Don't Shake the Wheel: Momentum-Aware Planning in End-to-End Autonomous Driving

Ziying Song[1,2], Caiyan Jia[1,2,⋆] , Lin Liu [1,2], Hongyu Pan[3], Yongchang Zhang[3], Junming Wang[3,7], Xingyu Zhang[3], Shaoqing Xu[4], Lei Yang[5], Yadan Luo[6,⋆]

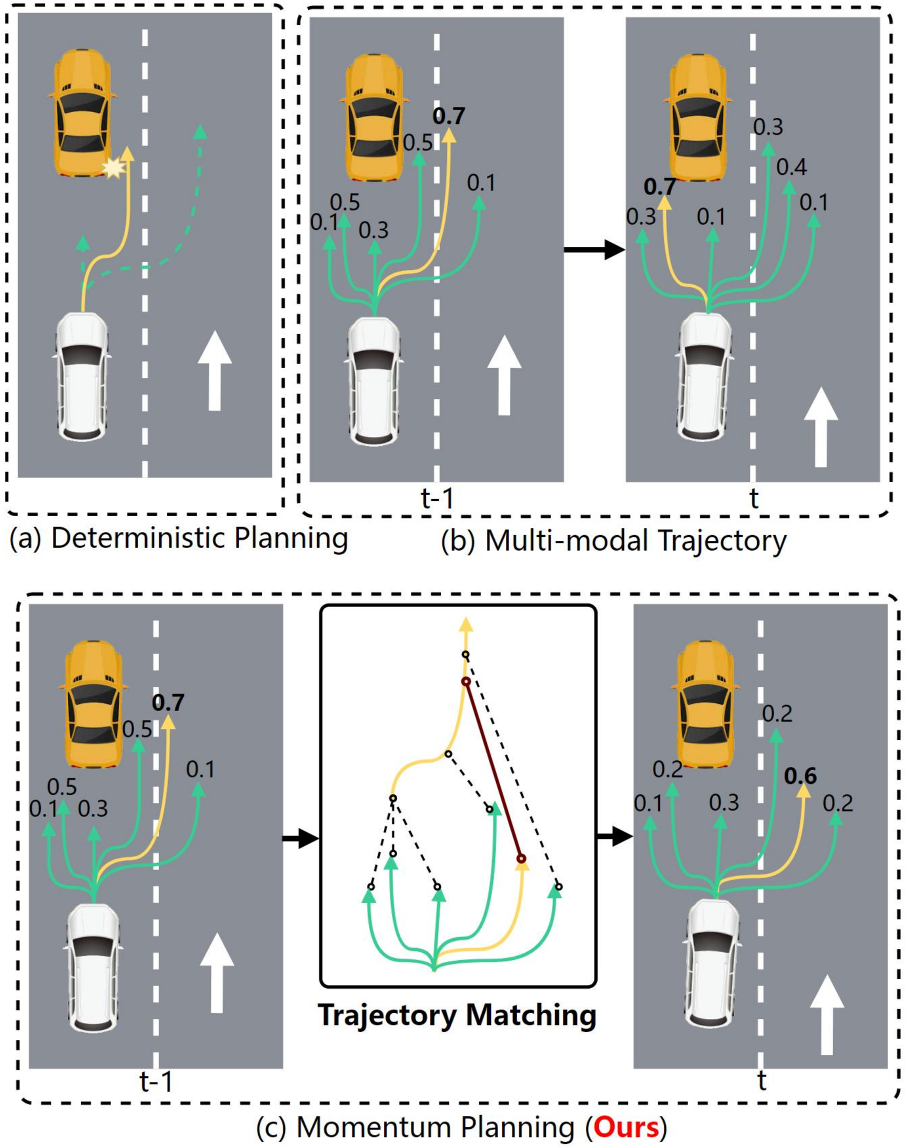
[1]Beijing Jiaotong University[2]Beijing Key Laboratory of Traffic Data Mining and Embodied Intelligence[3]Horizon Robotics [4]University of Macau [5]THU [6]The University of Queensland [7]HKU

https://github.com/adept-thu/MomAD

Arxiv

Github

Motivation



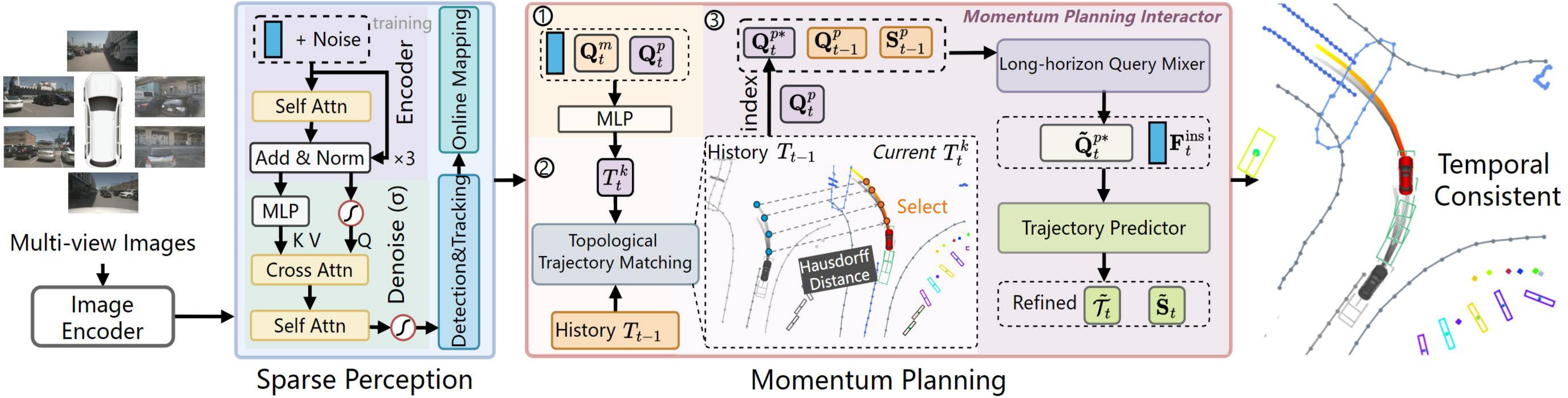
(a) Deterministic Planning (b) Multi-modal Trajectory

(c) Momentum Planning (**Ours**)

End-to-end autonomous driving frameworks enable seamless integration of perception and planning but often rely on one-shot trajectory prediction, which may lead to unstable control and vulnerability to occlusions in single-frame perception.

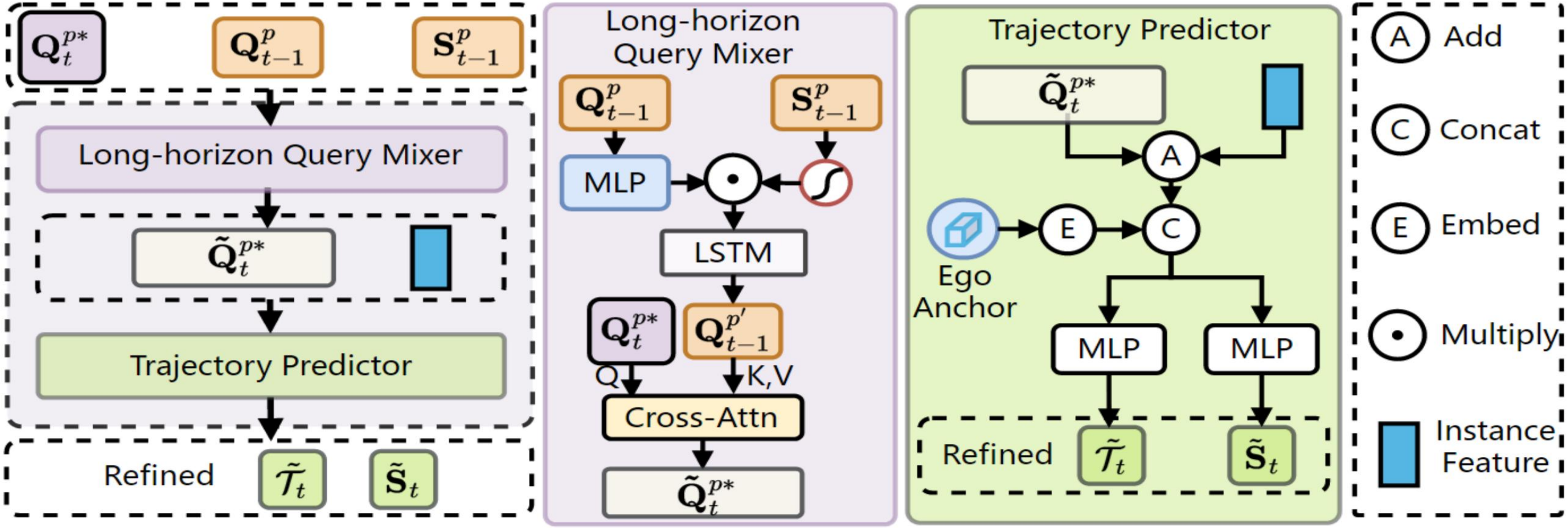
(a) Deterministic Planning [UniAD,VAD] predicts deterministic trajectories, but lacks action diversity, posing safety risks. (b) Multi-modal Trajectory Planning [VADv2,SparseDrive] selects the highest-scoring trajectory among the multi-modal trajectories, yet fails to ensure stability and consistency, having risks in vehicle trembling. (c) Momentum Planning leverages the trajectory and perception momentum to enhance current planning through historical guidance to overcome temporal inconsistency.

Method



The overall architecture of MomAD. MomAD, as a multi-modal trajectory end-to-end autonomous driving method, first encodes multi-view images into feature maps, then learns a sparse scene representation through a robust instance denoising via perturbation module, and finally performs a momentum planning through Topological Trajectory Matching (TTM) module and Momentum Planning Interactor (MPI) module to accomplish planning tasks. Our approach addresses critical challenges of stability and robustness in dynamic driving conditions.

Method



The illustration of Momentum Planning Interactor (MPI). MPI cross-attends a selected planning query with historical queries to expand static and dynamic perception files, resulting in an enriched query that improves long-horizon trajectory generation and reduces collision risks.

Experiments

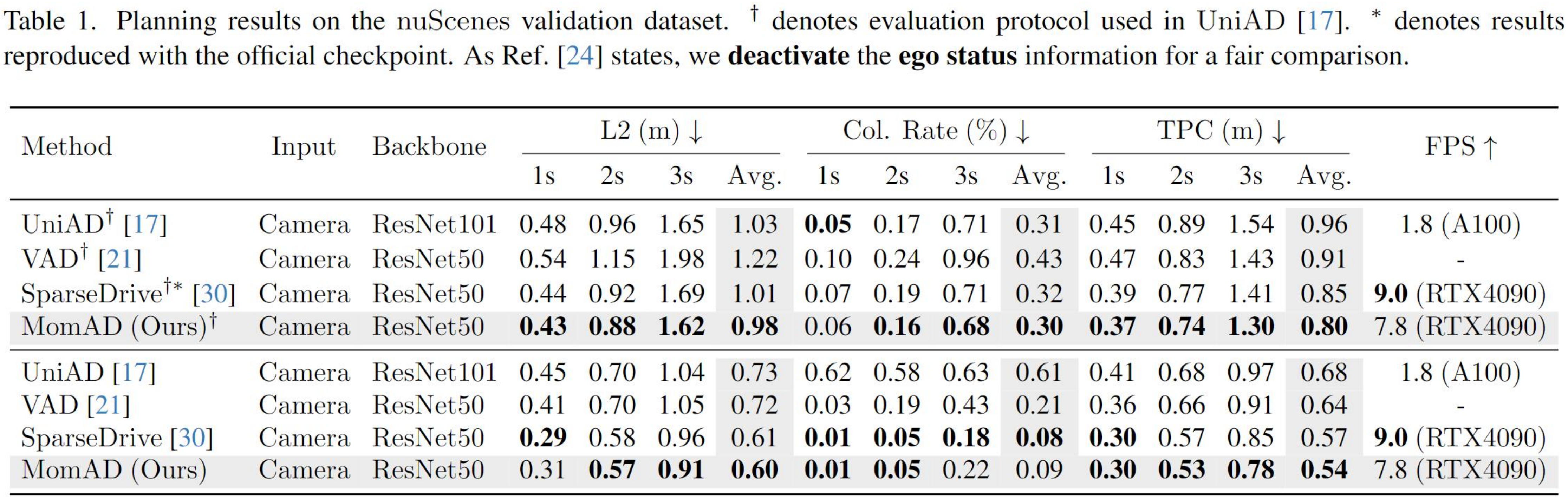


Table 1. Planning results on the nuScenes validation dataset. † denotes evaluation protocol used in UniAD [17]. * denotes results reproduced with the official checkpoint. As Ref. [24] states, we **deactivate** the **ego status** information for a fair comparison.

Method	Input	Backbone	L2 (m) ↓				Col. Rate (%) ↓				TPC (m) ↓				FPS ↑
			1s	2s	3s	Avg.	1s	2s	3s	Avg.	1s	2s	3s	Avg.	
UniAD† [17]	Camera	ResNet101	0.48	0.96	1.65	1.03	**0.05**	0.17	0.71	0.31	0.45	0.89	1.54	0.96	1.8 (A100)
VAD† [21]	Camera	ResNet50	0.54	1.15	1.98	1.22	0.10	0.24	0.96	0.43	0.47	0.83	1.43	0.91	-
SparseDrive†* [30]	Camera	ResNet50	0.44	0.92	1.69	1.01	0.07	0.19	0.71	0.32	0.39	0.77	1.41	0.85	**9.0** (RTX4090)
MomAD (Ours)†	Camera	ResNet50	**0.43**	**0.88**	**1.62**	**0.98**	0.06	**0.16**	**0.68**	**0.30**	**0.37**	**0.74**	**1.30**	**0.80**	7.8 (RTX4090)
UniAD [17]	Camera	ResNet101	0.45	0.70	1.04	0.73	0.62	0.58	0.63	0.61	0.41	0.68	0.97	0.68	1.8 (A100)
VAD [21]	Camera	ResNet50	0.41	0.70	1.05	0.72	0.03	0.19	0.43	0.21	0.36	0.66	0.91	0.64	-
SparseDrive [30]	Camera	ResNet50	**0.29**	0.58	0.96	0.61	**0.01**	**0.05**	**0.18**	**0.08**	**0.30**	0.57	0.85	0.57	**9.0** (RTX4090)
MomAD (Ours)	Camera	ResNet50	0.31	**0.57**	**0.91**	**0.60**	**0.01**	**0.05**	0.22	0.09	**0.30**	**0.53**	**0.78**	**0.54**	7.8 (RTX4090)

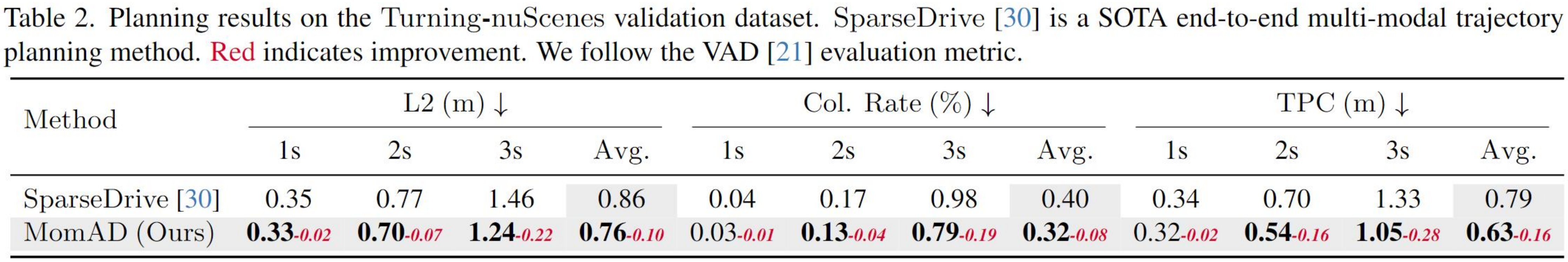
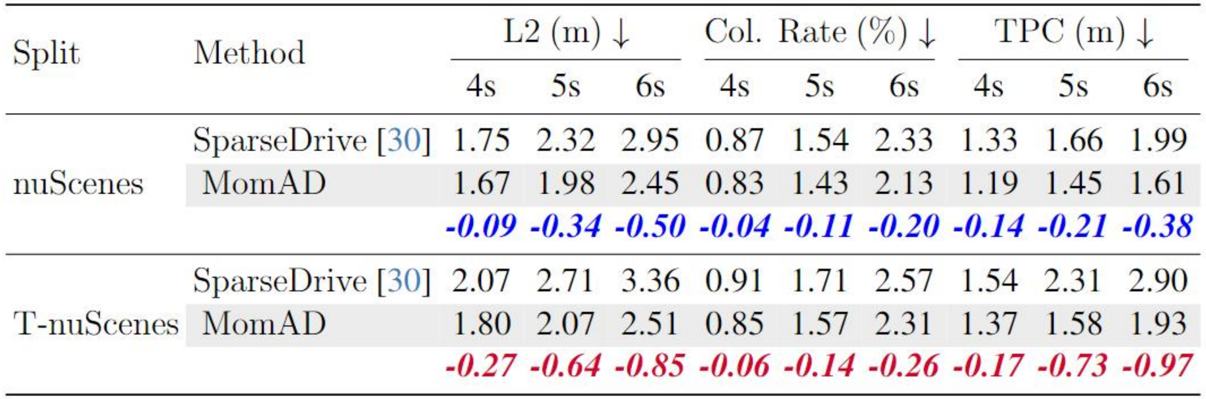


Table 2. Planning results on the Turning-nuScenes validation dataset. SparseDrive [30] is a SOTA end-to-end multi-modal trajectory planning method. Red indicates improvement. We follow the VAD [21] evaluation metric.

Method	L2 (m) ↓				Col. Rate (%) ↓				TPC (m) ↓			
	1s	2s	3s	Avg.	1s	2s	3s	Avg.	1s	2s	3s	Avg.
SparseDrive [30]	0.35	0.77	1.46	0.86	0.04	0.17	0.98	0.40	0.34	0.70	1.33	0.79
MomAD (Ours)	**0.33**$_{-0.02}$	**0.70**$_{-0.07}$	**1.24**$_{-0.22}$	**0.76**$_{-0.10}$	0.03$_{-0.01}$	**0.13**$_{-0.04}$	**0.79**$_{-0.19}$	**0.32**$_{-0.08}$	0.32$_{-0.02}$	**0.54**$_{-0.16}$	**1.05**$_{-0.28}$	**0.63**$_{-0.16}$

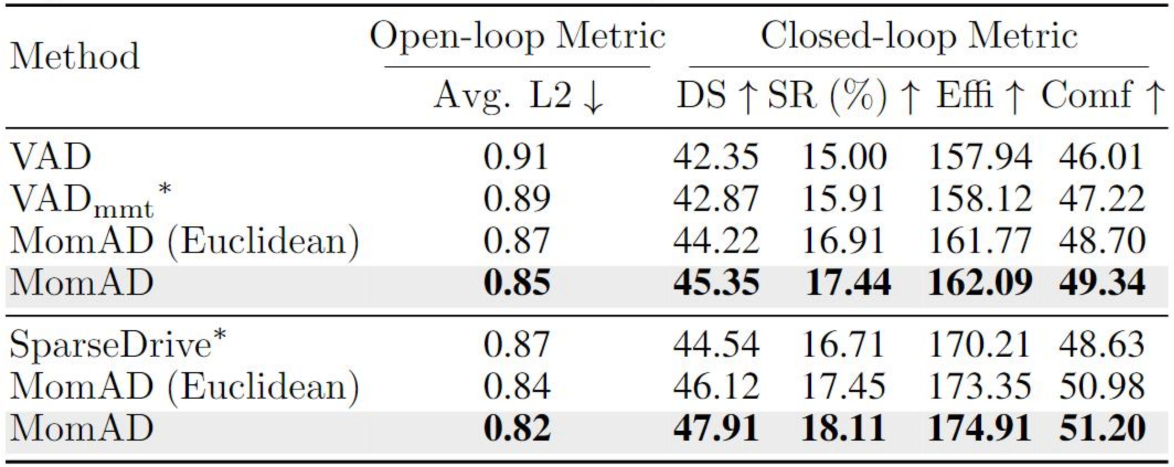
Experiments

Table 3. Long trajectory planning results on the nuScenes and Turning-nuScenes validation sets. We train models for 10 epochs for 6s-horizon prediction. T-nuScenes indicates the challenging Turning-nuScenes. We follow the VAD [21] evaluation metric.



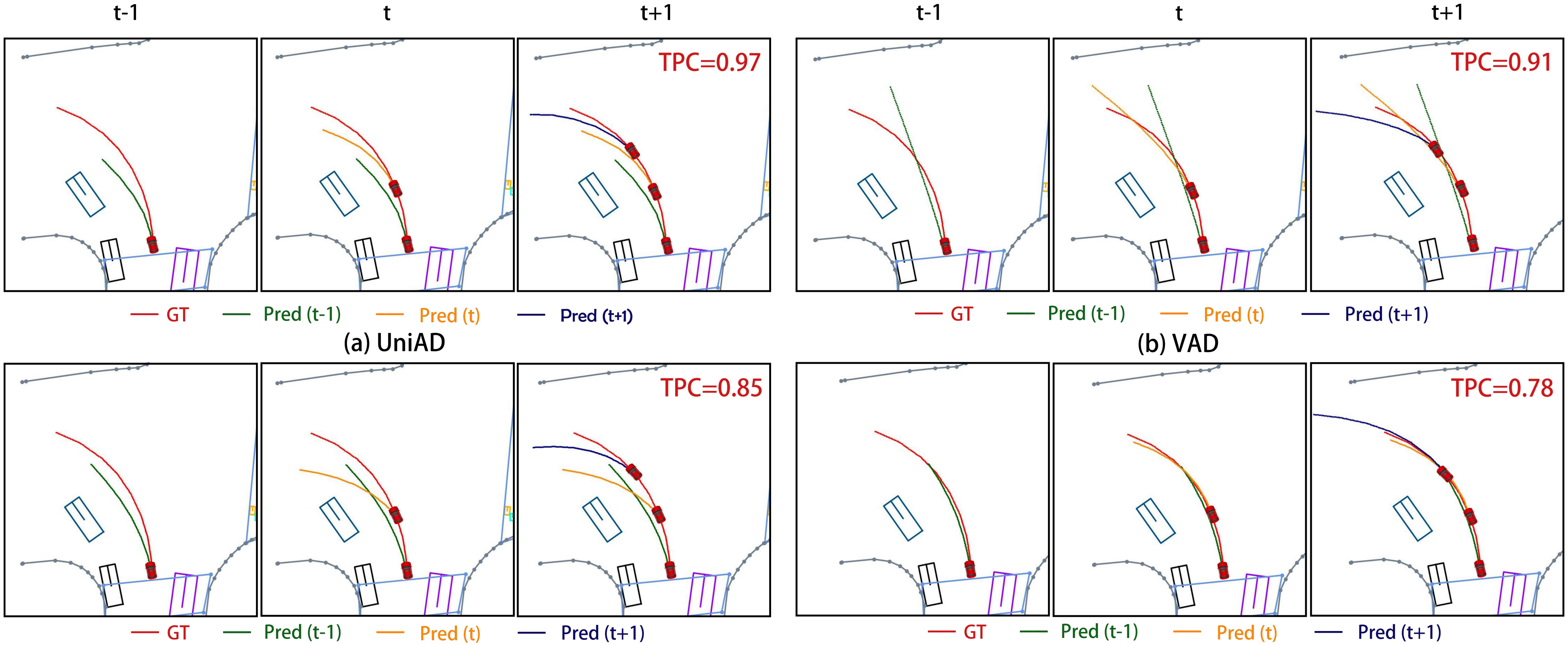
Split	Method	L2 (m) ↓			Col. Rate (%) ↓			TPC (m) ↓		
		4s	5s	6s	4s	5s	6s	4s	5s	6s
nuScenes	SparseDrive [30]	1.75	2.32	2.95	0.87	1.54	2.33	1.33	1.66	1.99
	MomAD	1.67	1.98	2.45	0.83	1.43	2.13	1.19	1.45	1.61
		-0.09	*-0.34*	*-0.50*	*-0.04*	*-0.11*	*-0.20*	*-0.14*	*-0.21*	*-0.38*
T-nuScenes	SparseDrive [30]	2.07	2.71	3.36	0.91	1.71	2.57	1.54	2.31	2.90
	MomAD	1.80	2.07	2.51	0.85	1.57	2.31	1.37	1.58	1.93
		-0.27	*-0.64*	*-0.85*	*-0.06*	*-0.14*	*-0.26*	*-0.17*	*-0.73*	*-0.97*

Table 4. Open-loop and Closed-loop results on Bench2Drive (V0.0.3) under base training set. 'mmt' refers multi-modal trajectory variant of VAD and * the re-implementation.



Method	Open-loop Metric	Closed-loop Metric			
	Avg. L2 ↓	DS ↑	SR (%) ↑	Effi ↑	Comf ↑
VAD	0.91	42.35	15.00	157.94	46.01
$\text{VAD}_{\text{mmt}}{}^*$	0.89	42.87	15.91	158.12	47.22
MomAD (Euclidean)	0.87	44.22	16.91	161.77	48.70
MomAD	**0.85**	**45.35**	**17.44**	**162.09**	**49.34**
SparseDrive*	0.87	44.54	16.71	170.21	48.63
MomAD (Euclidean)	0.84	46.12	17.45	173.35	50.98
MomAD	**0.82**	**47.91**	**18.11**	**174.91**	**51.20**

Experiments



(a) UniAD

(b) VAD

(c) SparseDrive

(d) MomAD (Ours)

Thanks !

Arxiv

Github