EgoFSD: Ego-Centric Fully Sparse Paradigm with Uncertainty Denoising and Iterative Refinement for End-to-End Self-Driving

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY
HKUST
商汤 sensetime
绝影 SenseAuto

Haisheng Su[1], Wei Wu[2], Zhenjie Yang[1], Isabel Guan[3,♡]

[1]Shanghai Jiao Tong University [2]SenseAuto [3]The Hong Kong University of Science and Technology
♡ Corresponding Author

Introduction

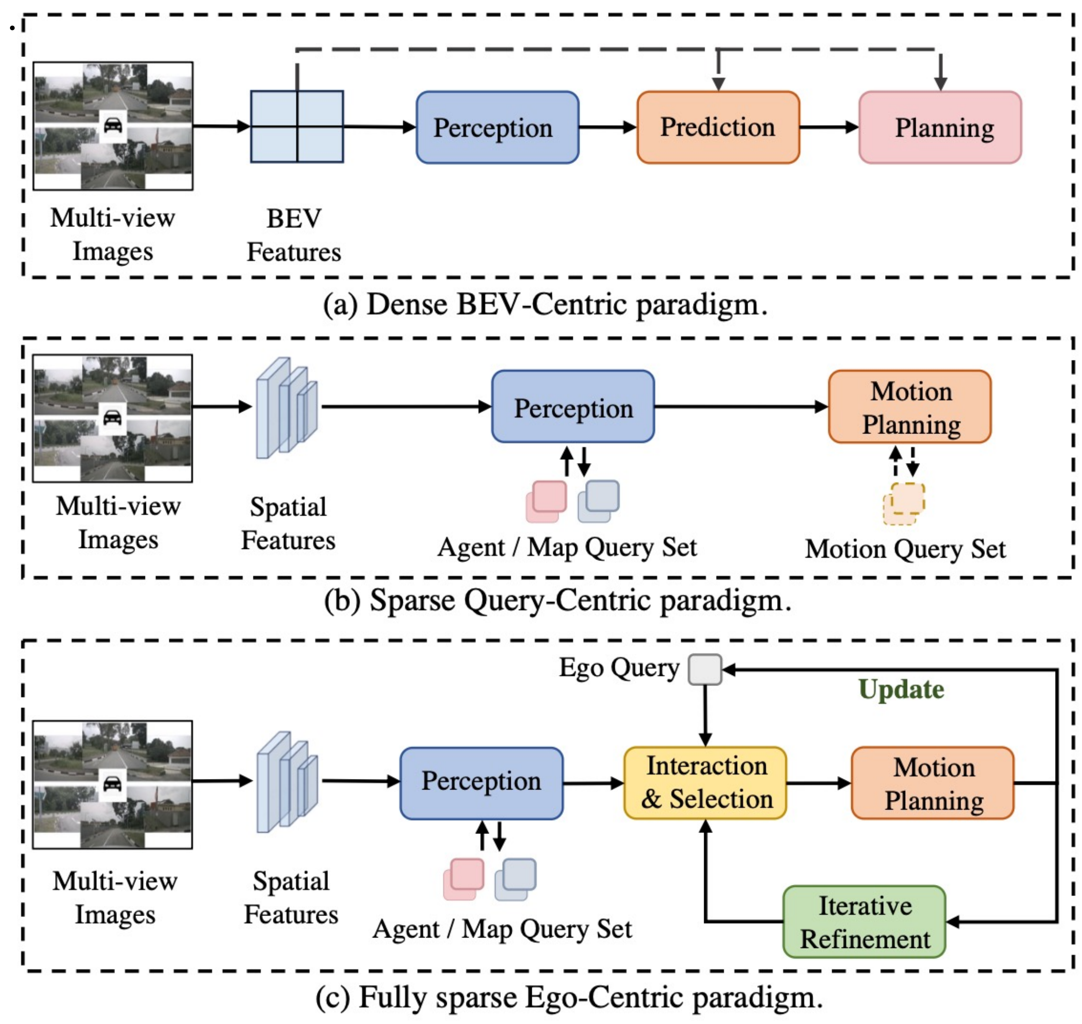
Motivation

Existing end-to-end driving systems lacking ego-centric designs still suffer from unsatisfactory performance and inferior efficiency, due to rasterized scene representation learning and redundant information transmission.

Contribution

- ✓ An **Ego-Centric Fully Sparse** paradigm for end-to-end self-Driving, named EgoFSD, without dense representation learning and redundant environmental modeling.
- ✓ A geometric prior through intention-guided attention, where the **Closest In-Path Vehicle / Stationary** (CIPV / CIPS) are gradually picked out through ego-centric cross attention and selection.
- ✓ Both **position-level** diffusion of interactive agents and **trajectory-level** denoising of ego-vehicle are adopted for uncertainty motion modeling.
- ✓ SOTA performance achieved on both **Open-loop** and **Closed-loop** planning evaluation.

Comparison of E2E-AD paradigms

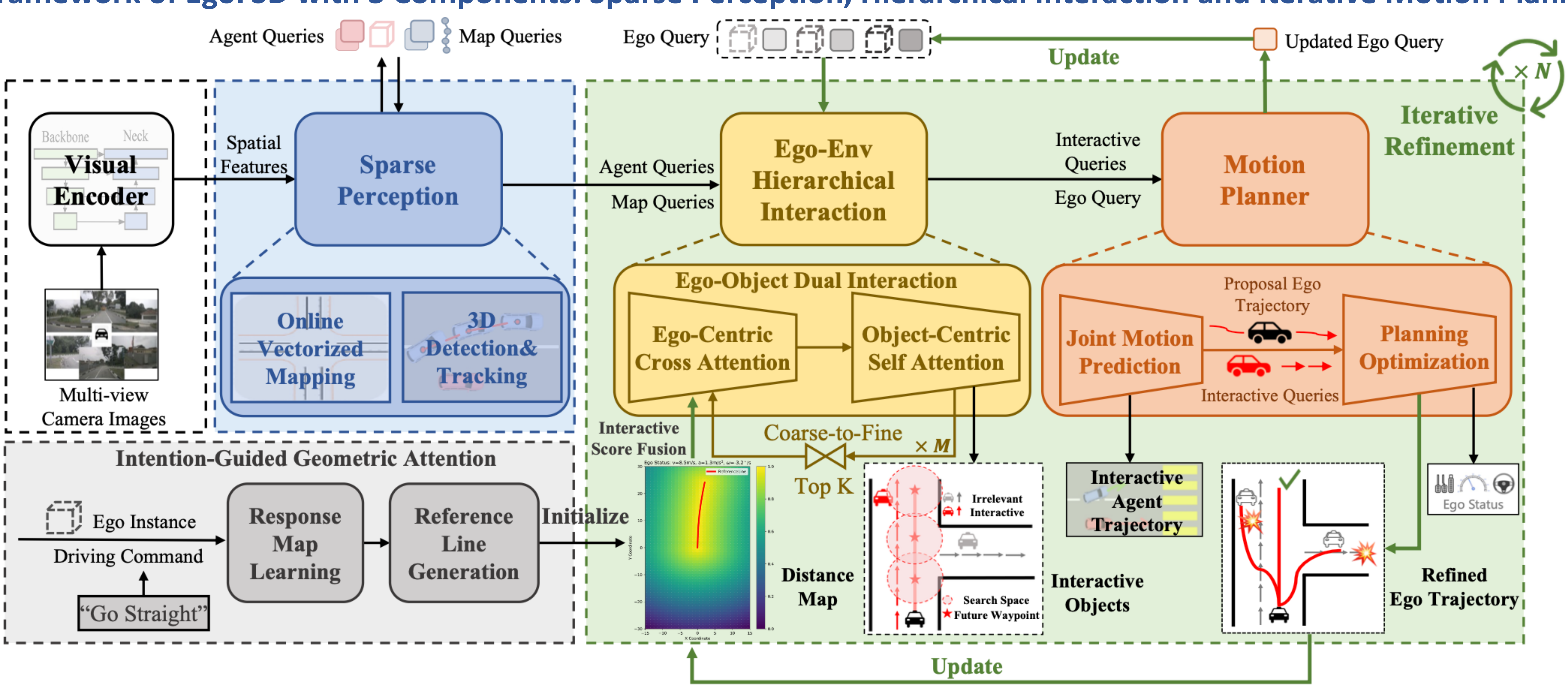


(a) Dense BEV-Centric paradigm.

(b) Sparse Query-Centric paradigm.

(c) Fully sparse Ego-Centric paradigm.

Framework of EgoFSD with 3 Components: Sparse Perception, Hierarchical Interaction and Iterative Motion Planner



Comprehensive Benchmarking of EgoFSD on both Open-Loop and Closed-Loop Planning Capability

Open-loop Planning Evaluation on nuScenes val dataset

EgoFSD-S can achieve great efficiency with 16.7 FPS, 9.3×and 3.7×faster than UniAD and VAD respectively.

Protocol	Method	Backbone	L2 (m) ↓ 1s	2s	3s	Avg.	Collision (%) ↓ 1s	2s	3s	Avg.	Latency (ms) ↓	FPS ↑
ST-P3 Metrics	FF* [10]	-	0.55	1.20	2.54	1.43	0.06	0.17	1.07	0.43	-	-
	EO* [20]	-	0.67	1.36	2.78	1.60	0.04	0.09	0.88	0.33	-	-
	ST-P3 [11]	EfficientNet-b4	1.33	2.11	2.90	2.11	0.23	0.62	1.27	0.71	628.3	1.6
	VAD-Base [17]	ResNet50	0.41	0.70	1.05	0.72	0.07	0.17	0.41	0.22	224.3	4.5
	SparseDrive-S [43]†	ResNet50	0.30	0.58	0.95	0.61	0.47	0.47	0.69	0.54	111.1	9.0
	EgoFSD-S (BEV)	ResNet50	**0.16**	**0.33**	**0.59**	**0.35**	**0.00**	**0.04**	**0.18**	**0.07**	67.7	14.8
SparseDrive Metrics‡	UniAD [12]†	ResNet101-DCN	0.45	0.70	1.04	0.73	0.62	0.58	0.63	0.61	555.6	1.8
	VAD-Base [17]†	ResNet50	0.41	0.70	1.05	0.72	0.03	0.19	0.43	0.21	224.3	4.5
	SparseDrive-S [43]	ResNet50	0.29	0.58	0.96	0.61	0.01	0.05	0.18	0.08	111.1	9.0
	SparseDrive-B [43]	ResNet101	0.29	0.55	0.91	0.58	0.01	**0.02**	**0.13**	**0.06**	137.0	7.3
	EgoFSD-S (BEV)	ResNet50	0.16	0.33	0.59	0.35	0.03	0.07	0.21	0.10	67.7	14.8
	EgoFSD-S (PV)	ResNet50	0.15	0.31	0.56	0.33	**0.00**	0.06	0.19	0.08	**59.8**	**16.7**
	EgoFSD-B (PV)	ResNet101	**0.12**	**0.28**	**0.52**	**0.30**	**0.00**	0.04	0.15	**0.06**	80.5	12.4

Closed-loop Planning Evaluation on Bench2Drive dataset

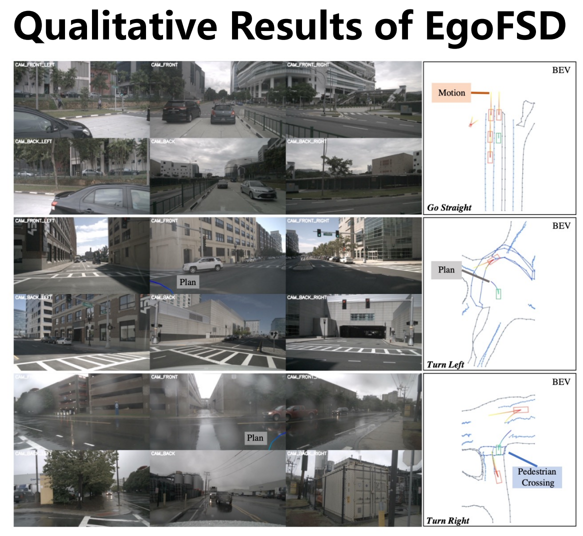
EgoFSD-S achieves both the lowest L2 error and best closed-loop performance with great efficiency, showcasing the superiority and generalizability of our proposed method.

Method	Open-loop Metric: Avg. L2 ↓ (m)	Closed-loop Metrics: Driving Score ↑	Success Rate ↑ (%)	Efficiency ↑
AD-MLP [45]	3.64	18.05	0.00	48.45
UniAD-Tiny [12]	0.80	40.73	13.18	123.92
UniAD-Base [12]	0.73	45.81	16.36	129.21
VAD [16]	0.91	42.35	15.00	157.94
GenAD [48]	-	44.81	15.90	-
MomAD [34]	0.82	44.54	16.71	170.21
EgoFSD-S	0.70	52.02	21.00	178.30
EgoFSD-B	**0.66**	**60.39**	**31.78**	**180.63**

Necessity of Geometric Prior

Object Selection	Geometric Attention	Planning L2 (m) ↓ 1s	2s	3s	Avg.	Planning Coll. (%) ↓ 1s	2s	3s	Avg.
100%	✗	0.27	0.47	0.74	0.49	0.10	0.21	0.37	0.22
Random (5%)	✗	0.28	0.49	0.79	0.52	0.08	0.17	0.38	0.21
Random (2%)	✗	0.33	0.57	0.87	0.59	0.18	0.30	0.51	0.33
0%	✗	2.25	3.75	5.26	3.75	2.82	5.42	6.39	4.88
Attn (5%)	✗	0.16	0.34	0.63	0.38	0.07	0.09	0.31	0.16
Attn (2%)	✗	0.16	0.34	0.61	**0.37**	0.06	0.08	0.24	**0.12**
Attn (2%)	Random	0.17	0.36	0.67	0.40	0.07	0.10	0.34	0.17
Attn (2%)	GroundTruth	0.14	0.23	0.33	0.23	0.07	0.08	0.10	0.07
Attn (2%)	✓	0.16	0.33	0.59	**0.35** (-5%)	0.00	0.04	0.18	**0.07** (-42%)

Qualitative Results of EgoFSD



Effect of Iterative Refinement stages

Number of Stages	Planning L2 (m) ↓ 1s	2s	3s	Avg.	Planning Coll. (%) ↓ 1s	2s	3s	Avg.
1	0.16	0.33	0.61	0.37	0.01	0.08	0.23	0.11
2	0.16	0.33	0.59	**0.35**	0.00	0.04	0.18	**0.07**
3	0.16	0.33	0.60	0.36	0.01	0.40	0.22	0.09
4	0.16	0.33	0.61	0.36	0.00	0.04	0.20	0.08

ICRA™ IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION
2026 VIENNA