Collaboration Helps Camera Overtake LiDAR in 3D Detection

Yue Hu[1], Yifan Lu[1], Runsheng Xu[2], Weidi Xie[1,3], Siheng Chen[1,3], Yanfeng Wang[3,1]

[1]Shanghai Jiao Tong University, [2]University of California, Los Angeles, [3]Shanghai AI Laboratory

Paper: https://arxiv.org/abs/2303.13560
Github: https://github.com/MediaBrain-SJTU/CoCa3D

CVPR 2023

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

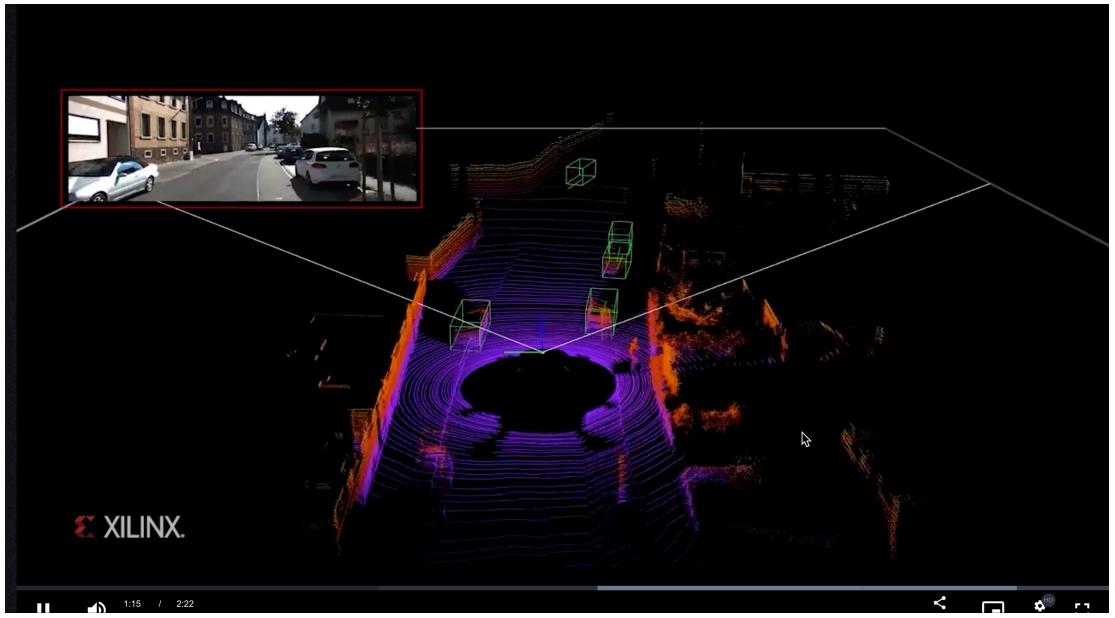
UCLA

Introduction

- 3D object detection

Camera-based: Telsa

Cheap while inferior performance
due to depth-lossy



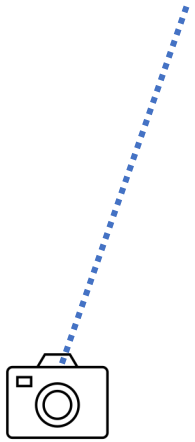
LiDAR-based: Waymo

Superior performance while
40times more expensive

Motivation

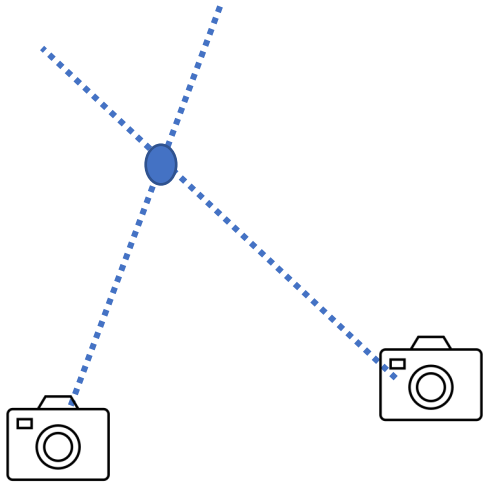
- Camera-only collaborative 3D object detection

Collaborative perception can allivate the ambiguiaty issues in single-camera and provide a cheap 3D detection solution.



Single-camera

Infinite depth candidates



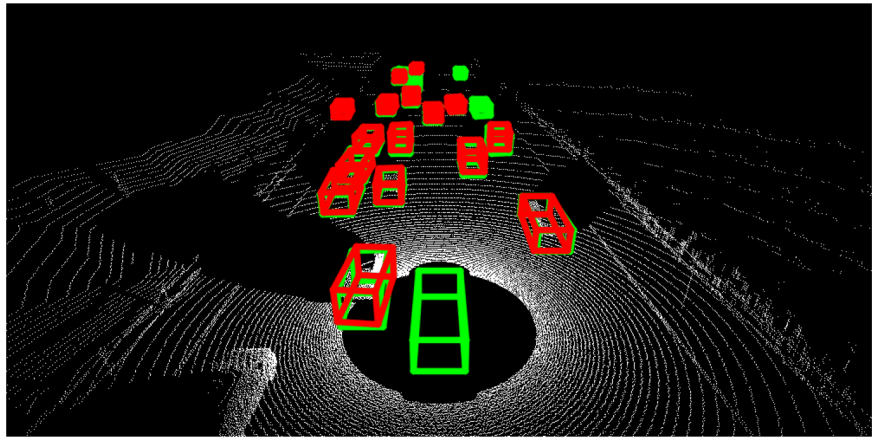
Collaborative cameras

Localize the correct depth candidate

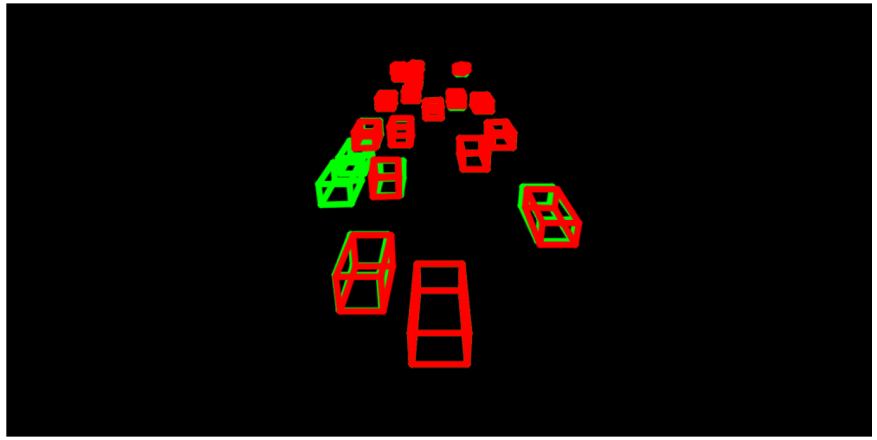
Motivation

- Camera-only collaborative 3D object detection

Collaborative perception can fundamentally solve the long range and occlusion issues in single-agent perception.



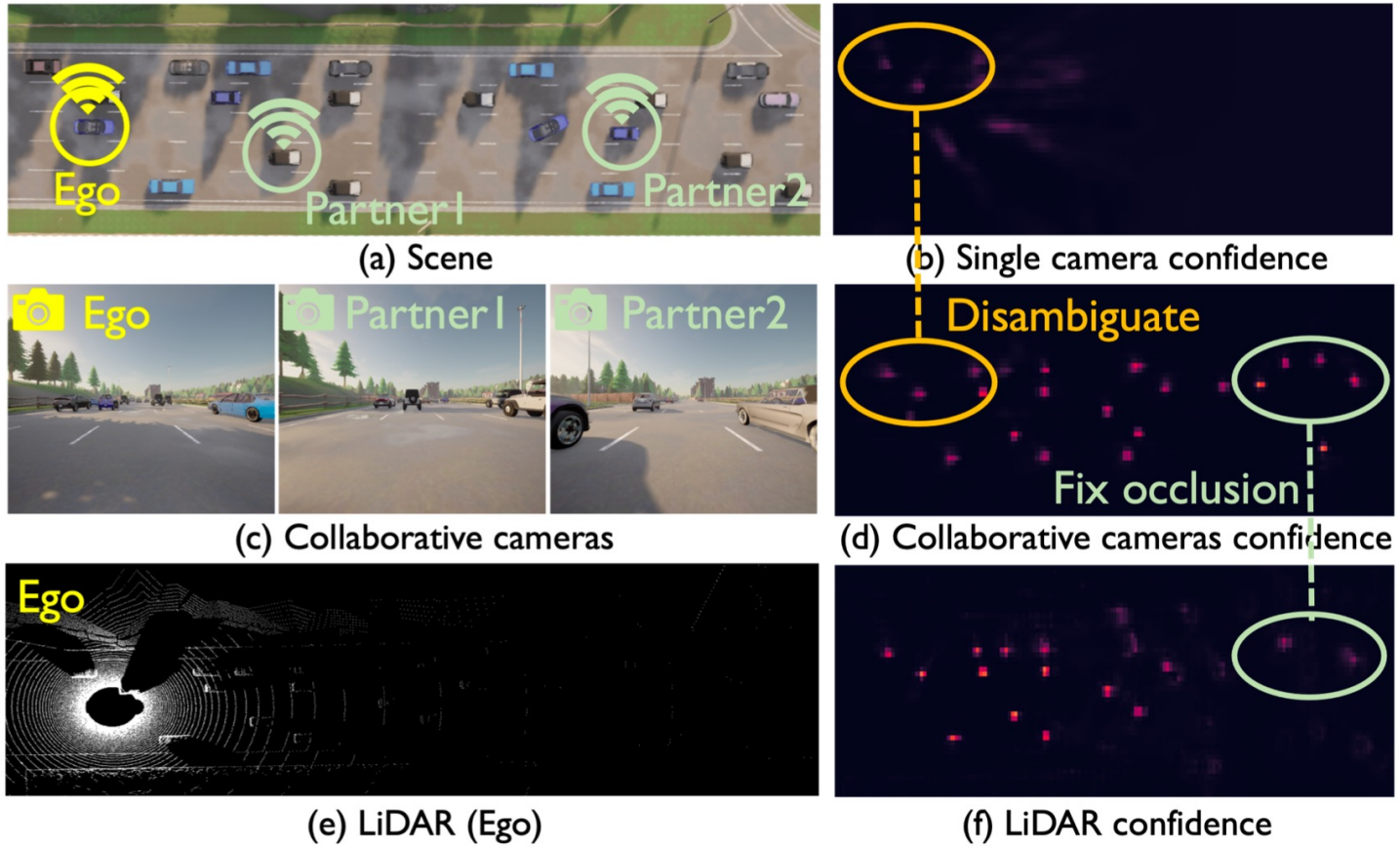
Single LiDAR



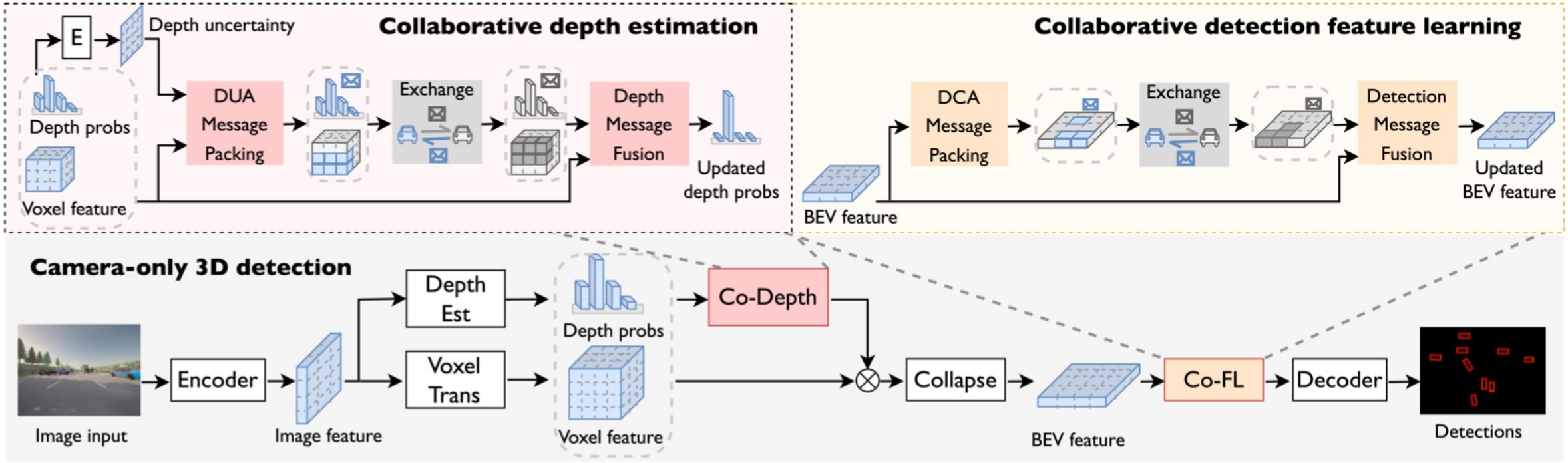
Collaborative cameras

Methodology – CoCa3D

- Core idea1: disambiguate single-view-estimated depth
- Core idea2: complement single-view occluded / long range regions



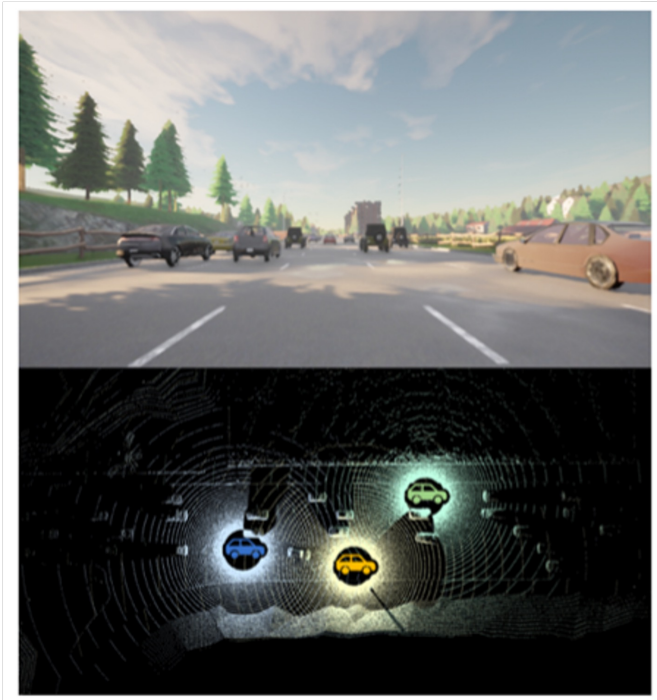
Methodology – CoCa3D



- Camera-only 3D detection detect 3D objects in the physical space based on the 2D camera inputs.
- Co-Depth **localize the correct depth candidate** through multi-view consistency.
- Co-DL **exchange complementary 3D detection** feature to achieve more holistic 3D detection.

Experiment – Dataset

Dataset	OPV2V+	DAIR-V2X[1]	CoPerception-UAVs+
View	Front (car)	Front (car)	Aerial
Data	Simulation	Real	Simulation
Agents	10	2	10



(a) OPV2V+

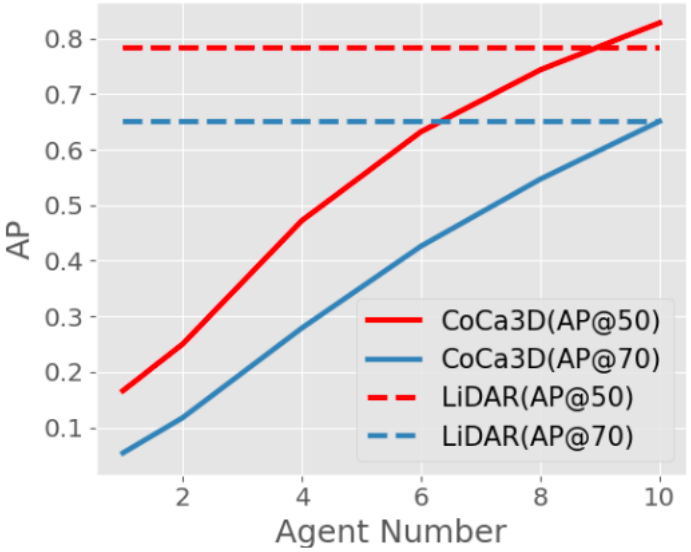
(b) DAIR-V2X

(c) CoPerception-UAVs+

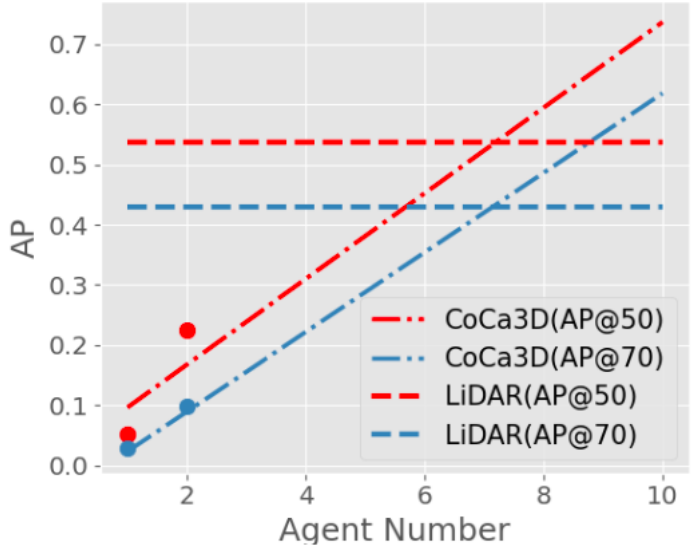
[1] Yu, Haibao et al. "DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection." CVPR (2022)

Experiment

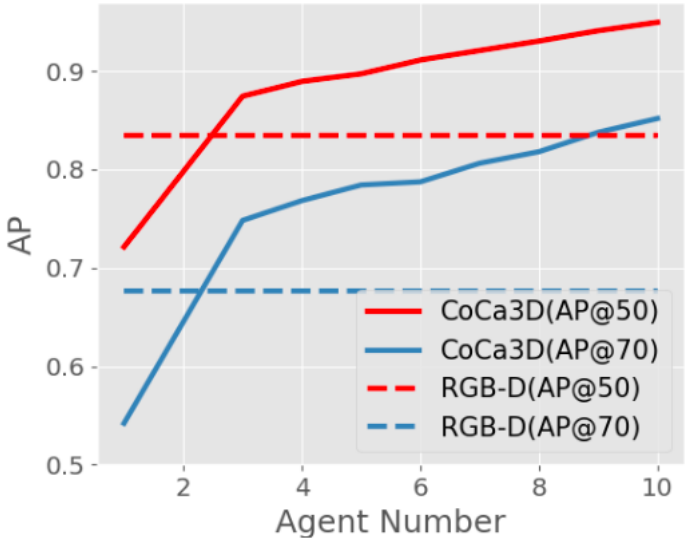
- Collaborative camera-only 3D detection overtakes LiDAR.



(a) OPV2V+



(b) DAIR-V2X



(c) CoPerception-UAVs+

Experiment

- CoCa3D significantly outperforms previous SOTAs, improves the SOTA performance by 30.60%/12.59%/ 44.21% on OPV2V+/CoPerception-UAVs+/DAIR-V2X

Method	OPV2V+			CoPerception-UAVs+			DAIR-V2X		
	AP@30	AP@50	AP@70	AP@50	AP@70	AP@80	AP@30	AP@50	AP@70
No Collaboration	0.2748	0.2041	0.0853	0.6956	0.4900	0.2309	0.0977	0.0524	0.0305
Late Fusion	0.6501	0.6198	0.5109	0.7206	0.5372	0.2597	0.2060	0.1078	0.0455
When2com (CVPR'20)	0.4853	0.4211	0.3737	0.8219	0.6705	0.4102	0.1957	0.0984	0.0459
V2VNet (ECCV'20)	0.6246	0.5042	0.3852	0.9093	0.7177	0.3804	0.1640	0.0847	0.0512
DiscoNet (NeurIPS'21)	0.7300	0.6009	0.4179	0.9054	0.7079	0.3564	0.1836	0.1262	0.0683
V2X-ViT (ECCV'22)	0.8346	0.6659	0.3946	0.9094	0.7143	0.3525	0.1862	0.1075	0.0490
Where2comm (NeurIPS'22)	0.8191	0.7089	0.4741	0.9102	0.7383	0.3676	0.1754	0.1025	0.0547
CoCa3D	**0.8642**	**0.8260**	**0.6675**	**0.9497**	**0.8502**	**0.5835**	**0.3522**	**0.2260**	**0.0985**

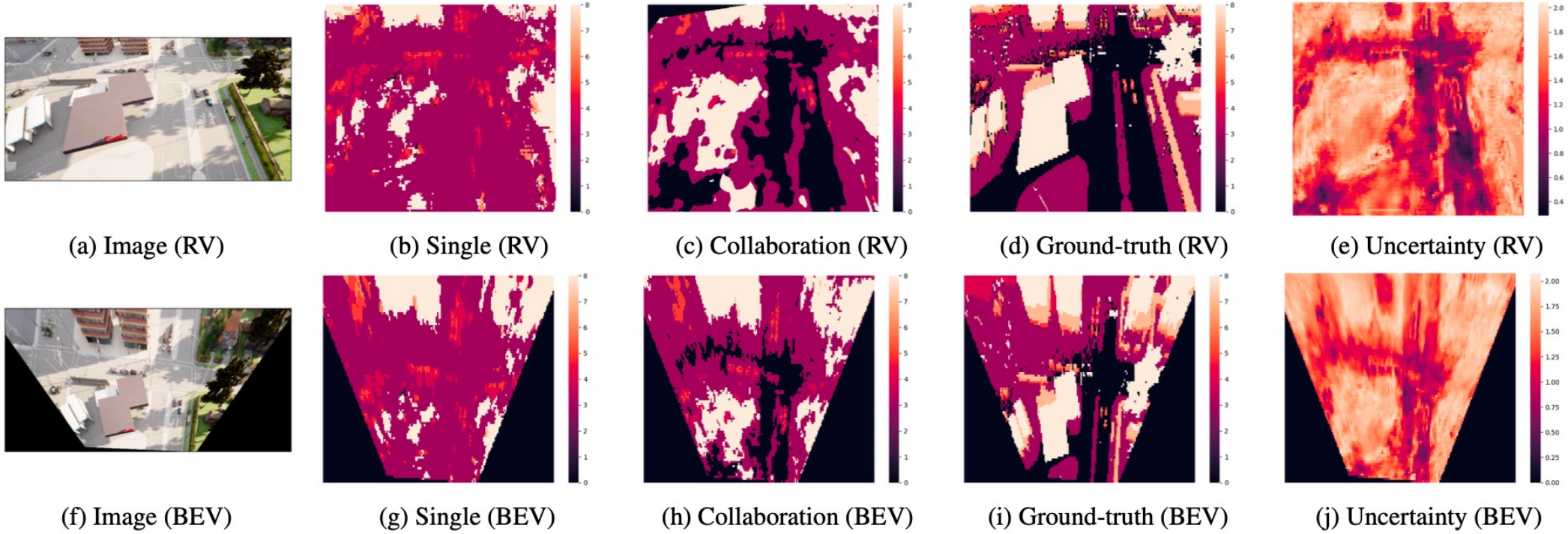
Experiment

- Co-Depth and Co-DL module substantially improves performance.

Co -Depth	Co -FL	OPV2V+ AP@30	OPV2V+ AP@50	OPV2V+ AP@70	CoPerception-UAVs+ AP@50	CoPerception-UAVs+ AP@70	CoPerception-UAVs+ AP@80
-	-	0.2748	0.2041	0.0853	0.7213	0.5421	0.2846
GT	-	0.3454	0.2553	0.0973	0.8347	0.6764	0.4120
-	✓	0.8201	0.7191	0.4756	0.9084	0.7256	0.4028
GT	✓	**0.9120**	**0.8805**	**0.7434**	**0.9505**	0.8398	0.5504
✓	✓	0.8642	0.8260	0.6675	0.9495	**0.8518**	**0.5849**

Experiment

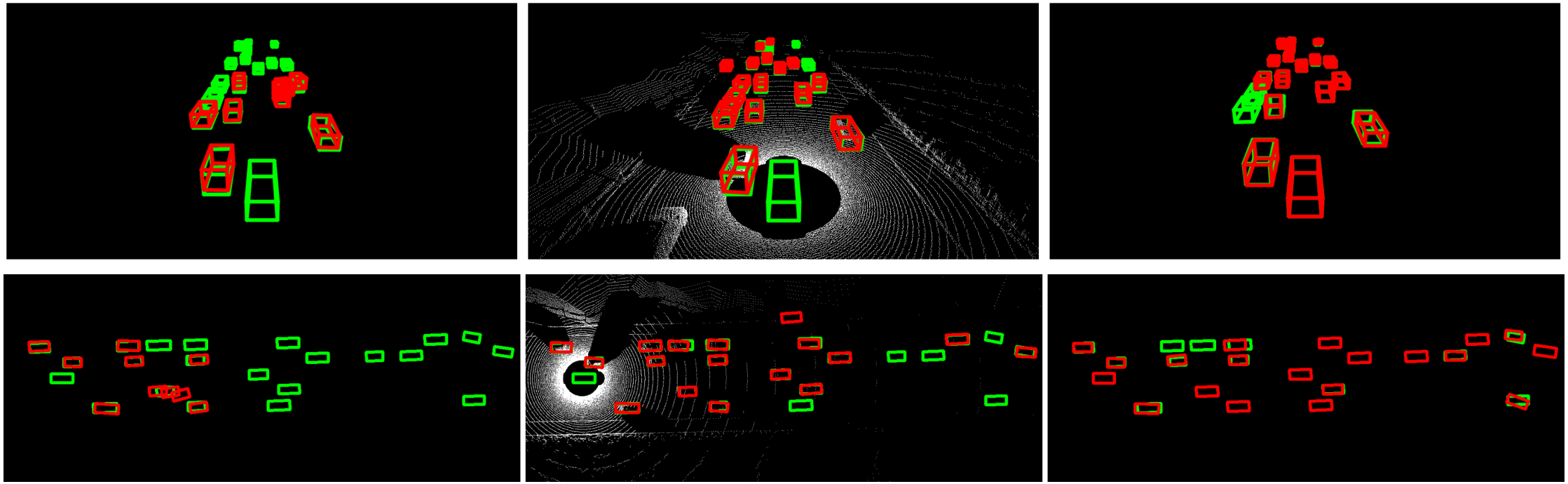
- Co-Depth outperforms single-agent depth estimation and approaches the ground truth depth.



(a) Image (RV) (b) Single (RV) (c) Collaboration (RV) (d) Ground-truth (RV) (e) Uncertainty (RV)

(f) Image (BEV) (g) Single (BEV) (h) Collaboration (BEV) (i) Ground-truth (BEV) (j) Uncertainty (BEV)

Experiment



Single-camera Single-LiDAR CoCa3D