V2X-Sim: A Virtual Collaborative Perception Dataset for Autonomous Driving

Yiming Li[1] Ziyan An[1] Zixun Wang[2] Yiqi Zhong[3] Siheng Chen[4,✉] Chen Feng[1,✉]
[1]New York University [2]Beihang University
[3]University of Southern California [4]Shanghai Jiao Tong University
yimingli@nyu.edu, sihengc@sjtu.edu.cn, cfeng@nyu.edu

(a) Intersection (b) RGB images (c) Point cloud

Figure 1: (a) **Intersection** for vehicle-to-everything (V2X) communication. (b) **RGB images** from four vehicles passing through the same intersection. (c) Bird's eye view (BEV) **point cloud** from vehicles and roadside infrastructure (each color represents an entity).

Abstract

Vehicle-to-everything (V2X), which denotes the collaboration between a vehicle and any entity in its surrounding, can fundamentally improve the perception in self-driving systems. As the individual perception rapidly advances, collaborative perception has made little progress due to the shortage of public V2X datasets. In this work, we present the V2X-Sim dataset, the first public large-scale collaborative perception dataset in autonomous driving. V2X-Sim provides: 1) well-synchronized recordings from roadside infrastructure and multiple vehicles at the intersection to enable collaborative perception, 2) multi-modality sensor streams to facilitate multi-modality perception, 3) diverse well-annotated ground truth to support various downstream tasks including detection, tracking, and segmentation. We seek to inspire research on multi-agent multi-modality multi-task perception, and our virtual dataset is promising to promote the development of collaborative perception before realistic datasets become widely available.

1. Introduction

The autonomous driving community has recently made great efforts in dataset construction to support research in this area, especially with perception and prediction [2–4, 11, 30, 37, 42]. Current efforts center around increasing the dataset scale [37], sensing modality [3], and downstream task diversity [2]. With the help of available datasets, researchers have proposed and validated novel methods to build more robust and efficient self-driving systems.

Notwithstanding the great progress in dataset construction, existing published datasets are all captured by single – rather than multiple – vehicles. This presents a gap in collaborative autonomous driving research. Vehicle-to-everything (V2X), which denotes the collaboration between a vehicle and other entities such as vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I), seeks to help self-driving vehicles see further, better and even see through occlusion, thereby fundamentally improving safety. According to the estimation of U.S. NHTSA [1], there would be a minimum of 13% reduction in traffic accidents if a V2V system were implemented, which means 439,000 fewer crashes every year.

To fill in the gap in current research, it is an imperative to develop a well-established dataset for collaborative autonomous driving settings. Given that building such a dataset in the real world can be costly and laborious, we build a virtual dataset to advance collaborative perception research. Specifically, we employ SUMO [20], a micro-traffic simulation, to produce numerically-realistic traffic flow, and CARLA [8], a widely-used open-source simulator for autonomous driving research, to retrieve the sensor streams from multiple vehicles located at the same intersection. Besides, we mount sensors on the traffic lights to empower the roadside to perceive the environment, and the sensor streams of both the vehicles and the roadside infrastructure are synchronized to ensure smooth collaboration. In addition, multi-modality sensor streams of different entities are recorded to enable cross-modality perception. Meanwhile, diverse annotations including bounding boxes, vehicle trajectories, and pixel-wise as well as point-wise semantics labels are provided to facilitate various downstream tasks. Our dataset will be public and may inspire research in multi-agent multi-modality multi-task perception before realistic data becomes readily available to the community. To summarize, contributions of this work are:

- We propose V2X-Sim, the first public collaborative perception dataset in autonomous driving.
- We provide multi-modality data from multiple agents to enable cross-modality perception.
- We provide diverse well-annotated ground truth to support various downstream tasks.

2. Related Work

Autonomous driving dataset. Since the pioneer dataset KITTI [11] was released, the autonomous driving community has been trying to increase the dataset comprehensiveness in terms of driving scenarios, sensor modalities, and data annotations. Regarding driving scenarios, current datasets covered crowded urban scenes [30], adverse weather conditions [32], night scenes [31], multiple cities [3] to enrich the data distribution. As for sensor modalities, nuScenes [3] collected data with Radar, RGB camera, and LiDAR in a 360° viewpoint; WoodScape [42] captured data with fisheye cameras; and A2D2 [12] provided extensive vehicle bus data including the steering wheel angle, throttle, and braking. Regarding data annotations, semantic labels in both images [7, 15, 29, 35] and point cloud [2, 14] were provided to enable semantic segmentation; 2D/3D box trajectories were offered [4, 9] to facilitate tracking and prediction. In summary, existing datasets generally emphasized the data comprehensiveness in single-vehicle situations, but ignored the multi-vehicle collaborative self-driving scenarios.

V2X system and dataset. By sharing information with other vehicles or the roadside infrastructure, V2X mitigates the shorting-comings of individual-vehicle perception and planning such as the limited sensing range and frequent occlusion. Previous research [18] developed an enhanced cooperative microscopic traffic model in V2X scenarios, and investigated the effect of V2X in traffic disturbance scenarios. [19] proposed a multi-modal cooperative perception system that provides see-through, lifted-seat, satellite and all-around views to drivers. More recently, [39] incorporated deep learning into the V2V system: multiple intelligent vehicles share the intermediate features output by the neural network to promote the vehicle's perception and prediction capability. As for the dataset, [5, 23, 41] simulated the V2V scenarios with different frames from KITTI [11]. Yet, they were unrealistic for not capturing the measurements at the same time. Some other works used a platoon strategy for data capture [6, 33], but they were biased because the observations were highly correlated with each other. [39] proposed V2V-Sim based on a high-quality LiDAR simulator [24]. Unfortunately, V2V-Sim does not include the V2I scenario and is not publicly available.

Synthetic dataset. Simulation environments can help generate large-scale datasets with well-annotated ground truth. Current literature on computer vision has exploited synthetic datasets in a wide array of tasks, *e.g.*, visual tracking [10, 28], semantic segmentation [22, 34], flow estimation [27], visual surveillance [38], visual odometry [34], 3D perception [40], multi-view stereo [21], and egocentric localization [17]. Synthetic datasets not only enable large-scale training through free and precise annotations, but also support the cutting-edge research before realistic data becomes readily available. An example of the latter application is the long-range sensor in [40]. Multiple prior works have proven that pre-training a model using synthetic data can improve the model's performance on the real data [16, 26, 26, 36]. To further optimize such usage of synthetic data, domain adaptation techniques [13, 25] are utilized. In this work, we use a fully open-sourced autonomous driving simulator, CARLA [8], to generate V2X-Sim.

Sensor	Description
V: 6 × RGB camera **I:** 4 × RGB camera	Each vehicle is equipped with 6 cameras. Each camera has a FoV of 70°, except for the back camera that has a FoV of 110°. Each roadside has 4 cameras looking diagonally downward at 35° with a 70° FoV. The image size is 1600×900.
V: 6 × Depth camera	Each vehicle has 6 depth cameras with the same setting as RGB cameras.
V: 6 × Semantic camera	Each vehicle has 6 semantic segmentation cameras with the same setting as RGB cameras.
V&I: 1 × BEV semantic camera	Each vehicle and roadside has one BEV semantic camera at the top, looking downward. Both the raw images (semantic tags encoded in the red channel) and the converted colored images are provided. The image size is 900×900.
V&I: 1 × LiDAR and Semantic LiDAR	We attach one LiDAR and one semantic LiDAR on top of the ego vehicle and the intersection center. Specs: 32 channels, 70m max range, 250,000 points per second, 20 Hz rotation frequency.

Table 1: Sensor specification of vehicle (**V**) and roadside infrastructure (**I**) in our V2X-Sim dataset. All the sensors are recorded at 5Hz.

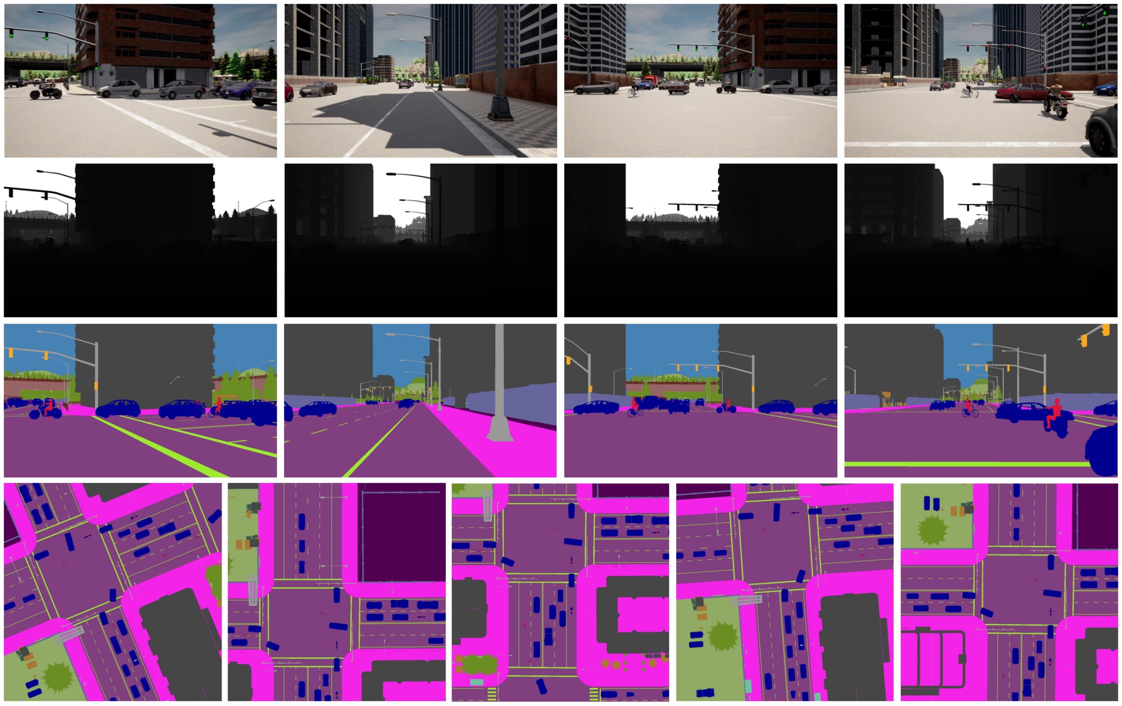

Figure 2: Example of multi-agent multi-modality perception. From top to bottom: RGB image, depth, semantic segmentation, and BEV semantic segmentation. From left to right are respectively four vehicles' recordings, except for the last row which appends an image of roadside in the last column.

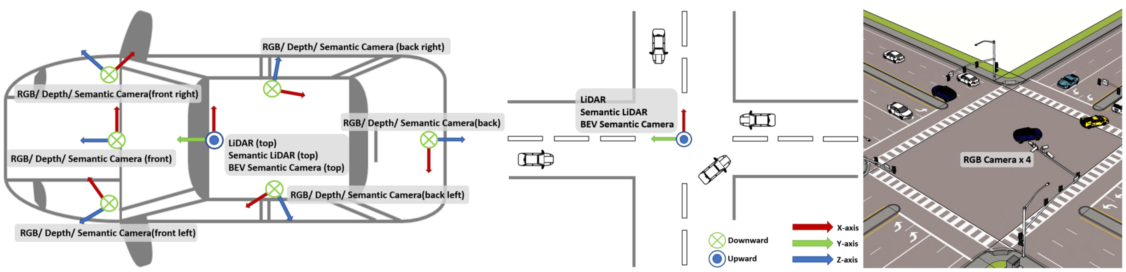

Figure 3: Sensor layout and coordinate systems.

3. V2X-Sim Dataset

3.1. Sensor suite of vehicle and roadside

Multi-modality sensing data is essential for robust perception. To ensure the comprehensiveness of our dataset, we equip each vehicle with a sensor suite based on CARLA. It is composed of RGB cameras, depth cameras, semantic segmentation cameras, BEV semantic segmentation cameras, LiDAR, and semantic LiDAR. Meanwhile, the road infrastructure is equipped with RGB cameras, BEV semantic segmentation cameras, LiDAR, and semantic LiDAR.

Sensor configuration. On both ego vehicles and roadside infrastructure, the camera and LiDAR cover 360° horizontally to enable full-view perception. Specifically, each ego-vehicle carries six RGB cameras following nuScenes configuration [3]; the roadside infrastructure is equipped with four RGB cameras toward four directions at the crossroad. Note that the BEV semantic segmentation camera is based on orthogonal projection while the ego-vehicle semantic segmentation camera uses perspective projection. Table 1 summarizes the detailed sensor specification.

Sensor layout and coordinate system. The overall sensor layout and coordinate system is shown in Fig. 3, and one example of multi-agent multi-modality perception is shown in Fig. 2. On both ego-vehicle and roadside infrastructure, LiDAR and semantic LiDAR, RGB/depth/semantic cameras are placed at the same location to obtain depth/semantics ground truth. The BEV semantic segmentation camera shares the same x, y position with LiDAR yet is placed higher to ensure a certain size of field of view. As for the roadside infrastructure, sensors are placed at random heights within a realistic range to enhance

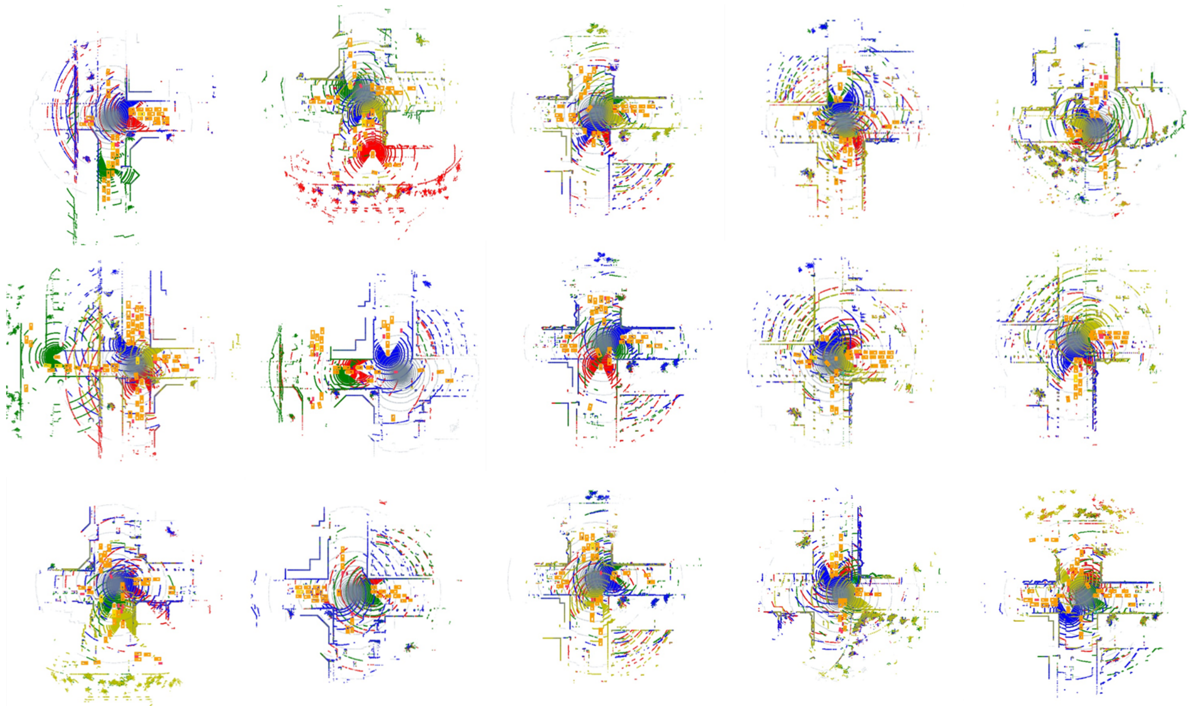

Figure 4: Visualizations of the bird's eye view point cloud from different scenes. Gray denotes the point cloud captured by the roadside LiDAR. Each color (except for gray) represents an vehicle, and the orange boxes denote the vehicles in the scene.

the diversity. Note that we invert the y-axis in CARLA and use right-hand coordinate system following nuScenes [3].

Diverse annotations. To assist downstream tasks including detection, tracking and semantic segmentation, we provide various annotations such as 3D bounding boxes, pixel-wise and point-wise semantic labels. Each box is defined by the location of its center in x, y, z coordinates, and its width, length, and height. Besides, there are totally twenty-three categories such as pedestrian, building, ground, *etc*. In addition, precise depth values are also provided for depth estimation.

3.2. CARLA-SUMO co-simulation

We consider it a realistic V2X scenario when multiple vehicles with their own routes are simultaneously located in the same geographical area, *i.e.*, an intersection. The roadside infrastructures are also empowered by sensing capability. We use CARLA-SUMO co-simulation for traffic flow simulation and data recording. Vehicles are spawned in CARLA via SUMO, and managed by the Traffic Manager. The script $spawn\_npc\_sumo.py$ provided by CARLA automatically generates a SUMO network in a certain town, and produces random routes to make vehicles roam around. Hundreds of vehicles are spawned in different towns ($Town03$, $Town03$ and $Town05$ that have cross junctions and multiple lanes per direction), and we record several log files, each with a length of five minutes. Then we read out 100 scenes from the log files at different intersections. Each scene includes a duration of 20 seconds, and we select $M(M = 2, 3, 4, 5)$ vehicles in a scene as the intelligent agents to share information with each other. See Fig. 4 for several example scenes.

3.3. Downstream tasks

Our dataset can not only support individual perception tasks such as 3D object detection, tracking, image-/point cloud-based semantic segmentation, depth estimation, but also enable collaborative perception like collaborative 3D object detection, tracking, and collaborative BEV semantic segmentation in urban driving scenes. We will provide a benchmark for the collaborative perception algorithms.

4. Conclusion

We propose V2X-Sim, the first virtual collaborative perception dataset in autonomous driving scenes based on CARLA simulator. By providing both multi-agent multi-modality sensor streams in realistic traffic flows and rich annotations, V2X-Sim can facilitate various perception tasks especially collaborative perception before realistic datasets become widely available. Our work seeks to inspire a variety of relevant research areas including but not limited to computer vision, multi-robot system, and deep learning.

References

[1] Vehicle-to-vehicle communication technology for light vehicles. https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/v2v_pria_12-12-16_clean.pdf, 2016. 1

[2] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, C. Stachniss, and Juergen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9296–9306, 2019. 1, 2

[3] H. Caesar, Varun Bankiti, A. Lang, Sourabh Vora, Venice Erin Liong, Q. Xu, A. Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11618–11628, 2020. 1, 2, 3, 4

[4] Ming-Fang Chang, J. Lambert, Patsorn Sangkloy, J. Singh, Slawomir Bak, Andrew T. Hartnett, De Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays. Argoverse: 3d tracking and forecasting with rich maps. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8740–8749, 2019. 1, 2

[5] Qi Chen, Sihai Tang, Q. Yang, and Song Fu. Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds. In *IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 514–524, 2019. 2

[6] Qi Chen, Ting Yuan, J. Hillenbrand, A. Gern, Tobias Roth, F. Kuhnt, Johann Marius Zöllner, J. Breu, Miro Bogdanovic, and Christian Weiss. Dsrc and radar object matching for cooperative driver assistance systems. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 1348–1354, 2015. 2

[7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, M. Enzweiler, Rodrigo Benenson, Uwe Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, 2016. 2

[8] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017. 2

[9] S. Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Han Zhao, Sabeek Pradhan, Yuning Chai, Benjamin Sapp, C. Qi, Yin Zhou, Zoey Yang, Aurelien Chouard, Pei Sun, Jiquan Ngiam, Vijay Vasudevan, Alexander McCauley, Jonathon Shlens, and Drago Anguelov. Large scale interactive motion forecasting for autonomous driving : The waymo open motion dataset. *ArXiv*, abs/2104.10133, 2021. 2

[10] Adrien Gaidon, Qiao Wang, Yohann Cabon, and Eleonora Vig. Virtualworlds as proxy for multi-object tracking analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4340–4349, 2016. 2

[11] Andreas Geiger, Philip Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012. 1, 2

[12] Jakob Geyer, Y. Kassahun, M. Mahmudi, Xavier Ricou, R. Durgesh, Andrew S. Chung, L. Hauswald, Viet Hoang Pham, Maximilian Mühlegg, S. Dorn, Tiffany Fernandez, M. Jänicke, S. Mirashi, Chiragkumar Savani, M. Sturm, O. Vorobiov, Martin Oelker, Sebastian Garreis, and P. Schuberth. A2d2: Audi autonomous driving dataset. *ArXiv*, abs/2004.06320, 2020. 2

[13] Judy Hoffman, Eric Tzeng, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 4068–4076, 2015. 2

[14] Qingyong Hu, Bo Yang, Sheikh Khalid, Wenna Xiao, A. Trigoni, and A. Markham. Towards semantic segmentation of urban-scale 3d point clouds: A dataset, benchmarks and challenges. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[15] Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, P. Wang, Yuanqing Lin, and Ruigang Yang. The apolloscape dataset for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1067–10676, 2018. 2

[16] Braden Hurl, K. Czarnecki, and Steven L. Waslander. Precise synthetic image and lidar (presil) dataset for autonomous vehicle perception. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 2522–2529, 2019. 2

[17] Hae-Gon Jeon, Sunghoon Im, Byeong-Uk Lee, François Rameau, Dong-Geol Choi, Jean Oh, In-So Kweon, and M. Hebert. A large-scale virtual dataset and egocentric localization for disaster responses. *IEEE transactions on pattern analysis and machine intelligence*, 2021. 2

[18] Dongyao Jia and D. Ngoduy. Enhanced cooperative car-following traffic model with the combination of v2v and v2i communication. *Transportation Research Part B-methodological*, 90:172–191, 2016. 2

[19] Seong-Woo Kim, B. Qin, Z. J. Chong, Xiaotong Shen, Wei Liu, M. Ang, Emilio Frazzoli, and D. Rus. Multivehicle cooperative driving using cooperative perception: Design and experimental validation. *IEEE Transactions on Intelligent Transportation Systems*, 16:663–680, 2015. 2

[20] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. Recent development and applications of sumo-simulation of urban mobility. *International journal on advances in systems and measurements*, 5(3&4), 2012. 2

[21] A. Ley, R. Hänsch, and O. Hellwich. Syb3r: A realistic synthetic benchmark for 3d reconstruction from images. In *ECCV*, 2016. 2

[22] Yen-Cheng Liu, Junjiao Tian, Nathan Glaser, and Z. Kira. When2com: Multi-agent perception via communication graph grouping. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4105–4114, 2020. 2

[23] Y. Maalej, Sameh Sorour, A. Abdel-Rahim, and M. Guizani. Vanets meet autonomous vehicles: A multimodal 3d environment learning approach. In *IEEE Global Communications Conference*, pages 1–6, 2017. 2

[24] Sivabalan Manivasagam, Shenlong Wang, K. Wong, Wenyuan Zeng, Mikita Sazanovich, Shuhan Tan, Binh Yang, Wei-Chiu Ma, and R. Urtasun. Lidarsim: Realistic lidar simulation by leveraging the real world. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages

11164–11173, 2020. 2
[25] Francisco Massa, Bryan C. Russell, and Mathieu Aubry. Deep exemplar 2d-3d detection by adapting from real to rendered views. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6024–6033, 2016. 2
[26] Maxim Maximov, Kevin Galim, and L. Leal-Taix'e. Focus on defocus: Bridging the synthetic to real domain gap for depth estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1068–1077, 2020. 2
[27] N. Mayer, Eddy Ilg, Philip Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4040–4048, 2016. 2
[28] Matthias Mueller, Neil G. Smith, and Bernard Ghanem. A benchmark and simulator for uav tracking. In *ECCV*, 2016. 2
[29] Gerhard Neuhold, Tobias Ollmann, S. R. Bulò, and P. Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *IEEE International Conference on Computer Vision (ICCV)*, pages 5000–5009, 2017. 2
[30] Abhishek Patil, Srikanth Malla, Haiming Gang, and Yi-Ting Chen. The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. *2019 International Conference on Robotics and Automation (ICRA)*, pages 9552–9557, 2019. 1, 2
[31] Quang-Hieu Pham, Pierre Sevestre, R. Pahwa, Huijing Zhan, C. H. Pang, Yuda Chen, A. Mustafa, V. Chandrasekhar, and Jie Lin. A*3d dataset: Towards autonomous driving in challenging environments. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2267–2273, 2020. 2
[32] Matthew Pitropov, D. Garcia, J. Rebello, Michael Smart, Carlos Wang, K. Czarnecki, and Steven L. Waslander. Canadian adverse driving conditions dataset. *The International Journal of Robotics Research*, 40:681 – 690, 2021. 2
[33] Z. Y. Rawashdeh and Z. Wang. Collaborative automated driving: A machine learning-based method to enhance the accuracy of shared information. In *International Conference on Intelligent Transportation Systems (ITSC)*, pages 3961–3966, 2018. 2
[34] Stephan R. Richter, Zeeshan Hayder, and V. Koltun. Playing for benchmarks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2232–2241, 2017. 2
[35] G. Ros, Laura Sellart, Joanna Materzynska, David Vázquez, and Antonio M. López. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3234–3243, 2016. 2
[36] Ahmad El Sallab, Ibrahim Sobh, Mohamed Zahran, and M. Shawky. Unsupervised neural sensor models for synthetic lidar data augmentation. *ArXiv*, abs/1911.10575, 2019. 2
[37] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, P. Tsui, J. Guo, Y. Zhou, Yuning Chai, Benjamin Caine, V. Vasudevan, Wei Han, J. Ngiam, Hang Zhao, A. Timofeev, S. Ettinger, Maxim Krivokon, A. Gao, Aditya Joshi, Y. Zhang, Jon Shlens, Zhi-Feng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2443–2451, 2020. 1
[38] Geoffrey R. Taylor, Andrew J. Chosak, and P. C. Brewer. Ovvv: Using virtual worlds to design and evaluate surveillance systems. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 2
[39] Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Binh Yang, Wenyuan Zeng, J. Tu, and R. Urtasun. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In *ECCV*, 2020. 2
[40] Xinshuo Weng, Yunze Man, Dazhi Cheng, Jinhyung Park, Matthew O'Toole, and Kris Kitani. All-In-One Drive: A Large-Scale Comprehensive Perception Dataset with High-Density Long-Range Point Clouds. *arXiv*, 2020. 2
[41] Zhongyang Xiao, Zhaobin Mo, Kun Jiang, and Diange Yang. Multimedia fusion at semantic level in vehicle cooperactive perception. In *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6, 2018. 2
[42] S. Yogamani, C. Hughes, J. Horgan, Ganesh Sistu, P. Varley, D. O'Dea, Michal Uriar, Stefan Milz, Martin Simon, Karl Amende, Christian Witt, Hazem Rashed, Sumanth Chennupati, Sanjaya Nayak, Saquib Mansoor, Xavier Perroton, and P. Perez. Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9307–9317, 2019. 1, 2