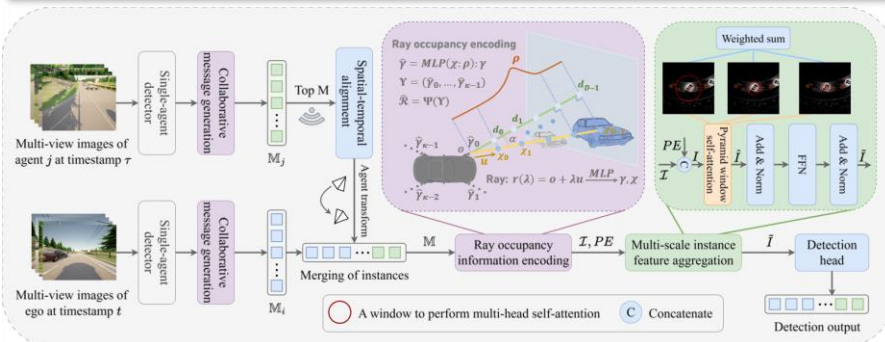
- ➤ The spatial-temporal alignment module enhances system robustness to latency by modeling motion.
- ➤ The ray occupancy information encoding module leverages multi-view information to mitigate depth estimation ambiguity
- ➤ The multi-scale instance feature aggregation facilitates effective interaction among instance features, promoting comprehensive and precise collaborative perception.
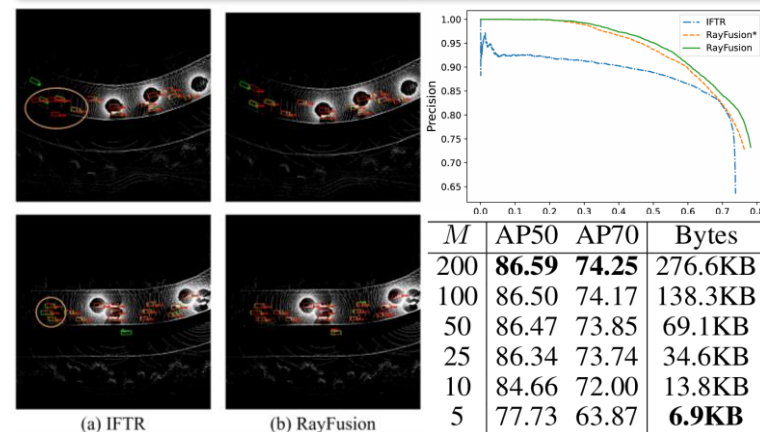
Benchmark comparison

Method	DAIR-V2X		V2XSet		OPV2V	
	AP50	AP70	AP50	AP70	AP50	AP70
No Collaboration	10.74	1.64	30.37	13.79	45.94	25.56
Late Fusion	18.57	5.16	51.41	25.59	77.62	51.92
V2VNet (ECCV'20)	15.26	2.97	59.54	39.00	79.06	57.59
V2X-ViT (ECCV'22)	15.84	3.07	59.14	41.23	78.41	58.38
Where2comm (NeurIPS'22)	16.03	3.67	61.69	43.96	77.14	58.60
CoBEVT (CoRL'22)	15.92	3.18	58.84	40.81	80.26	59.34
CoAlign (ICRA'23)	16.55	3.23	64.79	39.64	80.21	60.46
IFTR (ECCV'24)	20.51	7.90	**71.73**	49.67	85.56	66.04
CoSparse4D †(Baseline)	23.38	9.17	65.69	49.33	79.38	67.43
RayFusion †(Ours)	**26.29**	**11.54**	70.32	**53.14**	**86.59**	**74.25**

Robustness to noise

(a) Localization error std (m)

(b) Communication delay (ms)

Analysis

M	AP50	AP70	Bytes
200	**86.59**	**74.25**	276.6KB
100	86.50	74.17	138.3KB
50	86.47	73.85	69.1KB
25	86.34	73.74	34.6KB
10	84.66	72.00	13.8KB
5	77.73	63.87	**6.9KB**

(a) IFTR   (b) RayFusion

Contacts

Email: wangsh0111@zju.edu.cn
Add: No. 38, Zheda Road, Lingyin Subdistrict, Xihu District, Hangzhou, Zhejiang Province, China, 310007