AutoDrive-P^3: Unified Chain of Perception–Prediction–Planning Thought via Reinforcement Fine-Tuning

Yuqi Ye[†,1], *Zijian Zhang*[†,1], *Junhong Lin*[1], *Shangkun Sun*[1], *Changhao Peng*[1], *Wei Gao*[*,1]

[1]*School of Electronic and Computer Engineering, Peking University*

[†]*Core Authors*, * ✉ *gaowei262@pku.edu.cn*

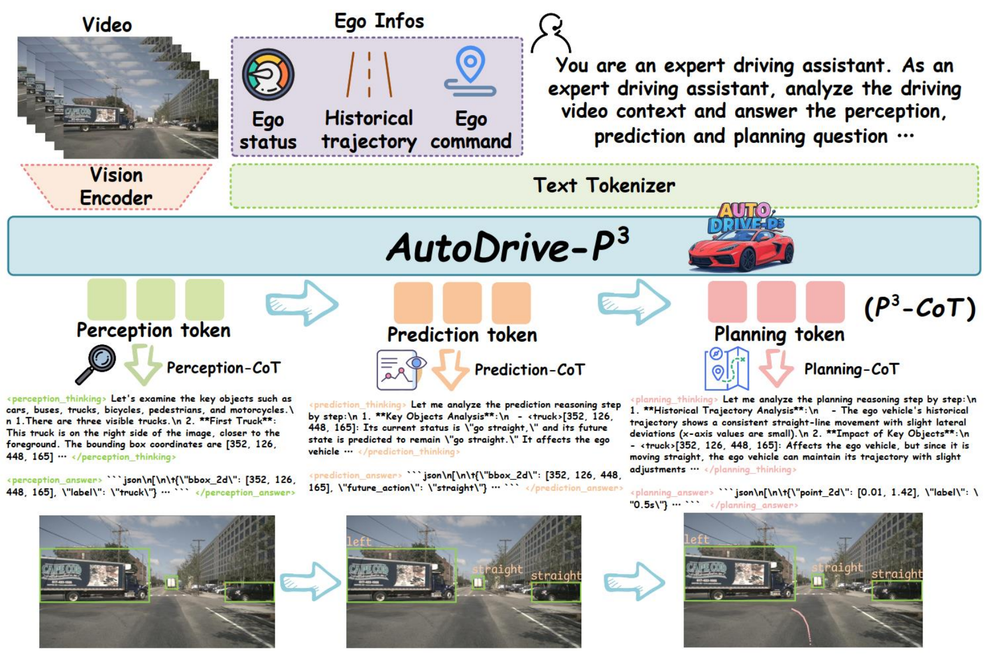
Spatial-Temporal AI Research Lab
Peking University

北京大学信息工程学院
School of Electronic and Computer Engineering
Peking University

Abstract

Vision-language models (VLMs) are increasingly being adopted for end-to-end autonomous driving systems due to their exceptional performance in handling long-tail scenarios. However, current VLM-based approaches suffer from two major limitations: 1) Some VLMs directly output planning results without chain-of-thought (CoT) reasoning, bypassing crucial perception and prediction stages which creates a significant domain gap and compromises decision making capability; 2) Other VLMs can generate outputs for perception, prediction, and planning tasks but employ a fragmented decision-making approach where these modules operate separately, leading to a significant lack of synergy that undermines true planning performance. To address these limitations, we propose AutoDrive-P^3, a novel framework that seamlessly integrates Perception, Prediction, and Planning through structured reasoning. We introduce the P3-CoT dataset to facilitate coherent reasoning and propose P^3-GRPO, a hierarchical reinforcement learning algorithm that provides progressive supervision across all three tasks. Specifically, AutoDrive-P3 progressively generates CoT reasoning and answers for perception, prediction, and planning, where perception provides essential information for subsequent prediction and planning, while both perception and prediction collectively contribute to the final planning decisions, enabling safer and more interpretable autonomous driving. Additionally, to balance inference efficiency with performance, we introduce dual thinking modes: detailed thinking and fast thinking. Extensive experiments on both open-loop (nuScenes) and closed-loop (NAVSIMv1/v2) benchmarks demonstrate that our approach achieves state-of-the-art performance in planning tasks.

AutoDrive-P^3 Pipeline



AutoDrive-P3 processes video and ego vehicle data through structured Perception-Prediction-Planning Chain-of-Thought (P^3-CoT) reasoning, generating interpretable step-by-step rationale and structured outputs for perception, prediction, and planning.

P^3-GRPO Pipeline

We first cold start the base model using P^3-CoT to make up for the gap between VLM and autonomous driving and learn the CoT answer format. Next we further enhance the VLM's reasoning capability across all three stages by applying the GRPO algorithm to the perception, prediction, and planning modules collectively, yielding the P^3-GRPO algorithm. Our approach employs a multicomponent reward function to guide the policy model toward generating accurate, coherent, and well-structured outputs through coordinated reinforcement learning across these cognitive layers.

Results

Table 1: **Performance comparison on nuScenes Benchmark.**

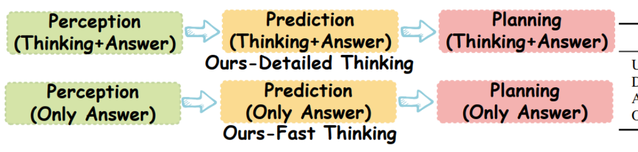
Method	L2 (m) ↓				Collision (%) ↓				VLM
	1s	2s	3s	Avg.	1s	2s	3s	Avg.	
Non-Autoregressive Methods									
ST-P3 (Hu et al., 2022)	1.33	2.11	2.90	2.11	0.23	0.62	1.27	0.71	-
VAD (Jiang et al., 2023)	0.17	0.34	0.60	0.37	0.07	0.10	0.24	0.14	-
Ego-MLP (Li et al., 2024d)	0.46	0.76	1.12	0.78	0.21	0.35	0.58	0.38	-
UniAD (Hu et al., 2023)	0.44	0.67	0.96	0.69	0.04	0.08	0.23	0.12	-
InsightDrive (Song et al., 2025)	0.23	0.41	0.68	0.44	0.09	0.10	0.27	0.15	-
Autoregressive Methods									
GPT-Driver (Mao et al., 2023)	0.20	0.40	0.70	0.44	0.04	0.12	0.36	0.17	GPT-3.5
DriveVLM (Tian et al., 2024)	0.18	0.34	0.68	0.40	0.10	0.22	0.45	0.27	Qwen2-VL-7B
OpenEMMA (Xing et al., 2025)	1.45	3.21	3.76	2.81	-	-	-	-	Qwen2-VL-7B
RDA-Driver (Huang et al., 2024a)	0.17	0.37	0.69	0.40	0.01	0.05	0.26	0.10	LLaVa-7B
OmniDrive (Wang et al., 2024)	**0.14**	**0.29**	0.55	**0.33**	0.01	0.04	0.27	0.11	LLava-7B
OpenDriveVLA (Zhou et al., 2025a)	**0.14**	0.30	0.55	**0.33**	0.02	0.07	0.22	0.10	Qwen2.5-VL-3B
AutoVLA (Zhou et al., 2025b)	0.25	0.46	0.73	0.48	0.07	0.07	0.26	0.13	Qwen2.5-VL-3B
AutoDrive-R² (Yuan et al., 2025)	0.35	0.49	0.62	0.49	-	-	-	-	Qwen2.5-VL-3B
AutoDrive-P^3 (Ours-Detailed)	0.15	0.30	**0.54**	**0.33**	**0.00**	**0.02**	**0.15**	**0.06**	Qwen2.5-VL-3B
AutoDrive-P^3 (Ours-Fast)	0.16	0.31	0.56	0.34	**0.00**	0.04	0.20	0.08	Qwen2.5-VL-3B

Table 2: **Performance comparison on NAVSIMv1 benchmark.**

Method	Image	Lidar	NC↑	DAC↑	EP↑	TTC↑	Comf↑	PDMS↑
Human	✗	✗	100.0	100.0	87.5	100.0	99.9	94.8
Constant Velocity	✗	✗	69.9	58.8	49.3	49.3	100.0	21.6
Ego Status MLP	✗	✗	93.0	77.3	62.8	83.6	100.0	65.6
VADv2 (Weng et al., 2024)	✔	✗	97.9	91.7	77.6	92.9	100.0	83.0
UniAD (Hu et al., 2023)	✔	✗	97.8	91.9	78.8	92.9	100.0	83.4
LTF (Prakash et al., 2021)	✔	✗	97.4	92.8	79.0	92.4	100.0	83.8
TransFuser (Prakash et al., 2021)	✔	✔	97.7	92.8	79.2	92.8	100.0	84.0
PARA-Drive (Weng et al., 2024)	✔	✗	97.9	92.4	79.3	93.0	99.8	84.0
LAW (Li et al., 2024a)	✔	✔	96.4	95.4	81.7	88.7	99.9	84.6
DRAMA (Yuan et al., 2024)	✔	✔	98.0	93.1	80.1	94.8	100.0	85.5
Hydra-MDP (Li et al., 2024b)	✔	✔	98.3	96.0	78.7	94.6	100.0	86.5
DiffusionDrive (Liao et al., 2025)	✔	✔	98.2	96.2	82.2	94.7	100.0	88.1
WoTE (Li et al., 2025)	✔	✔	98.5	96.8	81.9	94.9	99.9	88.3
AutoDrive-P^3 (Ours-Detailed)	✔	✗	**99.1**	97.4	**84.8**	96.5	**100.0**	**90.6**
AutoDrive-P^3 (Ours-Fast)	✔	✗	98.9	**97.7**	83.7	**96.6**	99.9	90.2

Table 3: **Performance comparison on NAVSIMv2 benchmark.**

Method	NC↑	DAC↑	DDC↑	TLC↑	EP↑	TTC↑	LK↑	HC↑	EC↑	EPDMS↑ False / True
Human	100.0	100.0	99.8	100.0	87.4	100.0	100.0	98.1	90.1	90.3 / 94.5
Ego Status MLP	93.1	77.9	92.7	99.6	86.0	91.5	89.4	98.3	85.4	64.0 / -
Transfuser (Prakash et al., 2021)	96.9	89.9	97.8	99.7	87.1	95.4	92.7	98.3	87.2	76.7 / 84.0
HydraMDP++ (Li et al., 2024b)	97.2	97.5	99.4	99.6	83.1	96.5	94.4	98.2	70.9	81.4 / -
DiffusionDrive (Liao et al., 2025)	98.2	96.2	**99.5**	**99.8**	87.4	97.3	**96.9**	**98.4**	**87.7**	84.7 / 88.2
WoTE (Li et al., 2025)	98.5	96.8	98.8	**99.8**	86.1	97.9	95.5	98.3	82.9	84.2 / 87.7
AutoDrive-P^3 (Ours-Detailed)	**99.1**	97.4	99.2	**99.8**	**88.0**	**98.7**	96.3	98.3	85.5	**86.2 / 89.9**
AutoDrive-P^3 (Ours-Fast)	98.9	**97.6**	98.9	**99.8**	86.8	98.5	95.4	98.3	80.6	85.2 / 88.7



Method	Avg. L2 (m) ↓	Avg. Collision (%) ↓	FPS (Hz) ↑
UniAD (23' CVPR)	0.69	0.12	1.8
DriveVLM (24' CoRL)	0.40	0.27	**2.4**
AutoVLA (Slow / Fast, 25' NeurIPS)	0.48 / 0.43	0.13 / 0.19	0.1 / 0.9
Ours (Detailed / Fast)	**0.33** / 0.34	**0.06** / 0.08	0.5 / 1.0

Figure 5: **Dual thinking modes and running time on nuScenes Benchmark.**

Table 5: **Ablation study on different training setting on nuScenes benchmark.**

Method	Group Size	History Traj.	Sensor Type	L2 (m) ↓				Collision (%) ↓			
				1s	2s	3s	Avg.	1s	2s	3s	Avg.
Ablation 1	4	✔	Video	0.17	0.32	0.65	0.38	0.01	0.06	0.30	0.13
Ablation 2	8	✗	Video	0.17	0.33	0.68	0.39	0.02	0.07	0.33	0.14
Ablation 3	8	✔	Image	0.16	0.32	0.61	0.36	0.01	0.05	0.26	0.12
P^3-GRPO	8	✔	Video	**0.15**	**0.30**	**0.54**	**0.33**	**0.00**	**0.02**	**0.15**	**0.06**