Driving on Unstructured Roads: A 3D Dataset (Supplementary Material)

1. Introduction

In the supplementary material, we show additional statistics on the proposed dataset classes and expansion of the results for both 3D object detection and multi-object tracking. Furthermore, the accompanying video and point cloud data with the supplementary material highlights the LiDAR data collection process of the vehicle for one of the raw data sequences available. We perform point cloud registration and odometry on the data and show the reconstructed environment as the ego-vehicle travels. We use the method provided in [1] for reconstruction of the scene and provide the point cloud file which was generated as output for the static background environment. A sample image of the trajectory and environment can be seen in fig 1.

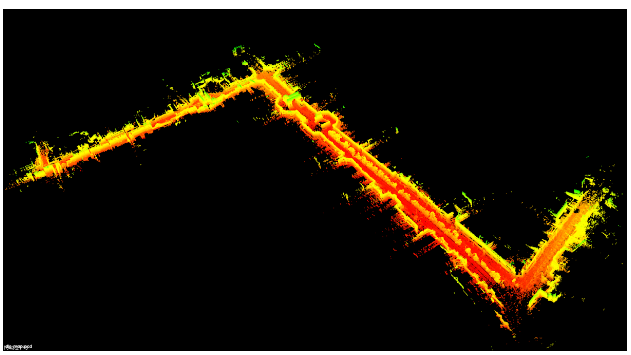

Figure 1. Generated trajectory from the LiDAR point clouds for one of the sequences of the dataset.

2. Sensor Configuration

In Table 2 we show the details of the sensors used for the current data collection. The ego-vehicle is equipped with 6 RGB cameras which provide high-resolution visual information from around the vehicle, the 64-channel LiDAR for dense point clouds and a GPS sensor for registering the location of vehicle. The information for the same has been discussed in Section 3.1 in the main paper.

3. Dataset Statistics

In addition to the fig. 6 and 7 in the paper, we provide the distribution of number of bounding boxes for each category in the dataset in figure 3. We also provide the distribution of distances of each annotated bounding box per category for the fraction of the frames in figure 2. These statistics provide a deeper understanding of the dataset structure and understanding of the experimental results.

Sensor	Qty.	Resolution	Configuration	Manufacturer/Model
LiDAR	1	64 channel (vertical) 1024 channel (horizontal)	10 Hz capture. XYZ, Intensity, Reflectivity, Range	Ouster OS1 sensor
Camera	6	2048 x 1536	BayerRG8 format 10 Hz capture	FLIR Blackfly S, C-mount
Lens	6	-	UC Series Fixed focal length 12/25mm	Edmund optics
GPS	1	-	G-Star IV BU-353-S4 sensor ~1Hz	GlobalSat

Table 1. Available sensors on-board the vehicle used for data collection. The description of each sensor and its configuration is provided in the dataset section. The resolution is mentioned wherever applicable. The arrangement of the sensors is highlighted in Fig. 5 in the main paper.

4. Additional Results

We discuss the extensions of the results from the experiments reported in the main paper in the following section for the tasks of 3D object detection and tracking and outline a few points towards the performance of the models.

4.1. 3D Object Detection

In continuation of the results reported in Table 1, 2, and 3 in the main paper, we show the expansion of results across all categories on each distance bucket for the models prepared in Table 7. While we still arrive at the conclusion that CenterPoint provides better mAP scores on the maximum cases, we observe that CenterPoint approach performs better for objects which are closer to the ego-vehicle and usually perform worse than other methods for the distance buckets which are far. This could be attributed to the fact that point cloud density per object decreases as we move far and that affects CenterPoint approach since it follows prediction of centers for each point for object detection.

Another interesting observation is that in SECOND architecture, we see better performance on categories which

Category	**AMOTA**	**AMOTP**	**Recall**	**MOTAR**	**MOTP**	**MOTA**	**lgd**	**tid**	**faf**
Bus	0.831	0.679	0.812	0.907	0.589	0.736	3.045	2.659	13.805
Car	0.641	0.726	0.667	0.787	0.518	0.521	3.422	2.035	44.806
Motorcycle	0.202	0.826	0.242	0.941	0.356	0.228	2.000	2.000	2.321
MotorcyleRider	0.507	0.735	0.496	0.801	0.320	0.390	5.027	2.585	36.410
Pedestrian	0.254	0.912	0.319	0.737	0.363	0.225	9.918	6.731	34.557
Scooter	0.250	0.494	0.323	1.000	0.092	0.323	0.000	0.000	0.000
ScooterRider	0.540	0.536	0.581	0.742	0.258	0.427	3.868	2.274	35.251
TourCar	0.796	0.433	0.848	0.821	0.351	0.692	2.877	1.034	48.866
Truck	0.701	0.635	0.675	0.903	0.403	0.607	5.108	2.676	17.796
Van	0.000	1.677	0.275	0.000	0.563	0.000	14.500	0.000	75.163
Overall	0.472	0.765	0.524	0.764	0.381	0.415	4.977	2.199	30.898

Table 2. Tracking (3D-MOT) results on the proposed dataset for Centerpoints method pre-trained with the nuScenes dataset.

Category	AMOTA	AMOTP	Recall	MOTAR	MOTA	MOTP	lgd	tid	faf
Bus	0.775	0.887	0.825	0.822	0.675	0.691	4.380	2.960	27.190
Car	0.641	0.775	0.691	0.766	0.525	0.558	3.373	2.115	51.056
Motorcycle	0.166	1.035	0.231	0.981	0.227	0.324	3.375	3.125	0.725
MotorcyleRider	0.480	0.730	0.520	0.759	0.383	0.337	4.781	2.204	45.556
Pedestrian	0.281	0.851	0.356	0.726	0.248	0.369	9.304	5.373	40.096
Scooter	0.383	0.447	0.361	1.000	0.361	0.122	2.750	2.750	0.000
ScooterRider	0.575	0.570	0.540	0.887	0.474	0.298	3.286	2.214	14.520
TourCar	0.780	0.443	0.808	0.840	0.673	0.350	3.848	1.201	43.006
Truck	0.671	0.634	0.730	0.760	0.553	0.451	4.628	2.395	42.224
Van	0.000	1.753	0.275	0.000	0.000	0.763	14.500	0.000	52.381
Overall	0.475	0.812	0.534	0.754	0.412	0.426	5.422	2.434	31.675

Table 3. Tracking (3D-MOT) results on the proposed dataset for Centerpoints method only trained on the proposed dataset.

won't get affected significantly when voxelized such as Cars and Buses. When the objects in Pedestrian category are voxelized, a significant amount of low-level information may be lost making the model prone to more errors. Hence, the performance gap in SECOND compared to both CenterPoint and PointPillars.

4.2. 3D Multi-Object Tracking

We report the same table from the main paper in Table 2, along with the results for 3D-MOT (Multi-Object Tracking) for the other detectors in Tables 3, 5, 4, and 6. We notice a lower performance in the Van category due to the low frequency of occurrence of the class in the dataset. We also observe the differences in the models based on the AMOTA and AMOTP scores. While the tracking method used for all the tables has been the same (SimpleTrack), we notice some differences in category specific performance in some of the models. For example, for the Pedestrian category, while the CenterPoint approach shows higher AP score compared to SECOND, we see that the SECOND approach reports better tracking results. This could be attributed to the fact that SECOND reports more false postive bounding boxes for the Pedestrian class, and due to the strict NMS (Non-maximal supression) threshold in the SimpleTrack, most of these are either removed of stabilized across frames, hence resulting in a minor improvement in performance. However, the overall AMOTA score for SECOND is still lower that CenterPoints due to the performance degradation in other categories. This is majorly due to the detection performance that objects with sparser points are not handled well with the SECOND approach.

We also note that the Van category in the PointPillars approach has been removed but still contributes to the result average. The category reports "NaN" performances due to the lack of required number of predictions and hence did not get allocated to any predicted tracklets. Furthermore, we observe the number of false alarms per frame (Faf) is the lowest for centerpoints pre-trained with nuscenes dataset. We further provide all the plots and metrics from the experiments in the accompanying directory in the supplementary data namely **tracking_results**. The results from these popular approaches show that there is significant scope for improvement in the benchmarks present in the proposed dataset and that current approaches are not best suited for

Category	AMOTA	AMOTP	Recall	MOTAR	MOTA	MOTP	lgd	tid	faf
Bus	0.655	1.037	0.672	0.919	0.614	0.683	8.708	7.604	10.045
Car	0.607	0.891	0.668	0.778	0.515	0.564	3.738	2.040	45.625
Motorcycle	0.214	1.305	0.237	0.874	0.206	0.317	5.750	4.417	4.671
MotorcyleRider	0.429	0.992	0.420	0.825	0.339	0.323	6.199	3.029	26.982
Pedestrian	0.379	0.870	0.406	0.763	0.304	0.366	7.136	4.907	40.123
Scooter	0.285	0.991	0.323	1.000	0.323	0.110	0.000	0.000	0.000
ScooterRider	0.447	1.070	0.461	0.838	0.379	0.273	8.667	5.009	17.997
TourCar	0.725	0.619	0.714	0.887	0.628	0.333	6.294	2.825	27.187
Truck	0.633	0.758	0.670	0.822	0.550	0.427	4.368	2.763	31.275
Van	0.000	1.840	0.175	0.000	0.000	0.720	16.500	0.000	103.460
Overall	0.437	1.037	0.474	0.771	0.386	0.412	6.736	3.259	30.736

Table 4. Tracking (3D-MOT) results on the proposed dataset for SECOND without any pre-training.

Category	AMOTA	AMOTP	Recall	MOTAR	MOTA	MOTP	lgd	tid	faf
Bus	0.722	0.856	0.716	0.909	0.649	0.577	7.771	6.688	11.765
Car	0.597	0.917	0.668	0.750	0.495	0.567	4.023	2.004	51.263
Motorcycle	0.164	1.301	0.215	0.903	0.194	0.324	3.500	3.500	3.327
MotorcyleRider	0.385	1.037	0.457	0.697	0.305	0.338	6.493	3.421	49.066
Pedestrian	0.350	0.863	0.398	0.687	0.268	0.355	7.507	5.035	51.287
Scooter	0.250	1.212	0.323	1.000	0.323	0.100	0.000	0.000	0.000
ScooterRider	0.419	1.107	0.435	0.858	0.370	0.260	8.769	4.962	14.952
TourCar	0.751	0.560	0.733	0.898	0.655	0.307	5.441	2.495	24.896
Truck	0.630	0.766	0.644	0.829	0.531	0.414	5.263	2.776	29.618
Van	0.000	1.665	0.275	0.000	0.000	0.513	14.500	0.000	66.942
Overall	0.427	1.028	0.486	0.753	0.379	0.375	6.327	3.088	30.312

Table 5. Tracking (3D-MOT) results on the proposed dataset for SECOND method pre-trained on the KITTI dataset.

a general approach, especially in cases with variations in traffic density such as Indian road scenarios. Through this dataset, we hope to provide a step in the positive direction to bridge this gap.

5. Dataset Samples

We further provide samples from the dataset such as the ones highlighted as interesting cases in figure 2 and 4 (main paper) to extend the visual understanding of the reader. We show samples with BEV (Bird-Eye-View) annotations and some of the corresponding camera images for some samples of interest in figure 4. Another set of image samples for specific classes are additionally provided in the supplementary material.

References

[1] Kenny Chen, Brett T. Lopez, Ali-akbar Agha-mohammadi, and Ankur Mehta. Direct lidar odometry: Fast localization with dense point clouds. *IEEE Robotics and Automation Letters*, 7(2):2000–2007, 2022.

Category	AMOTA	AMOTP	Recall	MOTAR	MOTA	MOTP	lgd	tid	faf
Bus	0.663	0.884	0.677	0.948	0.640	0.582	8.854	7.229	6.729
Car	0.585	0.911	0.641	0.761	0.484	0.565	4.168	2.475	46.987
Motorcycle	0.108	1.307	0.152	0.986	0.149	0.285	2.000	0.500	0.362
MotorcyleRider	0.338	1.097	0.407	0.705	0.275	0.367	6.960	3.337	43.115
Pedestrian	0.326	0.896	0.320	0.811	0.256	0.311	7.305	4.229	25.784
Scooter	0.250	1.277	0.323	1.000	0.323	0.118	0.000	0.000	0.000
ScooterRider	0.356	1.154	0.341	0.845	0.285	0.245	10.757	6.486	12.747
TourCar	0.724	0.618	0.737	0.876	0.639	0.334	5.200	2.470	30.785
Truck	0.561	0.919	0.569	0.847	0.479	0.410	9.029	5.676	23.003
Van	-	-	-	-	-	-	-	-	-
Overall	0.391	1.106	0.417	0.778	0.353	0.522	7.427	5.240	68.951

Table 6. Tracking (3D-MOT) results on the proposed dataset for the Pointpillar method.

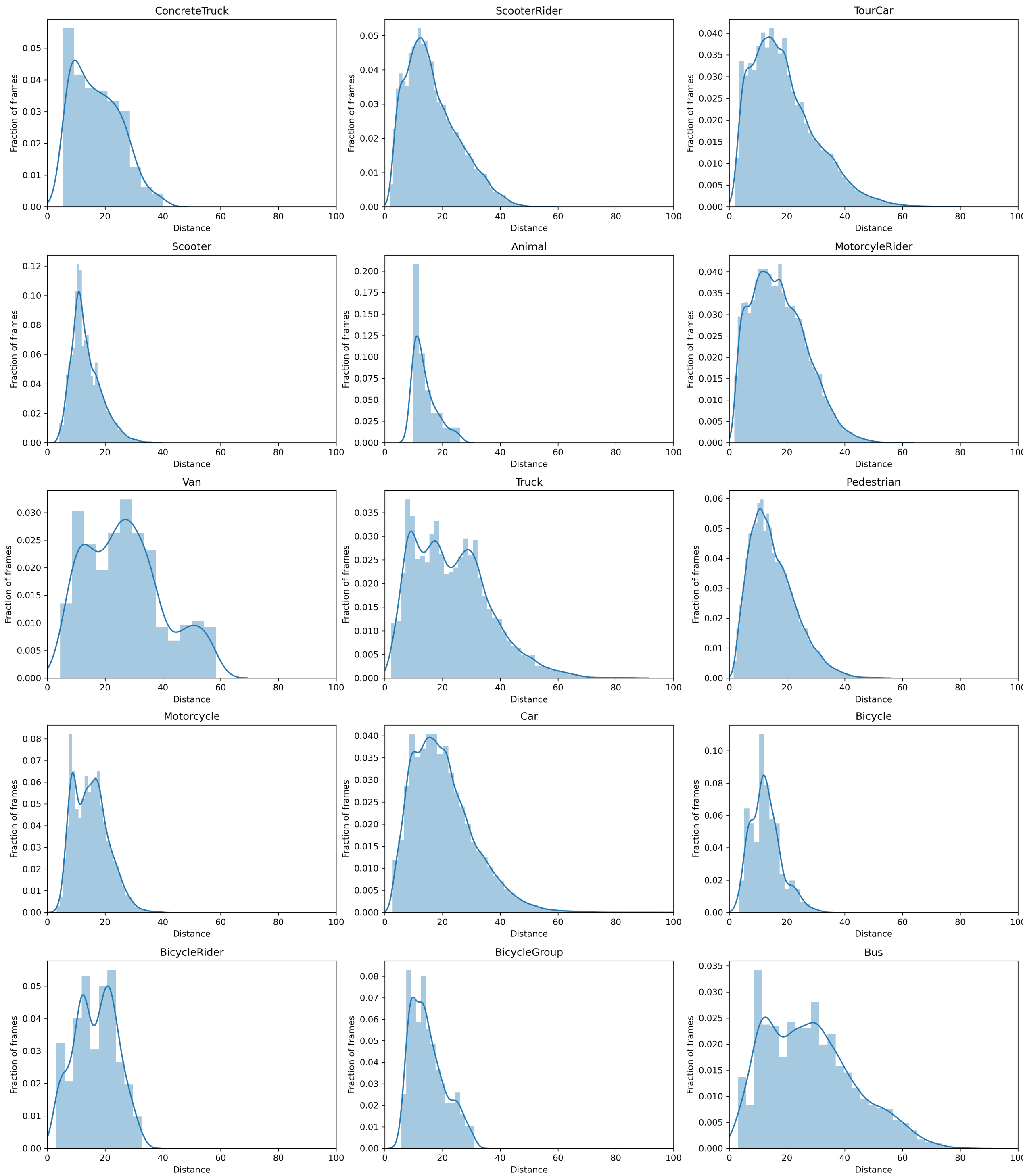

Figure 2. Distribution of distances of the annotated bounding boxes with respect to the fraction of frames in the dataset. The plots are category specific in the above figures.

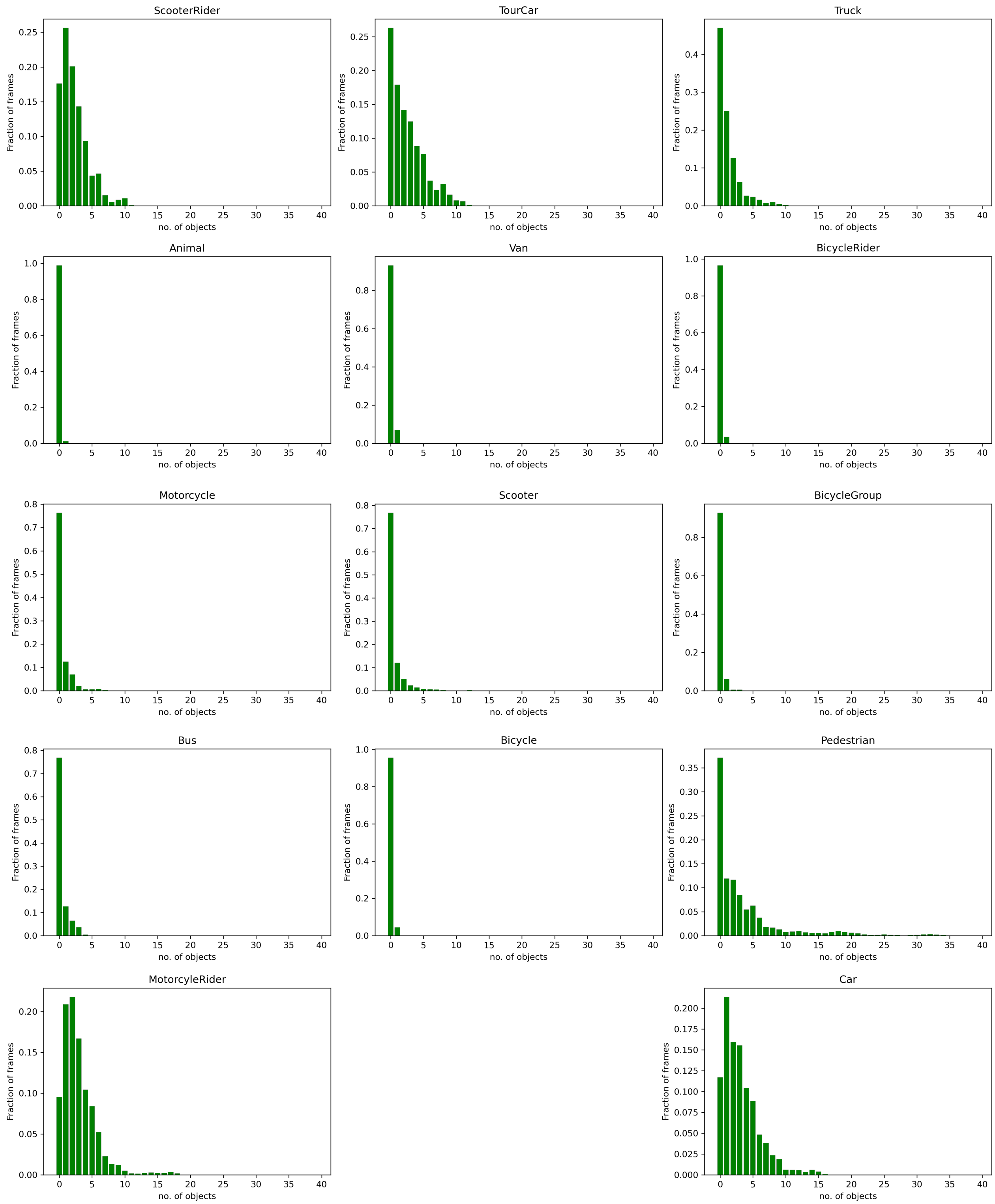

Figure 3. Distribution of bounding boxes annotated for each category in the dataset and the densities for number of boxes with respect to the fraction of frames in the dataset.

Category / Method	Distance	CenterPoint	CenterPoint (nuScenes)	SECOND	SECOND (KITTI)	PointPillar
Car	Overall	65.28	66.97	**68.89**	68.50	67.77
	0-10m	81.75	77.59	**84.79**	84.62	83.86
	10-25m	64.45	66.36	67.32	**67.94**	67.49
	>25m	18.14	23.15	25.07	23.94	**26.17**
Bus	Overall	59.09	**78.47**	59.12	49.69	43.70
	0-10m	76.55	**88.42**	82.43	67.41	54.83
	10-25m	60.09	**80.58**	56.22	47.84	43.10
	>25m	24.04	**32.24**	22.94	16.22	11.89
Truck	Overall	68.79	**72.18**	65.11	68.09	63.68
	0-10m	88.87	92.04	84.43	**93.44**	88.43
	10-25m	60.65	**66.22**	53.98	58.77	53.84
	>25m	23.43	23.63	**28.60**	24.16	24.96
Van	Overall	9.58	12.71	1.27	**15.77**	0.14
	0-10m	0.0	0.0	0.00	0.00	0.00
	10-25m	12.99	14.36	2.40	**19.85**	0.33
	>25m	0.0	0.0	0.00	0.00	0.00
TourCar	Overall	76.94	**77.40**	74.81	77.02	72.80
	0-10m	**88.94**	87.63	86.17	88.52	85.93
	10-25m	76.38	**77.17**	74.90	74.89	70.63
	>25m	33.85	40.44	**40.86**	42.69	39.37
Pedestrian	Overall	**28.60**	22.49	19.54	23.74	22.72
	0-10m	**44.89**	33.85	27.18	33.67	29.34
	10-25m	**24.39**	19.47	17.61	21.05	20.45
	>25m	3.48	4.48	**6.44**	5.58	5.45
Motorcycle	Overall	23.65	**25.28**	21.69	22.79	16.97
	0-10m	45.04	**47.28**	33.63	36.05	14.43
	10-25m	17.18	19.55	19.39	**20.19**	18.86
	>25m	4.72	**6.13**	3.54	3.48	3.56
Scooter	Overall	**42.36**	38.05	26.98	23.73	16.81
	0-10m	**40.39**	24.22	12.74	0.79	0.25
	10-25m	**42.75**	39.81	30.81	29.99	22.05
	>25m	**1.05**	0.51	1.00	0.00	0.00
MotorCycleRider	Overall	59.29	**61.48**	53.39	48.90	46.52
	0-10m	78.00	**79.40**	66.66	63.73	60.30
	10-25m	55.49	**57.88**	49.49	45.22	42.44
	>25m	12.66	**15.73**	13.11	11.26	12.09
ScooterRider	Overall	**66.33**	64.65	52.27	50.62	41.60
	0-10m	**76.36**	74.07	59.03	58.90	37.22
	10-25m	**68.18**	66.72	55.79	53.30	47.97
	>25m	14.62	**16.20**	8.88	7.39	10.22
mAP	Overall	49.99	**51.97**	44.31	44.89	39.27
	0-10m	**62.08**	60.45	53.71	52.71	45.46
	10-25m	48.26	**50.81**	42.79	43.90	38.72
	>25m	13.60	**16.25**	15.04	13.47	13.37

Table 7. Experimental results on proposed dataset with different popular methods. We report AP scores across different categories on the validation set. This table shows the results on all training categories with all distance buckets. The 'overall' distance metric ranges from 0-30m and is considered based on the distance distribution of objects present in the scene.

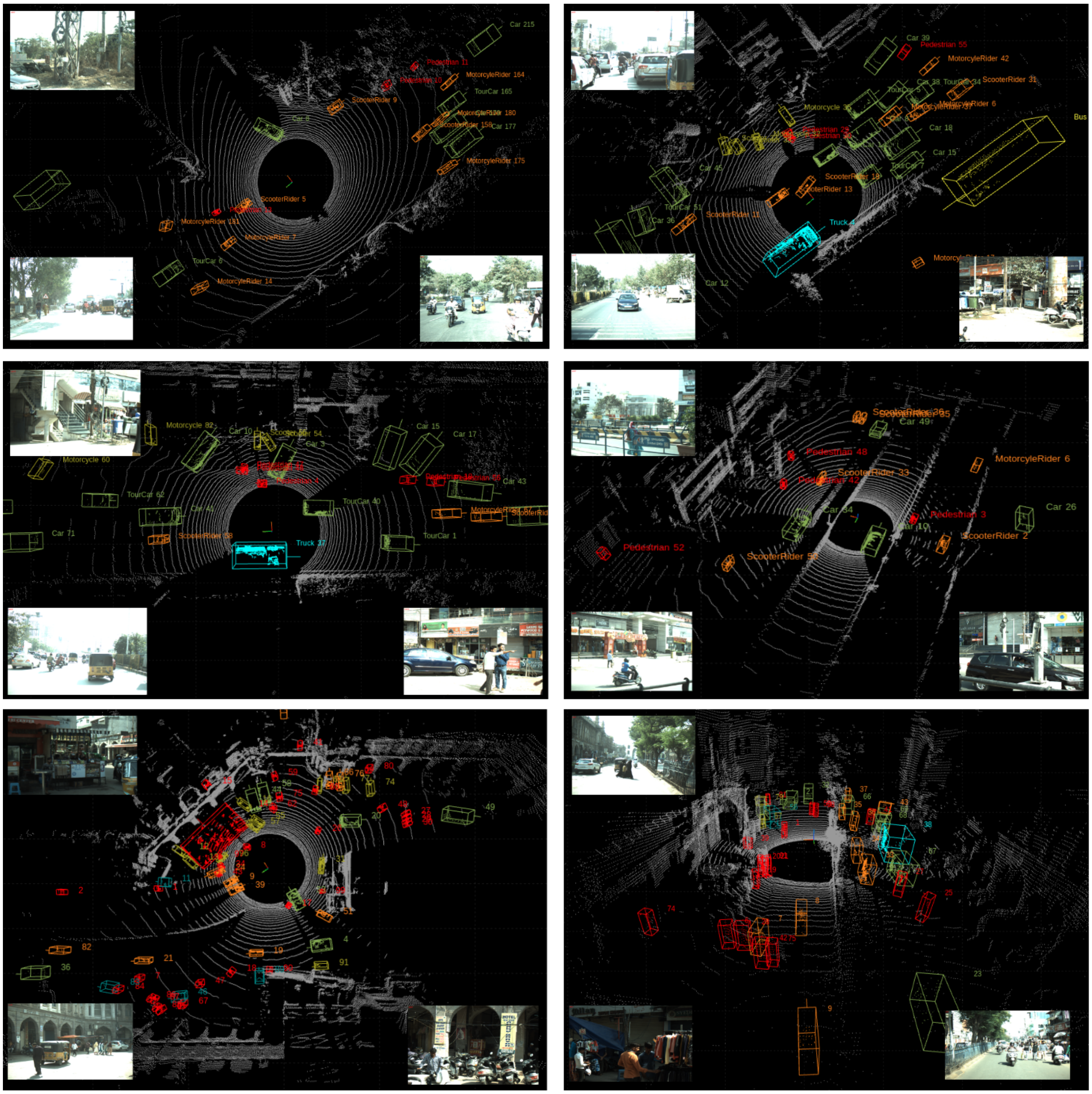

Figure 4. Some examples from the dataset showing different traffic scenarios, LiDAR data with annotations, and a sample of LiDAR point clouds projected on camera data.