

Supplementary Material

1. Datasets

Open-Loop. nuScenes is a challenging public dataset for autonomous driving evaluation, which consists of 28k total samples in a 22k/6k training/validation split. The objects in each scene are annotated with 3D bounding box, orientation, and vehicle speed information. This dataset comprises 1000 complex driving scenarios, each spanning approximately 20 seconds, with annotations at a frame rate of 2Hz. For evaluation, we employ the L2 error and collision rate (CR) metrics following the evaluation protocol established in [1] to ensure fair comparison with recent sota works.

Long-tail Scenario. We follow TOKEN [4] and manually construct a long-tail scenario validation set curated from nuScenes for comprehensive evaluation, including three scenarios: 1) executing 3-point turns; 2) resuming motion after a full stop; 3) overtaking parked cars through the oncoming lane. Details of each scenario are provided in Tab. 1.

Closed-Loop. Bench2Drive is a comprehensive evaluation protocol based on CARLA for evaluating abilities of end-to-end autonomous driving systems. We follow the standardized data partitioning and use 950 clips for training. For closed-loop evaluation, the benchmark provides 220 predefined short routes to assess dynamic planning capabilities in complex environments. Closed-loop evaluation employs five metrics: Driving Score (DS), Success Rate (SR), Efficiency, and Comfortness. Driving Score integrates route completion rate with multiplicative penalty factors for traffic violations. Success Rate quantifies the proportion of routes fully completed within predefined time constraints. Efficiency evaluates normalized speed performance relative to traffic compliance and route complexity. Comfortness measures ride smoothness through cumulative jerk and lateral acceleration integrals.

2. Implementation Details

CogAD plans a 3s future ego-trajectory without using any form of ego state or history information as input. The BEV perception range spans 60m longitudinally and 30m laterally,

Table 1. Long-tail.

Scenario	Scene ID	Frames Interval	Frames Number
3-point turn	scene-0778	frame 6-30	25
	scene-0921	frame 21-25	
Resume from stop	scene-0925	frame 19-23	
	scene-0968	frame 7-11	
	scene-0552	frame 13-17	
	scene-0917	frame 24-28	40
	scene-0221	frame 11-15	
	scene-1064	frame 21-25	
	scene-0331	frame 8-12	
Overtake	scene-0038	frame 4-33	
	scene-0271	frame 3-11	
	scene-0969	frame 14-33	102
	scene-0329	frame 3-33	
	scene-1065	frame 24-35	

with input images resized to 640×360 pixels. We use a 100×100 BEV feature map, 100×20 map queries, and 300 agent queries. Moreover, we set the number of intent anchors to 30 and the trajectory modes to 6. The feature dimension size is set to 256. We train CogAD using AdamW optimizer [3] and Cosine Annealing scheduler [2] with initial learning rate 4×10^{-4} and weight decay 0.01. CogAD is trained for 60 epochs on the nuScenes dataset and 6 epochs on the Bench2Drive dataset, utilizing 8 NVIDIA Tesla A100 GPUs with a total batch size of 32.

3. More Qualitative Results

We provide more visualization results to illustrate the effectiveness of CogAD on various driving scenarios as shown in Fig. 1, Fig. 2, Fig. 3, and Fig. 4.

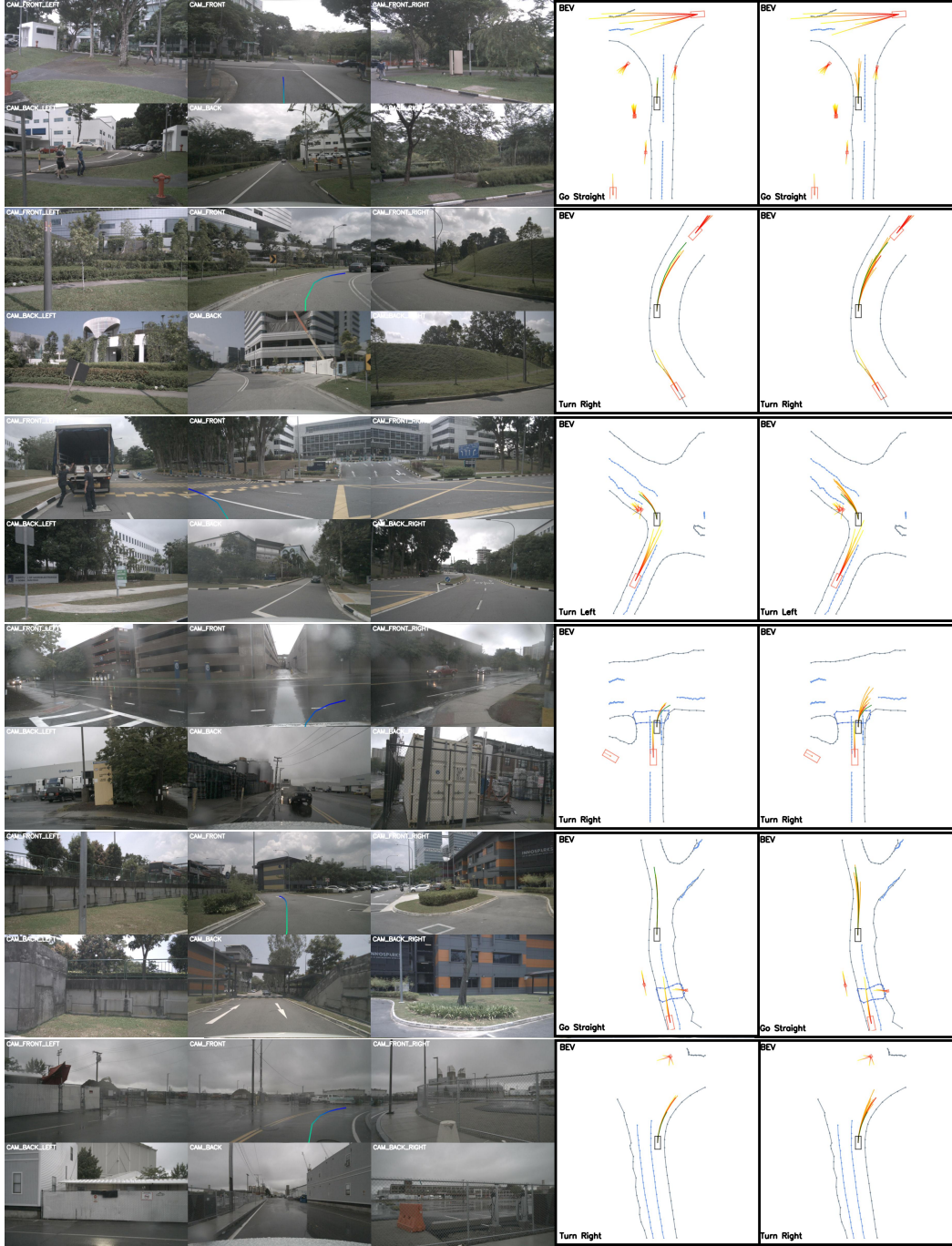


Figure 1. Qualitative results of CogAD on nuScenes. The 2nd column shows the Top-3 multi-mode trajectories of the highest-probability intent, with colors of red, orange, and yellow, respectively. The 3rd column displays the Top-10 multi-intent trajectories with the highest-probability mode. GT trajectory is drawn in green. CogAD demonstrates robust performance across diverse commands, scenarios, and weather conditions.

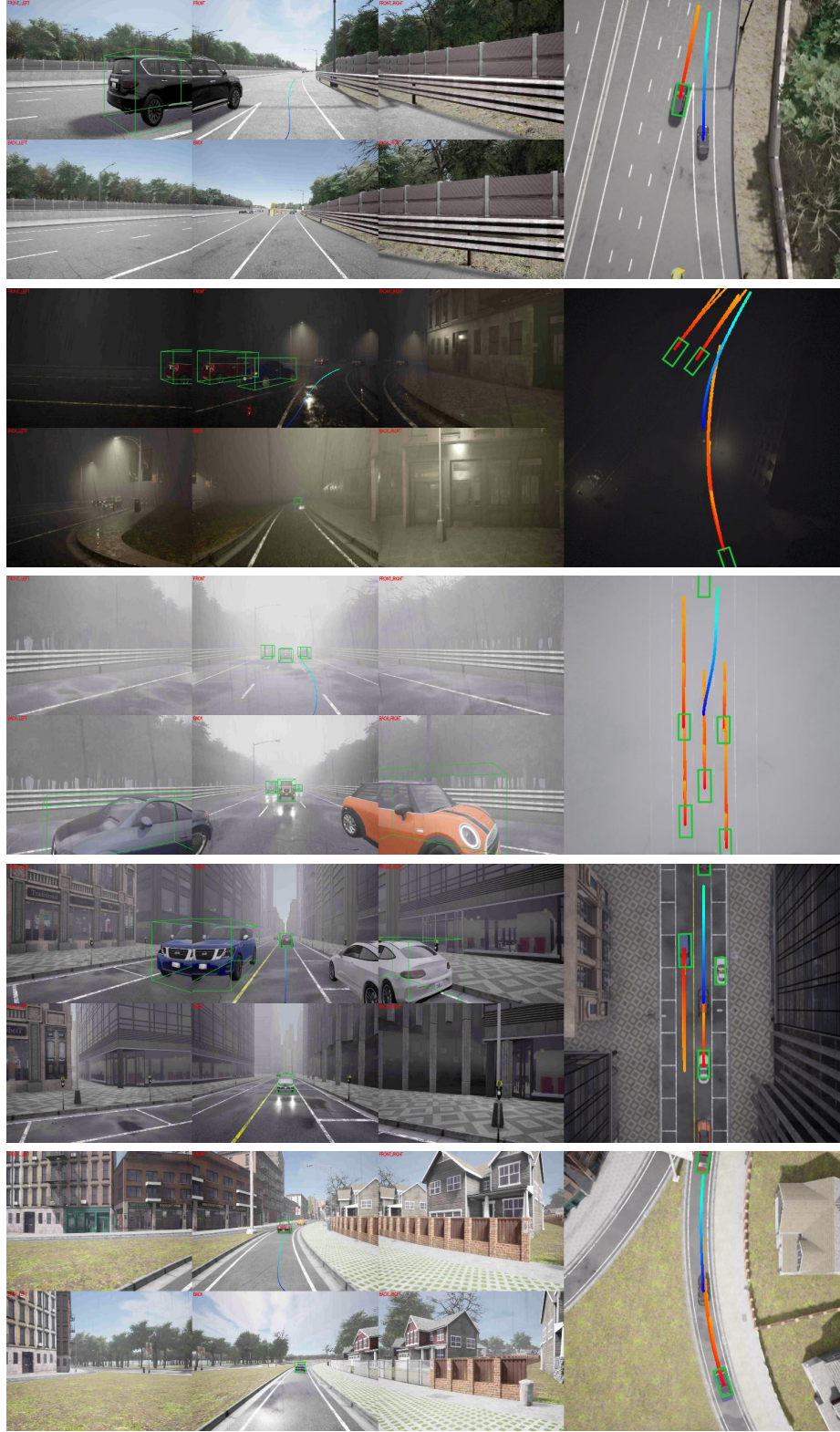


Figure 2. Qualitative results of CogAD on Bench2Drive closed-loop. In the rightmost column, the centrally positioned black vehicle denotes the ego vehicle. CogAD demonstrates robust performance across diverse commands, scenarios, and weather conditions.



(a) Turn left at the intersection in the rain



(b) Proceed through the intersection by following the lead car

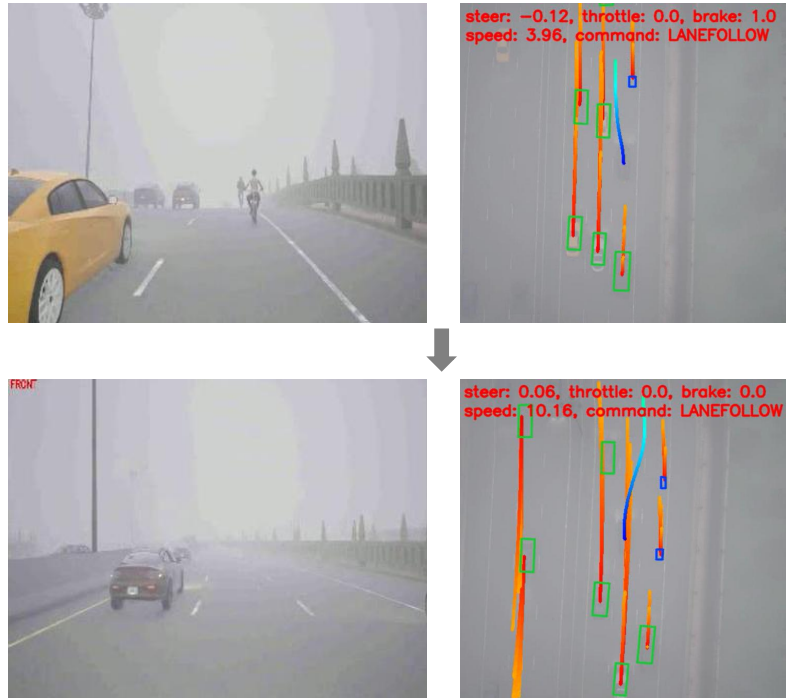


(c) Slow down and yield to merging traffic



(d) At night, slow down when approaching a vehicle

Figure 3. Qualitative results of CogAD on nuScenes.



(a) In rainy weather, change to the left lane to go around the bicycle, and then move back to the original lane.



(b) On a rainy night, change lanes to overtake the white car ahead.

Figure 4. Qualitative results of CogAD on Bench2Drive.

References

- [1] Bo Jiang, Shaoyu Chen, Qing Xu, Bencheng Liao, Jiajie Chen, Helong Zhou, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Vad: Vectorized scene representation for efficient autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8340–8350, 2023. [1](#)
- [2] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. [1](#)
- [3] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. [1](#)
- [4] Ran Tian, Boyi Li, Xinshuo Weng, Yuxiao Chen, Edward Schmerling, Yue Wang, Boris Ivanovic, and Marco Pavone. Tokenize the world into object-level knowledge to address long-tail events in autonomous driving. *arXiv preprint arXiv:2407.00959*, 2024. [1](#)