RCooper: A Real-world Large-scale Dataset for Roadside Cooperative Perception

Supplementary Material

A. Appendix

This supplementary document is organized as follows:

- Additional information and visualization of RCooper is shown in Appendix B.
- The visualization of roadside cooperative perception results is supplemented in Appendix C, offering an illustrative overview of the performance evaluation of our established benchmark.
- Some preliminary camera-based experiment results is shown in Appendix D.

B. Dataset Visualization

More visualizations of the proposed RCooper dataset are shown in Fig 1. These visualizations serve as a valuable supplementary to the discussion in Section 3.1, where two representative traffic scenes (i.e., intersection and corridor) are presented. Each figure within this context comprises an aggregated 3D LiDAR point cloud image and four camera images captured from distinct views. For the corridor scenes, the point clouds generated by two different LiDAR sensors are depicted in red and green, respectively. The infrastructure-side cooperation can effectively extend the sensing range to cover the whole corridor scene. For the intersection scenes, the point clouds generated by multiline LiDARs and MEMES LiDARs are denoted in red and green, respectively. The observations from multiple views serve to mitigate the challenges posed by occlusions in intricate intersection scenarios.

C. Perception Results Visualization

More visualizations of benchmark results are shown in this section.

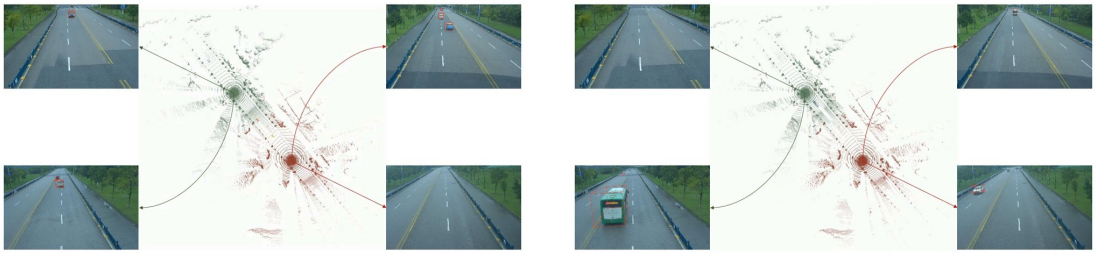
Typical perception benchmark results for corridor scenes (referred to as CoBEVT results) and intersection scenes (referred to as late fusion results) are illustrated in Fig 2 and Fig 3, respectively. Within each figure, alongside the LiDAR points and a camera image, we have delineated the predicted objects with red outlines and the ground truth objects with yellow outlines. For the corridor scenes (in Fig 2) and uncrowded intersection scenes (in Fig 3. a, b, c and d), the perception process proves to be highly effective. For the crowded intersection scenes (in Fig 3. e, f, g and h), the presence of occlusions and the inherent complexity of the scenarios introduce false negatives and false positives into the perception process. These challenges merit further investigation in future research endeavors.

D. Preliminary Camera-based Experiment Results

We provide some preliminary camera-based experiment results in this section. A single-view camera suffers from the inherent blind spot, which can be theoretically mitigated by roadside cooperative perception. Camera-based late-fusion results in both corridor and intersection scenes are reported in Tab. 1. It's worth noting that the data annotation of our datasets is based on point cloud data, and meanwhile the sensing range of camera group is smaller than that of LiDAR group, which means that the label may need filtering on demand considering camera group's view.

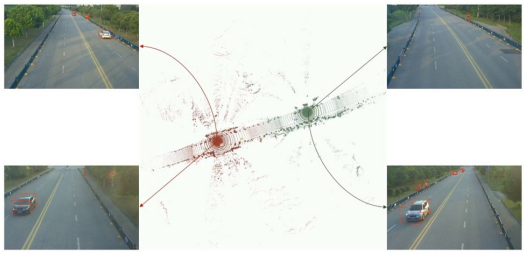
Scenario	3D@0.3	3D@0.5	BEV@0.3	BEV@0.5
Corridor	26.99	3.53	38.70	16.15
Intersection	11.32	2.41	12.19	7.89

Table 1. Camera-based cooperative detection AP metric (%).

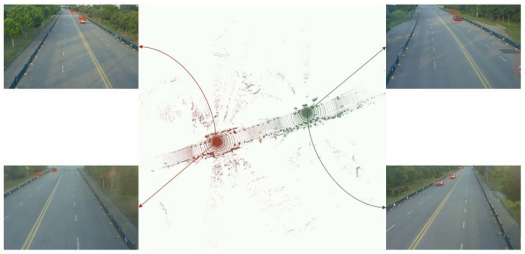


a. Corridor scene example 1

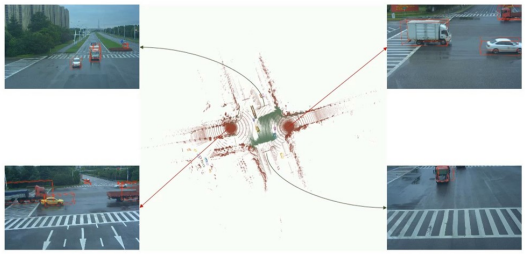
b. Corridor scene example 2



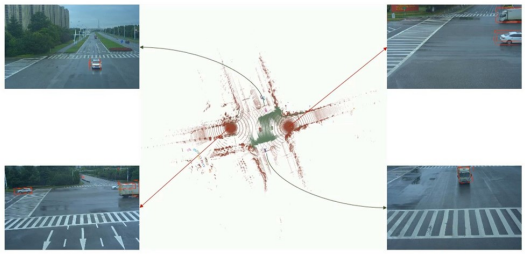
c. Corridor scene example 3



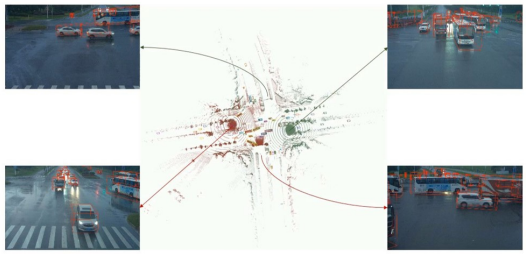
d. Corridor scene example 4



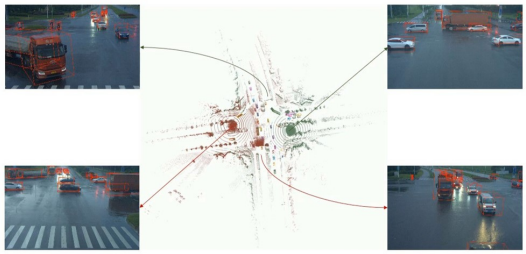
e. Intersection scene example 1



f. Intersection scene example 2

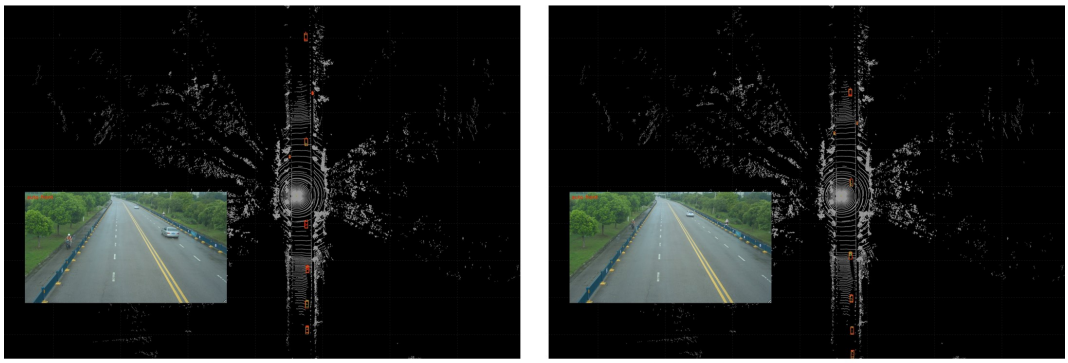


g. Intersection scene example 3



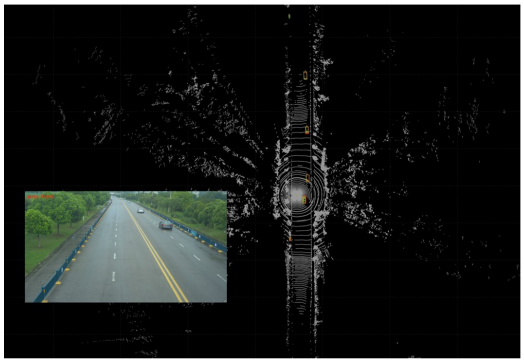
h. Intersection scene example 4

Figure 1. Visualization Examples of the RCooper Dataset.

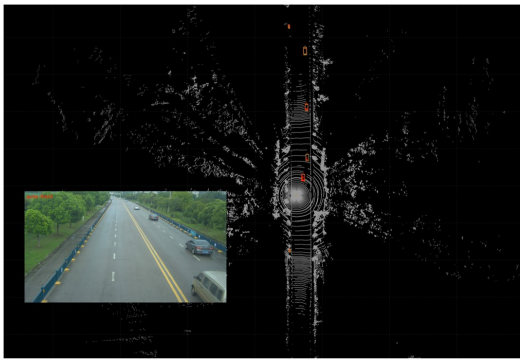


a. Corridor scene example 1

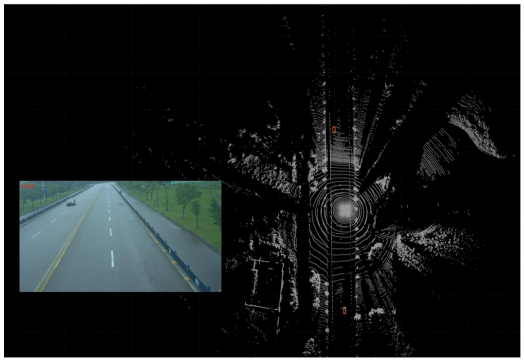
b. Corridor scene example 2



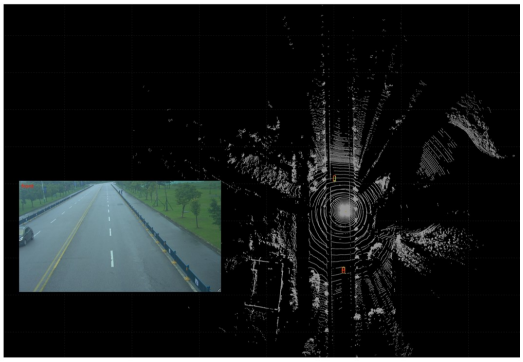
c. Corridor scene example 3



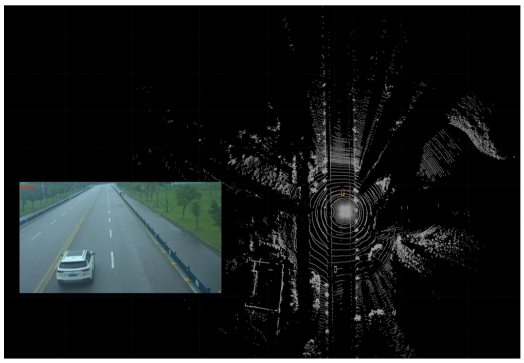
d. Corridor scene example 4



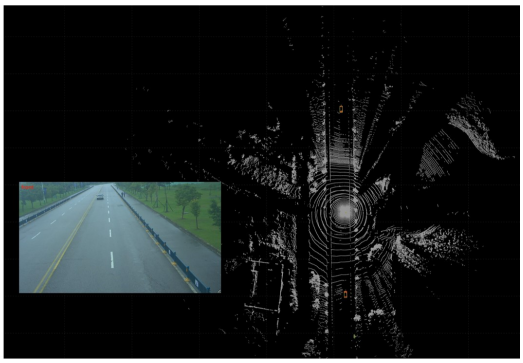
e. Corridor scene example 5



f. Corridor scene example 6

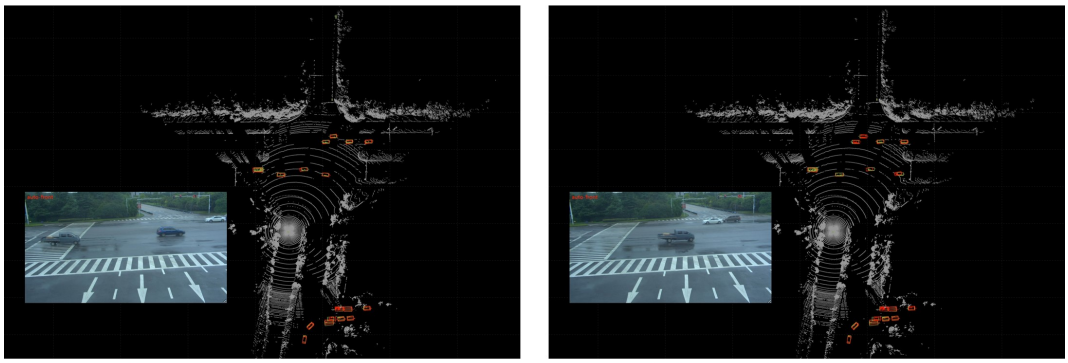


g. Corridor scene example 7



h. Corridor scene example 8

Figure 2. Visualization Examples of the Perception Results for Corridor Scenes. Red and yellow bounding boxes represent the prediction and ground truth, respectively.

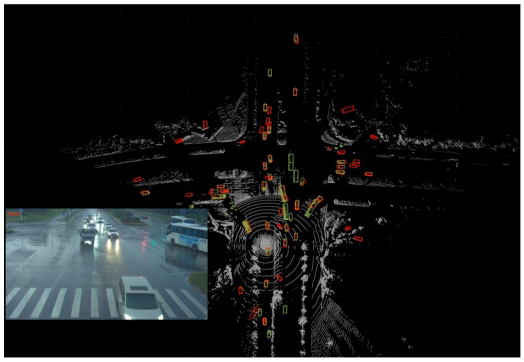


a. Intersection scene example 1

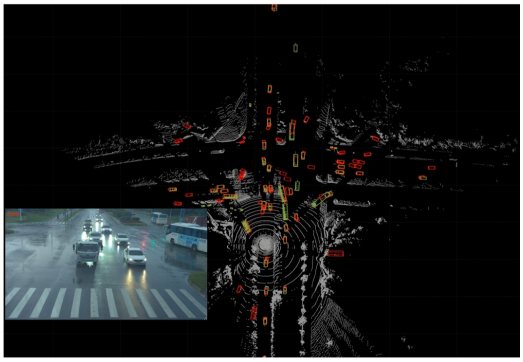
b. Intersection scene example 2

c. Intersection scene example 3

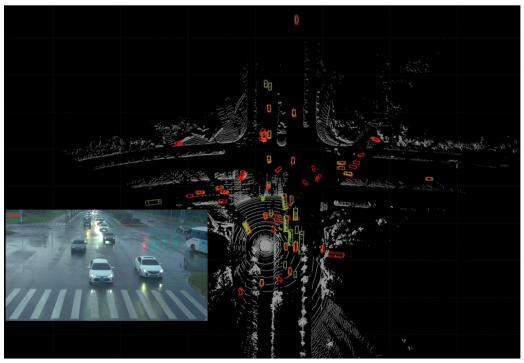
d. Intersection scene example 4



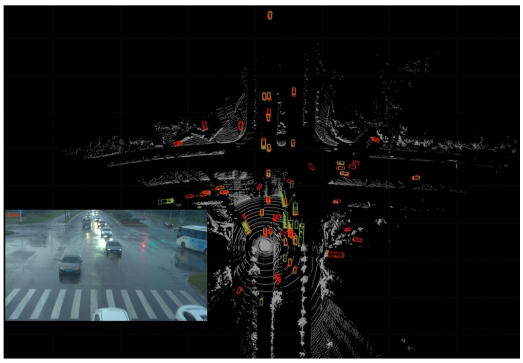
e. Intersection scene example 5



f. Intersection scene example 6



g. Intersection scene example 7



h. Intersection scene example 8

Figure 3. Visualization Examples of the Perception Results for Intersection Scenes. Red and yellow bounding boxes represent the prediction and ground truth, respectively.