PPAD: Iterative Interactions of Prediction and Planning for End-to-end Autonomous Driving
Supplementary Material

Zhili Chen[1†], Maosheng Ye[1], Shuangjie Xu[1], Tongyi Cao[2], and Qifeng Chen[1✉]

[1]HKUST [2]DeepRoute.AI
{zchenei, myeag, shuangjie.xu}@connect.ust.hk, tongyicao@deeproute.ai, cqf@cse.ust.hk

1 Metrics on Subtasks Evaluation

In Tab. 3 of the main paper, the detection evaluation metrics of mean Average Precision (mAP) and the final nuScenes Detection Score (NDS) are based on the nuScene 3D detection benchmark [1]. We followed [3,4] to evaluate map construction with mean average precision (mAP). In the motion forecasting subtask, we adopt the same metrics, minimum Average Displacement Error (minADE), minimum Final Displacement Error (minFDE), and Miss Rate (MR), as [2,3] to evaluate the performance of our approach.

2 Noisy Trajectory Training Details

During training, we create noisy trajectories from the ground truth motion labels. Each timestep of the noisy trajectories is considered as the starting position for each iteration of the prediction and planning processes. Under our PPAD framework, the model aims to learn to predict the correct next move by interacting with the surrounding agents and environments, neglecting the noisy starting positions. In practice, we create $N_{noisy} = 5$ noisy trajectories for the ego vehicle. We use the size of the ego vehicle (W_{ego}, H_{ego}) as the noise scales, corresponding to x and y trajectory axes. Then, we sample a random tensor from the uniform distribution with a dimension of $[N_{noisy}, T_{fut}, 2]$ ranging from $[-1, 1]$. We multiply the random tensor with the noise scales (W_{ego}, H_{ego}) and consider the resulting tensor as the noisy offsets. Finally, the noisy trajectories are obtained by adding the noisy offsets to the ground truth, which are offsetting from the ground truth centers by at most W_{ego} meters along the $x-$axis and H_{ego} meters along the $y-$axis.

We calculate the noisy planning losses between the resulting predictions from the noisy inputs with the ground truth. We then take the average of separately computed noisy planning losses as the $\mathcal{L}_{Plan}^{noisy}$ in Eqn. 8.

3 Additional Ablation Studies

Same as Sec. 4.3, the following experiments adhere to the progressive training pipeline as proposed in VAD [3], in which we first train the perception tasks for 48 epochs and then train the whole model for another 12 epochs to conduct different ablation studies for the PPAD.

Effect of Key Objects Attention Distance Ranges The hierarchical dynamic key objects attention is proposed to understand the the dynamic surrounding driving agents and the environments in a coarse-to-fine manner. Tab. 1 shows the ablation study, which was conducted by setting different distance ranges. We use the distance range set of $\{+\infty\ m, 15\ m, 7.5\ m\}$, which achieves the best performance in Tab. 1, as our model's default setting.

Distance Range	L2 (m) ↓				Collision (%) ↓			
	1s	2s	3s	Avg.	1s	2s	3s	Avg.
$\{+\infty\ m, 7.5\ m, 3.75\ m\}$	0.33	**0.56**	**0.87**	0.59	**0.08**	**0.12**	**0.35**	**0.19**
$\{+\infty\ m, 15\ m, 7.5\ m, 3.75\ m\}$	0.34	0.59	0.91	0.61	0.11	0.18	0.41	0.23
$\{+\infty\ m, 15\ m, 7.5\ m\}$	**0.31**	**0.56**	**0.87**	**0.58**	**0.08**	**0.12**	0.38	**0.19**

Table 1: The ablation study of different distance ranges for conducting the hierarchical dynamic key objects attention.

Effect of the Noisy Trajectory Number As shown in Tab. 2, we conducted the ablation study on training by introducing a different number of noisy trajectories. The Tab. 2 shows that the performance is the best when we utilize 5 noisy trajectories when training.

Num. Noisy Traj.	L2 (m) ↓				Collision (%) ↓			
	1s	2s	3s	Avg.	1s	2s	3s	Avg.
0	0.35	0.59	0.89	0.61	**0.08**	0.14	**0.32**	**0.18**
1	0.33	0.57	0.88	0.59	0.14	0.15	0.37	0.22
10	0.33	0.57	0.88	0.59	0.12	0.17	0.40	0.23
5	**0.31**	**0.56**	**0.87**	**0.58**	**0.08**	**0.12**	0.38	0.19

Table 2: The ablation study on training the model by utilizing a different number of noisy trajectories as predictions.

Effect of Planning Losses Weights The planning losses comprise the planning loss calculated from the PPAD's normal planning predictions and the planning loss of the predictions from the noisy trajectories. We conducted the ablation study on the weights of the planning losses in Tab. 3. Among different settings, we can balance the planning losses best when we set the $\zeta_1 = 0.6$ and $\zeta_2 = 0.4$.

Planning Loss Weights	L2 (m) ↓				Collision (%) ↓			
	1s	2s	3s	Avg.	1s	2s	3s	Avg.
$\zeta_1 = 0.4, \zeta_2 = 0.6$	0.35	0.60	0.93	0.62	0.09	0.17	0.43	0.23
$\zeta_1 = 0.8, \zeta_2 = 0.2$	0.33	0.57	0.88	0.59	0.52	0.64	0.93	0.70
$\zeta_1 = 0.6, \zeta_2 = 0.4$	**0.31**	**0.56**	**0.87**	**0.58**	**0.08**	**0.12**	**0.38**	**0.19**

Table 3: The ablation study on the weights ζ_1 and ζ_2 of the planning loss in Eqn. 8.

4 Additional Qualitative Results

Please refer to the video clips in the same zip file accompanied with this supplementary paper.

References

1. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11621–11631 (2020)
2. Gu, J., Hu, C., Zhang, T., Chen, X., Wang, Y., Wang, Y., Zhao, H.: Vip3d: End-to-end visual trajectory prediction via 3d agent queries. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5496–5506 (2023)
3. Jiang, B., Chen, S., Xu, Q., Liao, B., Chen, J., Zhou, H., Zhang, Q., Liu, W., Huang, C., Wang, X.: Vad: Vectorized scene representation for efficient autonomous driving. arXiv preprint arXiv:2303.12077 (2023)
4. Liao, B., Chen, S., Wang, X., Cheng, T., Zhang, Q., Liu, W., Huang, C.: Maptr: Structured modeling and learning for online vectorized hd map construction. arXiv preprint arXiv:2208.14437 (2022)