Making Large Language Models Better Planners with Reasoning-Decision Alignment

Zhijian Huang[1⋆], Tao Tang[1⋆], Shaoxiang Chen[2], Sihao Lin[3], Zequn Jie[2], Lin Ma[2], Guangrun Wang[4], Xiaodan Liang[1,5✉]

[1]Shenzhen Campus of Sun Yat-sen University [2]Meituan Inc.
[3]University of Technology Sydney [4]Sun Yat-sen University
[5]Research Institute of Multiple Agents and Embodied Intelligence, Peng Cheng Laboratory, Shenzhen, China

1 Additional Details

1.1 Implementation Details

We adapt LLaVa [2] as the foundation LLM in our RDA-Driver. To denote the final trajectory output format, we add special tokens "<SOT>" (Start of Trajectory) and "<EOT>" (End of Trajectory) at the beginning and end of the trajectory, respectively. In the perception data annotated by DriveLM [3], it is necessary to predict the specific coordinates in the image. However, due to the difficulty and lower priority of the coordinate prediction task, and to encourage the model to focus more on overall logical reasoning rather than excessively focusing on detailed predictions, we remove the coordinate part during the training process, retaining only the object's camera position in the image, *i.e.*, `<c4, CAM_FRONT>`. Finally, we train RDA-Driver with LoRA [1] strategy.

2 Additional Results

2.1 Few-shot learning

To assess the model's ability in logical reasoning and causal inference, we conducte few-shot experiments using 10%, 50%, and 100% of the training data, as shown in 1. Even with a small number of samples, *i.e.*, 10% of the training data, RDA-Driver demonstrate a competitive end-to-end decision-making capability, despite the full dataset containing only 4072 samples. As the amount of data increases, the model consistently achieves better performance.

3 Additional Visualizations

3.1 More Qualitative Results

As illustrated in Fig. 1, we presents examples showcasing the logical reasoning of RDA-Driver. By responding to CoT prompts related to perception, prediction,

⋆ Equal contribution.
✉ Corresponding author.

Training samples	L2 (m) ↓				Collision (%) ↓			
	1s	2s	3s	Avg.	1s	2s	3s	Avg.
10%	0.29	1.44	2.98	1.57	0.03	0.48	2.08	0.86
50%	0.25	0.79	1.66	0.89	0.00	0.18	1.18	0.45
100%	0.23	0.73	1.54	**0.80**	0.00	0.13	0.83	**0.32**

Table 1: Few-shot learning. With a small number of samples, RDA-Driver performs competitive results.

and behavior, RDA-Driver can consistently reason across various scenarios such as straight driving and turning, ultimately predicting accurate driving trajectory.

3.2 Video Demo

In addition to the figures, we have attached a video demo in the supplementary materials, which consists of hundreds of frames that provide a more comprehensive evaluation of our proposed approach.

Fig. 1: Examples of the logical reasoning and trajectory planning of RDA-Driver. The planned trajectory and gt trajectory are in green and red respectively.

References

1. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
2. Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning (2023)
3. Sima, C., Renz, K., Chitta, K., Chen, L., Zhang, H., Xie, C., Luo, P., Geiger, A., Li, H.: Drivelm: Driving with graph visual question answering. arXiv preprint arXiv:2312.14150 (2023)