Raw2Drive: Reinforcement Learning with Aligned World Models for End-to-End Autonomous Driving (in CARLA v2)

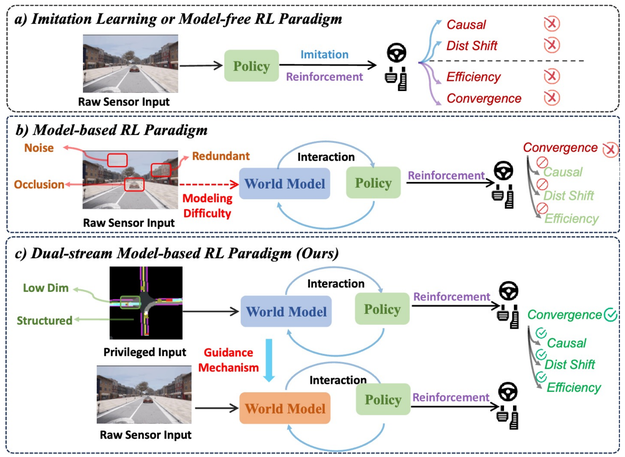
Zhenjie Yang, Xiaosong Jia†, Qifeng Li, Xue Yang, Maoqing Yao, Junchi Yan†
Shanghai Jiao Tong University, Fudan University, AgiBot
† Correspondence Author

NeurIPS 2025

Abstract

Reinforcement Learning (RL) can mitigate the causal confusion and distribution shift inherent in imitation learning (IL). However, applying RL to end-to-end autonomous driving (E2E-AD) remains an open problem for its training difficulty, and IL is still the mainstream paradigm in both academia and industry. Recently Model-based Reinforcement Learning (MBRL) have demonstrated promising results in neural planning; however, these methods typically require privileged information as input rather than raw sensor data. We fill this gap by designing Raw2Drive, a dual-stream MBRL approach.

Introduction

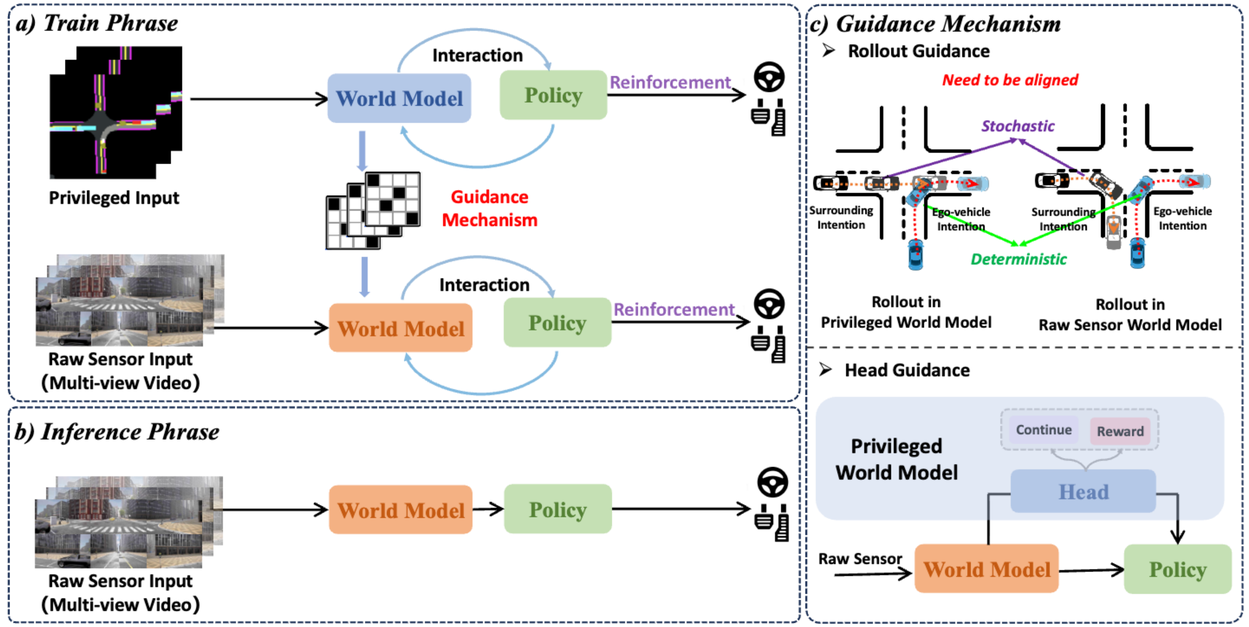


Comparison of different training paradigms in end-to-end autonomous driving. (a) Imitation Learning suffers from causal confusion and distribution shift. Model-free Reinforcement Learning faces efficiency problem and fails to converge. (b) Model-based Reinforcement Learning: There are no reported such works for raw sensor input E2E-AD as the raw data can be noisy and redundant, and Think2Drive assumes the privileged ground truth data is given, which cannot be directly applied in real-world AD. (c) In Raw2Drive, we propose the first feasible model-based reinforcement learning paradigm for end-to-end autonomous driving. By leveraging low-dimensional, structured privileged input, our approach guides the learning of a world model from raw sensor data, effectively addressing the issues outlined in (a) and (b).

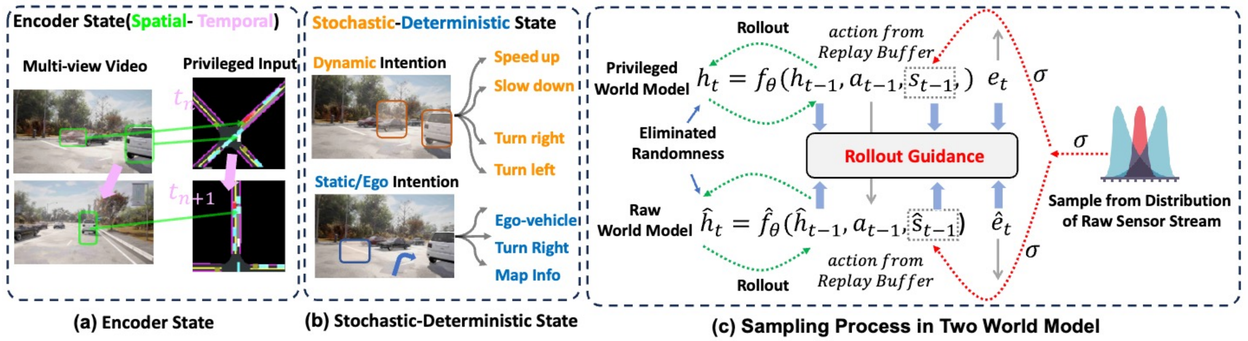
Our contribution:

1. Raw2Drive is the first MBRL framework for E2E-AD, i.e. from raw image input to planning, beyond existing IL or privileged input based RL approaches.·
2. Raw2Drive achieves state-of-the-art performance on the challenging CARLA v2 and Bench2Drive and surpasses IL methods by a large margin, validating the power of RL.·
3. We only use 64 H800 GPU days in total to deliver our final planner, and the cost can be further saved to 40 GPU days when Think2Drive is reused which dismisses our phase I training. In comparison, IL-based UniAD costs about 30 GPU days yet it only solves even 3~4 corner cases in CARLA v2.

Methodology



The Pipeline of Raw2Drive. (a) During training, we use privileged input to train the privileged world model and paired policy. Then, the privileged world model is used to guide the training of the raw sensor stream. (b) During inference, only raw sensor inputs are available, which aligns with real-world autonomous driving. (c) The guidance mechanism consists of two parts: (I) Rollout Guidance to ensure future modeling consistency; (II) Head Guidance to ensure the supervision for raw sensor policy is accurate and stable.



State Variables Aligned and Sampling Process in the Rollout Guidance. (a) The encoder state is aligned temporally and spatially. (b) The deterministic state and stochastic state is aligned to maintain dynamic and static intention consistency. (c) Eliminating Cumulative Errors Caused By Sampling. During the rollout process, when deducting next states, we only sample once for the stochastic state - from the distribution of the raw sensor stream. The sampled state is fed into both streams to eliminate the randomness and thus the alignment is more stable.

Results

Table 3: **Performance on Carla Official Town13 Validation and Devtest Benchmark**. *denotes expert feature distillation. As discussed in carla-garage [48] and Section 4.3, long routes evaluation in Leaderboard 2.0 can't reflect the actual driving performance [35].

Method	Venue	Scheme	Modality	Closed-loop Metric					
				DS ↑		RC (%) ↑		IS ↑	
				Validation	Devtest	Validation	Devtest	Validation	Devtest
AD-MLP [49]	Arxiv 2023	IL	State	0.00	0.00	0.00	0.00	0.00	0.00
UniAD-Base [9]	CVPR 2023	IL	Image	0.15	0.00	0.51	0.07	0.23	0.04
VAD [10]	ICCV 2023	IL	Image	0.17	0.00	0.49	0.06	0.31	0.04
DriveTrans [40]	ICLR 2025	IL	Image	0.85	0.68	1.42	2.13	0.33	0.35
TCP-traj* [14]	NeurIPS 2022	IL	Image	0.31	0.02	0.89	0.11	0.24	0.05
ThinkTwice* [16]	CVPR 2023	IL	Image	0.50	0.64	1.23	1.78	0.35	**0.43**
DriveAdapter* [17]	ICCV 2023	IL	Image	0.92	0.87	1.52	2.43	0.42	0.37
Raw2Drive (Ours)	-	RL	Image	**4.12**	**3.56**	**9.32**	**6.04**	**0.43**	0.42
Think2Drive [4]	ECCV 2024	RL Expert	Privileged*	43.8	56.8	78.2	98.6	0.73	0.92

Table 4: **Performance on Bench2Drive Multi-Ability Benchmark**. * denotes expert feature distillation. IL represents imitation learning. RL represents reinforcement learning. RL expert Think2Drive [4] uses privileged information for training.

Method	Venue	Scheme	Modality	Ability (%) ↑					
				Merging	Overtaking	Emergency Brake	Give Way	Traffic Sign	Mean
TCP-traj* [14]	NeurIPS 2022	IL	Image	8.89	24.29	51.67	40.00	46.28	34.22
AD-MLP [49]	Arxiv 2023	IL	State	0.00	0.00	0.00	0.00	4.35	0.87
UniAD-Base [9]	CVPR 2023	IL	Image	14.10	17.78	21.67	10.00	14.21	15.55
ThinkTwice* [16]	CVPR 2023	IL	Image	27.38	18.42	35.82	**50.00**	54.23	37.17
VAD [10]	ICCV 2023	IL	Image	8.11	24.44	18.64	20.00	19.15	18.07
DriveAdapter* [17]	ICCV 2023	IL	Image	28.82	26.38	48.76	**50.00**	56.43	42.08
DriveTrans [40]	ICLR 2025	IL	Image	17.57	35.00	48.36	40.00	52.10	38.60
Raw2Drive (Ours)		RL	Image	**43.35**	**51.11**	**60.00**	**50.00**	**62.26**	**53.34**
Think2Drive [4]	ECCV 2024	RL Expert	Privileged*	81.27	83.92	90.24	90.00	87.67	86.26

Table 5: **Results on Bench2Drive Closed-loop Benchmark.** *denotes expert feature distillation. * denotes expert feature distillation. IL represents imitation learning. RL represents reinforcement learning. RL expert Think2Drive [4] uses privileged information for training.

Method	Venue	Scheme	Modality	Closed-loop Metric			
				DS ↑	SR(%) ↑	Efficiency ↑	Comfort ↑
TCP-traj* [14]	NeurIPS 2022	IL	Image	59.90	30.00	76.54	18.08
AD-MLP [49]	Arxiv 2023	IL	State	18.05	0.00	48.45	22.63
UniAD-Base [9]	CVPR 2023	IL	Image	45.81	16.36	129.21	43.58
VAD [10]	ICCV 2023	IL	Image	42.35	15.00	157.94	**46.01**
ThinkTwice* [16]	CVPR 2023	IL	Image	62.44	31.23	69.33	16.22
DriveAdapter* [17]	ICCV 2023	IL	Image	64.22	33.08	70.22	16.01
GenAD [50]	ECCV 2024	IL	Image	44.81	15.90	-	-
DriveTrans [40]	ICLR 2025	IL	Image	63.46	35.01	100.64	20.78
MomAD [51]	CVPR 2025	IL	Image	44.54	16.71	170.21	48.63
Raw2Drive (Ours)	-	RL	Image	**71.36**	**50.24**	**214.17**	22.42
Think2Drive [4]	ECCV 2024	RL Expert	Privileged	91.85	85.41	269.14	25.97

Conclusion

We propose Raw2Drive, the first end-to-end model-based reinforcement learning method in autonomous driving. Our approach introduces a novel dual-stream architecture and a guidance mechanism to effectively enable MBRL learning. The approach is fulfilled by the careful treatment of the two world models for the raw and privileged information, respectively, and achieve the state-of-the-art performance on lately released benchmarks. We hope this work serves as a stepping stone toward exploring reinforcement learning for end-to-end autonomous driving.

Limitations: In our setting, the privileged input is ground truth bounding boxes and HD-Map. And in real-world autonomous driving, for industry, the ground truth bounding boxes and HD-Map can be from human annotation or advanced perception algorithms. Reinforcement learning in the real world is also a technical issue that would be solved by 3DGS or a diffusion-based simulator in the future. While CARLA remains the only viable closed-loop simulator for RL research at present, our work focuses on policy learning and introduces a dual-stream world model design that is conceptually decoupled from the specific simulator.