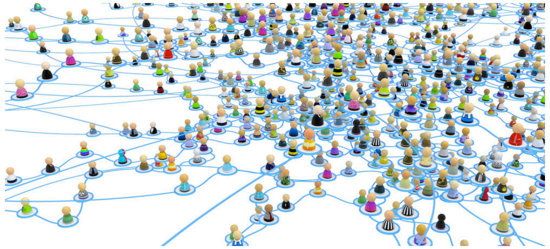
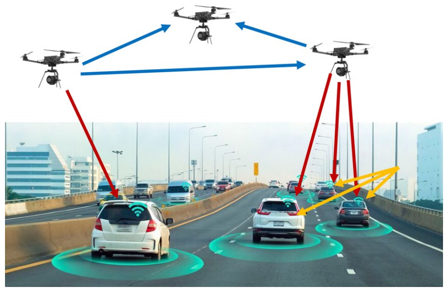
Towards Embodied Collective Intelligence in Real Physical World.

Individual intelligence



Social strategy



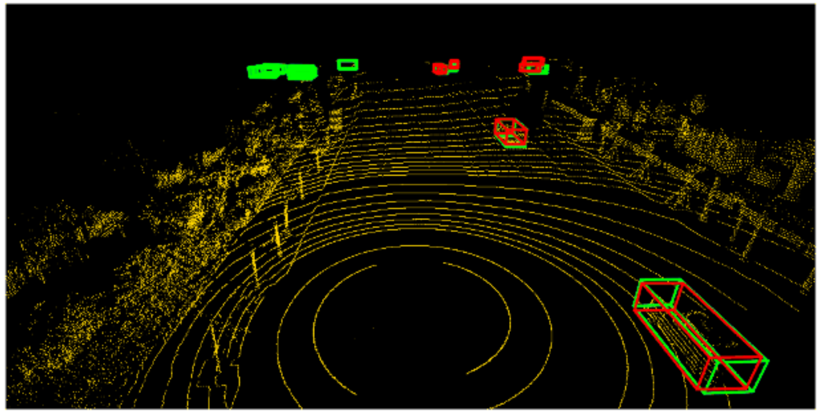
Group Intelligence

CoPerception in 2022-2023

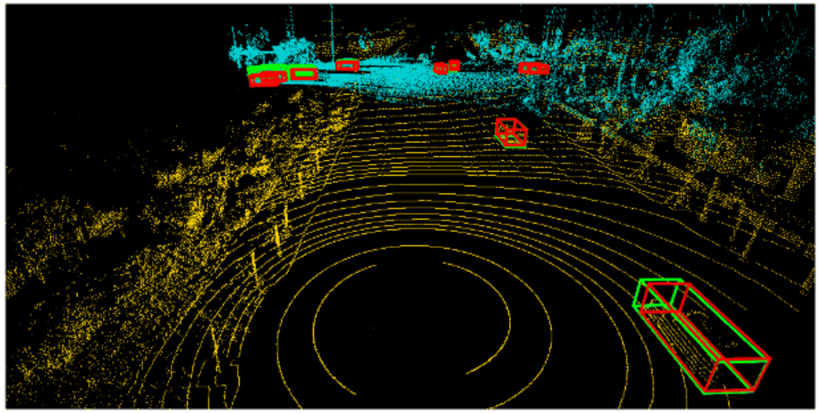
Introduction

- Autonomous system including robots, autonomous driving, humanoid, drones…… have been promoted rapidly thanks to the development of deep learning, sensors and semi-conductor technologies.
- In some areas, multi-robots system can collaborate to achieve a better performance. For example, for perception tasks, a cluster of robots can get a more comprehensive observation since they have different angle of view.

- Our research mainly focus on how to design the robust collaborative perception system.



Single-agent perception



Collaborative perception

Robust CoPerception

Communication Efficiency:
DiscoNet(NeurIPS2021), Where2comm (NeurIPS2022)

Communication Latency and Interruption
SyncNet (ECCV2022), CoBEVFlow (NeurIPS2023, under review)

Spatial-Temporal Alignment
CoAlign (ICRA2023),, FreeAlign (ICRA2024, under review),

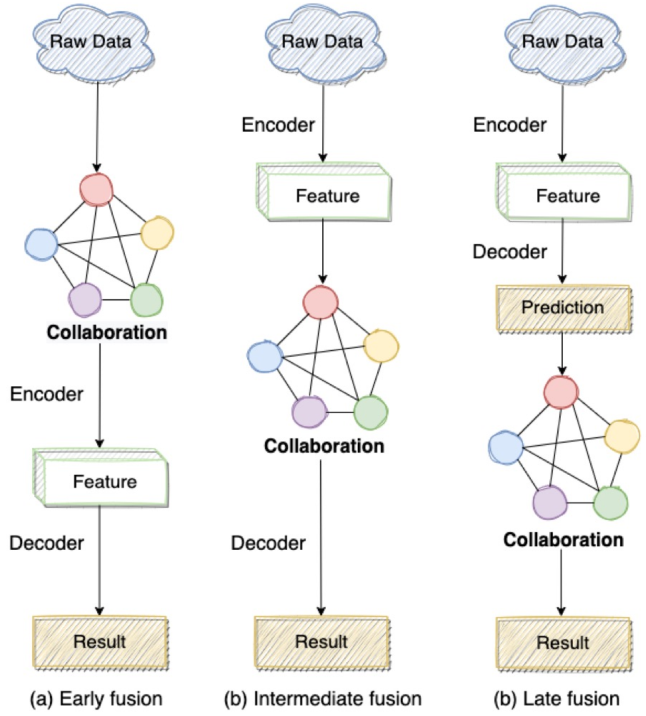
Perception heterogeneous
CoHeterogeneous

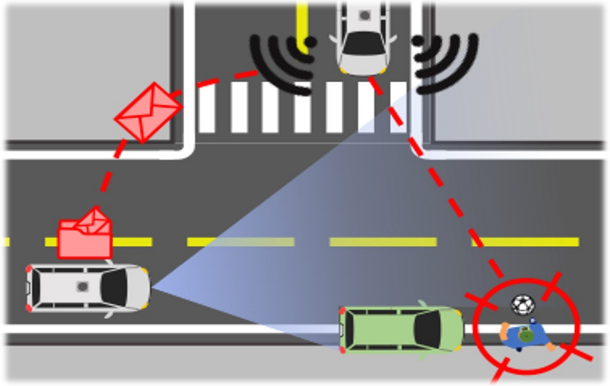
Security
Among Us

Introduction

For collaborative perception in V2V, V2X, or UAV scenarios, there are three critical issues:

1. **Communication efficient:** Bandwidth consumption
2. **Communication robust:** Latency and Interruption
3. **Spatial-temporal alignment:** Localization error, time-domain asynchronization

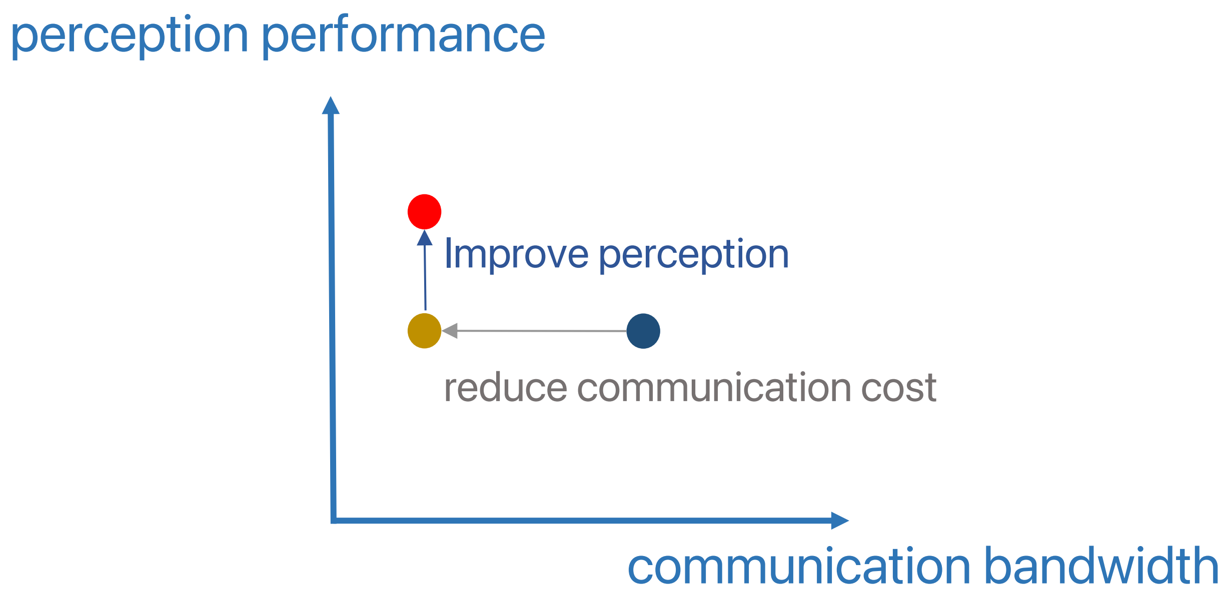




Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

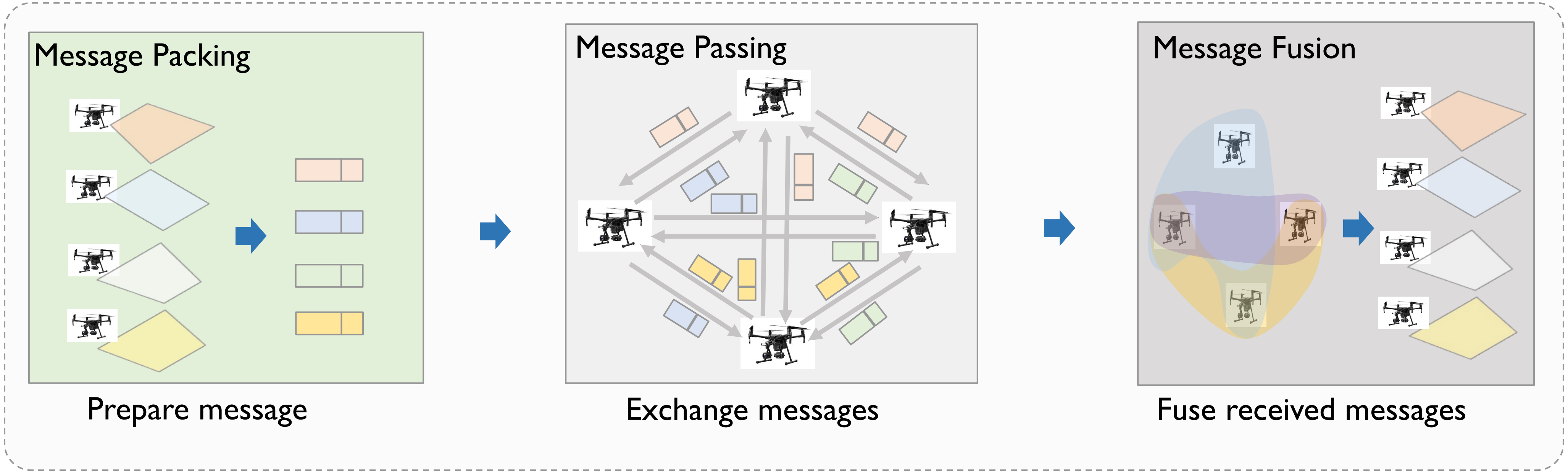
Motivation: Trade-off between perception performance and communication bandwidth.



Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

Motivation: Trade-off between perception performance and communication bandwidth.



What to collaborate?

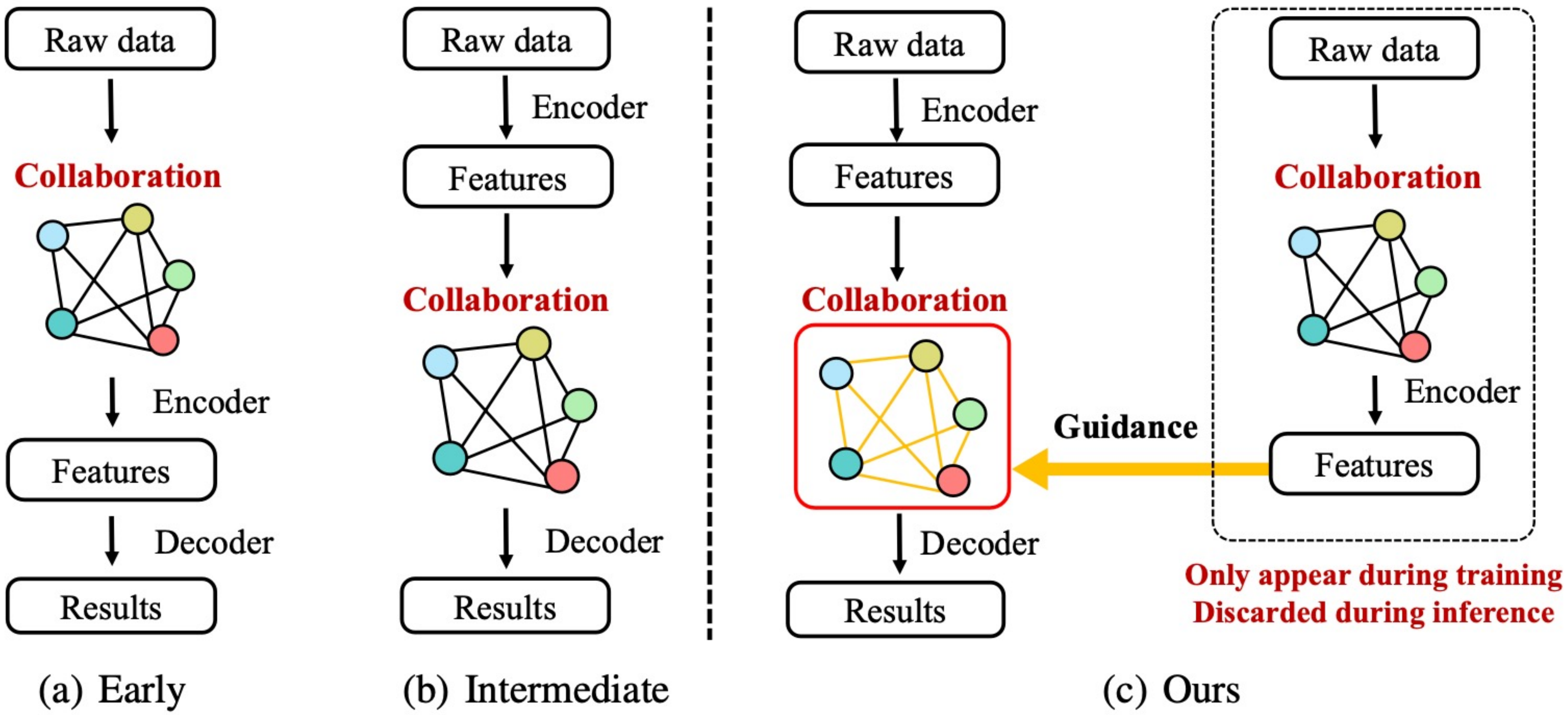
Who to collaborate?

How to fuse?

Communication efficient

Learning Distilled Collaboration Graph for Multi-Agent Perception

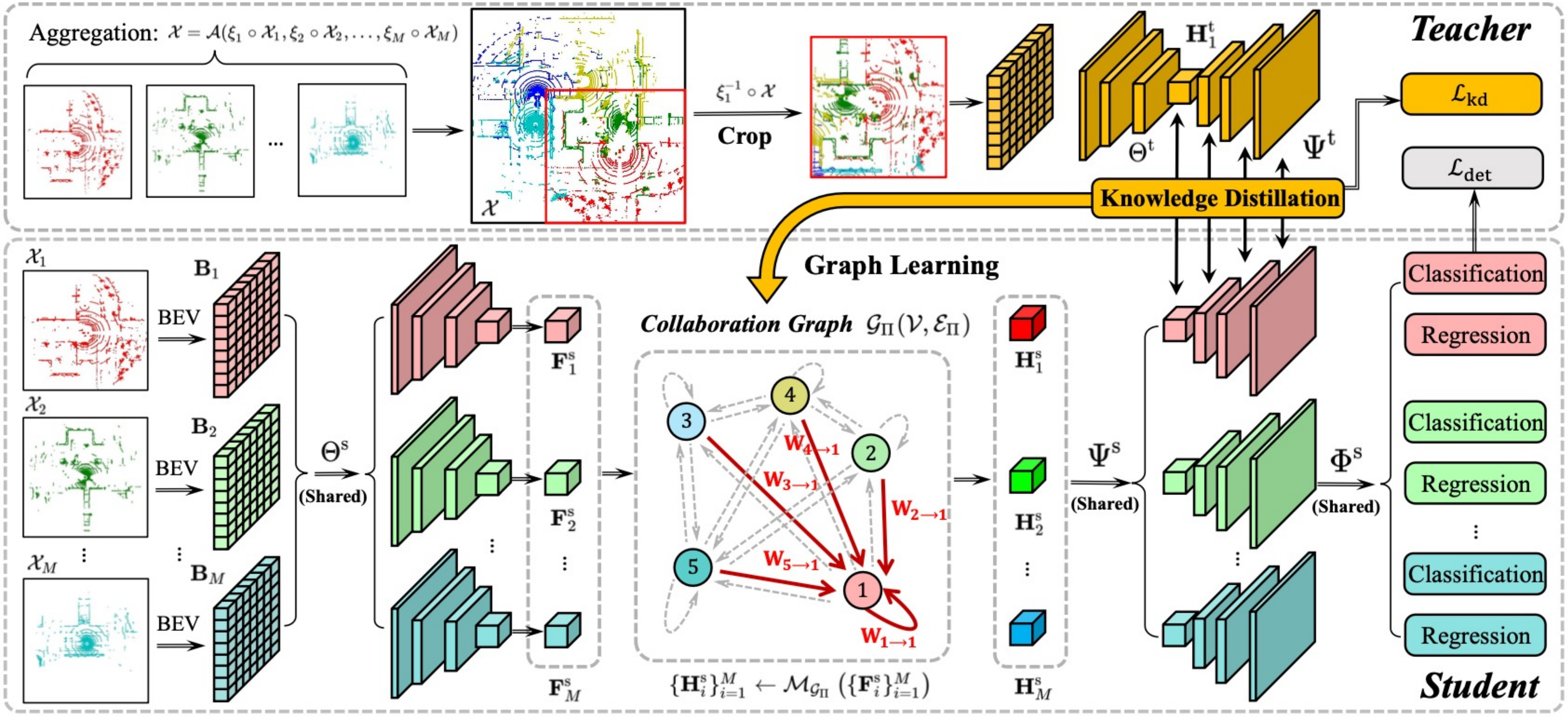
Motivation: Intermediate fusion/teacher-student framework



Communication efficient

Learning Distilled Collaboration Graph for Multi-Agent Perception

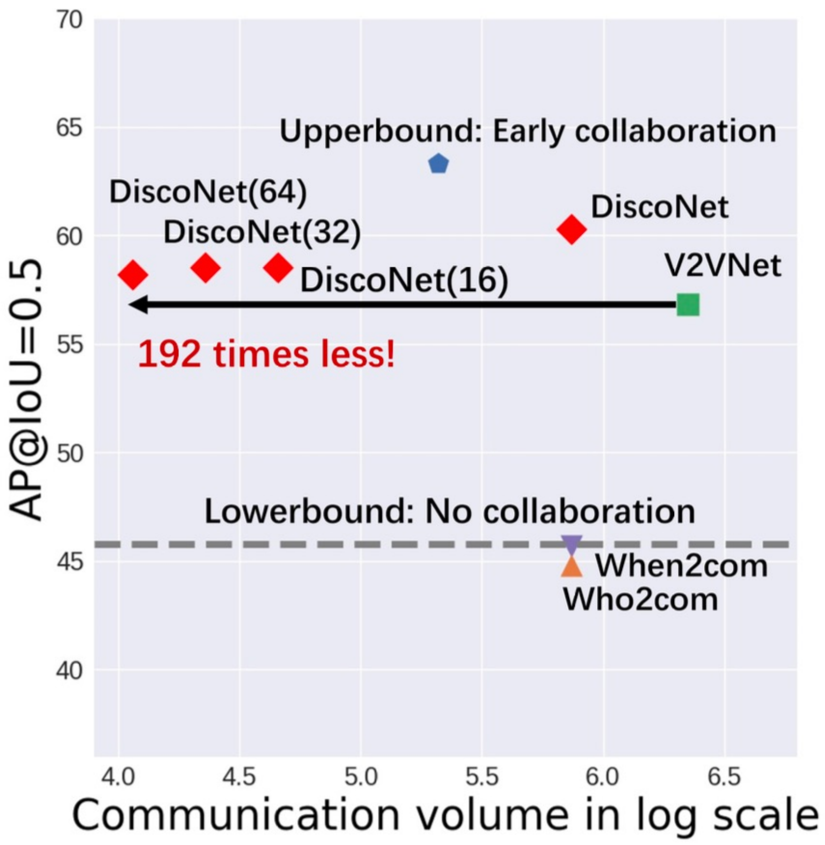
System Overview



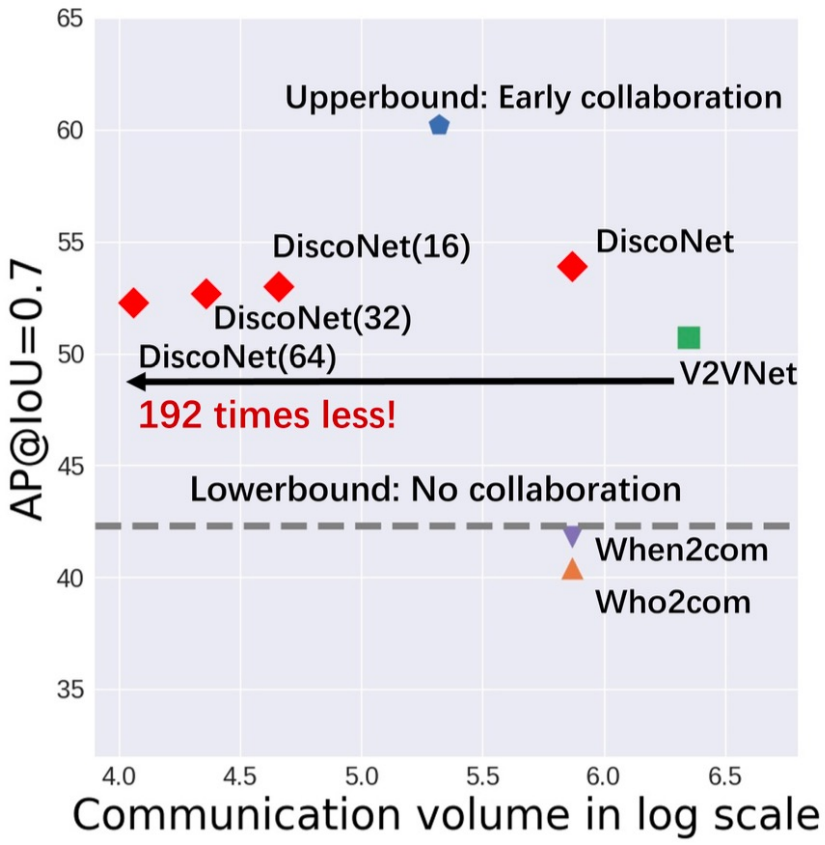
Communication efficient

Learning Distilled Collaboration Graph for Multi-Agent Perception

Experiments: Communication Efficiency



(a) Scatterplot in AP@IoU 0.5



(b) Scatterplot in AP@IoU 0.7

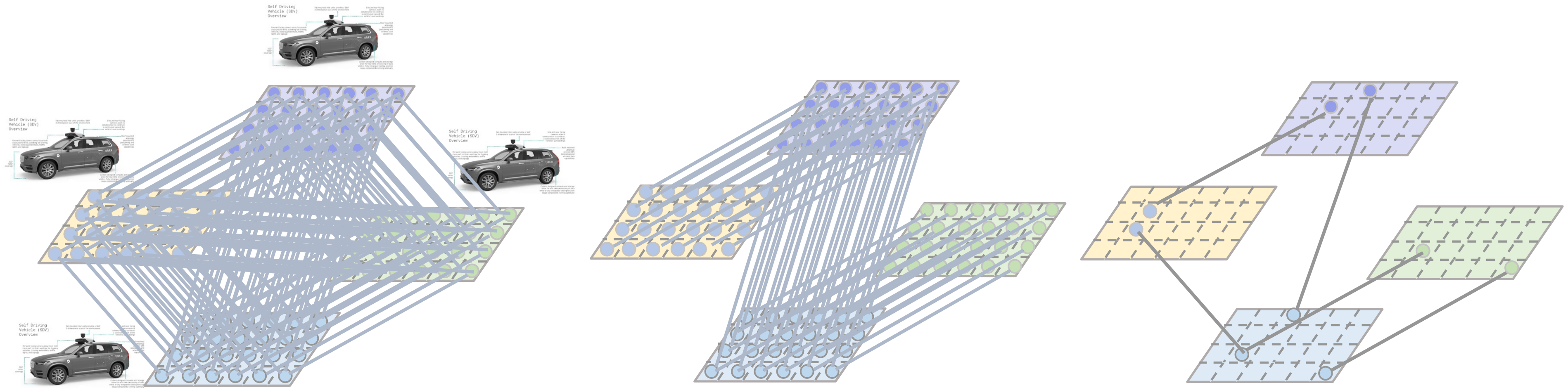
Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

Motivation: Trade-off between perception performance and communication bandwidth.

Messages should be spatially sparse, yet perceptually critical.

Fully connected | Agent-level partially connected | Spatial-decouple partially connected



V2VNet ECCV 2020
DiscoNet NeurIPS 2021
V2X-ViT ECCV 2022
CoBEVT CoRL 2022

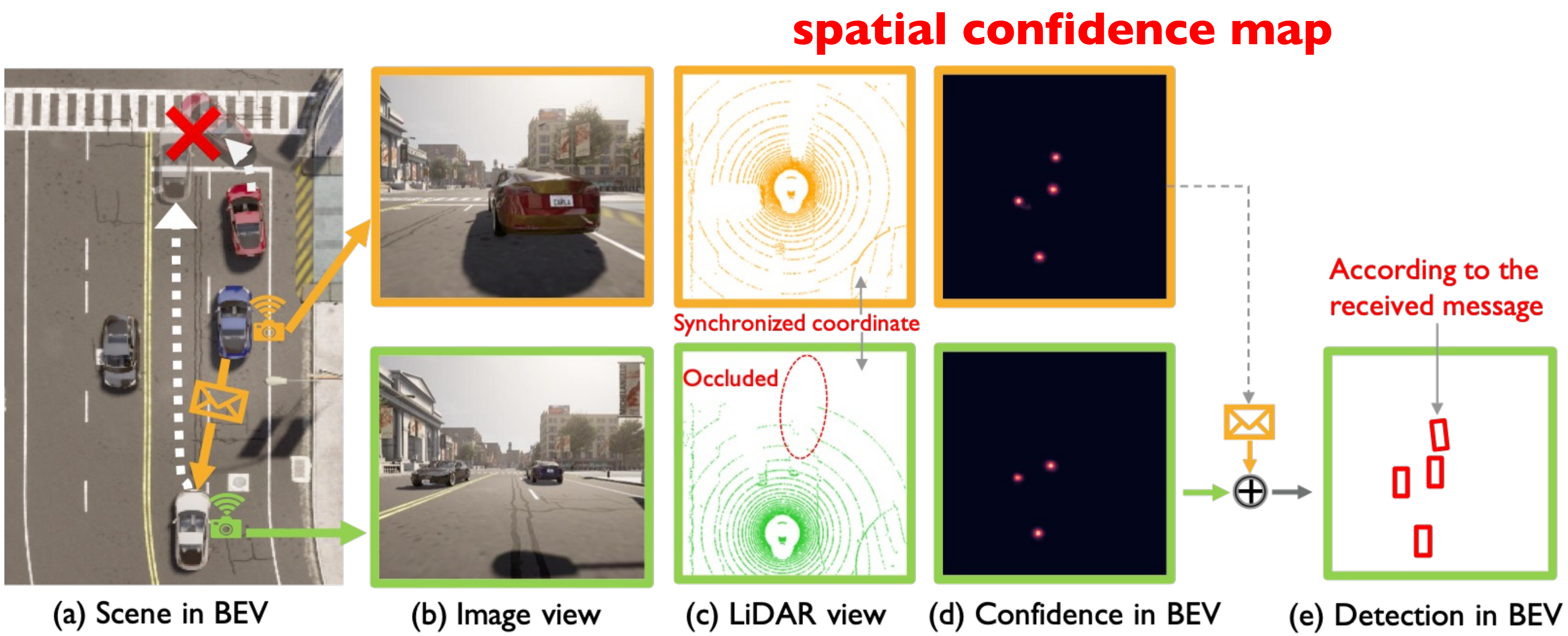
Who2com ICRA 2020
When2com CVPR 2020

Where2comm
NeurIPS 2022

Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

Core idea: Exploring spatial heterogeneity of perceptual information.

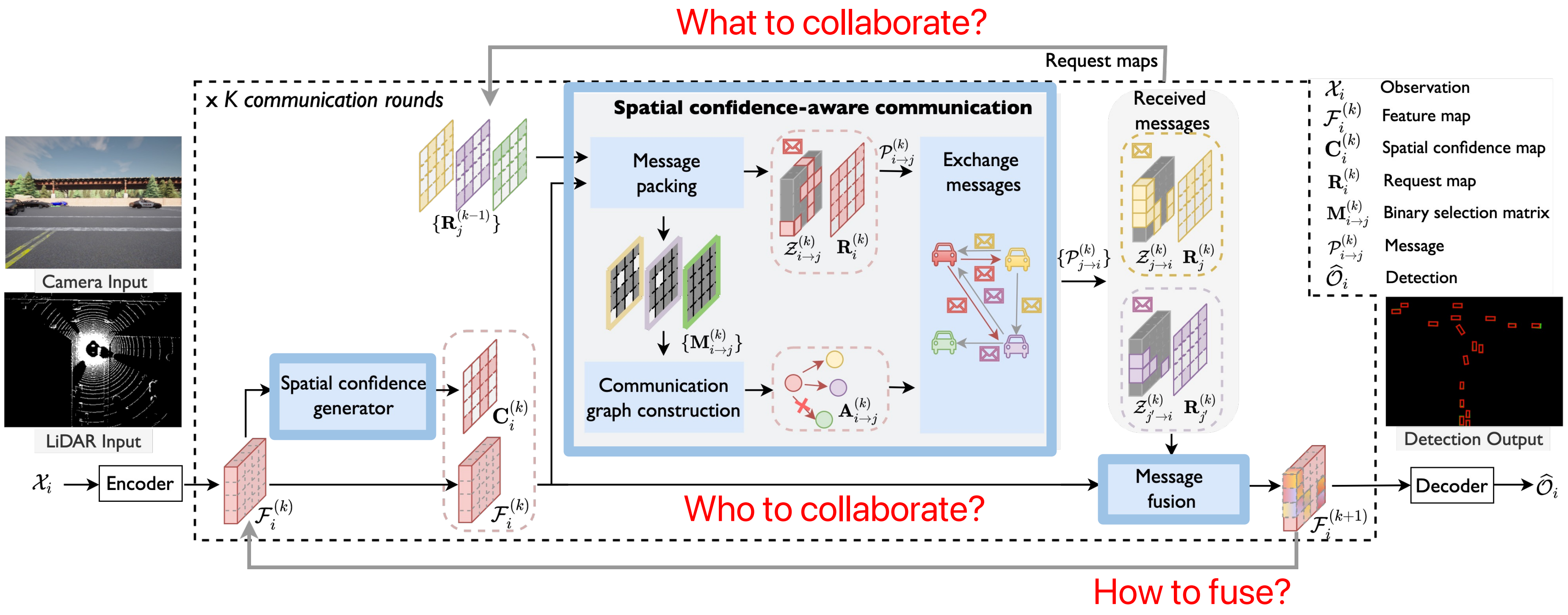


- Collaborative perception could contribute to safety-critical scenarios.
- The collision could be avoided when the blue car can share a message about the red car.

Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

Architecture: what, who and how.



- Generator produces a **spatial confidence map (SCM)** to indicate perceptually critical areas.
- Communication module leverages the SCM to decide **where and who to communicate**.
- Fusion module leverages the SCM as a prior to fuse all the messages via **multi-head attention**.

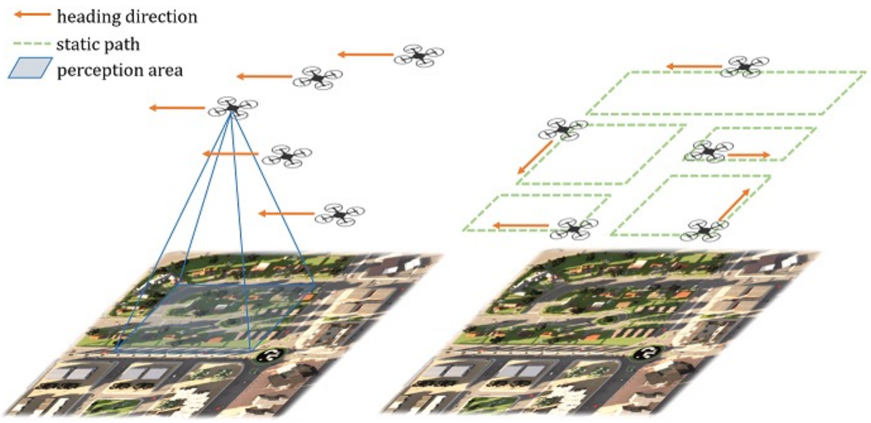
Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

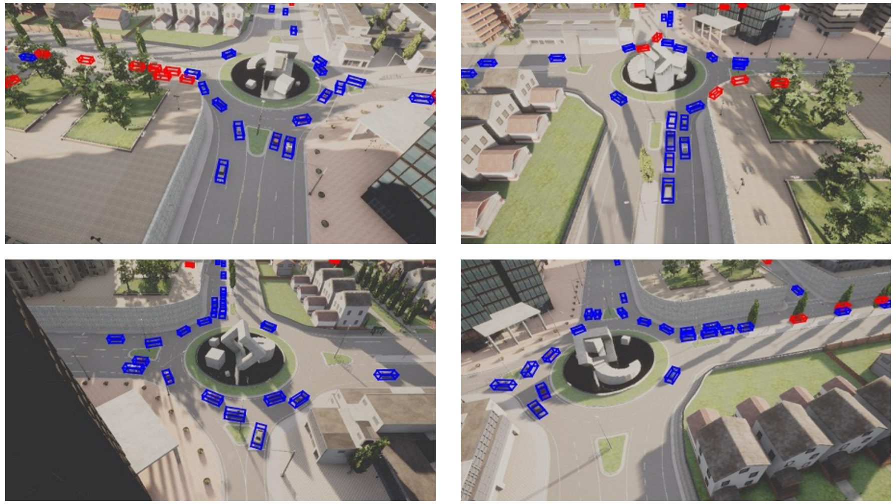
Datasets

Dataset	CoPerception-UAVs (NEW)	OPV2V[1]	V2X-Sim[2]	DAIR-V2X[3]
Modality	Camera-only	Camera-only	LiDAR	LiDAR
View	Aerial	Front (car)	Front (car)	Front (car)
Data	Simulation	Simulation	Simulation	Real

CoPerception-UAVs Dataset



(a) UAV swarm



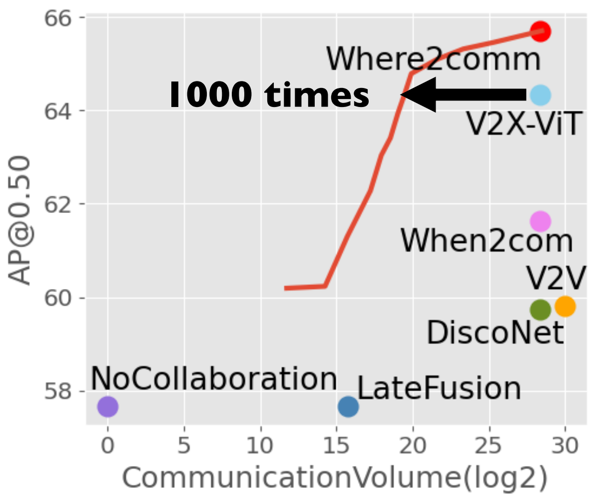
(b) Different views from 4 UAVs

[1] Xu, Runsheng et al. "OPV2V: An Open Benchmark Dataset and Fusion Pipeline for Perception with Vehicle-to-Vehicle Communication." (ICRA) (2022): 2583-2589.
[2] Li, Yiming et al. "V2X-Sim: Multi-Agent Collaborative Perception Dataset and Benchmark for Autonomous Driving." RAL (2022): 10914-10921.
[3] Yu, Haibao et al. "DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection." CVPR (2022)

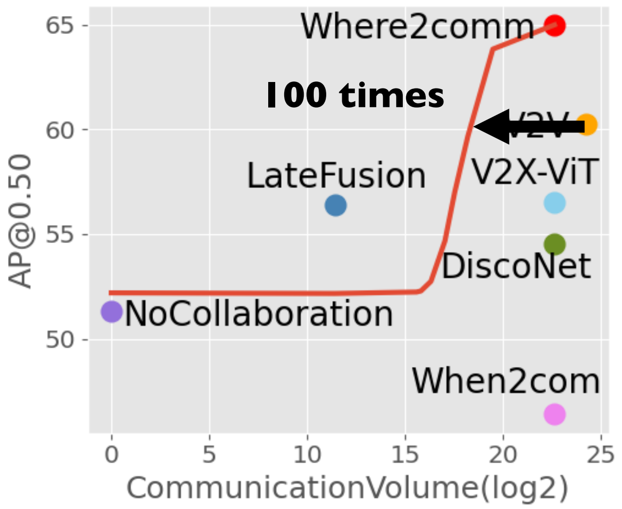
Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

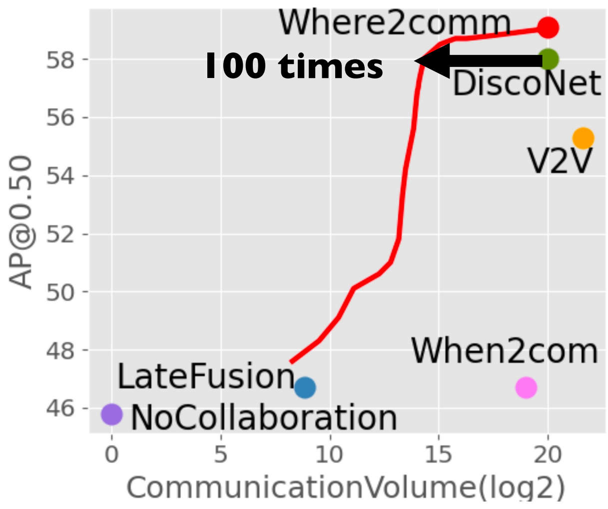
Architecture: what, who and how.



CoPerception-UAV



DAIR-V2X



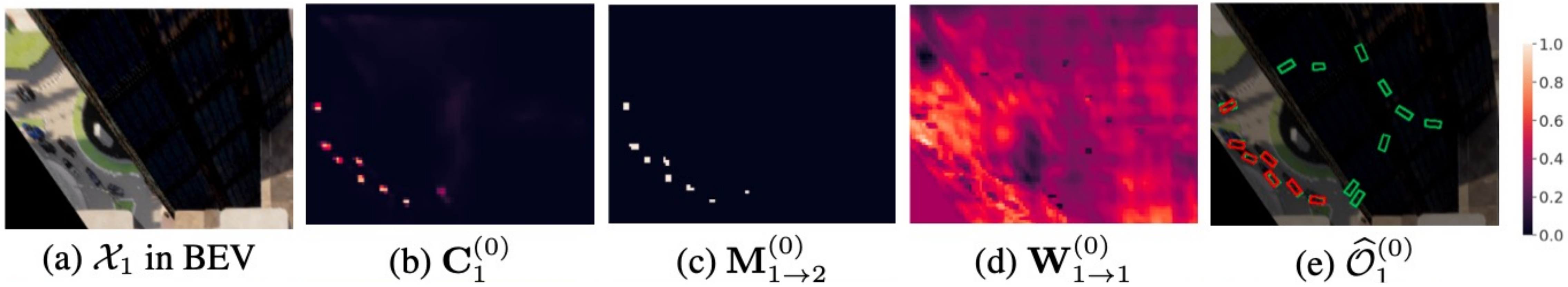
V2X-SIM

- Where2comm outperforms previous SOTA by 11.92%
- Where2comm adapts to various bandwidths while previous models only handle one predefined bandwidth

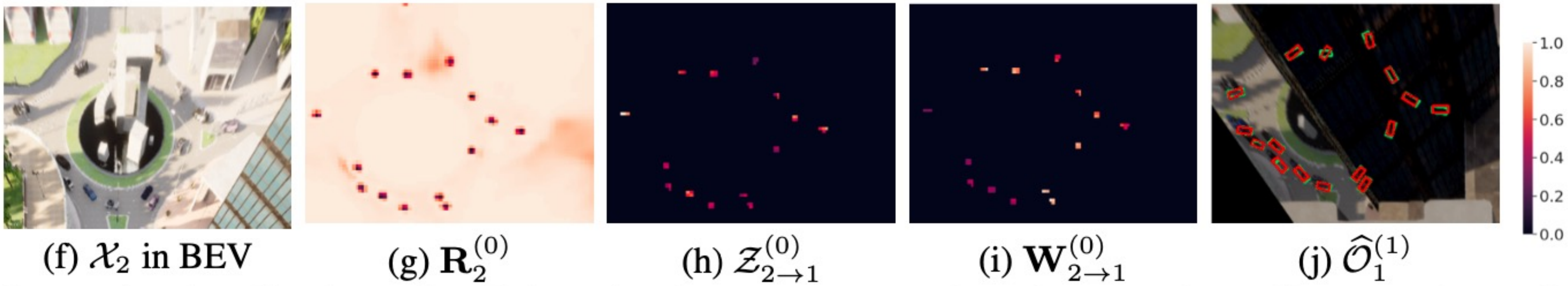
Communication efficient

Where2comm: Efficient Collaborative Perception via Spatial Confidence Maps

Qualitative evaluation



Warped image of drone 1 | Spatial confidence map | Selection matrix | Attention weight | Detections without collaboration

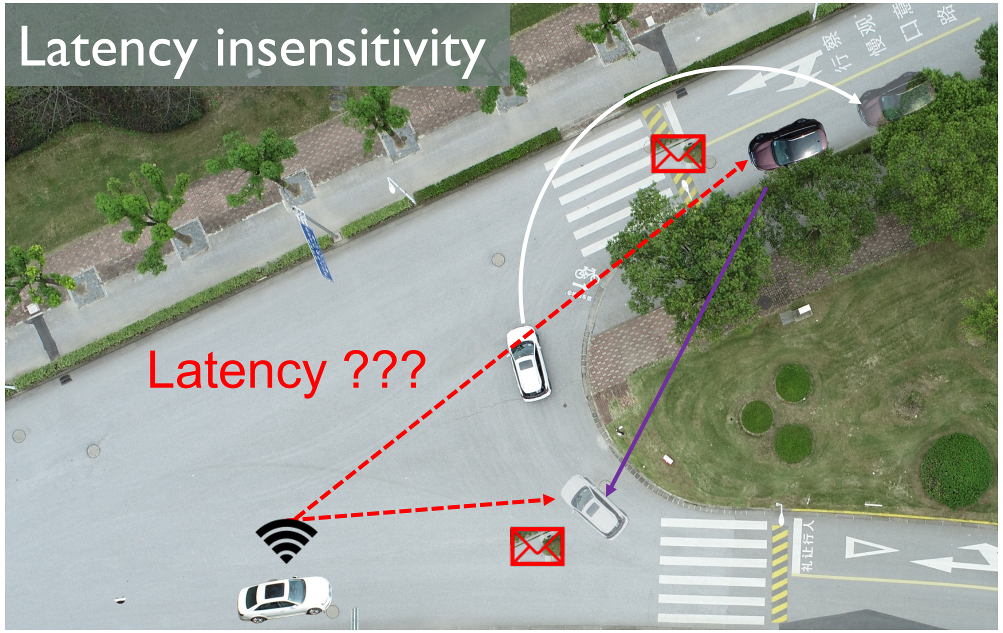


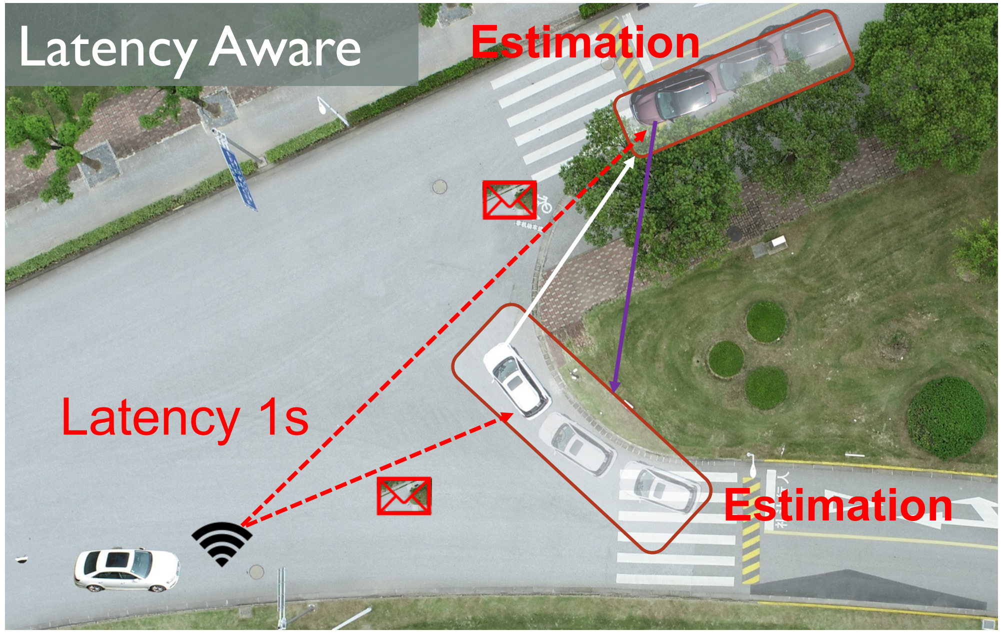
Warped image of drone 2 | Request map | Sparse feature map | Attention weight | Detections with collaboration

Communication Robust

Latency-aware collaborative perception

Motivation: A collision caused by latency.



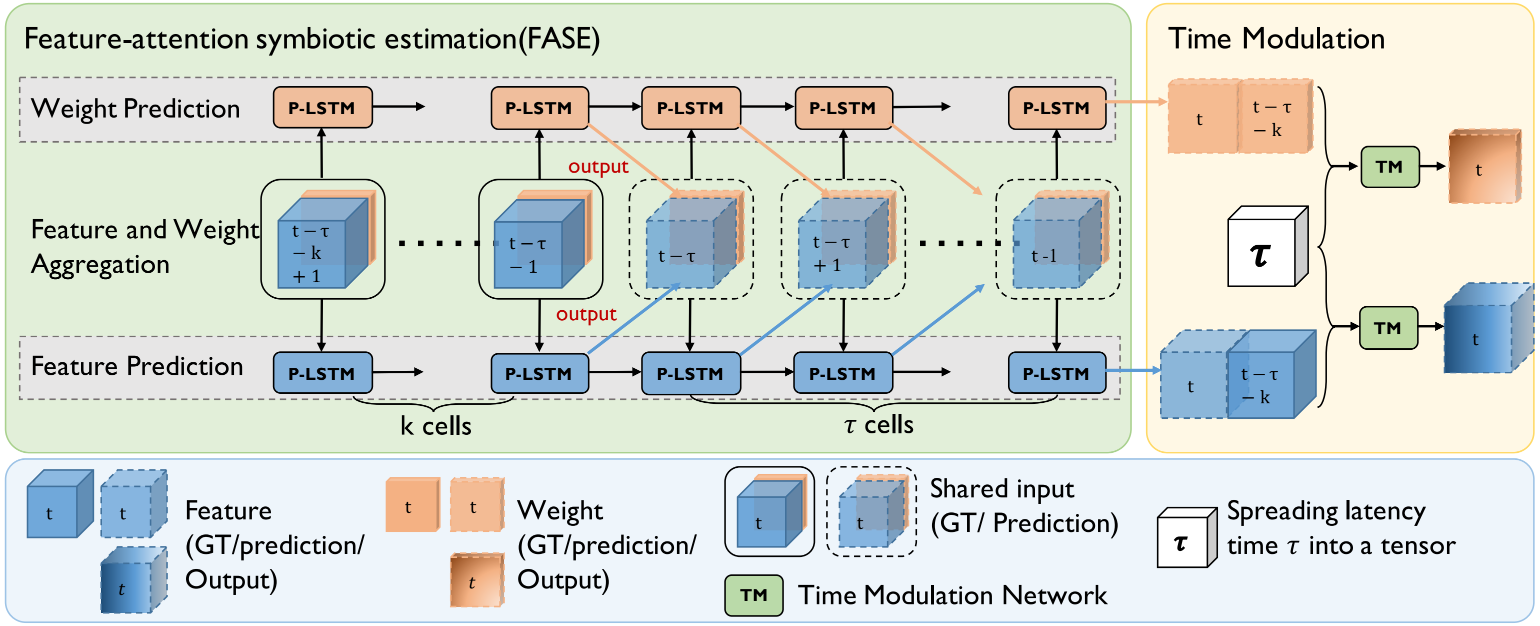


- Latency is ubiquitous in all kinds of communication systems.
- Latency compensation is essential for collaborative systems.

Communication Robust

Latency-aware collaborative perception

Compensation module: Knowledge distillation, multi-layer feature supervision.

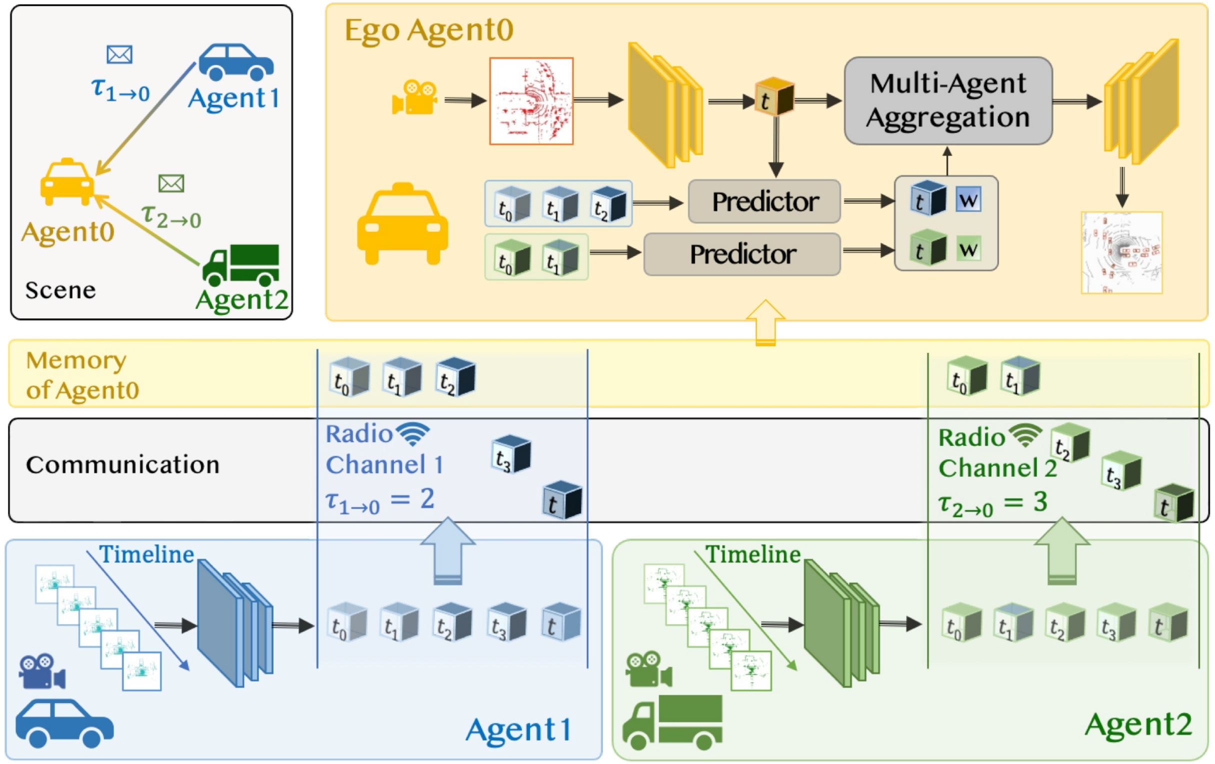


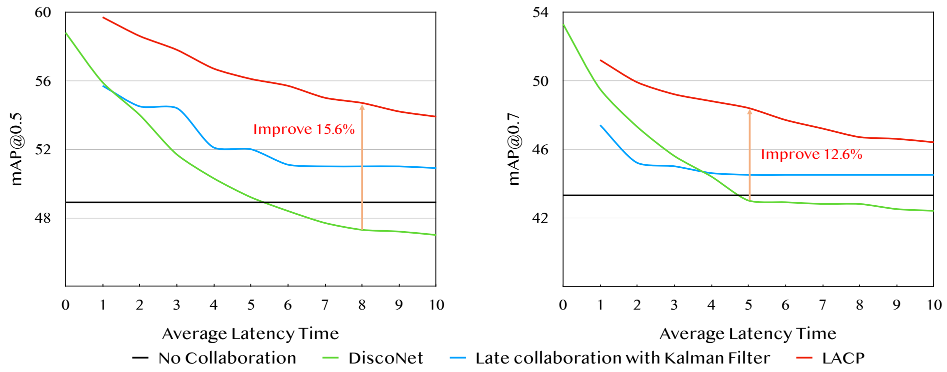
$$\mathcal{L} = \lambda_o \ell_{\text{output}} \left(\mathbf{Y}_i^{(t)}, \widetilde{\mathbf{Y}}_i^{(t)}\right) + \lambda_f \ell_{\text{fusion}} \left(\mathbf{H}_i^{(t)}, \widetilde{\mathbf{H}}_i^{(t)}\right) + \lambda_f \ell_{\text{feature}} \left(\mathbf{F}_i^{(t)}, \widetilde{\mathbf{F}}_i^{(t)}\right) + \lambda_w \ell_{\text{weight}} \left(\mathbf{W}_{j\to i}^{(t)}, \widetilde{\mathbf{W}}_{j\to i}^{(t)}\right)$$

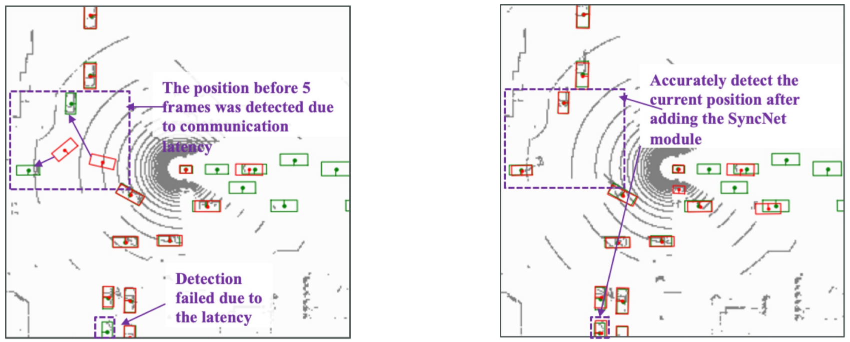
Communication Robust

Latency-aware collaborative perception

System overview: Keep sequential collaborative feature in memory leverage a compensation module.







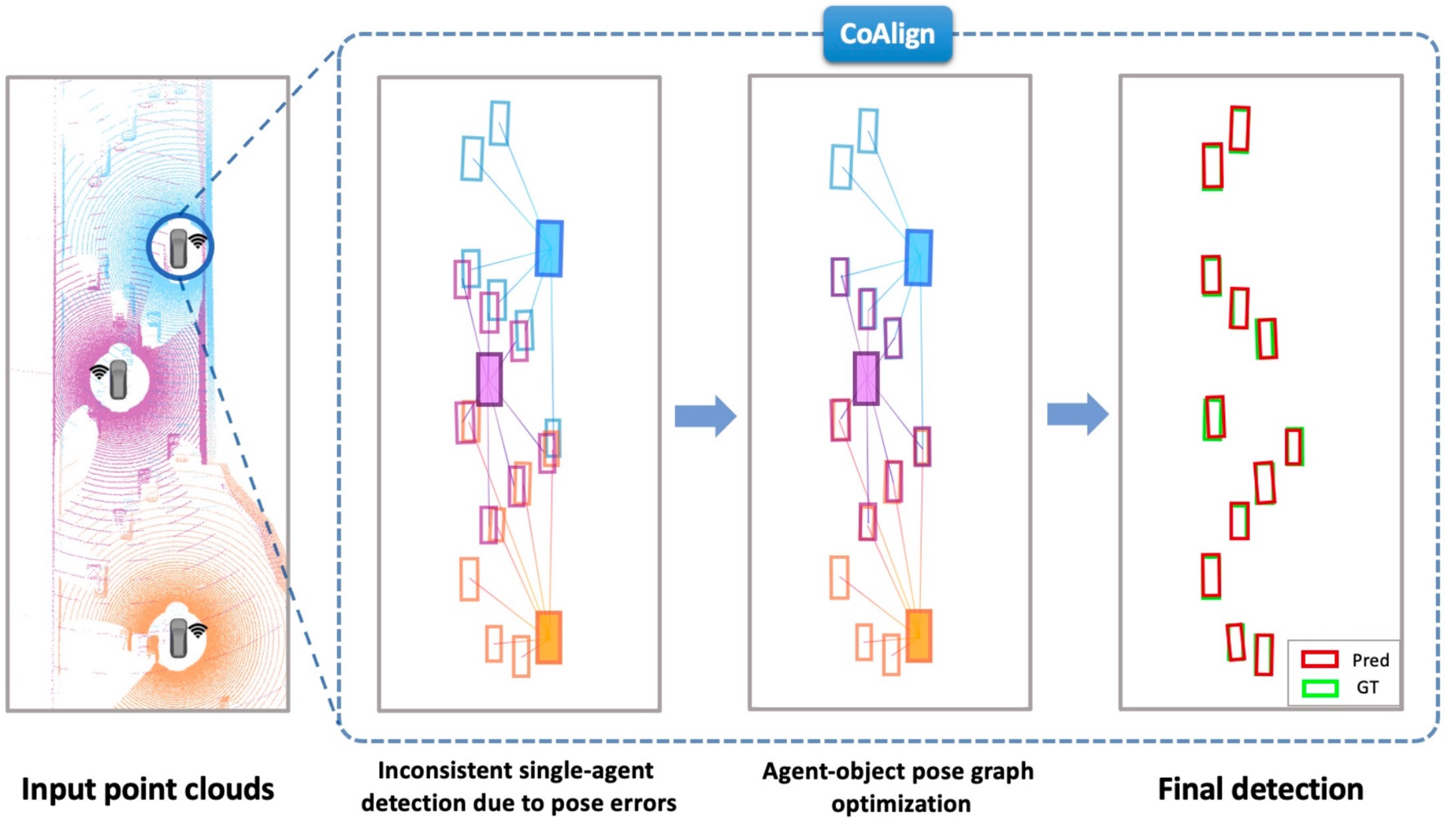
Latency-insensitivity

Latency-aware

Spatial-temporal alignment

Robust Collaborative 3D Object Detection in Presence of Pose Errors

Motivation: Accurate localization and synchronized clock is not always achievable

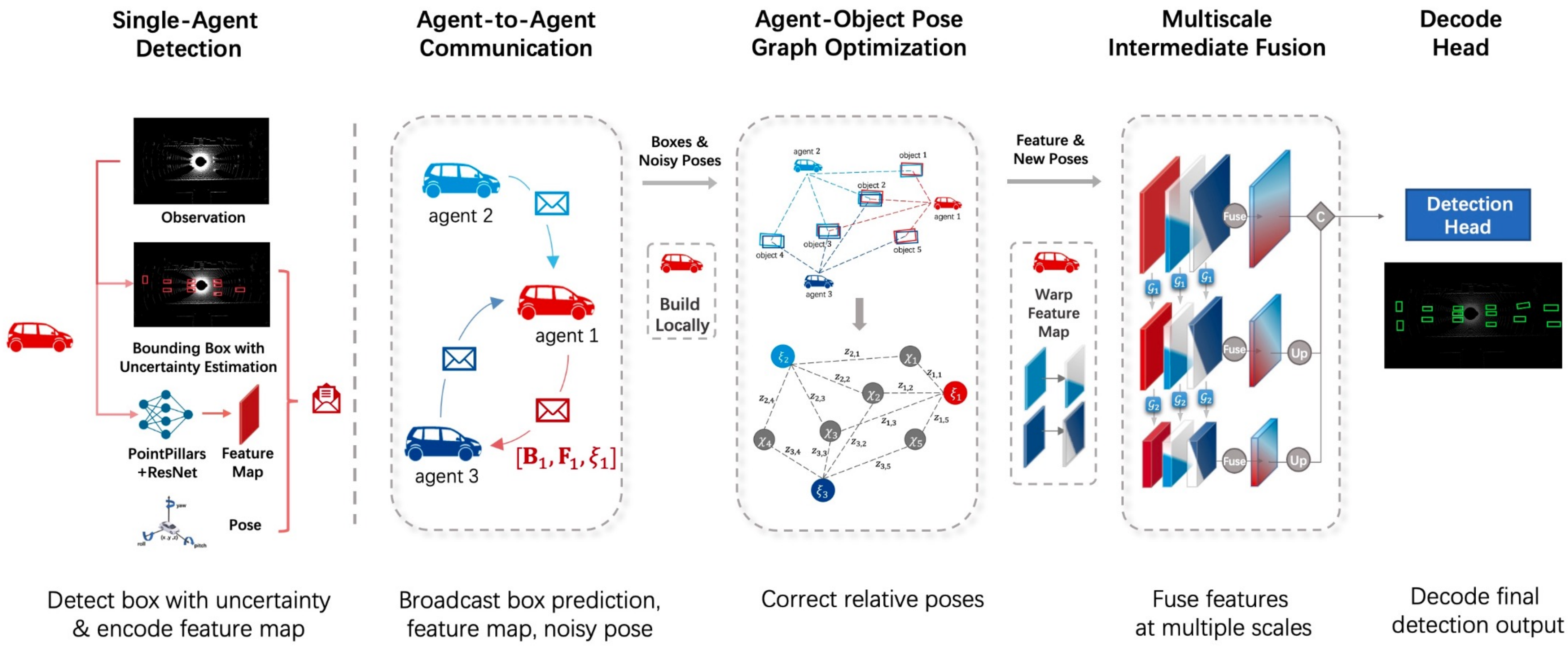


- Optimize the localization according to the correspondence of the bounding boxes
- Dealing with small noise(Require intersection between two corresponding boxes)

Spatial-temporal alignment

Robust Collaborative 3D Object Detection in Presence of Pose Errors

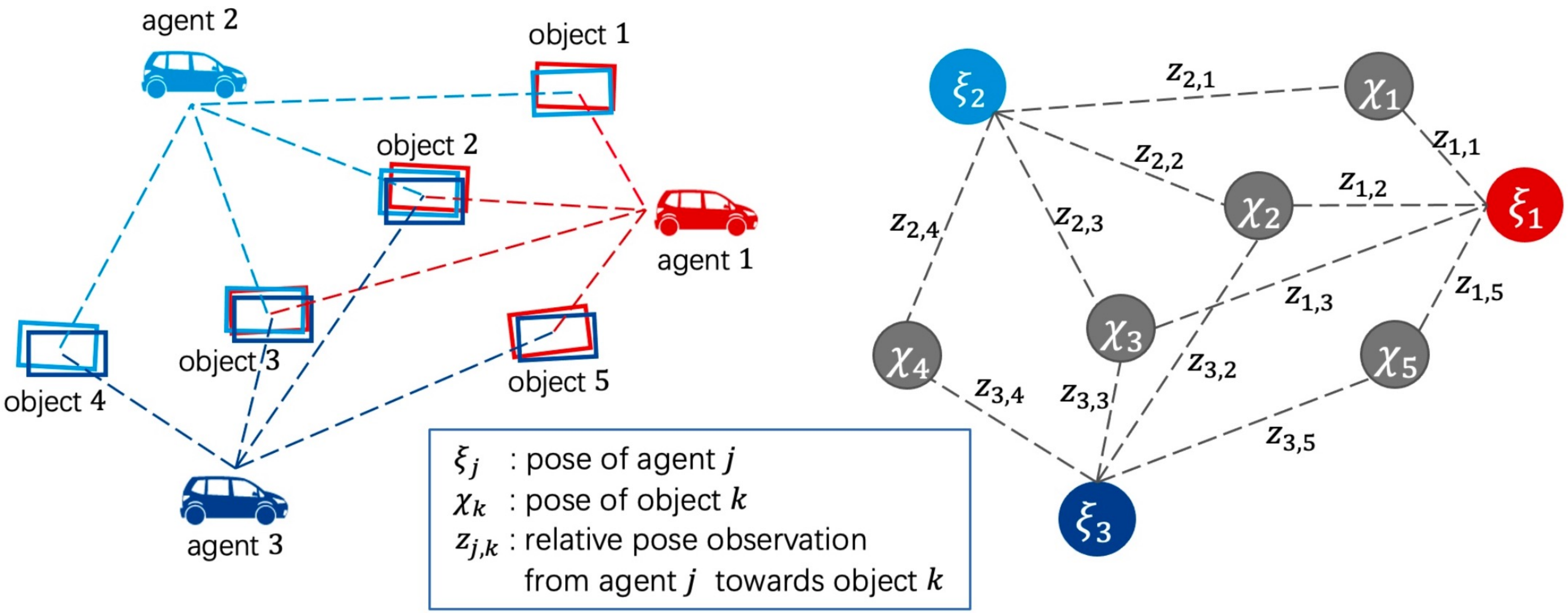
System Overview



Spatial-temporal alignment

Robust Collaborative 3D Object Detection in Presence of Pose Errors

Technique: Agent-object pose graph



(a) Agents and Objects' Boxes (b) Agent-Object Pose Graph

- Agent-object pose graph illustration.

Spatial-temporal alignment

Robust Collaborative 3D Object Detection in Presence of Pose Errors

Experiment Results

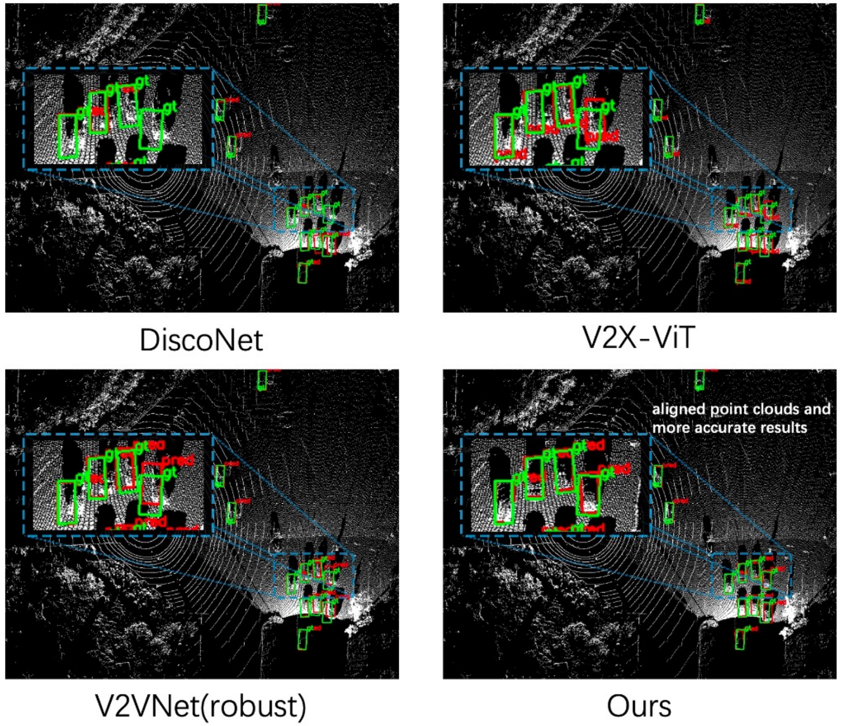
Dataset		OPV2V				V2X-Sim 2.0				DAIR-V2X			
Method/Metric		AP@0.5 ↑											
Noise Level $\sigma_t/\sigma_r(m/^\circ)$		0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6
w/o robust design	F-Cooper [8]	0.834	0.788	0.681	0.604	0.679	0.634	0.568	0.516	0.734	0.723	0.705	0.692
	V2VNet [9]	0.935	0.922	0.884	0.841	0.851	0.839	0.796	0.742	0.664	0.649	0.623	0.599
	DiscoNet [10]	0.916	0.906	0.884	0.862	0.785	0.775	0.748	0.708	0.736	0.726	0.708	0.697
	$\text{OPV2V}_{\text{pointpillar}}$ [5]	0.943	0.933	0.915	0.899	0.824	0.807	0.782	0.757	0.733	0.723	0.708	0.697
w/ robust design	MASH [15]	0.602	0.602	0.602	0.602	0.643	0.643	0.643	0.643	0.400	0.400	0.400	0.400
	FPV-RCNN [18]	0.858	0.817	0.591	0.419	0.870	0.835	0.654	0.480	0.655	0.631	0.580	0.581
	$\text{V2VNet}_{\text{robust}}$ [17]	0.942	0.938	0.929	0.918	0.840	0.836	0.811	0.778	0.660	0.655	0.646	0.636
	V2X-ViT [16]	0.946	0.942	0.931	0.914	**0.881**	**0.858**	0.808	0.759	0.704	0.700	0.689	0.678
	Ours	**0.966**	**0.962**	**0.958**	**0.945**	0.858	0.852	**0.822**	**0.796**	**0.746**	**0.738**	**0.720**	**0.700**

Method/Metric		AP@0.7 ↑											
Noise Level $\sigma_t/\sigma_r(m/^\circ)$		0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6
w/o robust design	F-Cooper [8]	0.602	0.504	0.412	0.376	0.489	0.434	0.379	0.362	0.559	0.552	0.542	0.538
	V2VNet [9]	0.740	0.686	0.586	0.504	0.769	0.726	0.673	0.634	0.402	0.388	0.367	0.350
	DiscoNet [10]	0.791	0.766	0.746	0.733	0.680	0.642	0.616	0.589	0.583	0.576	0.569	0.566
	$\text{OPV2V}_{\text{pointpillar}}$ [5]	0.827	0.804	0.780	0.765	0.672	0.651	0.632	0.625	0.553	0.547	0.540	0.538
w/ robust design	MASH [15]	0.198	0.198	0.198	0.198	0.384	0.384	0.384	0.384	0.244	0.244	0.244	0.244
	FPV-RCNN [18]	0.840	0.568	0.278	0.200	**0.838**	0.617	0.352	0.282	0.505	0.459	0.420	0.410
	$\text{V2VNet}_{\text{robust}}$ [17]	0.854	0.848	0.837	0.826	0.754	**0.743**	0.711	0.676	0.486	0.483	0.478	0.475
	V2X-ViT [16]	0.856	0.851	0.841	0.823	0.726	0.708	0.673	0.645	0.531	0.529	0.525	0.522
	Ours	**0.912**	**0.900**	**0.889**	**0.868**	0.765	0.742	**0.711**	**0.684**	**0.604**	**0.588**	**0.579**	**0.570**

Spatial-temporal alignment

Robust Collaborative 3D Object Detection in Presence of Pose Errors

Experiment Results



Collab-oration	Modules			AP@0.7 ↑			
	Agent-Object Pose Graph	Uncertainty	Intermediate Fusion	0.0/0.0	0.2/0.2	0.4/0.4	0.6/0.6
				0.703	0.703	0.703	0.703
		✓		0.730	0.730	0.730	0.730
✓			/	0.907	0.490	0.275	0.239
✓	✓		/	0.899	0.814	0.751	0.657
✓	✓	✓	/	0.903	0.818	0.758	0.672
✓			Single-scale	0.824	0.789	0.766	0.757
✓			Multi-scale	**0.914**	0.860	0.799	0.768
✓	✓		Multi-scale	0.910	0.897	0.886	0.865
✓	✓	✓	Multi-scale	0.912	**0.900**	**0.889**	**0.868**

TABLE III: Ablation studies on OPV2V dataset. All technique modules benefit 3D collaborative object detection. In Intermediate Fusion column, / is late fusion.

Lab and Scholars

UCLA

Mobility Lab (https://mobility-lab.seas.ucla.edu), Prof. Jiaqi Ma

Runsheng Xu

SJTU

CMIC (https://siheng-chen.github.io). Prof. Siheng Chen

Yue Hu

NYU

AI4CE Lab (https://ai4ce.github.io) , Prof. Chen Feng

Yiming Li