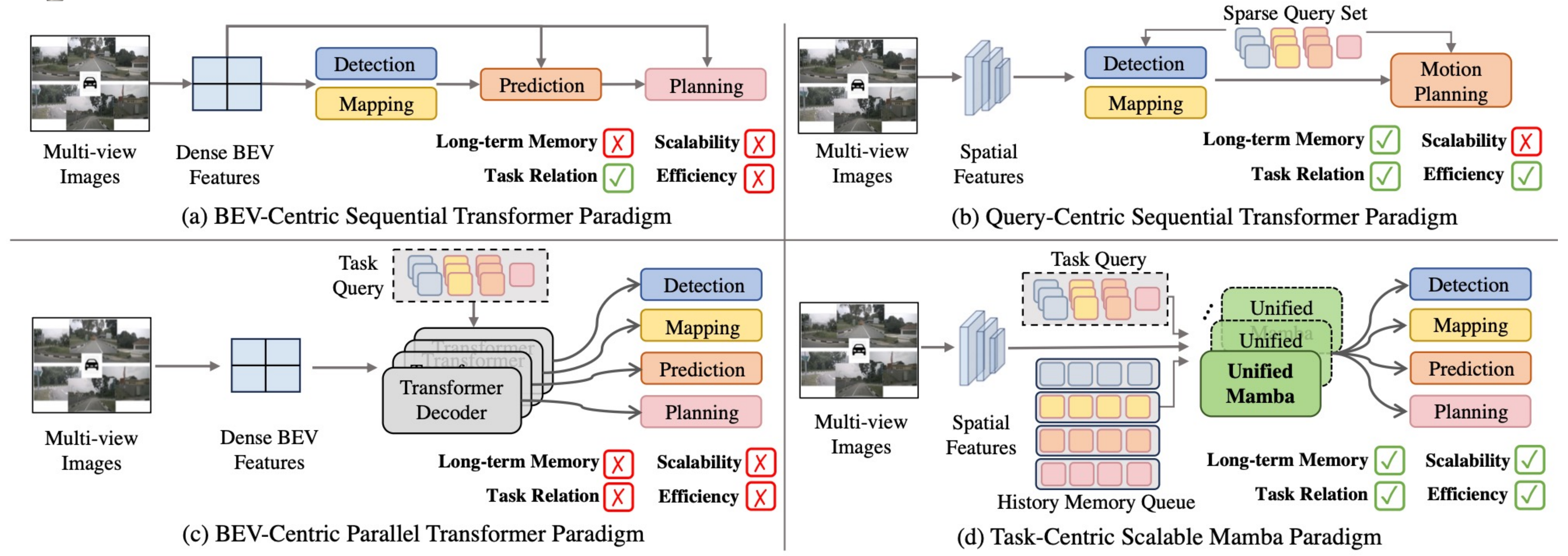
DriveMamba: Task-Centric Scalable State Space Model for Efficient End-to-End Autonomous Driving

Haisheng Su[1], Wei Wu[2], Feixiang Song[2], Junjie Zhang[2], Zhenjie Yang[1], Junchi Yan[1,3] ♡

[1]Shanghai Jiao Tong University [2]SenseAuto [3] Shanghai Innovation Institute ♡Corresponding Author

Motivation and Contribution of DriveMamba

Motivation



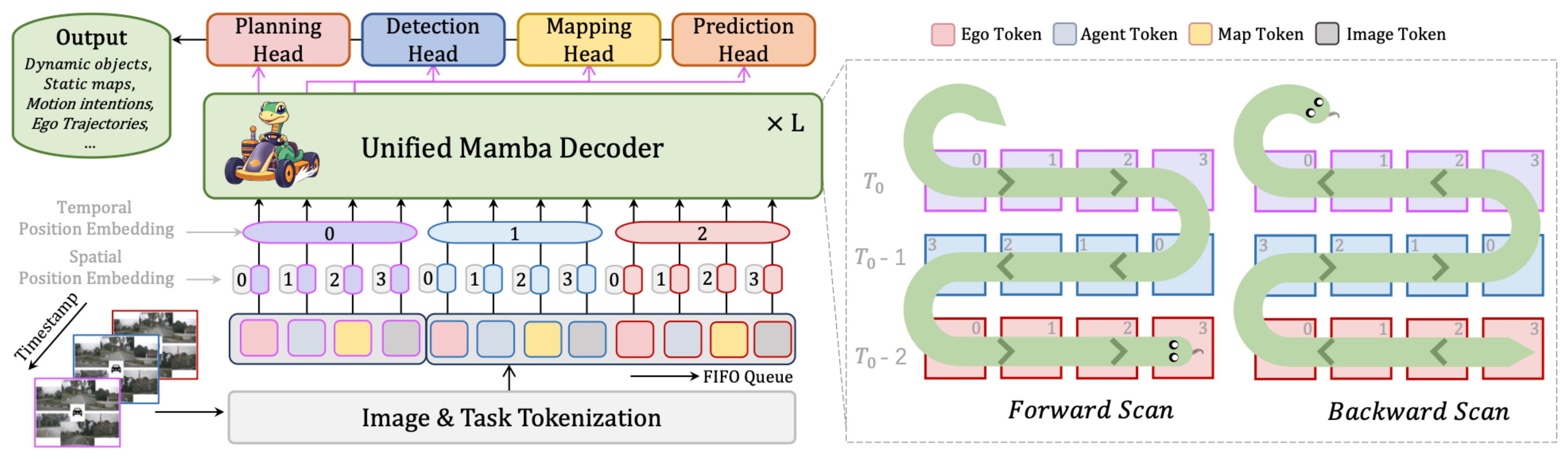
- ✓ Sequential paradigms suffer from information loss and cumulative errors and hinder the task synergies learning (e.g. diverse and flexible task relations).
- ✓ BEV-Centric paradigms are computational expensive for long-term and long-range modeling.
- ✓ Existing Query-Centric paradigms have limited scalability owing to quadratic-complexity attention and insufficient interaction order for downstream ego-planning.

Contribution:

- ✓ A Task-Centric paradigm for end-to-end autonomous driving, which integrates **view correspondence learning**, task relation modeling and long-term temporal fusion into a Unified framework with high efficiency and scalability.
- ✓ A Hybrid SpatioTemporal Scan method to capture task-related inter-dependencies from trajectory-guided "Local-to-Global" and "Spatial-to-Temporal" bidirectionally.
- ✓ SOTA performance achieved on both open-loop and closed-loop planning evaluation (e.g. nuScenes and B2D).

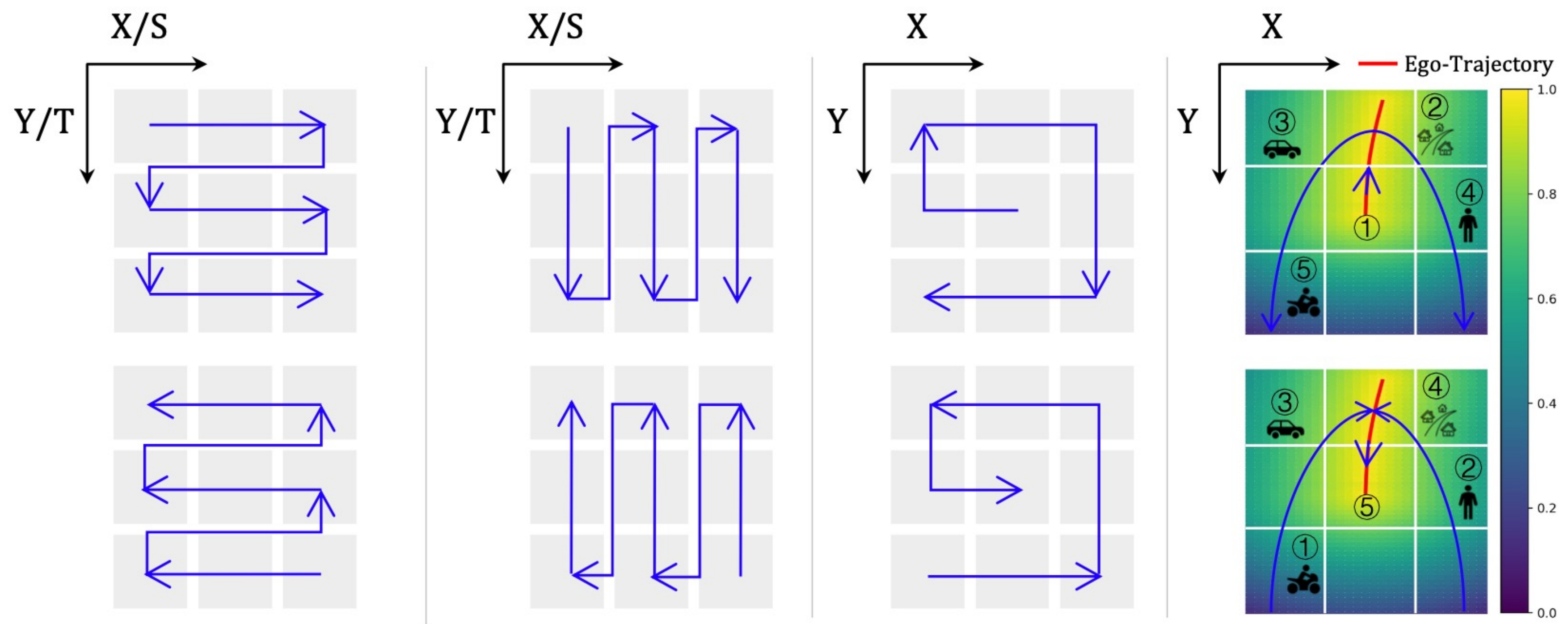
Design of DriveMamba

➢ Framework



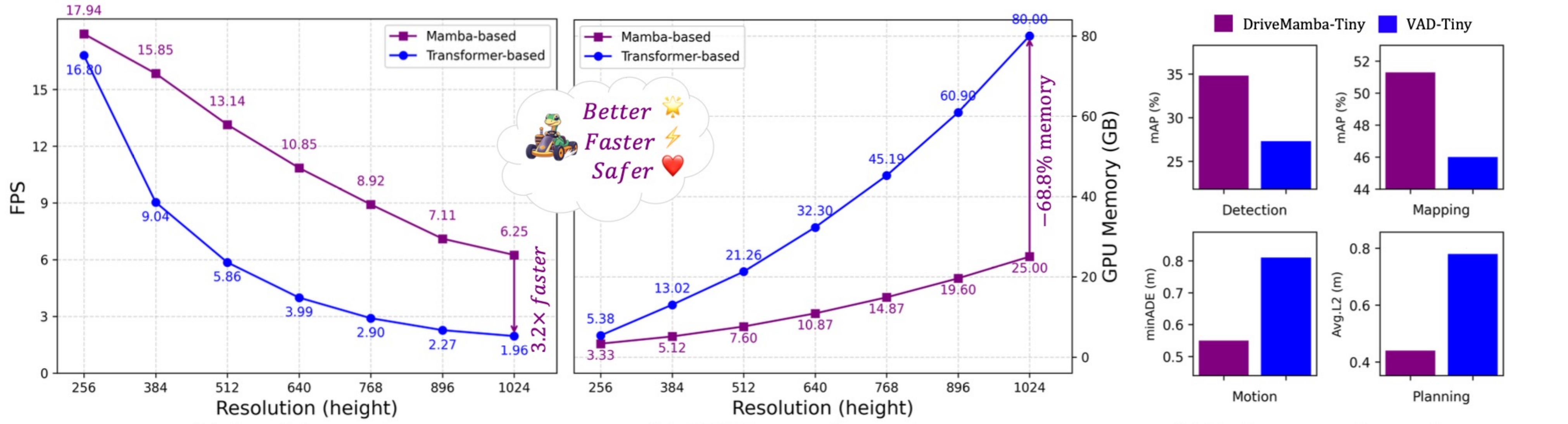
(a) *DriveMamba* (b) *Hybrid Spatiotemporal Scan*

➢ Different bidirectional scan methods



(a) *Horizontal/Spatial – First Bidirectional* (b) *Vertical/Temporal – First Bidirectional* (c) *Ego – L2G Bidirectional* (d) *Trajectory – L2G Bidirectional*

➢ Comparisons between Transformer-based E2E-AD and DriveMamba



(a) Speed Comparison (b) GPU Memory Comparison (c) Performance Comparison

Experimental Results

➢ Quantative Results on nuScenes dataset

Method	L2 (m) ↓ 1s	2s	3s	Avg.	Collision (%) ↓ 1s	2s	3s	Avg.	FPS ↑
ST-P3 [11]	1.33	2.11	2.90	2.11	0.23	0.62	1.27	0.71	1.6
OccNet [40]	1.29	2.13	2.99	2.13	0.21	0.59	1.37	0.72	-
UniAD [12]	0.48	0.74	1.07	0.76	0.12	0.13	0.28	0.17	1.8
VAD-Base [17]	0.41	0.70	1.05	0.72	0.07	0.17	0.41	0.22	4.5
BEVPlanner [22]	0.27	0.54	0.90	0.57	0.04	0.35	1.80	0.73	-
DriveMamba-Tiny (**Ours**)	0.25	0.42	0.66	0.44	0.12	0.09	0.24	0.15	**17.9**
DriveMamba-Small (**Ours**)	0.23	0.40	0.64	0.42	0.06	0.07	0.22	0.12	11.1
DriveMamba-Base (**Ours**)	0.22	0.40	0.63	0.41	0.05	0.06	**0.21**	0.11	6.1
DriveMamba-Large (**Ours**)	**0.21**	**0.38**	**0.62**	**0.40**	**0.01**	**0.04**	**0.21**	**0.09**	1.7
AD-MLP [49]‡	0.20	**0.26**	**0.41**	**0.29**	0.17	0.18	0.24	0.19	-
BEVPlanner++ [22]‡	**0.16**	0.32	0.57	0.35	**0.00**	0.29	0.73	0.34	-
SparseDrive-B [39]‡	0.29	0.55	0.91	0.58	0.01	**0.02**	**0.13**	**0.06**	7.3
ParaDrive [43]‡	0.25	0.46	0.74	0.48	0.14	0.23	0.39	0.25	4.2
DriveTransformer-Small [16]‡	0.19	0.33	0.66	0.39	0.01	0.07	0.21	0.10	10.7
DriveTransformer-Large [16]‡	**0.16**	0.30	0.55	0.33	0.01	0.06	0.15	0.07	4.5
DriveMamba-Tiny‡ (**Ours**)	0.19	0.35	0.59	0.38	0.07	0.09	0.24	0.13	**17.9**
DriveMamba-Small‡ (**Ours**)	0.18	0.33	0.56	0.36	0.03	0.05	0.17	0.08	11.1
DriveMamba-Base‡ (**Ours**)	0.17	0.33	0.54	0.35	0.02	0.04	0.17	0.07	6.1
DriveMamba-Large‡ (**Ours**)	**0.16**	0.30	0.52	0.33	**0.00**	0.03	0.16	**0.06**	1.7

➢ Quantative Results on Bench2Drive dataset

Method	Open-loop Metric: Avg. L2 (m) ↓	Closed-loop Metrics: Driving Score ↑	Success Rate (%) ↑	Efficiency ↑	Comfortness ↑	Latency (ms)
AD-MLP [49]	3.64	18.05	0.00	48.45	22.63	**3.0**
UniAD-Base [12]	0.73	45.81	16.36	129.21	43.58	663.4
VAD [17]	0.91	42.35	15.00	157.94	**46.01**	278.3
EgoFSD [37]	0.66	60.39	31.78	**180.63**	-	80.5
DriveTransformer-Large [16]	**0.62**	63.46	35.01	100.64	20.78	211.7
DriveMamba-Tiny (**Ours**)	0.77	53.54	27.27	137.88	20.55	55.8
DriveMamba-Small (**Ours**)	0.75	60.78	31.82	150.46	16.52	90.2
DriveMamba-Base (**Ours**)	0.71	65.50	36.82	158.33	21.51	164.3
DriveMamba-Large (**Ours**)	0.70	**66.82**	**37.73**	152.91	18.77	599.1

➢ Qualitative Results