Mixed Signals: A Diverse Point Cloud Dataset for Heterogeneous LiDAR V2X Collaboration

Katie Z Luo[*,1] **Minh-Quan Dao**[*,2,†] **Zhenzhen Liu**[*,1] **Mark Campbell**[1] **Wei-Lun Chao**[4] **Kilian Q Weinberger**[1]
Ezio Malis[2] **Vincent Frémont**[5] **Bharath Hariharan**[1] **Mao Shan**[3] **Stewart Worrall**[3] **Julie Stephany Berrio Perez**[3]
[1]Cornell University [2]Inria [3]University of Sydney [4]The Ohio State University [5]École Centrale de Nantes

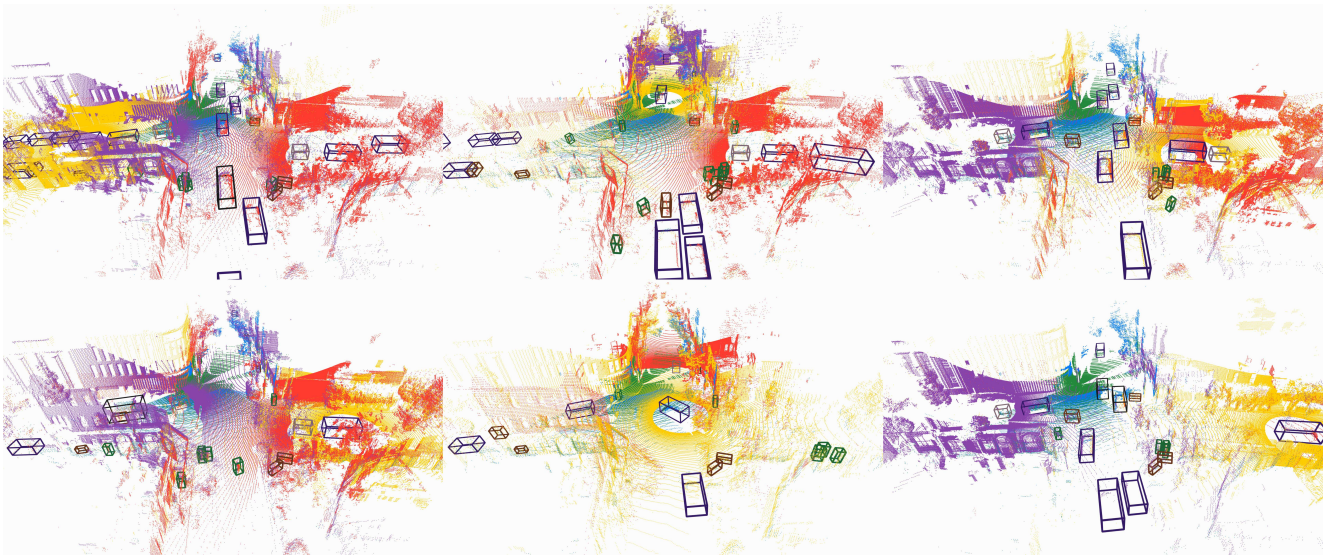


Figure 1. **Six samples of the scenes in the dataset.** The different agent's LiDAR point clouds are colored as follows: electric vehicle-001 (`EV-1`) in purple, electric vehicle-002 (`EV-2`) in red, urban vehicle (`Laser`) in yellow, and the RSU `DOME` and `TOP` LiDARs in green and blue, respectively. We also draw the annotated bounding boxes within the scene. Best viewed in color.

Abstract

Vehicle-to-everything (V2X) collaborative perception has emerged as a promising solution to address the limitations of single-vehicle perception systems. However, existing V2X datasets are limited in scope, diversity, and quality. To address these gaps, we present Mixed Signals, a comprehensive V2X dataset featuring 45.1k point clouds and 240.6k bounding boxes collected from three connected autonomous vehicles (CAVs) equipped with two different configurations of LiDAR sensors, plus a roadside unit with dual LiDARs. Our dataset provides point clouds and bounding box annotations across 10 classes, ensuring reliable data for perception training. We provide detailed statistical analysis on the quality of our dataset and extensively benchmark existing V2X methods on it. Mixed Signals is ***ready-to-use****, with precise alignment and consistent annotations across time and viewpoints. We hope our work advances research in the emerging, impactful field of V2X perception. Dataset details at* `https://mixedsignalsdataset.cs.cornell.edu/`.

[*]Denotes equal contribution.

[‡]This research is funded by the University of Sydney – Cornell University Ignition Grants/ Global Strategic Collaboration Awards.

1. Introduction

In recent years, driver assistance [19, 29] and autonomous driving [1, 47] technologies have advanced significantly, equipping vehicles with promising capabilities in perception [20, 44], planning [13, 15], and control [2, 8]. Most of these developments focus on single autonomous vehicle scenarios. Despite the advancements, such settings still face challenges in complex or unpredictable situations [42]. For instance, important traffic participants can be occluded from view, or sensors can fail unexpectedly. As autonomous vehicle deployment increases, new possibilities emerge to address these issues: multiple vehicles can communicate with each other and nearby infrastructure, enabling each vehicle to reliably detect road users even when its own sensors miss them by leveraging shared information. This approach is commonly referred as **vehicle-to-everything (V2X)** collaborative perception.

While single-vehicle perception datasets are abundant

Dataset	Hetero. Fleet	Location	Driving Side	# Roadside LiDARs	# CAV	# Point Clouds (K)	# 3D Boxes (K)	# Classes	# Vulnerable Classes	Track ID
V2X-Sim [23]	✗	CARLA (Sim.)	Right	1	5	10.0	26.6	1	0	✓
OPV2V [42]	✗		Right	0	2-7	11.4	232.9	1	0	✗
V2X-Set [41]	✗		Right	2-7	2-7	33.0	230.0	1	0	✗
DAIR-V2X-C‡ [45]	✗	China	Right	2	1	39.0	464.0	10	4	✗
V2X-Seq (SPD)‡ [46]	✗	China	Right	2	1	15.0	10.4	10	4	✓
RCooper‡ [12]	✗	China	Right	3	0	30.0	*N/A*	10	3	✗
HoloVIC‡ [26]	✗	China	Right	2	1	100.0	1800	3	2	✓
Open Mars [24]	✗	USA	Right	0	2-3	15.0	0	*N/A*	*N/A*	✗
V2V4Real [43]	Height	USA	Right	0	2	20.0	240.0	5	0	✓
V2X-Real [38]	Height	USA	Right	2	2	33.0	1200.0	10	2	✓
TUMTrafV2X [48]	✗	Germany	Right	2	1	2.0	30.0	8	3	✓
Mixed Signals	Height, Tilt	AUS	Left	2	3	45.1	240.6	10	4	✓

Table 1. **Comparison of Mixed Signals and existing V2X datasets.** To our best knowledge, Mixed Signals is the first dataset to include heterogeneous CAV LiDAR configurations, and also the first one that is collected in a left-hand driving country. It captures complex, real-world traffic scenarios and features a diverse range of traffic participants. Those marked with ‡ are valuable datasets, but are only accessible from certain geographical regions.

across diverse driving conditions [4, 5, 9, 11, 16, 18, 27, 28, 30, 33, 39], real-world V2X datasets remain limited in availability, diversity, and quality. Only a handful of publicly available V2X datasets exist [24, 38, 43, 48], with some of them accessible only within specific geographical regions [12, 26, 45, 46]. These data are collected exclusively from three right-hand traffic locations, overlooking the unique traffic dynamics in left-hand traffic countries which make up about a third of the world [40]. Furthermore, as collaborative perception becomes more widespread, it is valuable for vehicles equipped with different sensor configurations to communicate. However, in prior datasets, the connected autonomous vehicles (CAVs) share identical or very similar LiDAR configurations. Finally, as the V2X setting involves multiple agents and sensors, data collection and alignment present additional challenges. Often times, difficulty with pose estimation and faulty localization systems result in poor alignment (Figure 4). Such inaccuracies can lead to suboptimal performance for detector training [41].

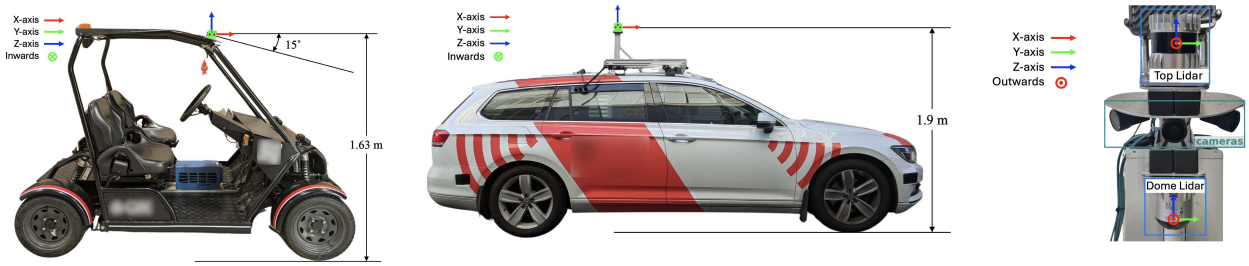
To address these limitations, we introduce the Mixed Signals dataset, designed to support diverse real-world V2X research scenarios with clean, high-quality data. Notably, Mixed Signals is the first V2X dataset that provides heterogeneous CAV LiDAR configurations in both position and orientation, and features a left-handing traffic country, Australia. The dataset includes 45.1k point clouds and 240.6k bounding boxes, collected from three CAVs equipped with two configurations of LiDAR sensors, along with a roadside unit with two LiDARs. It captures a diverse range of traffic participants across 10 different classes, including 4 vulnerable road user categories. Furthermore, compared to existing datasets, Mixed Signals offers significantly more precise alignment and consistent annotations across time and viewpoints. We emphasize that our dataset is ***ready-to-use***; a subset is provided in the supplementary materials, along with the corresponding video visualization showcasing the quality of our collected data and annotations. To summarize, our contributions are:

- We introduce the Mixed Signals dataset, a high quality, large-scale, publicly available V2X dataset created through careful processing and precise annotations.
- To the best of our knowledge, we are the first real-world V2X dataset that encompasses CAV LiDAR configurations that differ in both position and orientation, as well as left-hand traffic scenarios.
- We provide detailed analysis of the dataset's statistics, and conduct comprehensive benchmarking of existing V2X methods across various settings.

2. Related Works

While existing collaborative perception datasets have the same sensor setup for their CAVs, our dataset contains three vehicles with two different sensor configurations, including the height and tilt of LiDAR and the type of vehicle. This difference introduces heterogeneity to our fleet of vehicles, thus making our data more closely resemble the real-world collaboration deployment. To the best of our knowledge, we have the largest fleet of CAVs with the most diverse sensors of any prior works.

Vehicle-to-Everything Communication. One of V2X's objectives is to enhance the perception capabilities of CAVs, facilitating their deployment in urban environments. These areas usually have a high presence of Vulnerable Road Users (VRUs), which are defined as people not inside vehicles [32]. Despite this, VRUs are under represented in prior works. The three synthetic datasets made with CARLA



(a) Electric vehicle (EV). (b) Urban vehicle with a OS1-128 beams LiDAR. (c) Roadside Unit (RSU).

Figure 2. **Vehicles used for data collection.** (a) is a small electric vehicle outfitted with an OS1-128 beams LiDAR system. The LiDAR is mounted at a 15° angle relative to the vehicle's body and stands at a height of 1.63 meters. (b) is an urban vehicle equipped with an OS1-128 beam LiDAR system located at a height of 1.9 meters. (c) is the RSU which consists of two LiDARs: an OS1-64 beam (`TOP`) and an OSDome-128 (`DOME`) LiDAR mounted on a pole at the intersection at a height of 2.5 meters.

[10] and the real-world dataset V2V4Real [43] do not have VRUs. DAIR-V2X-C [45] and its extension V2X-Seq (SPD) [46] provide annotations for 4 VRU classes (pedestrian, bicyclist, tricyclist, and motorcyclist). However, the absence of details on class distribution in their publications make it hard to judge their VRU coverage. Additionally, restricted access to these datasets outside China limits their usability. TUMTrafV2X [48] annotates 3 VRU classes including pedestrian, bicycle, and motorcycle, which together account for only 24.6% of the total annotations. Such underrepresentation causes VRU detection to be overlooked in several collaborative perception studies [22, 36, 41, 42].

Real World Vehicle-to-Everything Datasets. The recent V2X-Real [38] has a large number pedestrian annotations, which is higher than annotations of the class car, and 3 other VRU classes (scooter, motorcycle, and bicycle). A drawback of this dataset for VRU detection evaluation is that its benchmark only accounts for pedestrians. Our dataset contains the highest number of VRU classes, including pedestrian, bicycle, portable personal mobility, and motorcycle. More importantly, these classes account for 50.3% of our dataset's total bounding boxes. Instead of selecting certain VRU classes for benchmarking, we group 4 VRU classes into 2 detection classes as in Section 3.4 to provide a better understanding of how different collaboration methods perform in detecting VRUs. We provide a detailed comparison of our dataset, Mixed Signals, with prior works in Table 1.

3. Mixed Signals Dataset

In this section, we describe the data collection process of the Mixed Signals dataset. We provide a devkit and our full dataset for download on our website: `https://mixedsignalsdataset.cs.cornell.edu/`.

3.1. Hardware

The data collection was carried out using three vehicles and a roadside unit.

Vehicles. The three vehicles included two small electric vehicles (EVs) and one urban vehicle, each equipped with OS1 128-beam LiDARs, as shown in Figure 2. The LiDAR on the urban vehicle is located horizontally with respect to the ground, while for the EV, the LiDAR is tilted downwards 15 degrees. We transformed both EVs' point clouds to have a horizontal reference frame as shown in Figure 2a. Although all the vehicles are equipped with the same type of LiDAR sensor, their configurations differ in terms of sensor position and orientation. This variation introduces additional complexity, creating a domain gap between the data collected from different vehicles.

Figure 3. **Geographical location of the roadside unit.**

Roadside Unit. The roadside unit is equipped with two different LiDAR sensors: an OS-Dome 128-beam for long-range detection and an OS1 64-beam LiDAR for detecting nearby objects. It was located at a fixed geographical position, 2.5 meters above the ground. The intersection where the roadside unit was installed experiences moderate vehicular traffic and features pedestrian crosswalks along with a bike lane that crosses the intersection. This setup allows us to capture diverse agents during data collection. The placement of the roadside unit is illustrated in Figure 3.

3.2. Data Acquisition

The data collection took place at the intersection between Abercrombie Street and Myrtle Street in Sydney, Austria, where the roadside unit is located. The vehicles recorded

LiDAR data for two hours during peak rush hour. Throughout this period, the three vehicles repeatedly passed through the intersection. This allowed them to capture interactions between the vehicles and other agents on the road, such as pedestrians, cyclists, and other vehicles.

3.2.1. Synchronization and Localization

Synchronization and localization are crucial for cross-sensor point cloud alignment. Our dataset employs proven techniques from robotics to achieve precise sensor synchronization and agent localization. The end result is superior point cloud alignment compared to previous V2X datasets (Figure 4). We describe the details below.

Synchronization refers to the temporal alignment of data streams, ensuring that synchronized sensors capture the same events simultaneously within their overlapping fields of view (FOV). We use GPS time to timestamp point clouds captured by our LiDARs at a frequency of 10 Hz. Even if two vehicles are GPS-synchronized, cross-sensor synchronization still needs to be considered. For example, since the LiDAR scans the environment in a rotating fashion, the data collected at different spatial locations are captured at slightly different moments. We defined data samples by setting a time window to match the closest available timestamps from each LiDARs. A maximum timestamp mismatch of 50 milliseconds between point clouds was set to achieve minimal spatial discrepancies. For additional details, refer to Appendix C.1.

Localization, i.e., estimating vehicle position relative to a global reference frame, is one of the most critical tasks for CAV. To overcome inherent problems of Global Navigation Satellite System (GNSS) in urban environments, we use dense and accurate point cloud maps [31] as references for our localization algorithm. Both the vehicles and the roadside units are localized within a common reference frame, referred to as the map_frame, which serves as the origin of our map. The localization algorithm employs a scan-matching technique [3] to estimate the vehicles' poses within this map, achieving a maximum positioning error of 15 cm and a heading error of 0.4 degrees. This allows for consistent spatial alignment between the vehicles and the roadside infrastructure. The vehicles' localization estimates their positions within the map_frame, while the roadside unit is static. We leave details about map construction and usage in Appendix C.2.

3.2.2. Scene Selection

In total, 37 scenes –each consisting of a 30-second snippet– were carefully selected for inclusion in the dataset due to their rich diversity of vehicles, pedestrians, and cyclists. The primary goal was to capture various vehicles and vulnerable road users. These scenes encompass a broad spectrum of interactions, including between different types of vehicles and between vehicles and vulnerable road users. The selected scenes feature intersections of the FOV of the LiDARs of 3 vehicles and the RSU. Among 37 scenes of our dataset, we select 33 scenes for training and 4 scenes of testing. The size of the training set and test set are 9553 and 1164 data samples, respectively. Our selection ensures that there is no temporal overlap between the training set and test set and among scenes of the test set.

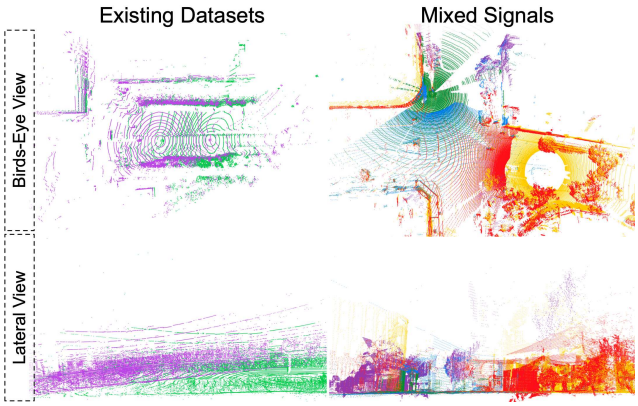


Figure 4. **Localization and synchronization quality of Mixed Signals and existing datasets.** Different colors correspond to different sensors. In the lateral view, existing datasets visually exhibit vertical inconsistencies, where one point cloud is tilted due to localization errors. In contrast, point clouds in Mixed Signals are all accurately aligned.

3.3. Dataset Annotation

The task of 3D object detection for autonomous vehicles requires annotations in the form of 3D bounding boxes, usually parameterized by the center location, three dimensions (length, width, height), and rotation (represented as a quaternion). To generate such annotations for each data sample, we first aggregate the point clouds of every agent in the coordinate of the roadside unit's top (`TOP`) LiDAR to focus the annotators' attention to the intersection of interest. Then, professional annotators from FlipSideAI [34] employ the SegmentsAI [35] annotation tool to label objects and localize them with a 3D bounding box. Classes labeled belong to 10 categories, consisting of: car, truck, pedestrians, bus, electric vehicle, trailer, motorcycle/bike, bicycle, portable personal mobility, and emergency vehicle. Figure 1 depicts the annotations applied to the dataset, where each object is enclosed within a cuboid.

Annotations. Our annotation process involved cycles of monitoring, reviewing, and adjusting labels to meet defined quality objectives. This allows Mixed Signals dataset to extend the quality of the pioneering datasets in the field, which are generally labeled by lay annotators, as shown in Figure 5. Here, we reproject the bounding box of a vehicle, as observed from other sensors, back onto its coordinate frame

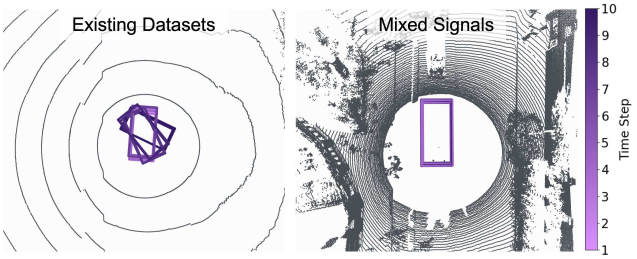


Figure 5. **Label quality of Mixed Signals and existing datasets.** We aggregate labels of an object across a entire snippet. Labels in Mixed Signals are consistent across time steps and viewpoints.

to visualize label consistency. Details of the class descriptions and labeling instructions are presented in Appendix Sec. B. While agents in our dataset are synchronized at 10 Hz, we sample keyframes at 1 Hz for manual annotation. To obtain annotations in a non-key frame, we linearly interpolate the pose of annotations of its closest preceding and succeeding keyframes based on their timestamp.

Category Labels. The Mixed Signals dataset categories consist of road agents in 10 categories of vehicle types and pedestrians including: Car, Truck, Emergency Vehicle, Bus, Motorcycle, Motorized Bike, Portable Personal Mobility Vehicle, (traditional) Bicycle, Electric Vehicle, Trailer, and Pedestrian. Detailed definition of each category can be found in the appendix.

3.4. Dataset Analysis

Statistics. In our benchmark, we group 10 categories into 3 detection classes according to Table 2. Figure 7 shows the distribution of annotations of three classes with respect to their polar coordinate in the coordinate system of `TOP`. Figure 8 shows the distribution of dimensions and yaw angle of annotations of three classes. Figure 6 shows the number of annotations of each class in the training set and test set. Figure 9 analyzes track lengths in the training and test set. For both splits, most tracks are under 10 seconds. This is due to the dynamic and typical speeds at the intersection environment. A sharp peak at 30 seconds indicates the presence of static objects detected primarily by the RSU for the entire sequence duration. Figure 10 depicts the aggregation of point clouds from 5 agents and ground truth annotations in the coordinate system of `TOP` during a 4-second time span, which amounts to 40 time steps. The consistent pose of static objects and the smooth trajectory of dynamic objects visually demonstrate the quality of our annotation.

4. Proposed Tasks and Benchmarks

Our dataset includes multiple agents and annotations in the form of 3D bounding boxes with track IDs. This enables the development of methods for various collaborative perception tasks, such as object detection, tracking, and motion forecasting. Given the importance of object detection in autonomous driving, we focus on collaborative detection tasks in the main text and report preliminary tracking benchmark results in Appendix B.2.

Detection Class	Annotation Classes
Vehicle	car, truck, emergency vehicle, bus, electric vehicle, trailer
Bike	motorbike, bicycle, portable personal mobility
Pedestrian	pedestrian

Table 2. **Definition of detection classes.** The Mixed Signals dataset includes 10 fine-grained annotation classes for traffic participants, organized into 3 broader detection classes.

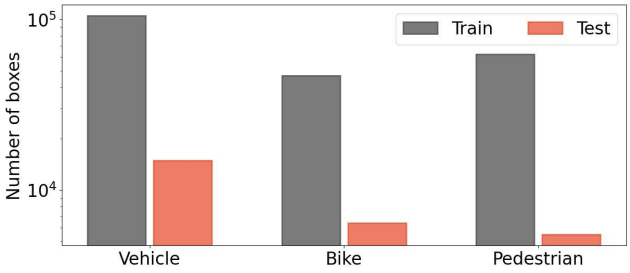


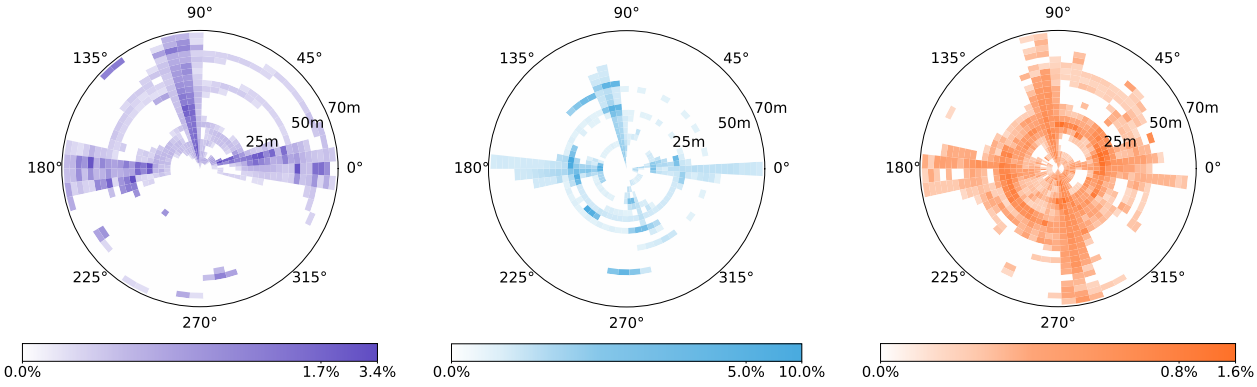
Figure 6. **Number of objects by class.** The y-axis is in log scale.

4.1. Definition of Tasks

We define two tasks that are distinguished by the collaboration setting: *Collaborative Object Detection* and *Single-Vehicle Object Detection enhanced by communication to RSU*, which we describe in the following sections.

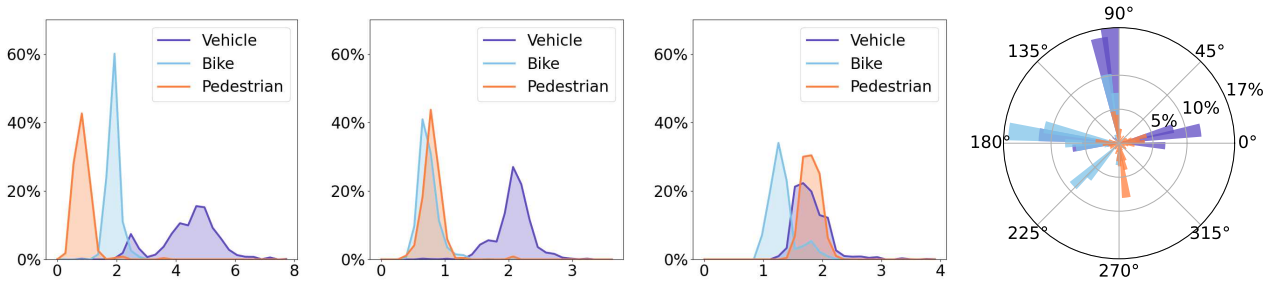
Collaborative Object Detection. This is the classical collaborative object detection task [22, 36], where every connected agent (i.e., vehicles and RSUs) uses a shared model to (i) extract features from their point clouds, (ii) generate messages to send to other agents, and (iii) fuse the features of their point clouds with messages received from others. The goal is to detect every visible object in a region of interest. We define visibility by comparing the number of LiDAR points contained within an object's bounding box to a threshold. In this task, these LiDAR points are sourced from any agents present within the region of interest.

Object Detection Enhanced by Communication to RSU. This task assumes that the RSU model is designed and trained by a different provider than the one responsible for the CAVs' models. In this task, the RSU model is pre-trained in the single-vehicle detection setting to detect objects visible to its LiDARs. After the pre-training process, the RSU model is fixed. CAVs in the proximity of the RSU receive messages from the RSU to enhance their detection



(a) Vehicle (b) Bike (c) Pedestrian

Figure 7. **Distribution of annotated object locations.** Locations are shown in polar coordinates relative to the RSU `TOP` sensor.



(a) Length (m) (b) Width (m) (c) Height (m) (d) Yaw Angle

Figure 8. **Distribution of bounding box dimensions and yaw angles.** Vehicles exhibit a wide range of sizes.

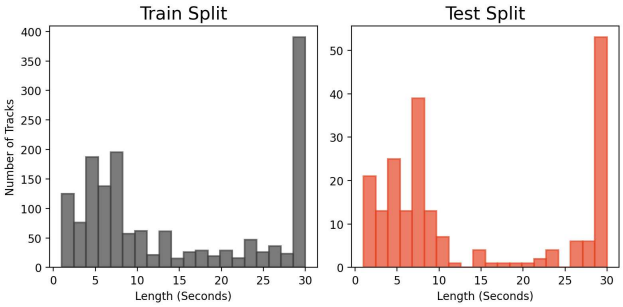


Figure 9. **Distribution of track lengths.** The peak at 30 seconds corresponds to static objects.

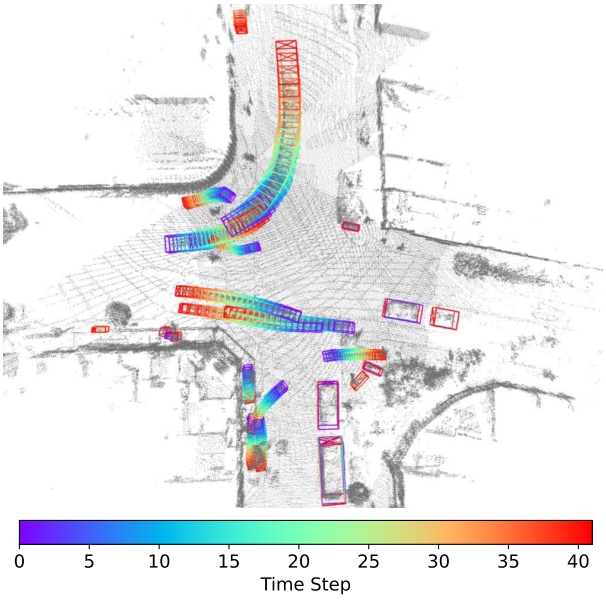


Figure 10. **Visualization of object tracks in Mixed Signals.** Dynamic objects display smooth trajectories, while static objects maintain consistent poses over time, highlighting the high quality of our annotations.

capabilities. The objective is to detect all objects in a region of interest that are visible to either the CAV or the RSU.

The differences between this task and *Collaborative Object Detection* are twofold. First, there is no communication among connected vehicles in this task, making it similar to Vehicle-to-Infrastructure (V2I) detection [38, 45, 48]. Second, instead of having a single model shared among all connected agents like prior works on V2I collaboration, we have one model for the CAVs and another independent model for the RSU. This introduces a different challenge, as the CAV's model must adapt to messages from the RSU, which may contain domain gaps due to differences in model architecture, types of LiDAR, and viewpoints.

	Vehicle AP@		Bike AP@		Pedestrian AP@		Avg. Bandwidth (MB)
	IOU 0.5	IOU 0.7	IOU 0.5	IOU 0.7	IOU 0.3	IOU 0.5	
No Fusion	0.42	0.42	0.19	0.19	0.47	0.41	0.00
Early Fusion (Adapting [43])	0.24	0.24	–	–	–	–	7.79
Early Fusion	0.65	0.65	0.65	0.65	0.74	0.67	7.79
Attentive Fusion [42]	0.82	0.82	0.71	0.71	0.74	0.68	5.26
V2V-Net [36]	0.72	0.72	0.69	0.69	0.42	0.32	4.19
F-Cooper [6]	0.75	0.75	0.68	0.68	0.72	0.65	15.31
V2X-ViT [41]	0.84	0.84	0.71	0.70	0.77	0.70	19.36
V2V-AM [21]	0.83	0.83	0.79	0.79	0.69	0.60	16.78
where2comm [14]	0.77	0.77	0.74	0.74	0.31	0.18	16.78
Laly Fusion [7]	0.61	0.61	0.68	0.68	0.69	0.62	0.11
Late Fusion (Adapting [43])	0.12	0.12	–	–	–	–	0.11
Late Fusion	0.43	0.43	0.56	0.56	0.57	0.48	0.11

Table 3. **Benchmarking results for the *Collaborative Object Detection* task.** All fusion methods outperform the No Fusion baseline, highlighting the advantage of collaborative perception. Each fusion method involves trade-offs between detection performance and communication bandwidth overhead. Models adapted from a premier R.H.S. V2V dataset [43] are shown in gray.

		Vehicle AP@		Bike AP@		Pedestrian AP@	
		IOU 0.5	IOU 0.7	IOU 0.5	IOU 0.7	IOU 0.3	IOU 0.5
EV-1 + RSU	No Fusion (EV-1 only)	0.33	0.33	0.28	0.28	0.37	0.30
	No Fusion (RSU only)	0.22	0.22	0.20	0.19	0.26	0.22
	Attentive Fusion	0.53	0.53	0.60	0.59	0.57	0.45
	V2V-Net	0.46	0.46	0.47	0.47	0.32	0.21
	Late Fusion	0.29	0.29	0.43	0.43	0.52	0.41
EV-2 + RSU	No Fusion (EV-2 only)	0.33	0.33	0.16	0.16	0.08	0.05
	No Fusion (RSU only)	0.24	0.24	0.20	0.19	0.26	0.23
	Attentive Fusion	0.56	0.56	0.56	0.56	0.40	0.27
	V2V-Net	0.52	0.52	0.49	0.48	0.27	0.18
	Late Fusion	0.41	0.41	0.49	0.49	0.43	0.31
Laser + RSU	No Fusion (Laser only)	0.30	0.30	0.32	0.32	0.46	0.44
	No Fusion (RSU only)	0.17	0.17	0.18	0.18	0.25	0.22
	Attentive Fusion	0.71	0.71	0.66	0.65	0.58	0.50
	V2V-Net	0.63	0.63	0.55	0.54	0.37	0.27
	Late Fusion	0.46	0.46	0.52	0.51	0.66	0.57

Table 4. **Benchmarking results for the *Object Detection Enhanced by Communication to RSU* task.** Communication between the agent and RSU generally improves performance compared to single-agent perception. Performance varies across agents with different sensor configurations, suggesting future research opportunities to develop methods that work effectively with diverse sensor types.

4.2. Benchmark

Evaluation Settings. Since the annotations are made in the coordinate system of `TOP`, we define the region of interest for the two detection tasks as the range $[-51.2, 51.2]$ meters along both the x and y axes of this coordinate system. For evaluation, we transform objects detected by each agent into this coordinate system. The visibility threshold is set to 5 points. Since timestamp mismatches and localization errors are inherent in real-world applications and consequently present in our dataset, we do not artificially introduce them into the messages exchanged among connected agents (something that is often done in synthetic datasets [41, 42]). We measure object detection performance using Average Precision (AP). Detected objects are matched with ground truth based on their Intersection over Union (IoU)

in the bird's-eye view plane. A detection and a ground truth object are considered a match if their IoU exceeds thresholds of 0.3, 0.5, or 0.7. In addition to AP, we measure the bandwidth consumption of each collaborative method to gauge their practicality. The total bandwidth consumption is calculated by multiplying the number of agents in the collaboration network by the size of the message each agent sends. While the number of agents is not dependent on the collaboration method of choice, the message size is. We report the bandwidth consumption by averaging the size of the messages that agents send, measured in Megabytes (MB). While some intermediate collaboration methods [22, 36, 41] employ specialized compressing algorithms to reduce the message size, other methods [7, 25, 42] do not. For fair comparison, we report uncompressed sizes.

Methods. Our benchmark covers three conventional collaboration frameworks, namely Early fusion, Intermediate fusion [6, 14, 21, 36, 41, 42], and Late fusion, and the recent *Laly* fusion [7]. We detail the benchmarking methodology specifics in the appendix.

4.3. Results

4.3.1. Collaborative Object Detection

We show the benchmark of the *Collaborative Object Detection* task in Table 3. The results in this table clearly demonstrate the advantage of collaboration perception over single-agent perception, as all fusion methods largely outperform No Fusion on every class. The comparison of three conventional fusion methods, including Early, Intermediate, and Late, shows that a higher precision is attained at the cost of a larger bandwidth consumption. In contrast, *Laly* fusion achieves comparable precision on Bike and Pedestrian compared to Early Fusion and Intermediate Fusion while consuming an order magnitude less bandwidth. The high performance at less bandwidth of *Laly* fusion coupled with its simplicity make this method a strong candidate for real-world deployment. However, we note that there is still ample room for improvement, particularly among the VRUs, suggesting the need for future algorithm design.

Domain Gap from prior R.H.T. Datasets. To illustrate the domain gap covered by Mixed Signals, we directly adapt an early-fusion and a late-fusion model trained on the right-hand traffic (R.H.T.) dataset, V2V4Real [43], onto our dataset (grayed-out rows in Table 3). Performance degraded significantly, with vehicle headings predicted incorrectly, indicating a learned prior from traffic flow (Figure 11). This highlights a substantial domain gap due to left-hand traffic and sensor modality differences, underscoring Mixed Signals' unique contribution to the V2X perception landscape.

4.3.2. Detection via Communication to RSU

Table 4 presents the performance of different fusion methods on *Object Detection Enhanced by Communication to RSU* task. In this setting, detector training is more challenging, as each vehicle-centric detector must adapt to a frozen RSU detector. Nevertheless, results show that communication with RSU is still advantageous, as evidenced by the substantial performance improvement over the No Fusion baselines. Furthermore, the performance of the Laser car is better than the performance of the two EVs. This is because the LiDAR of the Laser car has a 360-degree coverage of its surroundings. On the other hand, the tilted angle of the LiDAR on the two EVs makes the region behind them unobservable. The LiDARs on the two EVs do not capture the intensity information, resulting in a domain gap between their features and those from the RSU. These observations point to future research directions for developing methods that could work well with diverse sensor configurations.

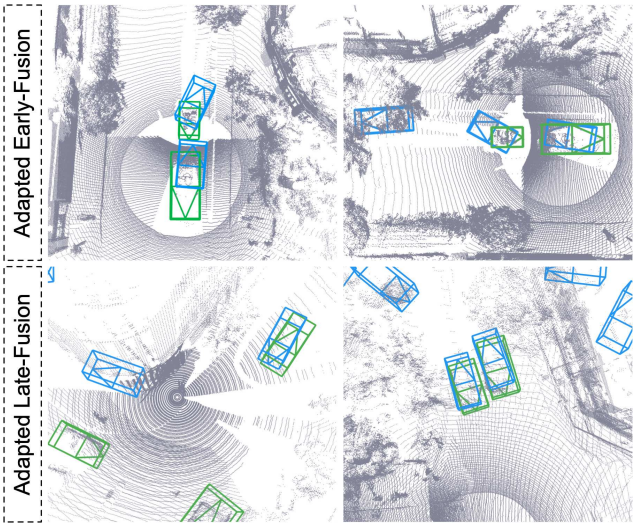


Figure 11. **Detection visualization** from adapting models trained on V2V4Real [43] into Mixed Signals. Ground truth bounding boxes are shown in green and predicted detections in blue, with heading indicated by the triangle. Observe that predicted heading directions are often aligned to priors learned in R.H.T. driving.

5. Discussion and Conclusion

Our work presents the Mixed Signals V2X dataset, created through careful data selection, sensor synchronization and localization, and a strong investment in high quality annotations. To the best of our knowledge, our dataset is the first to support heterogeneous sensor configurations with varying positions and orientations, collected in an out-of-domain left-hand traffic country, Australia, providing a diverse dataset addition to the field. We hope that the release of our dataset will facilitate research into complex and realistic settings for V2X perception. Future directions of research include studying communication protocols that ensure both fast transmission and directed communication that targets salient information.

Acknowledgement

We thank Runsheng Xu and Hao Xiang for their insightful discussions and support throughout the early stage of this project. This research is funded by University of Sydney – Cornell University Ignition Grants/Global Strategic Collaboration Awards, National Science Foundation (IIS-2107161), and the New York Presbyterian Hospital. Minh-Quan Dao is funded by ANNAPOLIS project managed by the French National Agency for Research (ANR-21-CE22-0014), and Katie Luo by AAUW American Dissertation Fellowship.

References

[1] National Highway Traffic Safety Administration. Research on connected vehicle technology. In *online*. https://www.nhtsa.gov/sites/nhtsa.gov/files/2024-02/research-connected-vehicle-technology-report-to-Congress-021524.pdf, accessed 29/09/2024. 1

[2] Alexander Amini, Igor Gilitschenski, Jacob Phillips, Julia Moseyko, Rohan Banerjee, Sertac Karaman, and Daniela Rus. Learning robust control policies for end-to-end autonomous driving from data-driven simulation. *IEEE Robotics and Automation Letters*, 5(2):1143–1150, 2020. 1

[3] P. Biber and W. Strasser. The normal distributions transform: a new approach to laser scan matching. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, pages 2743–2748 vol.3, 2003. 4

[4] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020. 2

[5] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8748–8757, 2019. 2

[6] Qi Chen, Xu Ma, Sihai Tang, Jingda Guo, Qing Yang, and Song Fu. F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds. In *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, pages 88–100, 2019. 7, 8

[7] Minh-Quan Dao, Julie Stephany Berrio, Vincent Frémont, Mao Shan, Elwan Héry, and Stewart Worrall. Practical collaborative perception: A framework for asynchronous and multi-agent 3d object detection. *IEEE Transactions on Intelligent Transportation Systems*, 2024. 7, 8, 1

[8] Xuan Di and Rongye Shi. A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to ai-guided driving policy learning. *Transportation research part C: emerging technologies*, 125:103008, 2021. 1

[9] Carlos A Diaz-Ruiz, Youya Xia, Yurong You, Jose Nino, Junan Chen, Josephine Monica, Xiangyu Chen, Katie Luo, Yan Wang, Marc Emond, et al. Ithaca365: Dataset and driving perception under repeated and challenging weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21383–21392, 2022. 2

[10] 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. 3

[11] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 2

[12] Ruiyang Hao, Siqi Fan, Yingru Dai, Zhenlin Zhang, Chenxi Li, Yuntian Wang, Haibao Yu, Wenxian Yang, Yuan Jirui, and Zaiqing Nie. Rcooper: A real-world large-scale dataset for roadside cooperative perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22347–22357, 2024. 2

[13] Shengchao Hu, Li Chen, Penghao Wu, Hongyang Li, Junchi Yan, and Dacheng Tao. St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning. In *European Conference on Computer Vision*, pages 533–549. Springer, 2022. 1

[14] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems*, 35:4874–4886, 2022. 7, 8

[15] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17853–17862, 2023. 1

[16] Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, and Ruigang Yang. The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 954–960, 2018. 2

[17] Minwoo Jung, Wooseong Yang, Dongjae Lee, Hyeonjae Gil, Giseop Kim, and Ayoung Kim. Helipr: Heterogeneous lidar dataset for inter-lidar place recognition under spatial and temporal variations, 2023. 3

[18] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang, and V. Shet. Lyft level 5 av dataset 2019. urlhttps://level5.lyft.com/dataset/, 2019. 2

[19] Muhammad Qasim Khan and Sukhan Lee. A comprehensive survey of driving monitoring and assistance systems. *Sensors*, 19(11):2574, 2019. 1

[20] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12697–12705, 2019. 1

[21] Jinlong Li, Runsheng Xu, Xinyu Liu, Jin Ma, Zicheng Chi, Jiaqi Ma, and Hongkai Yu. Learning for vehicle-to-vehicle

cooperative perception under lossy communication. *IEEE Transactions on Intelligent Vehicles*, 8(4):2650–2660, 2023. 7, 8

[22] Yiming Li, Shunli Ren, Pengxiang Wu, Siheng Chen, Chen Feng, and Wenjun Zhang. Learning distilled collaboration graph for multi-agent perception. *Advances in Neural Information Processing Systems*, 34:29541–29552, 2021. 3, 5, 8

[23] Yiming Li, Dekun Ma, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng. V2x-sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. *IEEE Robotics and Automation Letters*, 7(4):10914–10921, 2022. 2

[24] Yiming Li, Zhiheng Li, Nuo Chen, Moonjun Gong, Zonglin Lyu, Zehong Wang, Peili Jiang, and Chen Feng. Multiagent multitraversal multimodal self-driving: Open mars dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22041–22051, 2024. 2

[25] Yifan Lu, Quanhao Li, Baoan Liu, Mehrdad Dianati, Chen Feng, Siheng Chen, and Yanfeng Wang. Robust collaborative 3d object detection in presence of pose errors. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4812–4818. IEEE, 2023. 8

[26] Cong Ma, Lei Qiao, Chengkai Zhu, Kai Liu, Zelong Kong, Qing Li, Xueqi Zhou, Yuheng Kan, and Wei Wu. Holovic:large-scale dataset and benchmark for multi-sensor holographic intersection and vehicle-infrastructure cooperative. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, page 22129–22138. IEEE, 2024. 2

[27] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017. 2

[28] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Xiaodan Liang, Yamin Li, Chaoqiang Ye, Wei Zhang, Zhenguo Li, Jie Yu, Chunjing Xu, et al. One million scenes for autonomous driving: Once dataset. 2021. 2

[29] Jaswanth Nidamanuri, Chinmayi Nibhanupudi, Rolf Assfalg, and Hrishikesh Venkataraman. A progressive review: Emerging technologies for adas driven solutions. *IEEE Transactions on Intelligent Vehicles*, 7(2):326–341, 2021. 1

[30] Matthew Pitropov, Danson Evan Garcia, Jason Rebello, Michael Smart, Carlos Wang, Krzysztof Czarnecki, and Steven Waslander. Canadian adverse driving conditions dataset. *The International Journal of Robotics Research*, 40 (4-5):681–690, 2021. 2

[31] Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Rus Daniela. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5135–5142. IEEE, 2020. 4, 3

[32] National Road Safety Strategy. Who are vulnerable road users? https://www.roadsafety.gov.au/nrss/fact-sheets/vulnerable-road-users, 2024. 2

[33] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020. 2

[34] Flipside.ai Development Team. Flipside.ai: data labeling for computer vision. https://flipside.ai/, 2024. 4

[35] Segments.ai Development Team. Segments.ai: Multi-sensor data labeling platform for robotics and autonomous vehicles. https://segments.ai/, 2024. 4

[36] Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Bin Yang, Wenyuan Zeng, and Raquel Urtasun. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 605–621. Springer, 2020. 3, 5, 7, 8

[37] Xinshuo Weng, Jianren Wang, David Held, and Kris Kitani. 3D Multi-Object Tracking: A Baseline and New Evaluation Metrics. *IROS*, 2020. 2

[38] Hao Xiang, Zhaoliang Zheng, Xin Xia, Runsheng Xu, Letian Gao, Zewei Zhou, Xu Han, Xinkai Ji, Mingxi Li, Zonglin Meng, et al. V2x-real: a largs-scale dataset for vehicle-to-everything cooperative perception. *arXiv preprint arXiv:2403.16034*, 2024. 2, 3, 6

[39] Pengchuan Xiao, Zhenlei Shao, Steven Hao, Zishuo Zhang, Xiaolin Chai, Judy Jiao, Zesong Li, Jian Wu, Kai Sun, Kun Jiang, et al. Pandaset: Advanced sensor suite dataset for autonomous driving. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 3095–3101. IEEE, 2021. 2

[40] Jiawei Xu, Kun Guo, Xiaoqin Zhang, and Poly ZH Sun. Left gaze bias between lht and rht: a recommendation strategy to mitigate human errors in left-and right-hand driving. *IEEE Transactions on Intelligent Vehicles*, 2023. 2

[41] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European conference on computer vision*, pages 107–124. Springer, 2022. 2, 3, 7, 8

[42] 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, 2, 3, 7, 8

[43] Runsheng Xu, Xin Xia, Jinlong Li, Hanzhao Li, Shuo Zhang, Zhengzhong Tu, Zonglin Meng, Hao Xiang, Xiaoyu Dong, Rui Song, et al. V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13712–13722, 2023. 2, 3, 7, 8

[44] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11784–11793, 2021. 1

[45] Haibao Yu, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, Hanyu Li, Xing Hu, Jirui

Yuan, et al. Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21361–21370, 2022. 2, 3, 6

[46] Haibao Yu, Wenxian Yang, Hongzhi Ruan, Zhenwei Yang, Yingjuan Tang, Xu Gao, Xin Hao, Yifeng Shi, Yifeng Pan, Ning Sun, et al. V2x-seq: A large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5486–5495, 2023. 2, 3

[47] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443–58469, 2020. 1

[48] Walter Zimmer, Gerhard Arya Wardana, Suren Sritharan, Xingcheng Zhou, Rui Song, and Alois C Knoll. Tumtraf v2x cooperative perception dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22668–22677, 2024. 2, 3, 6

Mixed Signals: A Diverse Point Cloud Dataset for Heterogeneous LiDAR V2X Collaboration

Appendix

In this appendix material, we include: 1) extra details about the Mixed Signals dataset and the provided code devkit, 2) annotation details and instructions given to annotators, and 3) additional sensor details. We include an additional dataset teaser video in the dataset website* that we encourage readers to watch.

A. Data and Devkit

Please see https://mixedsignalsdataset.cs.cornell.edu/ for the dataset download instructions and the provided devkit. Below, we add a brief description of the devkit and visualize a dataset sample.

A.1. Devkit Description

We provide a separate devkit and additionally integrate our dataset into the framework OpenCOOD [42], which offers the implementation of various state-of-the-art collaborative perception methods. As OpenCOOD only provides single-class models, we adapt its implementation of Early, Intermediate, and Late Fusion models to detect three classes, including vehicles, bikes, and pedestrians. We added detection heads of 1-by-1 convolution layers to existing architectures to achieve this. In addition, we add the recent *Laly* fusion [7] to this framework. Every model in our benchmark uses PointPillar [20] as the backbone. Interested readers can refer to our devkit† and code release‡ and extended OpenCOOD integration for further details on architectures and training settings.

A.2. Sample Data

Figure A1 shows an example of the collected data, where the points are colour-coded to represent the different LiDARs. The dataset aims to replicate realistic urban scenarios that reflect the complexities of real-world implementations by using multiple vehicles with diverse sensor configurations and a roadside unit. Real-world deployments of autonomous vehicles on streets incorporate LiDARs, which are becoming more affordable. Roadside infrastructure, such as roadside units, is also gaining popularity for traffic monitoring and data analytics, now often equipped with LiDAR, traffic light timing information, and communication systems to enhance robustness and applicability.

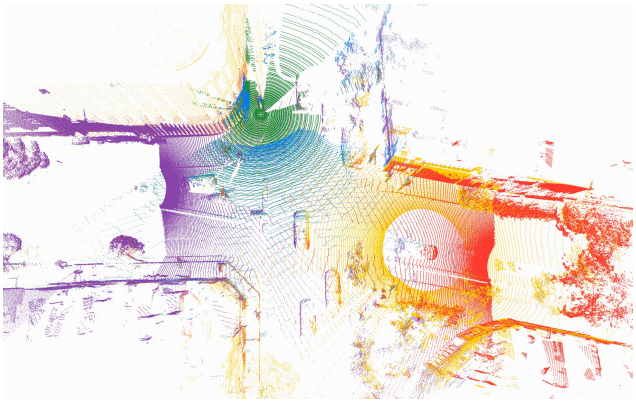


Figure A1. **Top-down view of the data collected at the location.** LiDAR point clouds are colored by the vehicle and RSU that collected them, consisting of the 3 vehicle agents (red, yellow, and purple) and the Top and Dome LiDAR sensors of the RSU (green, blue). Best viewed in colour.

Our dataset consists of LiDAR point clouds, which offer the advantage of not capturing identifiable information like faces or license plates, thus preserving data privacy. This contrasts with camera images, which often require post-processing to anonymize sensitive details, potentially affecting data quality. Our dataset includes tracking IDs for each bounding box, and this information will be released alongside this paper. Benchmarks will be made available at a later date.

Intensity Distributions. Figure A2 shows LiDAR intensity distributions from RSU `TOP`, `DOME`, and `Laser` car sensors. `DOME` and `TOP` sensors record higher intensities because there is a large number of static objects (e.g., buildings, traffic lights) near them. In contrast, the `Laser` car sensor presents a smoother decline in intensity values because of its location on the vehicle, which allows the detection of objects at greater distances. EV-1 and EV-2 sensors do not capture intensity readings. Therefore, a uniform approach to utilizing intensity values across all agent models would be inadequate.

B. Annotation Instructions

We provide the instructions given to the Segments.ai§ annotators in the attached material, titled "*Spec Document - Multi-sensor labeling*" at the bottom of the appendix. We selected to invest in the quality of the annotations, applying

*https://sites.coecis.cornell.edu/mixedsignals/#introvid

†https://github.com/quan-dao/mixed-signals-devkit

‡https://github.com/acfr/Mixed-Signals-Dataset

§https://segments.ai/

Category	Definition
Car	Includes passenger vehicles such as sedans, hatchbacks, SUVs, and coupes that are designed primarily for the transportation of passengers.
Truck	Encompasses larger vehicles primarily used for transporting goods and materials. This category includes pickup trucks, delivery trucks, and heavy-duty trucks.
Emergency Vehicle	Vehicles designated for emergency response, including ambulances, fire trucks, police cars, and other vehicles equipped with sirens and emergency lights.
Bus	Large motor vehicles designed to carry numerous passengers. Buses include city transit buses, school buses, and intercity coaches. They usually have designated routes and schedules.
Motorcycle Motorized Bike	Two-wheeled motor vehicles, including motorcycles and motorized bikes. This category also covers scooters and mopeds.
Portable Personal Mobility Vehicle	Small, lightweight vehicles designed for personal mobility, including electric scooters, hoverboards, and segways.
Bicycle	Human-powered, pedal-driven vehicles with two wheels. Bicycles include standard bikes, mountain bikes, and road bikes. This category include motorized bicycles or electric bikes.
Electric Vehicle	Refers to small, golf car-like vehicles used for data collection purposes.
Trailer	Non-motorized vehicles designed to be towed by a motor vehicle.
Pedestrian	Individuals traveling on foot. This category includes people walking or running.

Table A1. **Definitions of the annotation classes.**

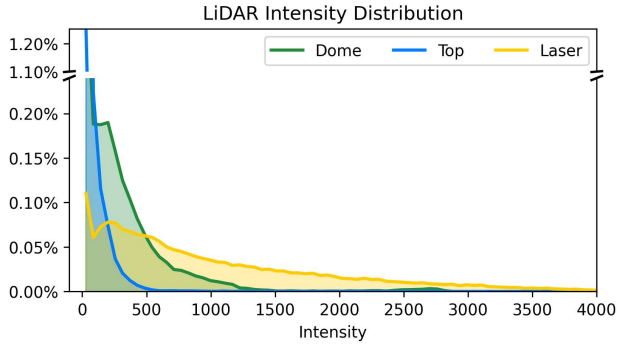


Figure A2. **Distribution of LiDAR intensities from RSU `TOP`, `DOME`, and `Laser` car sensors.** Each sensor shows different intensity ranges and distributions. `EV-1` and `EV-2` LiDAR sensors do not have intensity readings.

rigorous quality control measures to guarantee accurate and consistent labeled data, minimizing errors, and maintaining high standards.

B.1. Definitions of the Annotation Classes

The Mixed Signals dataset categories road agents in different vehicle types and pedestrians. Categories such as "Car" and "Truck" encompass common passenger and large transport vehicles, while "Emergency Vehicle" covers ambulances, fire trucks, and police cars, highlighting their importance in urban scenarios. "Bus" labels are designated for large passenger vehicles typically used in public transportation. The dataset also distinguishes between "Motorcycle" and "Motorized Bike," and "Portable Personal Mobility Vehicle," which includes modern personal transport devices like electric scooters and hoverboards. Traditional "Bicycle" labels account for both standard and electric bikes. Labels for "Electric Vehicle" and "Trailer" ensure that smaller, often data-collection vehicles and towable units are accurately represented. Finally, we labeled humans as "Pedestrians". In Table A1, we provide the definitions of the 10 fine-grained annotation classes in the Mixed Signals dataset. The breakdown of the fine-grain classes into the benchmarked classes can be found in the main text.

B.2. Track Annotations and Multi-agent Tracking

We benchmark the performance of the planned tracking task for our dataset. The Mixed Signals dataset has labels for track ID's, as seen in Figure 10 of the main text. We hope to include and benchmark tracking methods as an additional task which is supported by our dataset. We report some initial benchmarking results on the AB3DMOT tracking method [37] in Table A2. For further details about track labels, please explore the data itself; a distribution of the tracks are in Figure 9 of the main text.

B.3. Annotation Details

The annotation process for this multi-sensor dataset involves handling joint scenes and synchronization discrep-

Category	sAMOTA	AMOTA	AMOTP
Vehicle	89.6	43.1	63.3
Pedestrian	76.6	32.6	42.8

Table A2. Tracking performance for AB3DMOT with V2X-ViT detections on the Mixed Signals validation split.

ancies between sensors. Due to time synchronization, fast-moving objects might appear slightly offset across the data collected from different sensors. To address these discrepancies, annotators were instructed to prioritize the roadside unit point cloud for bounding box creation, following a set hierarchy. When there is a mismatch, bounding boxes should be aligned with the point cloud in the following order: roadside unit, EVs, and the urban vehicle. For example, if there is a difference between the roadside unit and the vehicles' point cloud, the bounding box should only be fitted around the roadside unit points. This ensures consistency in object localization across frames despite synchronization lags.

C. Sensor Details

C.1. Hardware and Synchronization Details

Sensor	**Agent**	**Range***	**Channels**	**Vertical FOV**
Ouster OS1-128	Vehicles	170 m	128	45
Ouster OS1-64	RSU	100 m	64	45
Ouster OS Dome	RSU	45 m	128	180

*Based on 80% Lambertian reflectivity in the sensors' official datasheets.

Table A3. **Hardware specifications.**

Synchronization is especially important in dynamic environments, as any introduced time shifts can lead to positional inconsistencies, resulting in multiple detections of the same object. The sensors in our multi-agent system were timestamped using GPS time as a common reference, and sensor details are provided in Table A3. Rotational LiDARs continuously scan the environment in 360 degrees, thus, different portions of the surroundings are captured at slightly different times during a full rotation. When vehicles are in motion, their positions and orientations change dynamically between LiDARs sweeps. The maximum time gap for matching sensor readings between 10 Hz rotational sensors is 50 ms. Since sensors rotate fully in 100 ms, angular positions differ by at most 180 degrees. If the time difference between readings were larger than 50 ms, it would be matched with the next or previous rotation instead. As shown in the original manuscript, precise sensor synchronization, robust multi-agent localization, and clearly defined annotation protocols produced high-quality data association across all sensors.

C.2. Localization

Localization is one of the most critical tasks for CAV, estimating their position relative to a global reference frame. One of the most commonly used sensors for localization is the Global Navigation Satellite System (GNSS). GNSS offers access to a satellite constellation that provides global positioning via triangulation. However, despite its widespread use, GNSS has several drawbacks, particularly in urban environments. Its accuracy can be reduced in urban canyons, where tall buildings block or reflect signals, leading to degraded positioning accuracy. To overcome this problem, we use dense and accurate point cloud maps [31] as references for our localization algorithm.

C.3. Definition of Heterogeneity in Sensor Suite

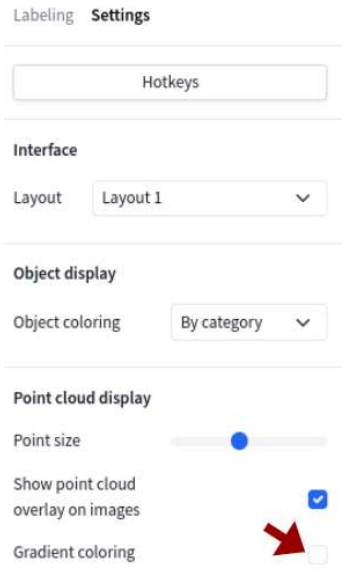
Heterogeneity in our context refers to the variability between LiDAR sensors and platform geometry within a single dataset. Heterogeneity can appear in multiple forms [17]; our dataset represents it in five LiDARs that span three models, each mounted in four configurations. In line with the feedback, Tab. 1 of the original manuscript has been updated accordingly. Our dataset demonstrates a realistic setting where collaborative agents have different LiDAR models and position them in different configurations.

Segments.ai

Spec Document - Multi-sensor labeling

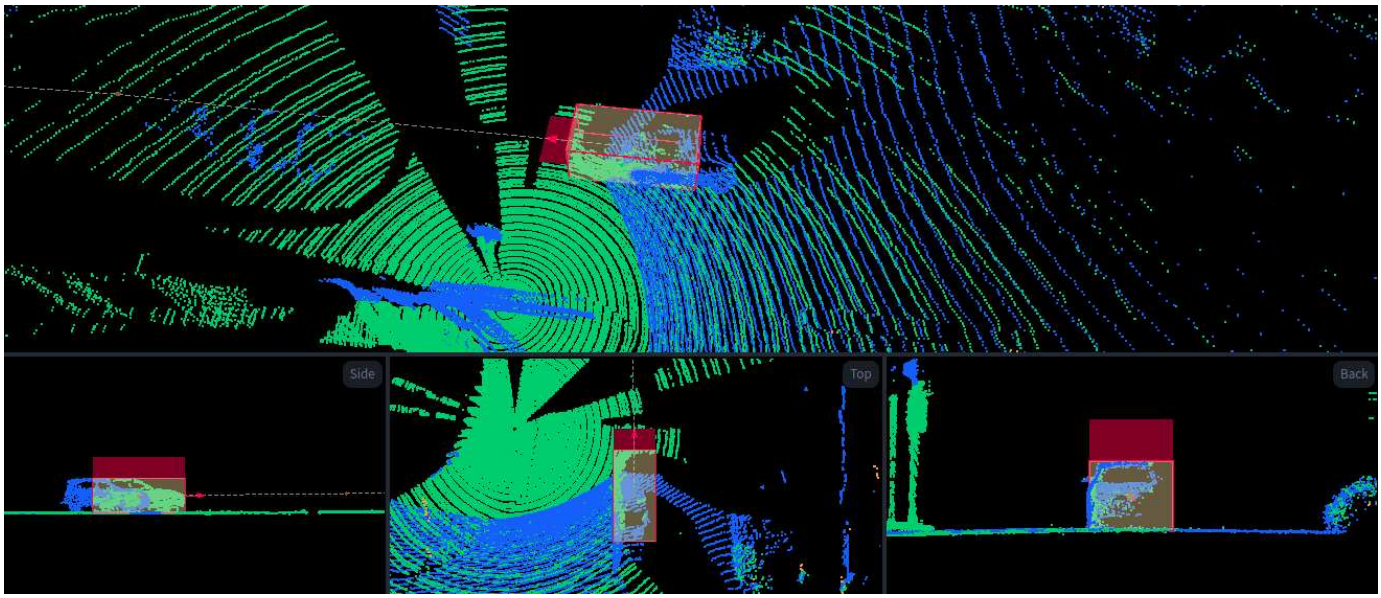
Additional resources

Getting Started
Begin by ensuring the point clouds are ***not*** visualized with the default gradient coloring (and to disable the color-by-gradient). The data is colored by the sensors they are collected from, and should be colored green, blue, orange, yellow, and pink. To do so, follow the instructions:
1. Click on the `Settings` tab of the control panel
2. Scroll to `Point cloud display` and un-check the `Gradient coloring` tab
Everything is correct if there are only 5 colors (green, blue, orange, yellow, and pink) displayed.



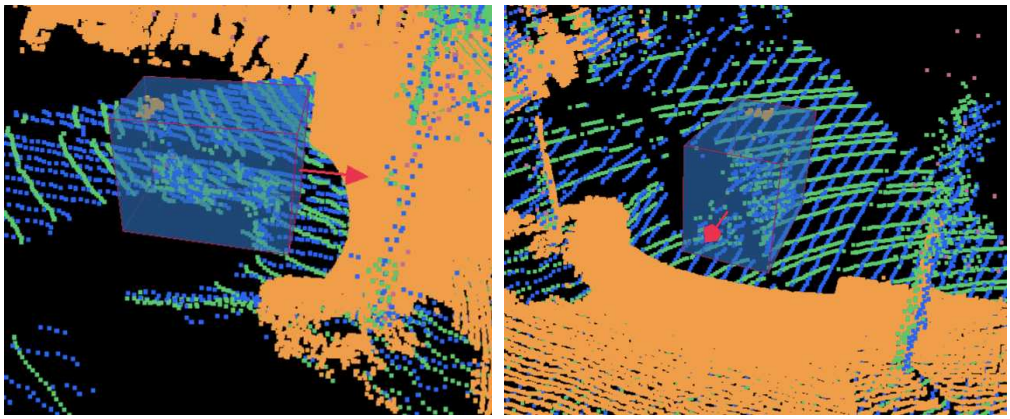
Discrepancy Resolution

Due to time synchronization issues, a single object may have different scans (discrepancy between sensors). This may occur for fast moving vehicles. When this happens, fit a bounding box ***onto a single sensor***, prioritizing the point cloud colors in the order of: green, yellow, orange, pink, and (lastly) blue. For example, if there is a discrepancy between the blue points and green points, ignore the blue points and fit a bounding box only on the green points.



Special Vehicle Cases

We have 2 electric data-collection vehicles (EV). If they show up in other sensors, label them as ***Electric Vehicle (EV)*** category. They look like small golf cars.



Link to external labeling guidelines	N/A (see above)
Link to reference labeled dataset	https://mobility-lab.seas.ucla.edu/v2v4real/#dataannotation
Link to initial unlabeled dataset	TBD

Sensor information

3D sensors

Type of 3D sensors *E.g. Velodyne, ...* *(Add lines where needed)*	How many sensors?	File format? *More info*	Point clouds in local or world reference system?	In case of local reference system, ego poses available?	Number of frames per sequence + Sampling rate?	Typical file size / point quantity per frame?	Comments
One single point cloud composed of multiple Ouster Lidar point clouds	3-6	PCD	Word reference	TBD	40 - 41 frames per sequence (20 s/seq): sampling rate 2 hz		
			TBD	TBD			

Type of task

3D labeling

	Type of labeling	Comments
☑	3D cuboids	3D bounding box on point cloud data
☐	3D segmentation	
☐	3D polylines	
☐	3D polygons	
☐	3D keypoints	

Labeling rules

General labeling rules

Rule	Answer	Default assumption going forward without explicit answer from customer	Comments
Maximum distance to label objects?	Everything	Everything	See ability to set a visual radius for taskers here
Other zones that do not require labeling?	Label everything	Label everything *including e.g. car parks*	
What to do with unclear objects/areas?	Do not label	Do not label	
What to do with reflections?	Do not label	Do not label	
How to cope with groups of individual	Label each individual	Label each individual	

instances?			
Are there any specific rules to adverse weather conditions / nighttime / etc.?	Same as day time	Same as day time	

Specific rules for cuboid labeling of instance classes

Type	Rule	Answer	Default assumption going forward without explicit answer from customer	Comments
Position	Can there be some overlap between cuboid & ground plane, or should cuboids be leveled with the ground plane where applicable?	Slight overlap is OK	Slight overlap is OK	
Rotation & Headi...	Label only yaw, or yaw & pitch & roll?	Yaw, roll & pitch	Yaw, roll & pitch	*Please note that labeling also pitch & roll can reduce throughput with up to a factor 2*
Rotation & Headi...	What should the yaw direction/heading be of an object?	Main face of object	Main face of object	
Rotation & Headi...	What should the yaw direction/heading be of a faceless object, e.g. a cone?	TBD Do Not Label Faceless Categories	TBD	Do not label "cones" or other faceless object categories
Dimensions	Is there a minimum size of cuboid?	No	No	
Dimensions	Should cuboids be labeled with default dimensions depending on their category?	No	No	
Dimensions	Can the dimensions of a cuboid change throughout a sequence?	Yes No for: vehicles and rigid objects	Yes *when needed*	
Dimensions	What should the dimension of a cuboid be based on?	Realistic size, namely ... *[copied over] In order of importance: default dimensions / on the available 3D points elsewhere / on the reference images*	Realistic size, namely ... *In order of importance: default dimensions / on the available 3D points elsewhere / on the reference images*	*Visible 3D points: only using the visible 3D points in the current frame, disregarding the realistic size* *Realistic size: e.g. based on the available 3D points elsewhere in the sequence, on the available reference images or on the provided default dimensions*
Dimensions	How tight should a cuboid be?	Loose As tight as possible, but can be a bit larger than the object itself. Ensure all parts are within the cuboid, see *"Extremities" sections*	Loose	*Very tight: in each frame, there should not be any space between the outer points and the cuboid* *Loose: the cuboid can be a bit larger than the object itself*
Occlusion	Should an object be labeled if it is only visible on 3D sensors and not on 2D sensors?	Yes	Yes	

Occlusion	**Should an object be labeled if it is only visible on 2D sensors and not on 3D sensors?**	Minimum of 10 points	Minimum of 10 points	
Occlusion	**Should an object be labeled with the same track ID if it re-enters a scene or becomes unoccluded again?**	Yes	Yes	
Extremities	**Should fixed extremities / protruding parts be included in the cuboid?**	Custom ... ***Include**: side mirrors, larger protruding parts such as bonnet* ***Exclude**: small protruding parts such as antenna*	Custom ... ***Include**: side mirrors, larger protruding parts such as bonnet* ***Exclude**: small protruding parts such as antenna*	
Extremities	**Should variable/articulating extremities be included in the current cuboid, or be annotated with separate cuboids?**	Custom ... **Include:** rider, people in/on vehicles, non-vehicle objects on trucks **Exclude:** vehicles on trucks	Custom ... ***Include**: rider, cars on trailers, people in/on vehicles* ***Exclude** (only when a relevant category is available): person trolley, car trailer, ...*	
Extremities	**Include relational tracker for variable extremities?**	No	No	*If yes: if a car has object ID 10, and a trailer is attached to this car, the trailer will be annotated separately and receive a relational attribute with value 10*
Issues/exceptions	**Can cuboids overlap?**	Only in reasonable cases	Only in reasonable cases	*Reasonable cases: articulating objects (turning truck with trailer attached, ...), bicycles standing very close to each other, ...*
Issues/exceptions	**What to do with unclear objects?**	Use 'unsure' category	Use 'unsure' category	
Issues/exceptions	**In case of bad calibration, how should the cuboid be fitted? should the cuboid be fitted to the 3D point cloud or rather to the most confident reference images?**	Fit to 3D	Fit to 3D	
Issues/exceptions	**In case of bad ego poses and data drift, what should happen?**	Fit on each frame	Fit on each frame	

Specific rules for 3D polygon/polyline labeling

Type	Rule	Answer	Default assumption going forward without explicit answer from customer	Comments

Temporality	**Can nodes be added/removed from polygons/polylines after initialization?**	No	No	*Note: currently not yet compatible with interpolation (below); on short-term roadmap*
Temporality	**Should interpolation be enabled?**	Yes	Yes	*Note: currently not yet compatible with adding/removing nodes (above); on short-term roadmap*

Categories

Category	Label instances?	To be labeled across all sensors & tasks?	What does it include?	What does it exclude?	Comments
Car	Yes	No, only 3D	• Sedan • SUV • Minivan • All personal/recreational vehicles	• *Speed bumps* • *Raised crosswalks*	
Truck	Yes	No, only 3D			
Emergency Vehicle	Yes	No, only 3D	• Police Cars • Ambulance		
Bus	Yes	No, only 3D			
Motorcycle/Motorized Bike	Yes	No, only 3D	• Motorcycles • Electric/Motorized Bike • Scooters		
Portable Personal Mobility Vehicle	Yes	No, only 3D	• Segway • Moped		
Bicycle	Yes	No, only 3D			
Pedestrian	Yes	No, only 3D	• Pedestrians crossing • Pedestrians with strollers (label a single box for both)		
Electric Vehicle (EV)	Yes	No, only 3D	• Data collection golf car vehicle		
Trailer	Yes	No, only 3D	• Trailer attached to vehicle • Wagon • Other vehicles attached to a pickup-truck/utility vehicles		

Attributes

Frame-level attributes

N/A

Attribute	Type	Options	Description	Comments

	TBD			

Object-level attributes

N/A

Attribute	Type	Applicable categories	Options	Description	Comments
Motion State	Select box	All categories (Car, Truck, Emergency Vehicle, Bus, Motorcycle, Bike, Pedestrian, Portable Vehicle…)	In-Motion / Static	Select if the object is in motion during the frame. Either select that it is in-motion (currently moving) or static (waiting, parked, stopped, etc)	

Edge cases

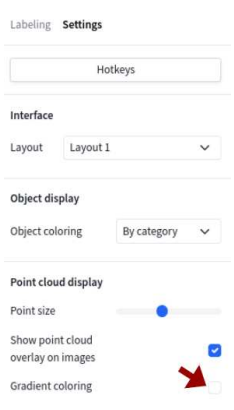
Edge Case	How to handle it	Example
Extreme		

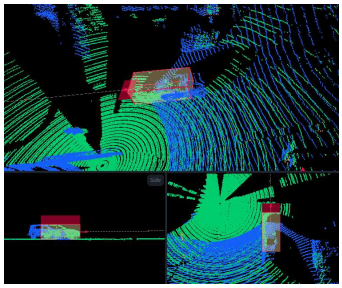
Version history of labeling specification document

N/A

Date	Version	Changes

Additional Q&A

Date	Question status	Question & assumptions	Reference Image/ Link	Answer
03/02/2024	Assumption	I do not see only 4 colors, and instead see the default gradient point color scheme. How do I toggle off the gradient coloring?		Click on the "Settings" tab, and make sure "Gradient coloring" is **not** selected. See *Additional Resources*: Getting Started instructions.

03/02/2024	Assumption	There are time synchronization issues, where fast moving objects are captured differently by different colored point clouds. How do I label these?		Fit a tight bounding box around the point cloud of a single color, prioritizing in the order of: 1. Green 2. Yellow 3. Orange 4. Pink 5. Blue In this example, only draw a bounding box around the green points. See *Additional Resources*: Discrepancy Resolution.