JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Planning-oriented Autonomous Driving

Yihan Hu*, Jiazhi Yang*, Li Chen*, Keyu Li*,

Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, Lewei Lu, Xiaosong Jia, Qiang Liu, Jifeng Dai, Yu Qiao, Hongyang Li

[1] Shanghai AI Lab, [2] Wuhan University, [3] SenseTime Research

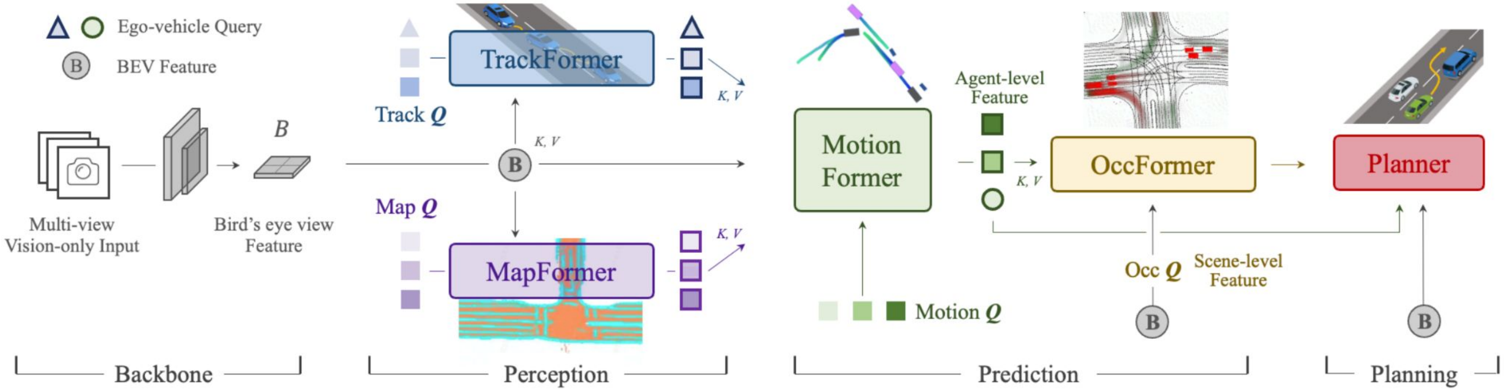
Poster: THU-AM-131

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

商汤
sensetime

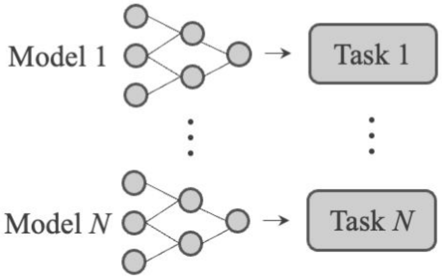
Brief Overview

- **Planning-oriented Philosophy**: An end-to-end autonomous driving framework in pursuit of safe-planning, facilitated with various well-organized AD tasks.
- **Unified Query** design: *Queries* as interfaces to coordinate all tasks in training and transmit upstream knowledge to planner.
- **SOTA Performance** with vision-only input: SOTAs on all tasks, even surpassing LiDAR-based framework in planning.

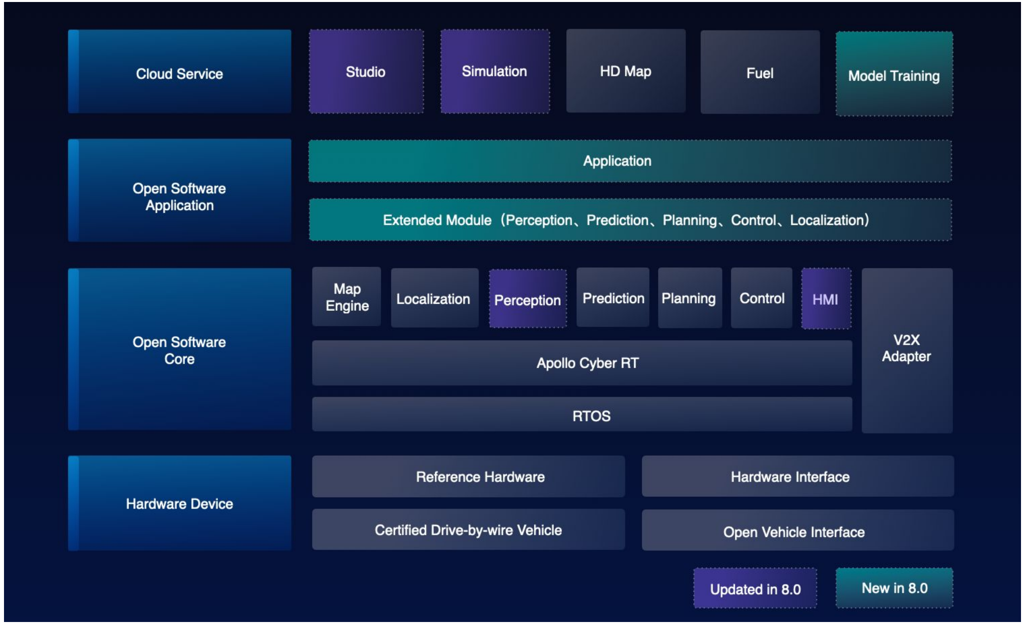


Background - The Design Space of AD Systems

(a) Standalone Models



- Typical **industrial** solutions
- Independent teams for module developments ✅
- Severe error accumulation and feature misalignment ❌

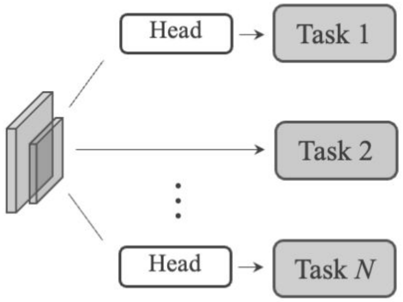


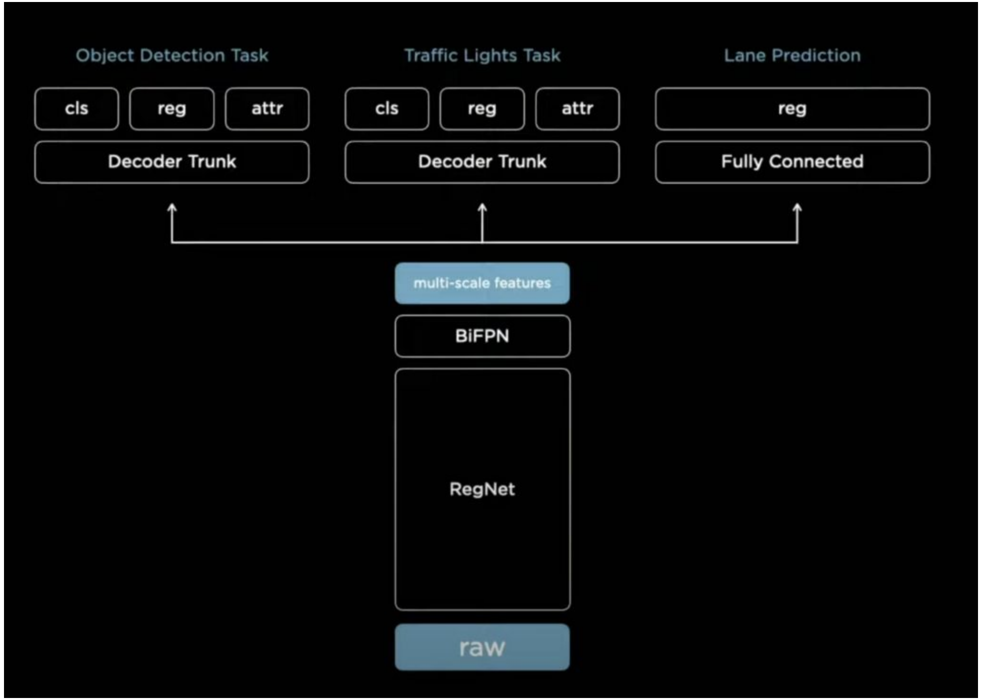
Apollo 8.0
https://github.com/ApolloAuto/apollo

Background - The Design Space of AD Systems

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

(b) Multi-task Framework



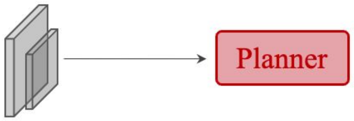


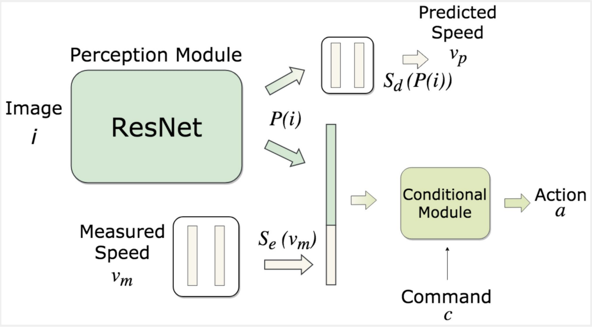
- **Shared feature extraction** for multiple tasks
- Easily extended to more tasks, and saving compute ✅
- Inefficient tasks' coordination ❌

Tesla AI Day 2021

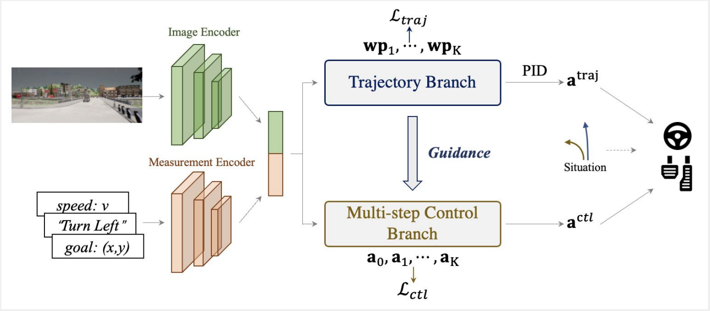
Background - The Design Space of AD Systems

(c.1) End-to-end Framework - Vanilla Solutions





CILRS, ICCV 2019. TRI

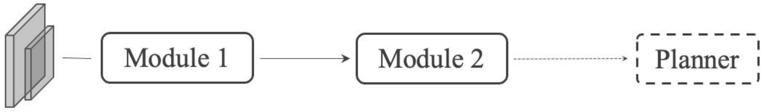


TCP, NeurIPS 2022. SH AI Lab

- **Directly learn planning** from sensor inputs, no intermediate tasks involved
- Simple network design with good performance in carla simulator ✅
- Deficient in interpretability ❌

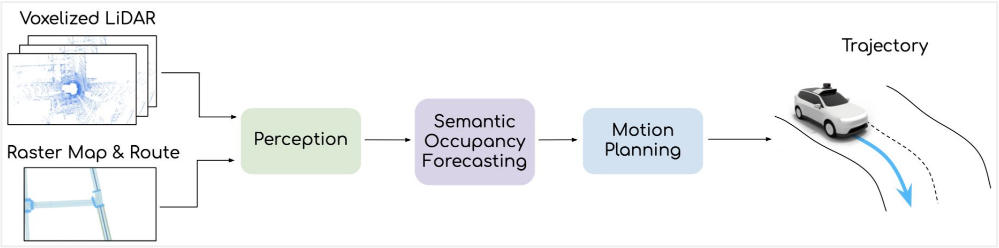
Background - The Design Space of AD Systems

(c.2) End-to-end Framework - Explicit Design

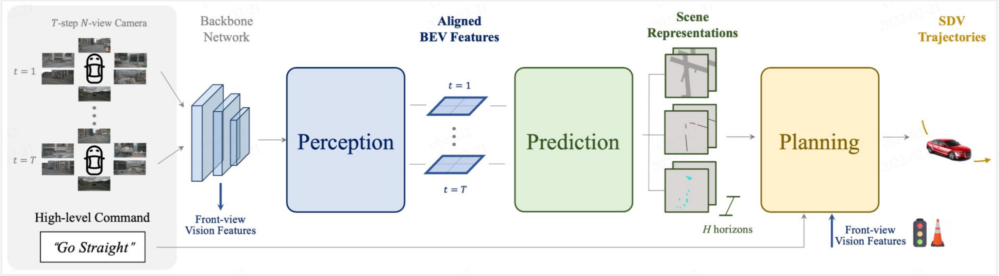


- Introducing **intermediate tasks** to assist planning
- Better interpretability ✅
- Some crucial components are missing ❌

Design	Approach	Perception			Prediction		Plan
		Det.	Track	Map	Motion	Occ.	
(c.2)	PnPNet† [50]	✓	✓		✓		
	ViP3D† [30]	✓	✓		✓		
	P3 [72]					✓	✓
	MP3 [11]			✓		✓	✓
	ST-P3 [37]			✓		✓	✓
	LAV [15]	✓		✓	✓		✓



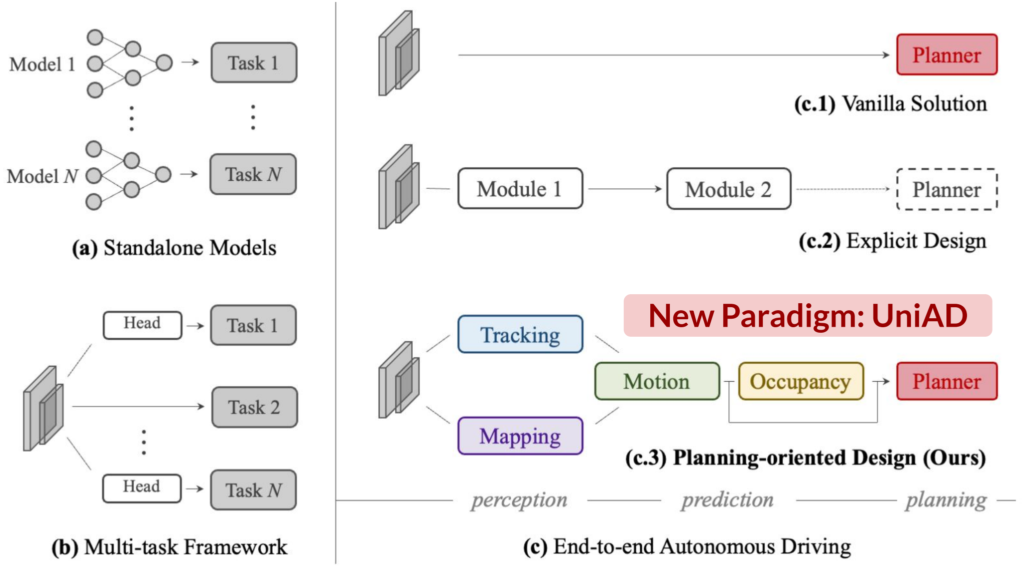
P3, ECCV 2020. Uber



ST-P3, ECCV 2022. SH AI Lab

Motivation - Towards Reliable Planning

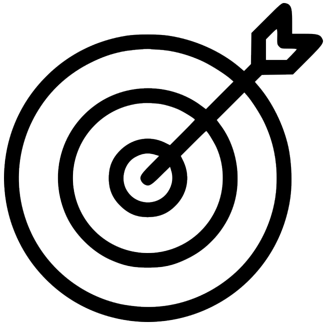
Planning-oriented Design (Ours)



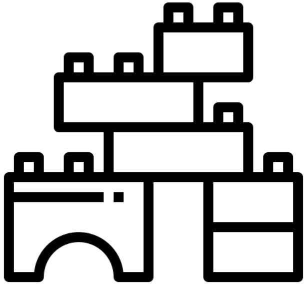
Design	Approach	Perception			Prediction		Plan
		Det.	Track	Map	Motion	Occ.	
(b)	NMP [101]	✓			✓		✓
	NEAT [19]			✓			✓
	BEVerse [105]	✓		✓		✓	
(c.1)	[14, 16, 78, 97]						✓
(c.2)	PnPNet† [57]	✓	✓		✓		
	ViP3D† [30]	✓	✓		✓		
	P3 [82]					✓	✓
	MP3 [11]			✓		✓	✓
	ST-P3 [38]			✓		✓	✓
	LAV [15]	✓		✓	✓		✓
(c.3)	**UniAD** (ours)	✓	✓	✓	✓	✓	✓

- **Jointly optimizing** five essential AD tasks to facilitate planning ✅
- **Efficient tasks' coordination** with **unified queries** as interfaces ✅
- Diverse **knowledge** from upstream tasks is **transmitted** to planner ✅

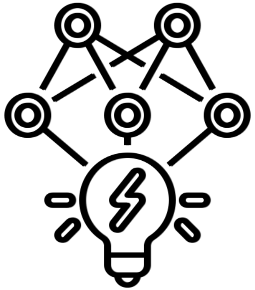
UniAD - Overview



Which tasks?



How to construct?

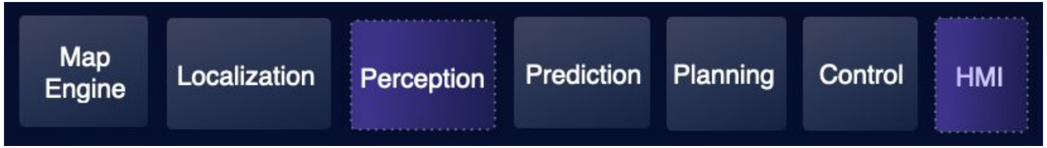


How to train?

UniAD - Which Tasks?

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- **Learn from industrial system**



Apollo 8.0
https://github.com/ApolloAuto/apollo

Perception

- Object **Detection**
- Object **Tracking**
- Semantic Segmentation
- **Online Mapping**
- Depth Estimation
- Traffic light / sign Detection
- …

Track agents behavior

Map (lane) guidance

Prediction

- **Motion** Forecasting
- Intention Prediction
- **Occupancy** Prediction

Interaction with environment

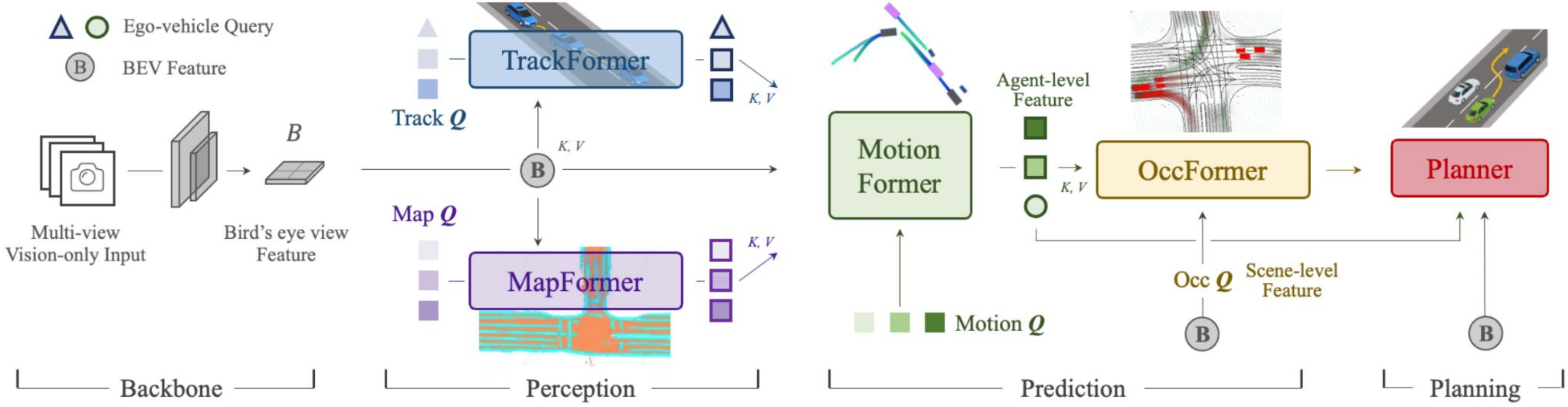
Find free space

Safe planning

UniAD - How to Construct?

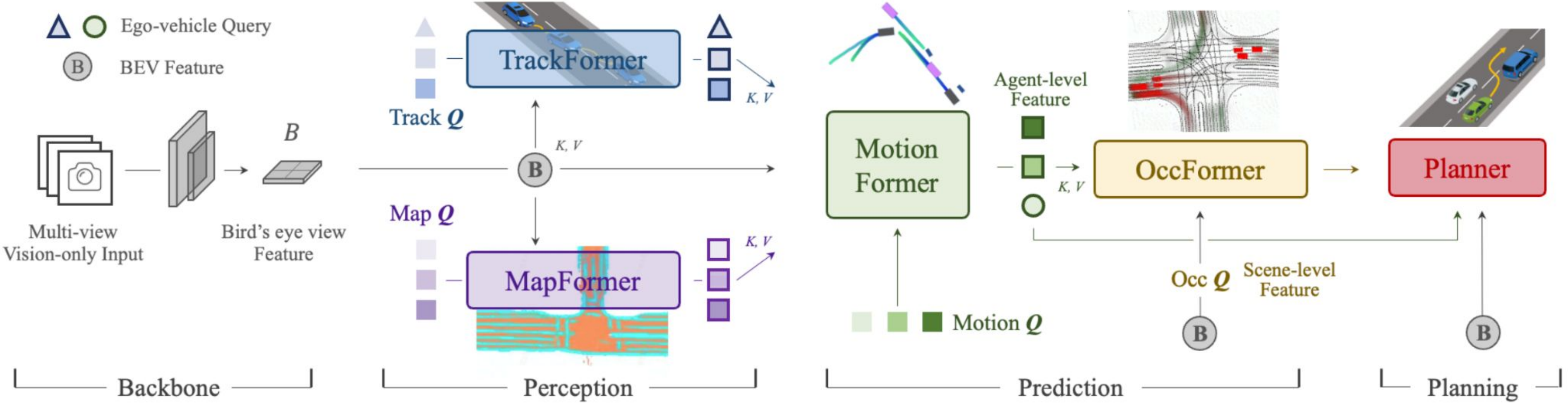
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- Track agents across time
- Complex interactions
- Long-term trajectory
- Sdc query
- Collision optimization



- Learning map elements as guidance
- Occupancy with agent-wise knowledge

UniAD - How to Construct?



- Track Q : one query for **one agent**
- Map Q : one query for **one map element**

Unified Query

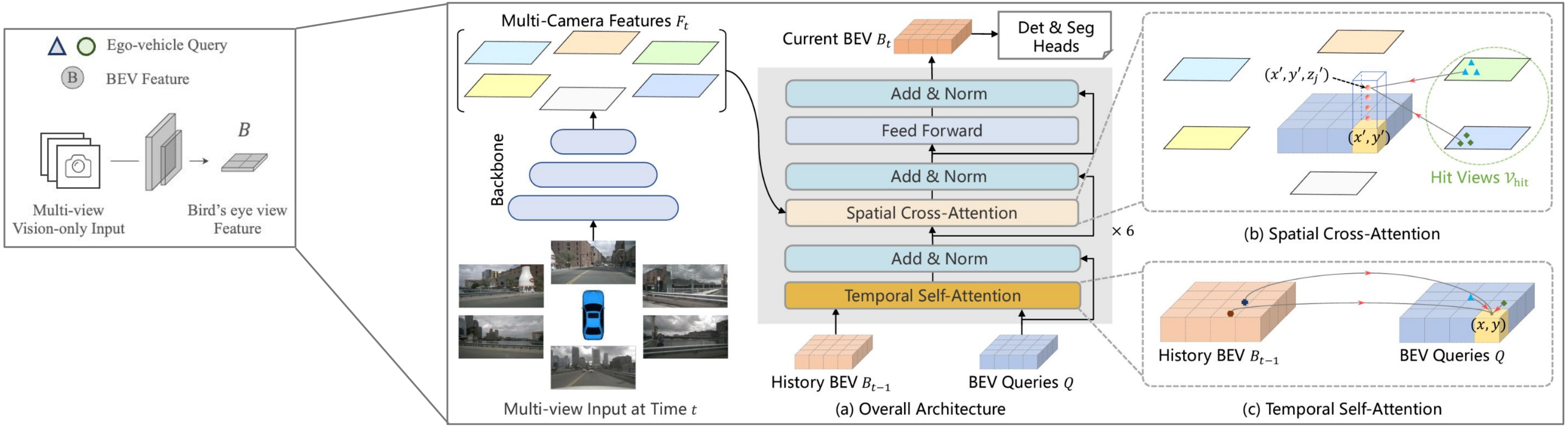
- Motion Q : one query for **one trajectory**
- Occ Q : one query for **one BEV grid**

Each tasks module is a **transformer-based** structure, with:

- ***Attention*** mechanisms model complex relations in the scene
- ***Queries*** interact between tasks modules and transmit upstream knowledge to final planner.

UniAD - Perception

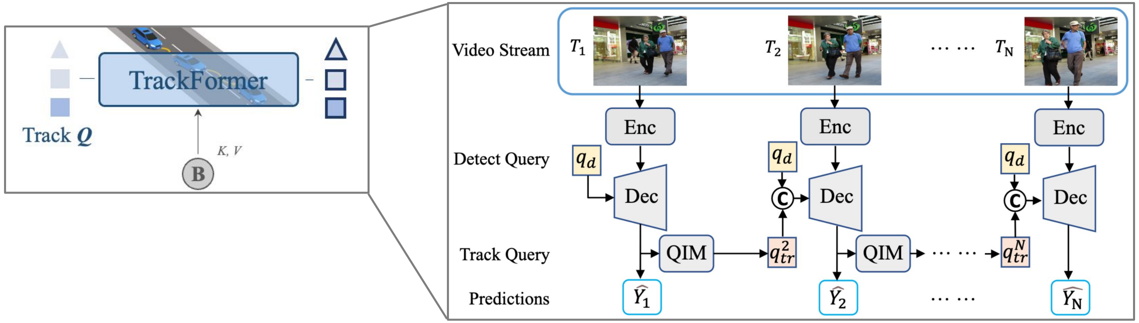
BEV Encoder - BEVFormer (ECCV 2022)



- BEVFormer is a strong BEV encoder with effective **spatial and temporal** feature extraction
- You can easily replace it with other advanced BEV encoders

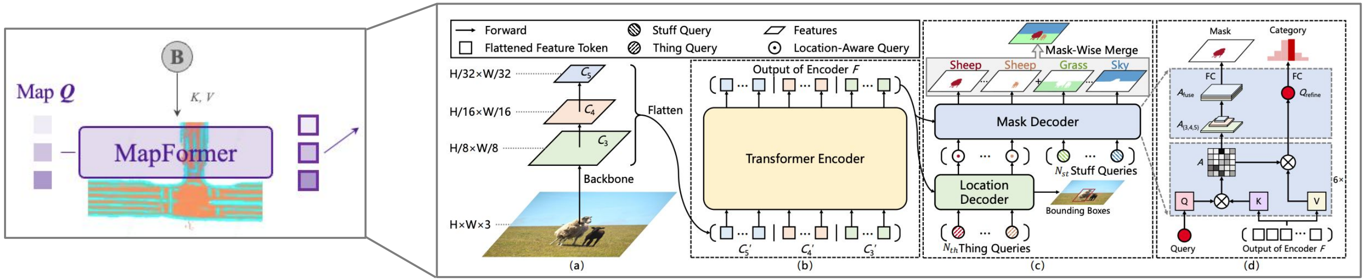
UniAD - Perception

TrackFormer - Modified from MOTR (ECCV 2022)



- **End-to-end trainable** tracking without post-association

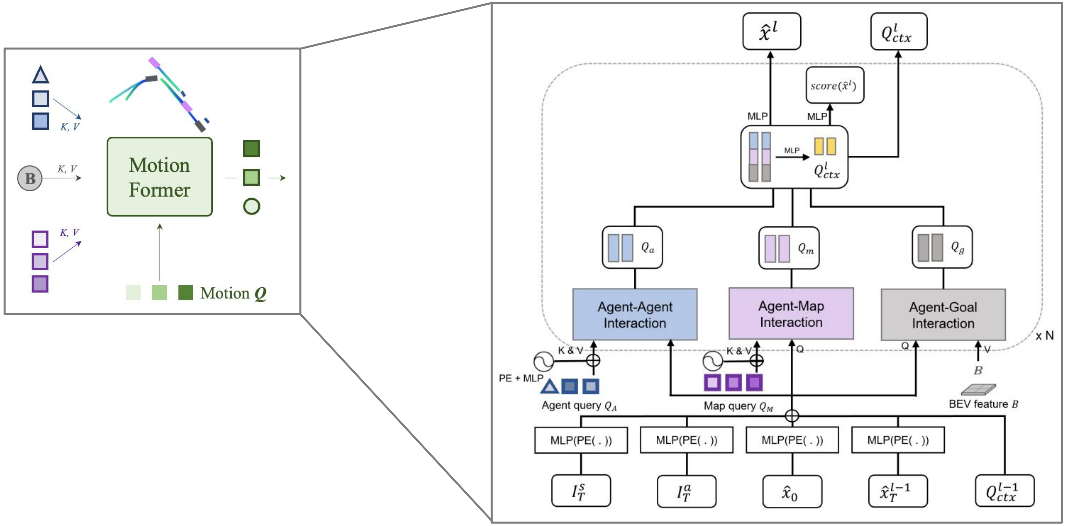
MapFormer - Modified from Panoptic SegFormer (CVPR 2022)



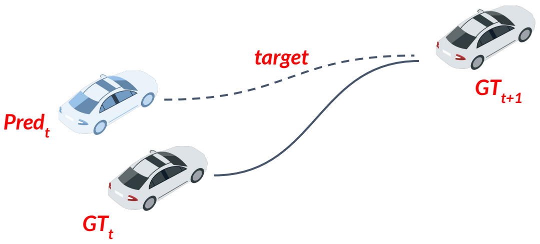
- Each query represents a map element

UniAD - Prediction

MotionFormer (Ours)



Non-linear Optimization

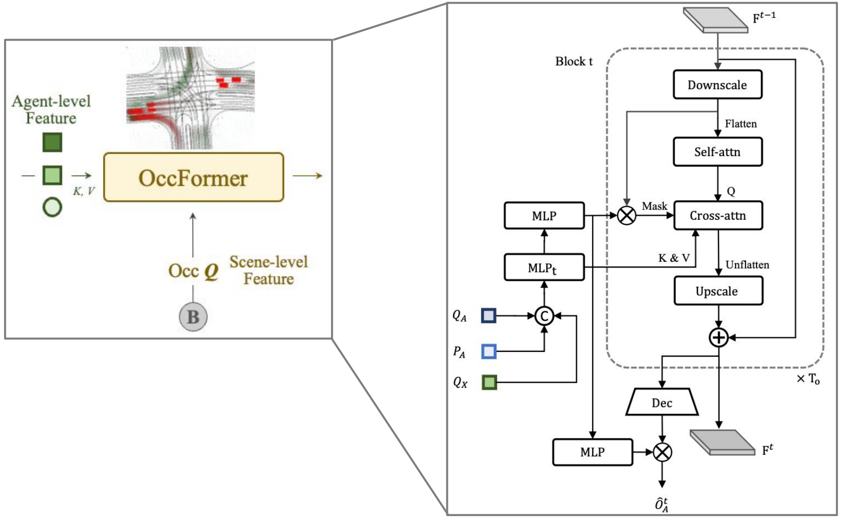


- Various **relation modelings** via attentions:
  - Agent-agent, agent-map, agent-goal point
- **Non-linear optimization**: Adjust the ground-truth trajectory according to upstream predictions

UniAD - Prediction

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

OccFormer (Ours)

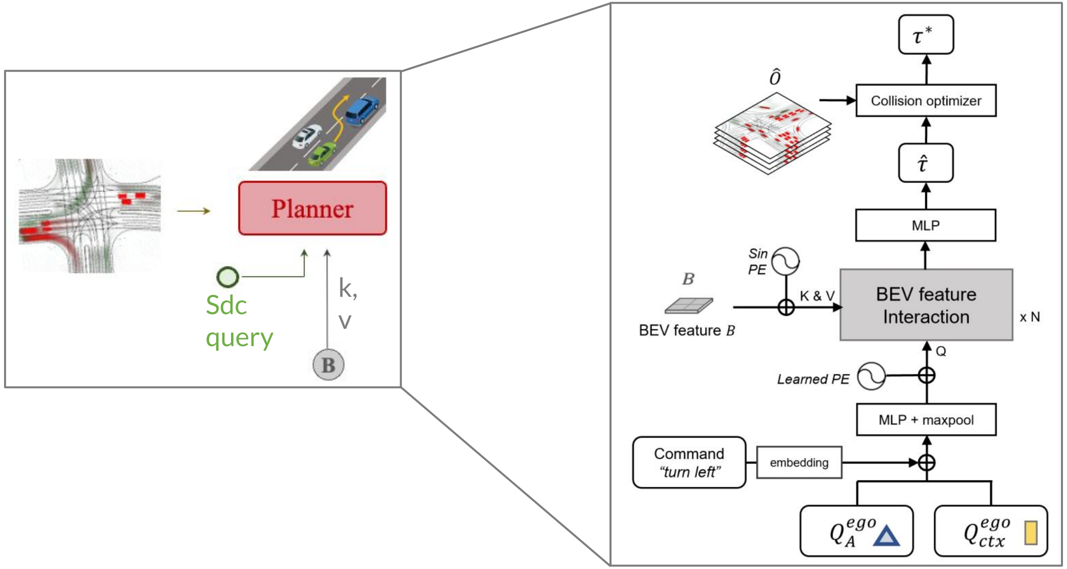


- **Encode agent-wise knowledge** into the scene representation
- Predicted **occupancy as attention mask** for concentrating the interactions between the agents and their corresponding BEV features.

UniAD - Planning

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Planner (Ours)



- **Sdc query:** consistently modeling self-car in TrackFormer and MotionFormer, and is passed to Planner
- **Collision optimization:** Steer the predicted trajectories clear of occupied areas to avoid potential collisions

UniAD - How to Train?

Two phases training. Perception stage + End-to-end stage

- The stabilized perception capability helps the end-to-end stage **converge faster**

Shared matching. Matching results of tracking reused in motion and occupancy

- Consistent learning of agent identities

UniAD - Results

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Ablations of Tasks' Coordinations

ID	Modules					Tracking			Mapping		Motion Forecasting			Occupancy Prediction				Planning	
	Track	Map	Motion	Occ.	Plan	AMOTA↑	AMOTP↓	IDS↓	IoU-lane↑	IoU-road↑	minADE↓	minFDE↓	MR↓	IoU-n.↑	IoU-f.↑	VPQ-n.↑	VPQ-f.↑	avg.L2↓	avg.Col.↓
0*	✓	✓	✓	✓	✓	0.356	1.328	893	0.302	0.675	0.858	1.270	0.186	55.9	34.6	47.8	26.4	1.154	0.941
1	✓					0.348	1.333	791	-	-	-	-	-	-	-	-	-	-	-
2		✓				-	-	-	**0.305**	0.674	-	-	-	-	-	-	-	-	-
3	✓	✓				0.355	1.336	785	0.301	0.671	-	-	-	-	-	-	-	-	-
4			✓			-	-	-	-	-	0.815	1.224	0.182	-	-	-	-	-	-
5	✓		✓			0.360	1.350	919	-	-	0.751	1.109	0.162	-	-	-	-	-	-
6	✓	✓	✓			0.354	1.339	820	0.303	0.672	0.736(-9.7%)	1.066(-12.9%)	0.158	-	-	-	-	-	-
7				✓		-	-	-	-	-	-	-	-	60.5	37.0	52.4	29.8	-	-
8	✓			✓		0.360	**1.322**	809	-	-	-	-	-	62.1	38.4	52.2	32.1	-	-
9	✓	✓	✓	✓		0.359	1.359	1057	0.304	**0.675**	**0.710**(-3.5%)	**1.005**(-5.8%)	**0.146**	**62.3**	39.4	**53.1**	32.2	-	-
10					✓	-	-	-	-	-	-	-	-	-	-	-	-	1.131	0.773
11	✓	✓	✓		✓	**0.366**	1.337	889	0.303	0.672	0.741	1.077	0.157	-	-	-	-	1.014	0.717
12	✓	✓	✓	✓	✓	0.358	1.334	**641**	0.302	0.672	0.728	1.054	0.154	**62.3**	**39.5**	52.8	**32.3**	**1.004**	**0.430**

Conclusion:

- **ID. 4-6:** Track & Map → Motion
- **ID. 7-9:** Motion ↔ Occupancy
- **ID. 10-12:** Motion & Occupancy → Planning

UniAD - Results

Planning

Method	L2(m)↓ 1s	2s	3s	Avg.	Col. Rate(%)↓ 1s	2s	3s	Avg.
NMP† [88]	-	-	2.31	-	-	-	1.92	-
SA-NMP† [88]	-	-	2.05	-	-	-	1.59	-
FF† [36]	0.55	1.20	2.54	1.43	0.06	0.17	1.07	0.43
EO† [42]	0.67	1.36	2.78	1.60	0.04	0.09	0.88	0.33
ST-P3 [37]	1.33	2.11	2.90	2.11	0.23	0.62	1.27	0.71
UniAD	**0.48**	**0.96**	**1.65**	**1.03**	**0.05**	**0.17**	**0.71**	**0.31**

†: LiDAR-based

Even outperforms LiDAR-based counterparts

UniAD - Results

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Multi-object Tracking

Method	AMOTA↑	AMOTP↓	Recall↑	IDS↓
Immortal Tracker† [82]	0.378	1.119	0.478	936
ViP3D [30]	0.217	1.625	0.363	-
QD3DT [35]	0.242	1.518	0.399	-
MUTR3D [91]	0.294	1.498	0.427	3822
UniAD	**0.359**	**1.320**	**0.467**	**906**

Motion Forecasting

Method	minADE(m)↓	minFDE(m)↓	MR↓	EPA↑
PnPNet† [50]	1.15	1.95	0.226	0.222
ViP3D [30]	2.05	2.84	0.246	0.226
Constant Pos.	5.80	10.27	0.347	-
Constant Vel.	2.13	4.01	0.318	-
UniAD	**0.71**	**1.02**	**0.151**	**0.456**

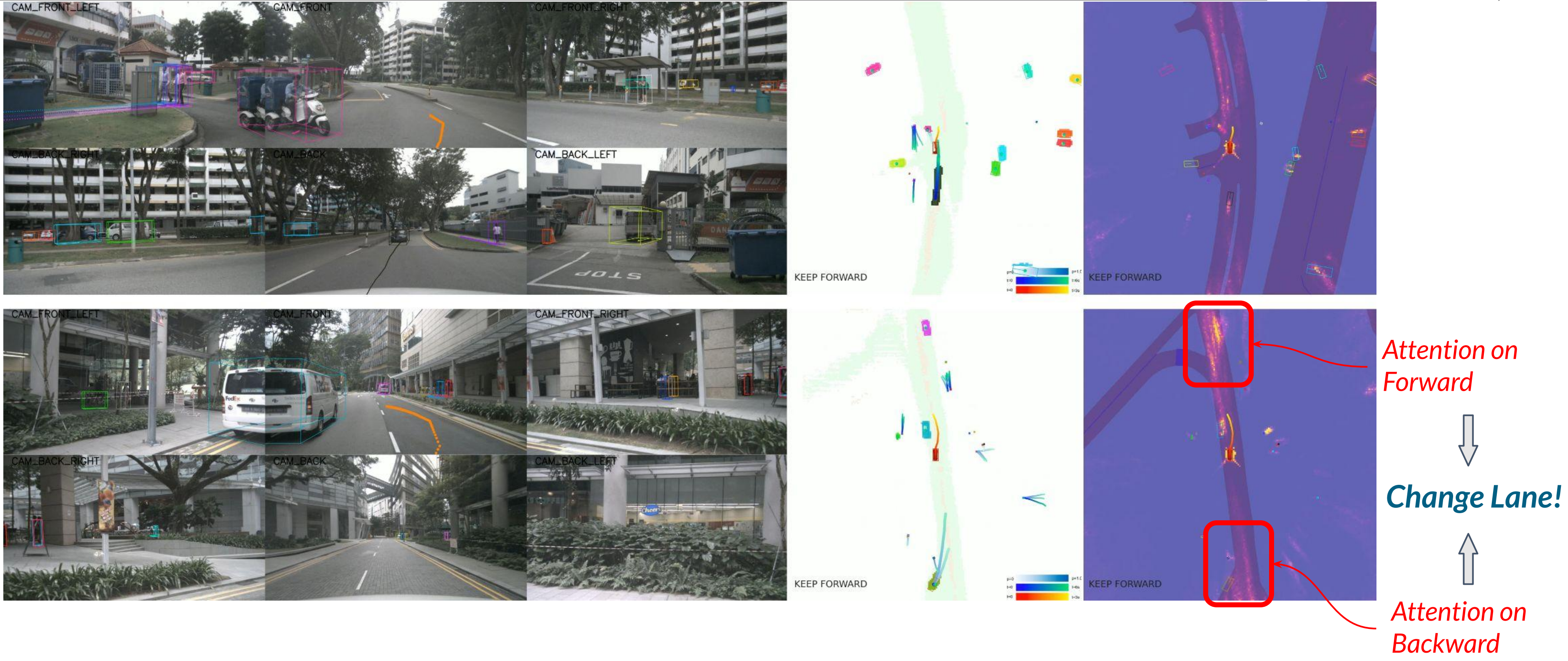
Mapping

Method	Lanes↑	Drivable↑	Divider↑	Crossing↑
VPN [63]	18.0	76.0	-	-
LSS [66]	18.3	73.9	-	-
BEVFormer [48]	23.9	**77.5**	-	-
BEVerse† [92]	-	-	**30.6**	**17.2**
UniAD	**31.3**	69.1	25.7	13.8

Occupancy Prediction

Method	IoU-n.↑	IoU-f.↑	VPQ-n.↑	VPQ-f.↑
FIERY [34]	59.4	36.7	50.2	29.9
StretchBEV [1]	55.5	37.1	46.0	29.0
ST-P3 [37]	-	38.9	-	32.1
BEVerse† [92]	61.4	**40.9**	54.3	**36.1**
UniAD	**63.4**	40.2	**54.7**	33.5

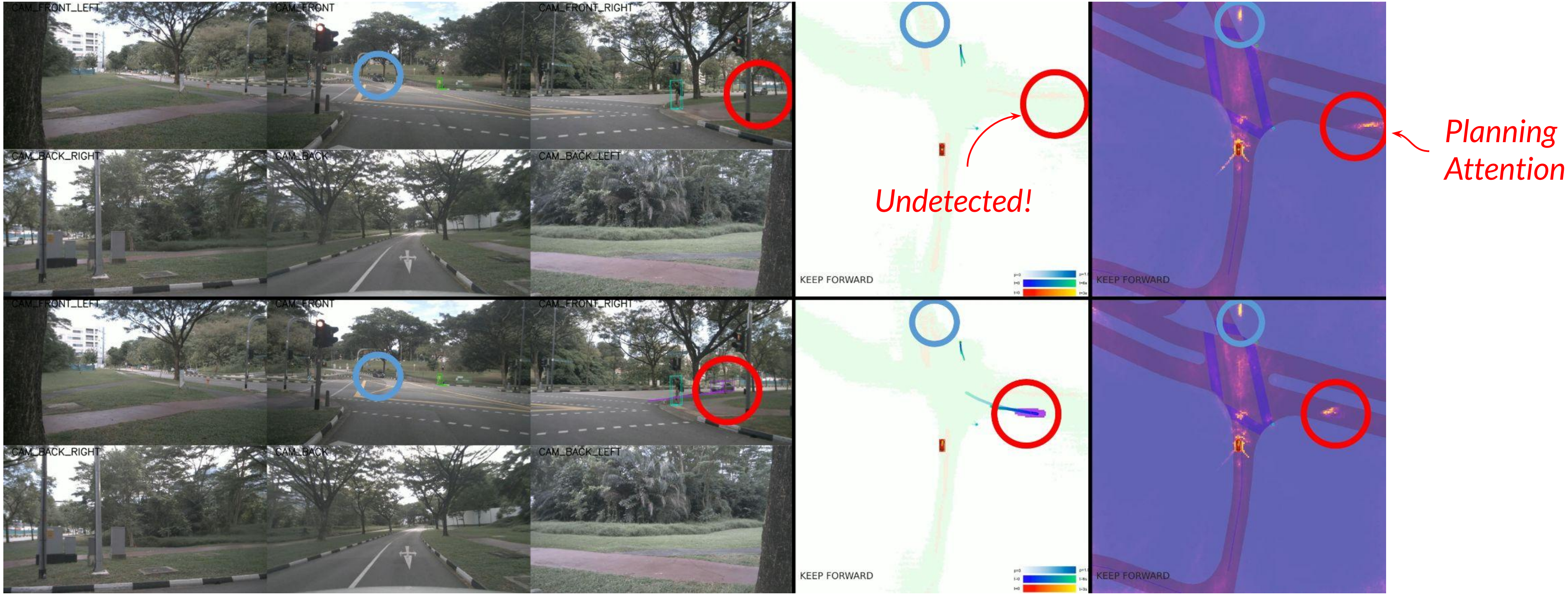
UniAD - Visualizations



Planner attends to crucial areas in complex scenes

UniAD - Recover from Upstream Errors

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



Planner could still attend to 'undetected' regions/objects

Planning-oriented Autonomous Driving

Thanks

Arxiv **OpenDriveLab** **Code**

Poster: THU-AM-131

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory
商汤 sensetime