

# Collaboration Helps Camera Overtake LiDAR in 3D Detection

Yue Hu<sup>1</sup> Yifan Lu<sup>1</sup> Runsheng Xu<sup>2</sup> Weidi Xie<sup>1,3</sup> Siheng Chen<sup>1,3\*</sup> Yanfeng Wang<sup>3,1</sup>

<sup>1</sup>Cooperative Medianet Innovation Center, Shanghai Jiao Tong University

<sup>2</sup>University of California, Los Angeles <sup>3</sup>Shanghai AI Laboratory

<sup>1</sup>{18671129361, yifan.lu, weidi, sihengc, wangyanfeng}@sjtu.edu.cn <sup>2</sup>{rxx3386}@ucla.edu

## 1. Appendix

Here, we start with dataset details, including generation and qualitative samples, then give out more implementation details.

### 1.1. Datasets

#### 1.1.1 OPV2V+

**Data generation.** We extend the original OPV2V [8] with more collaborative agents (10). Our OPV2V+ is co-simulated by OpenCDA [7] and CARLA [1]. Figure. 3 shows the simulation environment. OpenCDA provides the driving scenarios which ensure the agents drive smoothly and safely, including the vehicle’s initial location and moving speed. CARLA provides the maps, and weather and controls the movements of the agents. We replay the simulation logs of OPV2V and equip more vehicles with camera and depth sensors. Figure. 5 shows the four views (front, left, right, back) of the same agent. Figure. 6 shows a randomly selected data sample with 10 collaborative agents, the collected front view images in the same timestamp.

**Agent distribution.** We provide more statistical analysis of the distance distribution between agents to objects. 1) The distance between objects and their closest agents decreases as the number of agents increases, see Fig. 1a. Given 10 agents, this distance is mostly within 20m, see Fig. 1b. 2) Distribution of agents is uniform w.r.t objects, instead of the field of view (0-280m). Fig. 1c shows that the distribution of agents is same with all objects.

#### 1.1.2 CoPerception-UAVs+

**Data generation.** We extend the original CoPerception-UAVs [2] with more collaborative agents (10). Our CoPerception-UAVs+ is co-simulated by AirSim [6] and CARLA [1]. We use CARLA to generate complex simulation scenes and traffic flow, and AirSim to simulate drones flying in the scene and taking images. And we carefully design the drones’ flying route to ensure safety as more agents

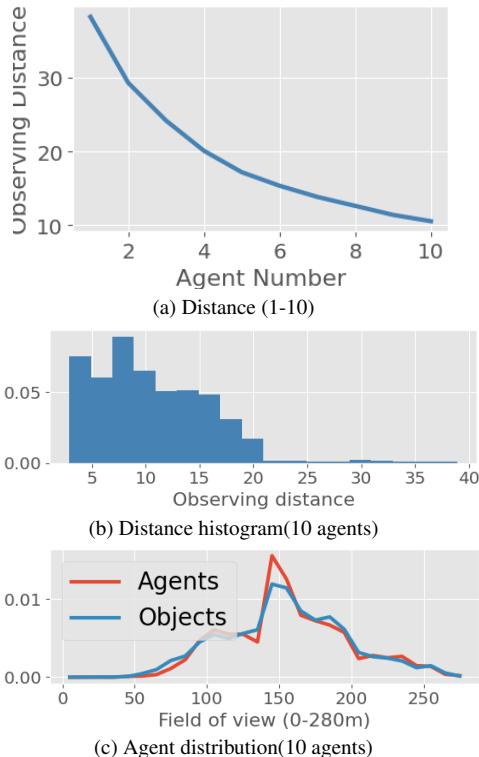


Figure 1. Object distance distribution.

increase the collision possibility. Figure. 4 shows the simulation environment. For CoPerception-UAVs+, we simulate more UAVs in AirSim and additionally equip depth sensors for each UAV at the same coordinate with the camera sensor. Figure. 7 shows a randomly selected data sample.

## 1.2. Implementation details

For the camera-only 3D object detection for cars, we implement the detector following the LSS [4] and CaDDN [5] for OPV2V+ and DAIR-V2X datasets. We uniformly space the depth into 50 categories. For the training strategy, we first train the single agent detector for 50 epochs with an initial learning rate of  $1.5e-3$ , and decay by 0.1 at epoch 30. Then we load the single pre-trained model and train the

\*Corresponding author.

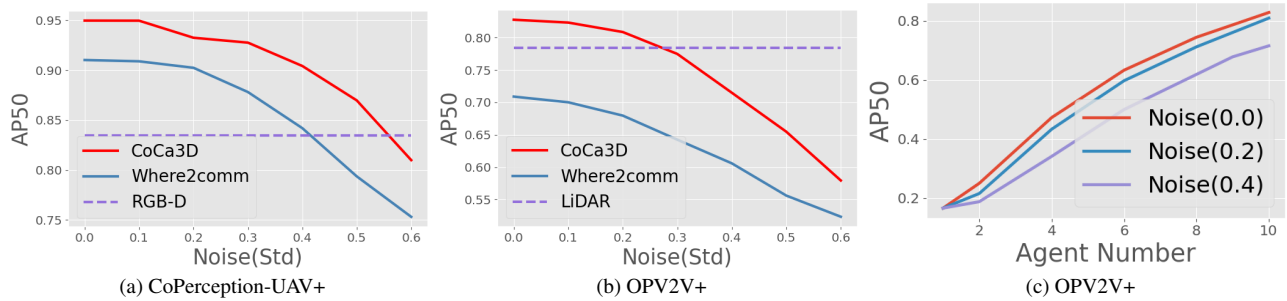


Figure 2. CoCa3D is as robust as SOTA where2comm (NeurIPS 22) to pose errors.



Figure 3. OPV2V+ is co-simulated by OpenCDA [7] and CARLA [1].



Figure 4. CoPerception-UAVs+ is co-simulated by CARLA [1] and AirSim [6].

whole model with collaboration for another 20 epochs with a learning rate of  $1e-3$ .

For the camera-only 3D object detection for drones, we implement the detector following the 3D aerial object detection DVDET [3]. We uniformly space the depth into 10 categories. For the training strategy, we train the model for 140 epochs with an initial learning rate of  $5e-4$ , and decay by 0.1 at epoch 80 and 120.

### 1.3. Robustness to pose error

In the paper, we assume collaborative agents' poses are accurate as practical agents should have strong self-localization ability. We further assess the model's ro-

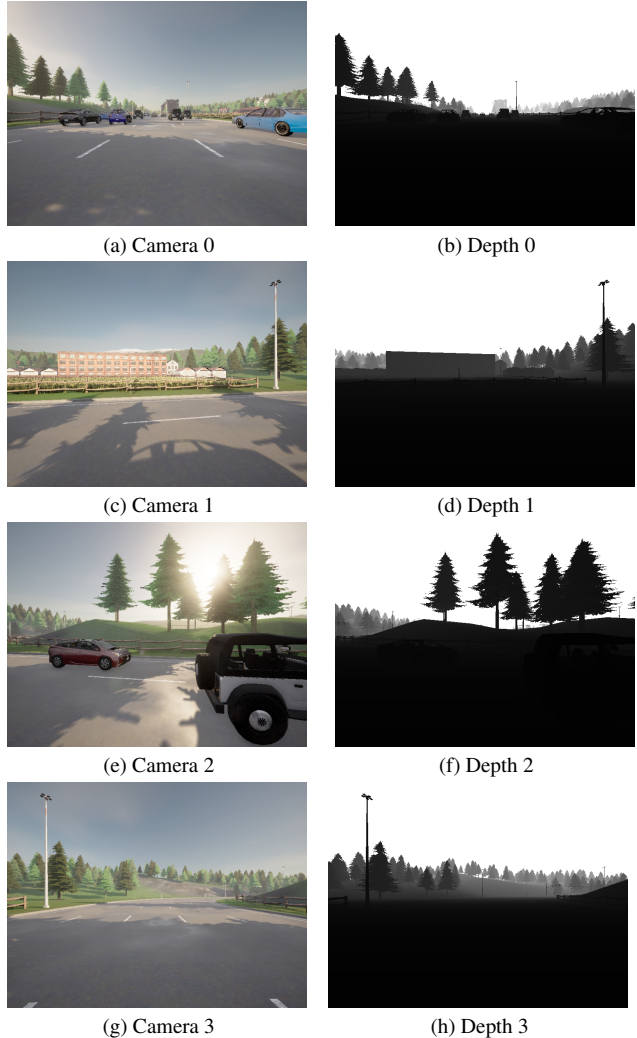


Figure 5. Each agent is equipped with 4 cameras and 4 depth sensors in OPV2V+.

bustness to agent pose errors. Encouragingly, **CoCa3D still performs well even pose errors appear**. Following the same pose-error setting in Where2comm [2] (Gaussian noise with 0 mean and 0m-0.6m standard deviation), our experiments validate CoCa3D's robustness and find: 1) CoCa3D outperforms Where2comm under vari-

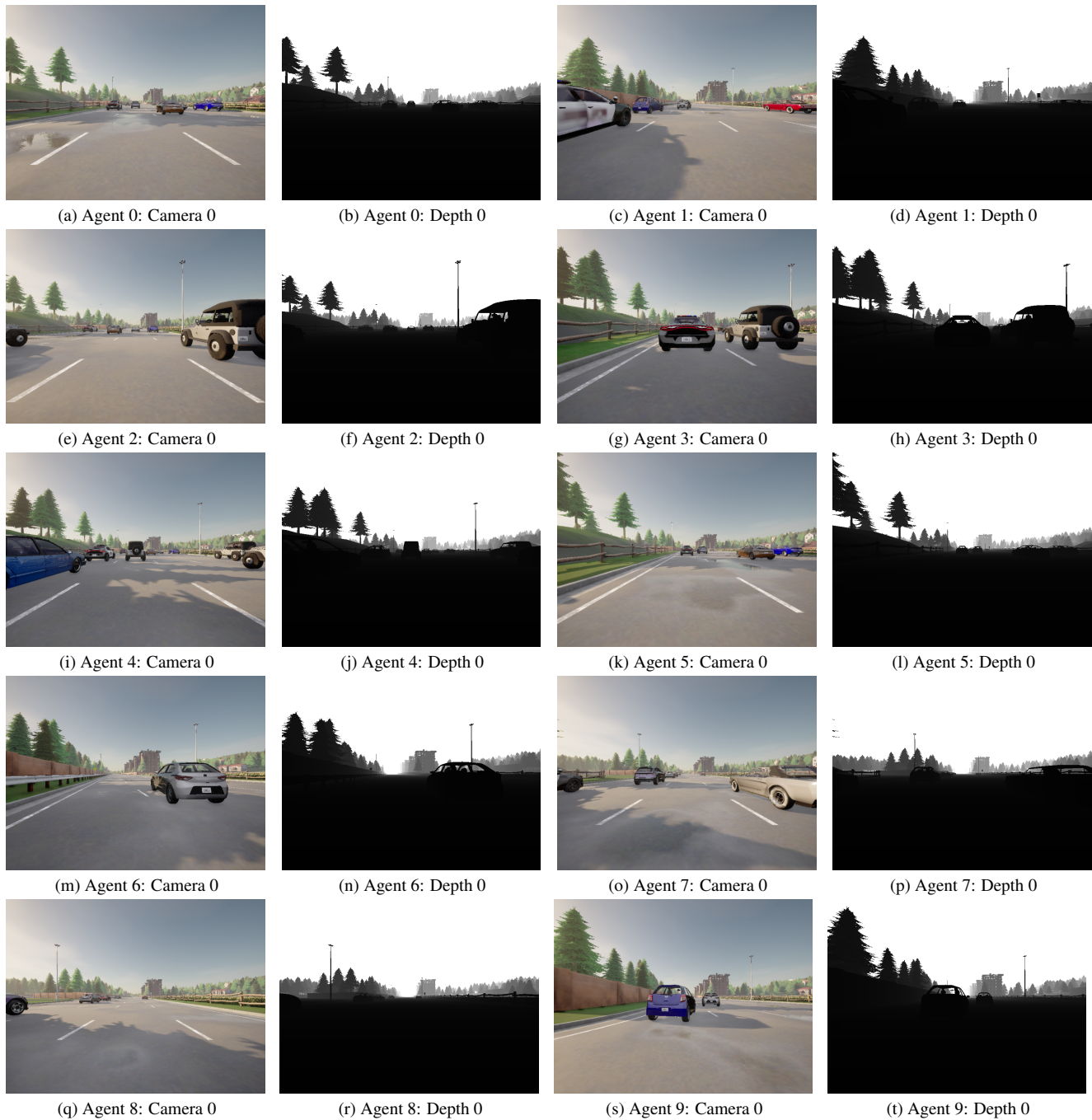


Figure 6. Data sample with 10 agents of OPV2V+.

ous pose errors, see Fig. 2a and 2b. 2) CoCa3D still outperforms LiDAR under large pose error 0.5m/0.2m on Co-UAV+/OPV2V+. 3) CoCa3D’s performance steadily increases with agent number even with pose errors, see Fig. 2c. Moreover, CoCa3D can integrate customized alignment methods, such as [1,2], to further tackle pose errors.

## References

- [1] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 1, 2
- [2] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in Neural*

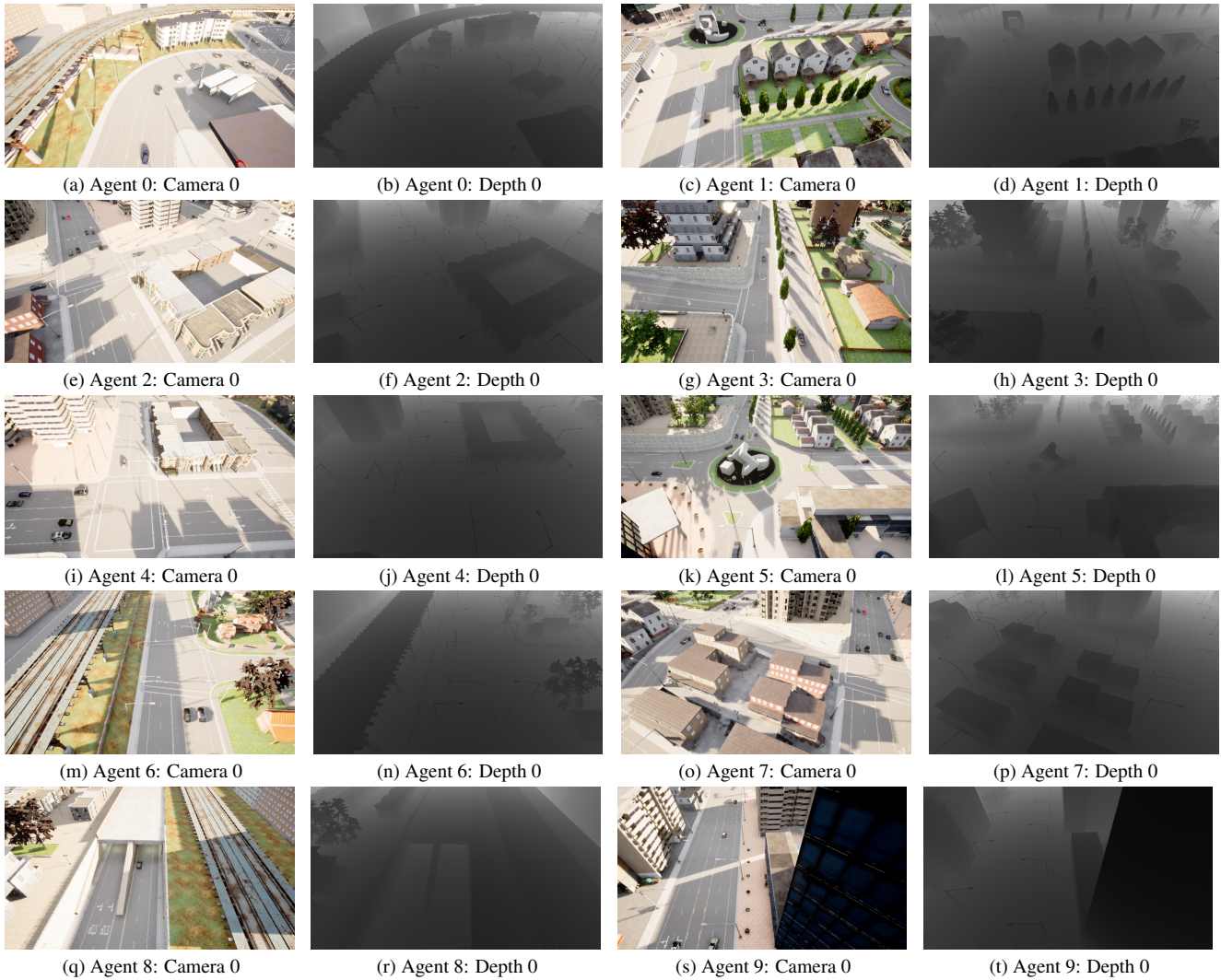


Figure 7. Data sample with 10 agents of CoPerception-UAVs+.

- Information Processing Systems*, 2022. 1, 2
- [3] Yue Hu, Shaoheng Fang, Weidi Xie, and Siheng Chen. Aerial monocular 3d object detection. *IEEE Robotics and Automation Letters*, 8:1959–1966, 2022. 2
- [4] Jonah Philion and Sanja Fidler. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In *European Conference on Computer Vision*, pages 194–210. Springer, 2020. 1
- [5] Cody Reading, Ali Harakeh, Julia Chae, and Steven L. Waslander. Categorical depth distribution network for monocular 3d object detection. *CVPR*, 2021. 1
- [6] Shital Shah, Debadepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and service robotics*, pages 621–635. Springer, 2018. 1, 2
- [7] Runsheng Xu, Yi Guo, Xu Han, Xin Xia, Hao Xiang, and Jiaqi Ma. Opencda: an open cooperative driving automation framework integrated with co-simulation. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 1155–1162. IEEE, 2021. 1, 2
- [8] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2583–2589. IEEE, 2022. 1