Appendix for V2X-Seq Dataset and Benchmarks

Abstract

In this appendix, we provide further details on the V2X-Seq dataset. Specifically, Section 1 outlines the sensor deployment and intersection layout in detail. Section 2 describes the trajectory collection and mining process. Finally, in Section 3, we have included selected examples from the released dataset along with their visualizations.

1. Sensor Deployment and Intersection Layout

We selected 28 urban traffic intersections in Beijing and deployed sensors at these locations. The layout of these selected intersections can be seen in Figure 1.

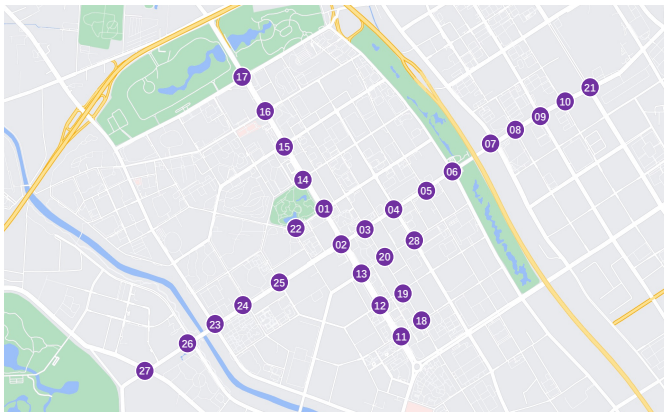


Figure 1. Layout of 28 Deployed Urban Intersections.

We deploy 4∼6 pairs of 300-beam LiDAR and high-resolution cameras for each intersection. These infrastructure sensors can fully cover the intersection areas. We provide the configuration of infrastructure sensor deployment in Fig. 2. We deploy one 40-beam LiDAR and six high-quanlity cameras for the self-driving vehicle. We provide the vehicle sensors deployment in Fig. 3.

2. Trajectory Collecting and Mining

In this section, we explain the process of building the trajectory forecasting dataset, with a particular focus on the 50,000 cooperative-view scenarios.

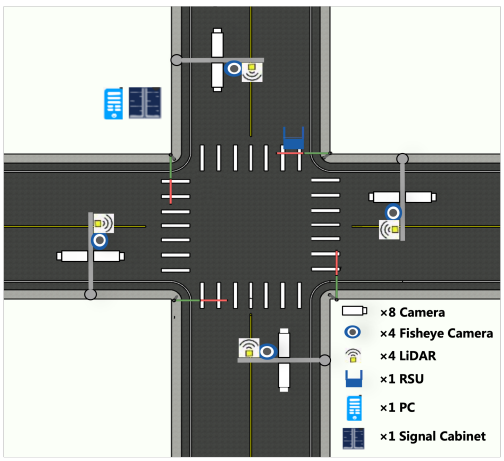


Figure 2. The Infrastructure Sensor Deployment.

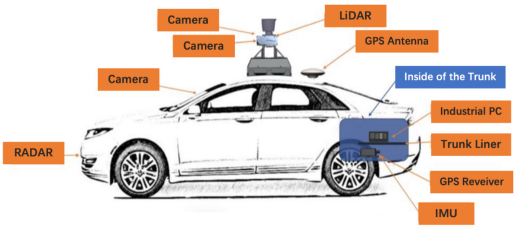


Figure 3. The Vehicle Sensor Deployment.

To collect the sensor data, we drove our self-driving vehicle through sensor-equipped areas, collecting both infrastructure-side and vehicle-side sensor data over a period of 672 hours. This data was saved every three minutes. We then only input this infrastructure images and vehicle sensor data into trained 3D object detection and tracking models to generate trajectory sequences consisting of 3D boxes, each with a class attribute from 8 categories and a unique trajectory ID. These boxes were uniformly transformed from local coordinates into world coordinates, resulting in the trajectory sequences repository. Finally, we mined interesting trajectory segments from the repository to create about 50,000 cooperative-view scenarios.

The trajectory mining process consisted of several steps, including scene fragmentation, scene selection, trajectory fusion, trajectory scoring, filtering, and attribute creation.

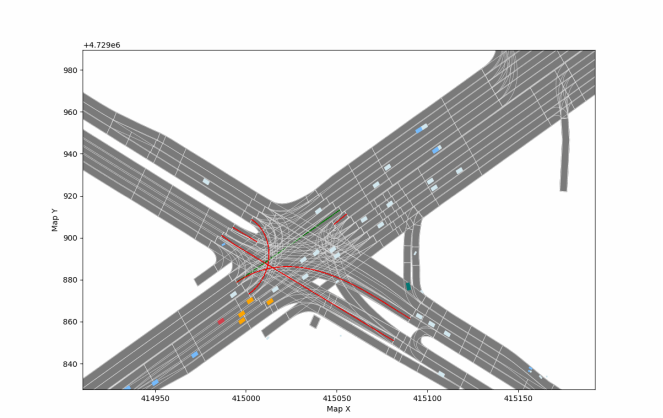
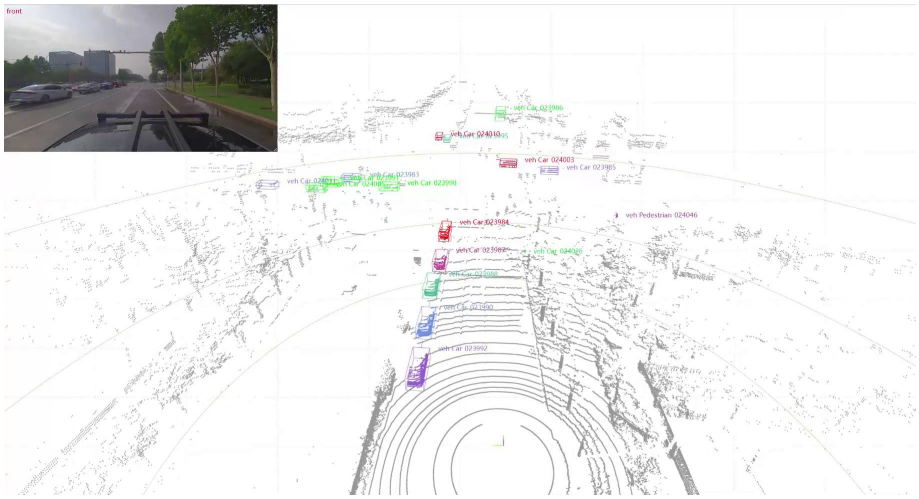


Figure 4. Screenshot of an interesting scenario from the Trajectory Forecasting Dataset with Cooperative-view. The green point denotes the ego vehicle. The red point denotes the first target agent, while the orange points denote the second to fifth target agents. The light blue boxes represent the objects generated from the vehicle side, while the dark blue boxes represent the complementary objects with infrastructure data. The red lines indicate the red traffic light for the located lanes.

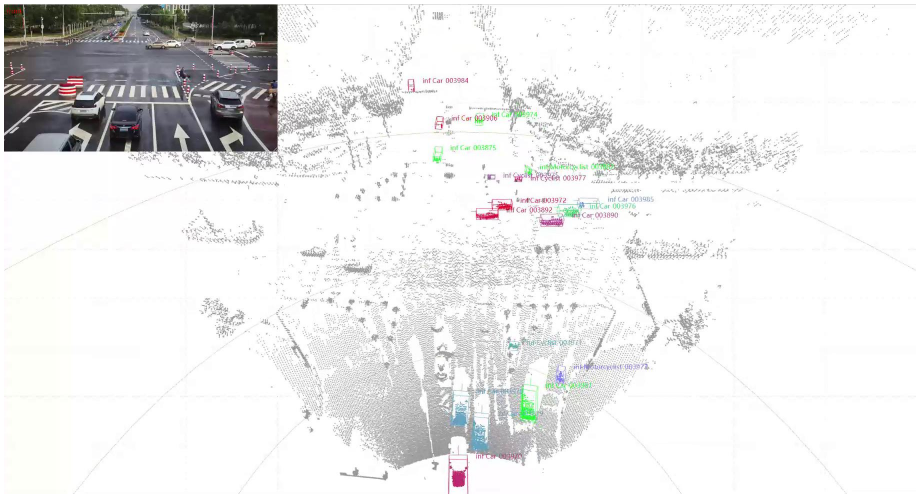
First, we fragmented the infrastructure-side and vehicle-side sequences into 10-second segments, with 5 seconds of overlap between adjacent segments. Next, we selected the infrastructure-side and vehicle-side segments that corresponded to the same equipped intersection to form segment pairs. We then fused the infrastructure-side and vehicle-side trajectories, generating cooperative-view trajectories for each segment pair. We followed the method of generating cooperative annotation for VIC3D tracking, but filtered out any matching with low scores and directly discarded them. Each cooperative trajectory was connected to the origin-view trajectory IDs. Next, we assigned a score to each cooperative trajectory based on various factors such as turning, speeding up, slowing down, lane changing, and completion. Finally, we kept 50,000 sequences with high-score trajectories. We set one to five trajectory in each segment as the target agent type, and these trajectories were located within a certain range of the ego vehicle. Other trajectories were set as other agent types. Additionally, we mined more trajectories from the trajectory repository to increase the size of the infrastructure-side and vehicle-side sequences to about 80,000 each.

3. Visual Example for V2X-Seq

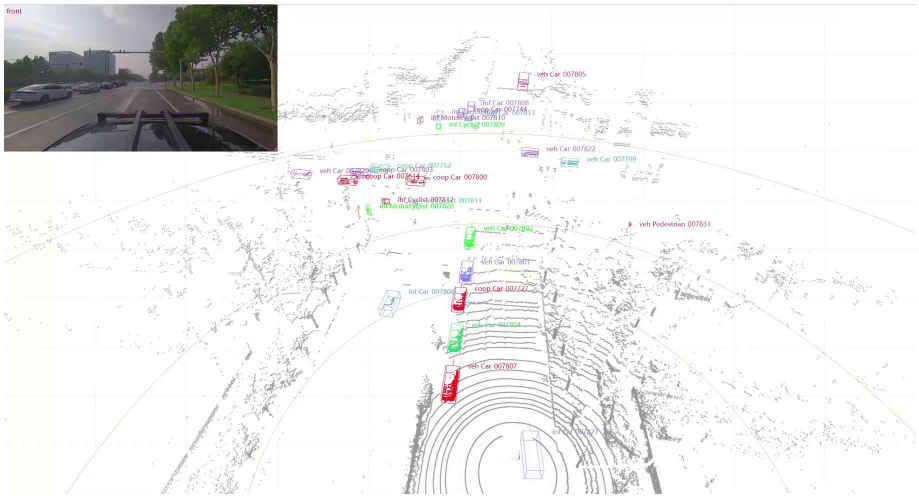
We present a interesting scenario that were mined from the trajectory forecasting dataset in Figure 4. Additionally, we provide a visualization example for the sequential perception dataset in Figure 5.



((a)) Vehicle-view Visualization.



((b)) Infrastructure-view Visualization.



((c)) Cooperative-View Visualization. We visualize the cooperative annotation on vehicle-side images and point clouds. Each 3D box is marked with a label indicating whether it is from the infrastructure-side or vehicle-side annotations. The cooperative view allows us to see more objects and obtain a more comprehensive understanding of the scene.

Figure 5. Visualization Example of the Sequential Perception Dataset. Each 3D box is marked with its attribute and tracking ID.