OpenDriveLab

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+

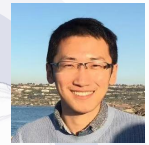
*equal contribution +project lead


Yihan


Jiazhi


Li


Keyu


Hongyang

Poster: THU-AM-131

arXiv: https://arxiv.org/abs/2212.10156

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

Shanghai AI Laboratory | 上海人工智能 实验室

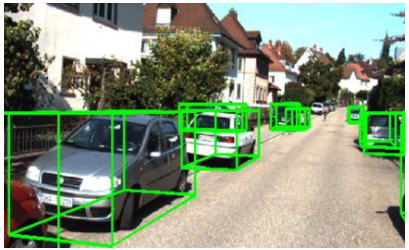
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Planning-oriented Autonomous Driving

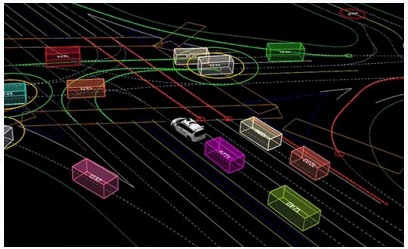
Background and Motivation

Background - Autonomous Driving (AD) Systems

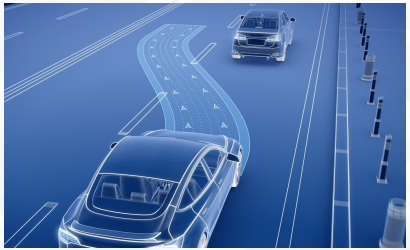
Perception → (Bounding boxes) → Prediction → (Waypoints) → Planning → (Trajectory)



What are around?



How will they go in the future?

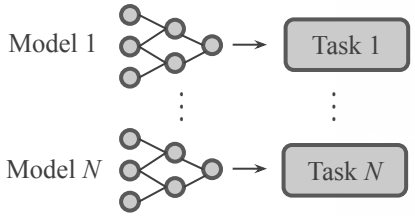


Where should I go?

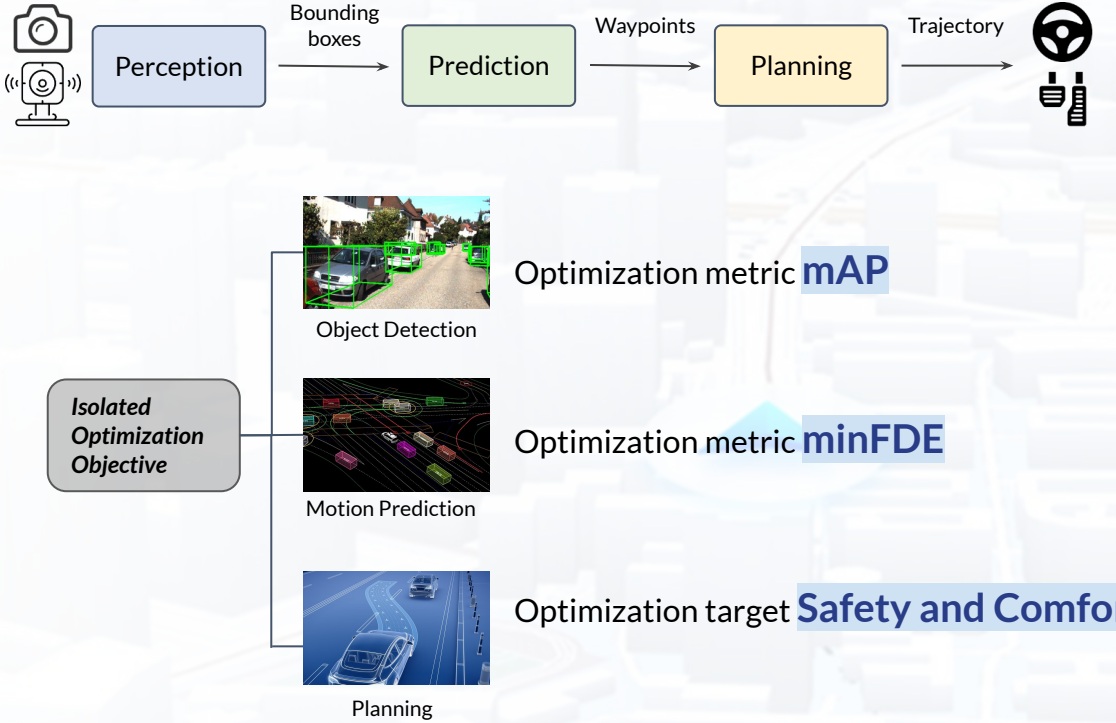
Photos credit to DARPA 2007 Urban Challenge, Waymo, Cruise, and other online resources.

Background - Design Options for Autonomous Driving (AD) Systems

(a) Standalone Models

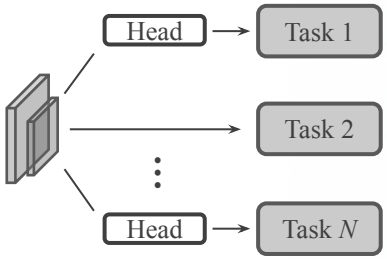


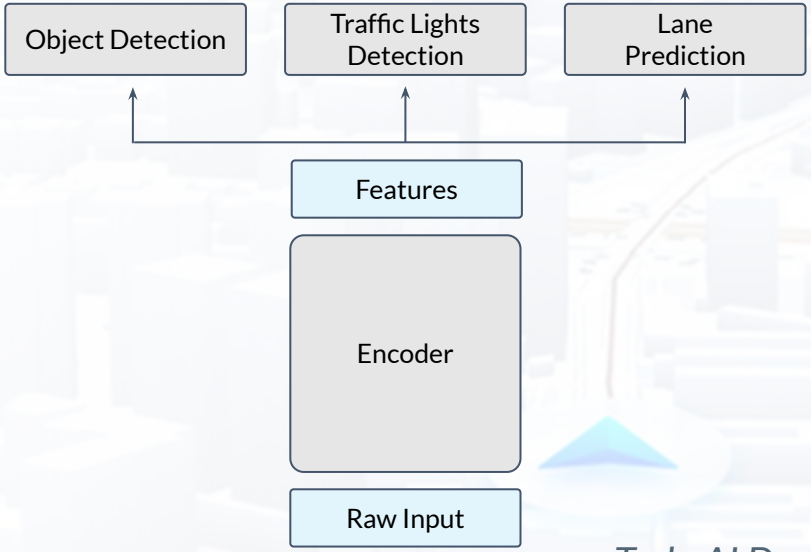
- Typical **Industry** solutions
- ✅ Independent teams for module developments
- ❌ Severe error accumulation and feature misalignment



Background - Design Options for Autonomous Driving (AD) Systems

(b) Multi-task Framework





- **Shared feature** for multiple tasks

✅ • Easily extended to more tasks, Compute-efficient

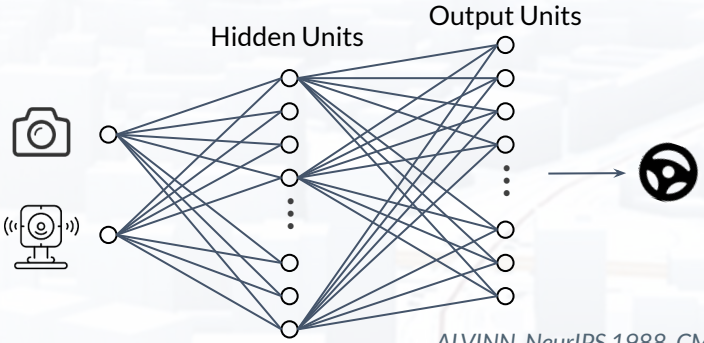
❌ • Feature misalignment, “negative transfer”

credit to Tesla AI Day 2021

TESLA XPENG WAYMO

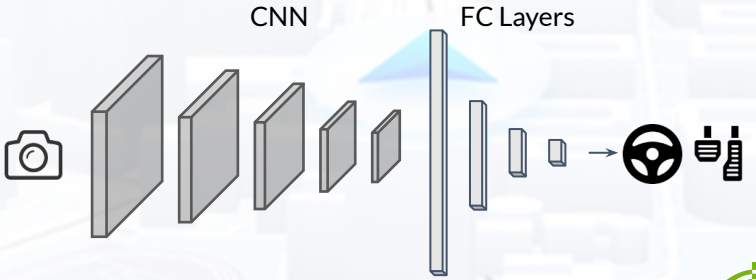
Background - Design Options for Autonomous Driving (AD) Systems

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



ALVINN, NeurIPS 1988. CMU

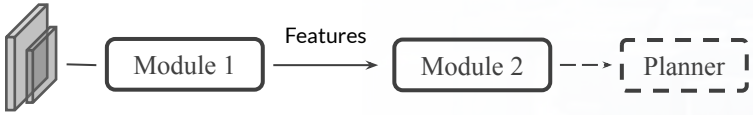
- **Direct policy learning** from sensor inputs, bypassing intermediate tasks
- ✅ Simple design with good performance in the simulator
- ❌ Deficient in interpretability



DAVE-2, arXiv 2016. Nvidia

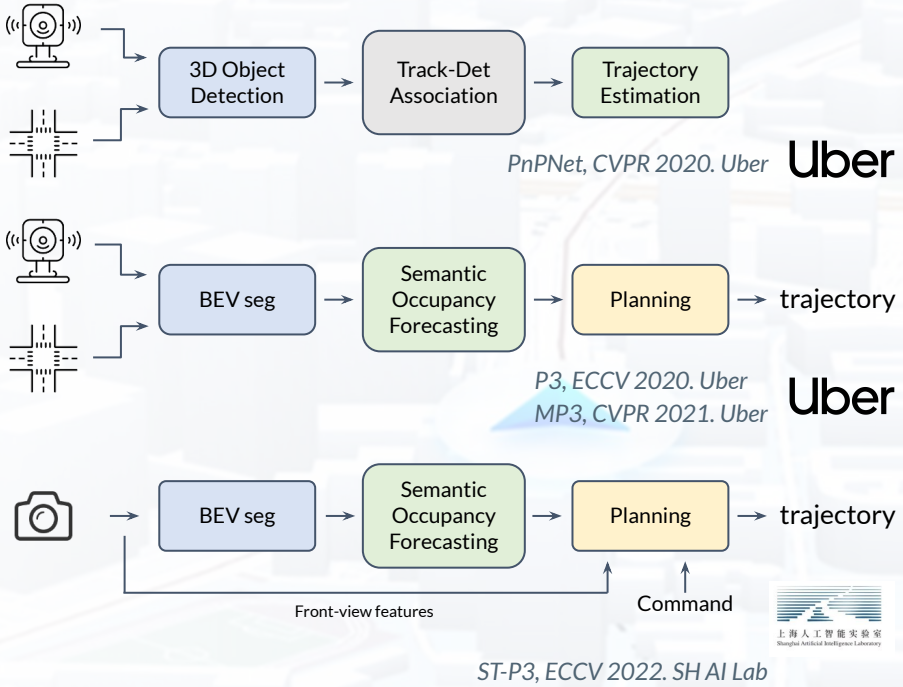
Background - Design Options for Autonomous Driving (AD) Systems

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



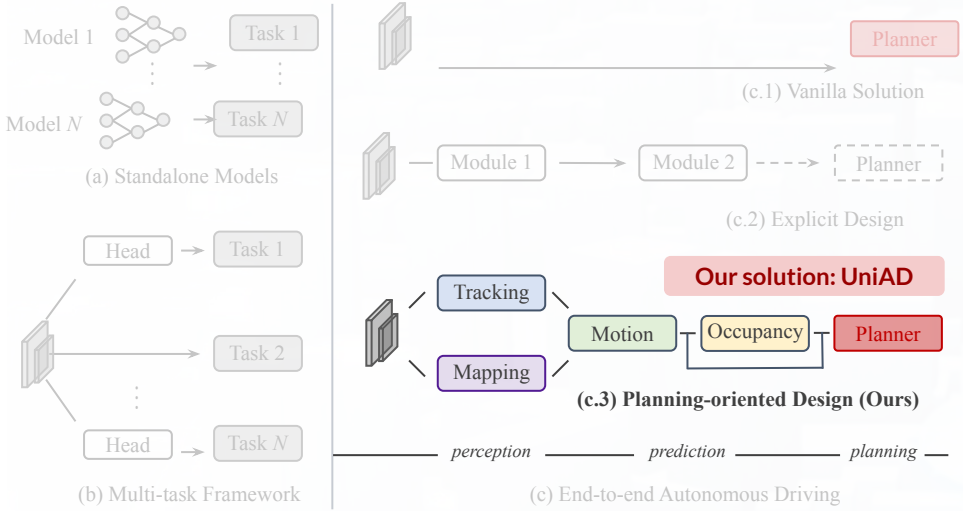
- Introducing **intermediate tasks** to assist planning
- ✅ Better interpretability (e.g. Bird's-eye-view, BEV)
- ❌ Lack some crucial components[1]

1. *The necessities of each component is mentioned in Appendix.*



Motivation- Towards Reliable Planning

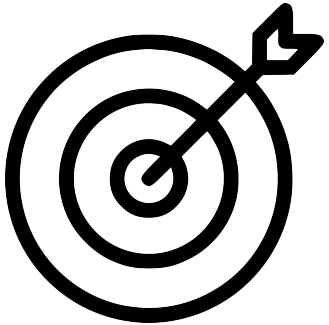
Ours: **Planning-oriented** Autonomous Driving



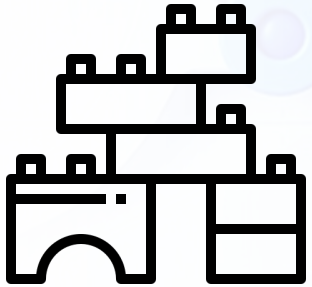
What do we want:

- ✅ **Unify** full-stack AD tasks
- ✅ **Tasks are coordinated towards an optimal planner**

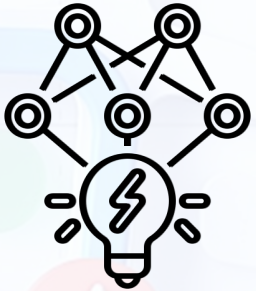
UniAD - Overview



Which tasks?



How to construct?



How to train?

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Planning-oriented Autonomous Driving

Delving into Details

UniAD - Which Tasks?

Perception System

- Object **Detection**
- Object **Tracking**
- **Online Mapping**
- ...

Track agents behavior

Guide with map (lane)

Prediction System

- **Motion** Forecasting
- **Occupancy** Prediction
- ...

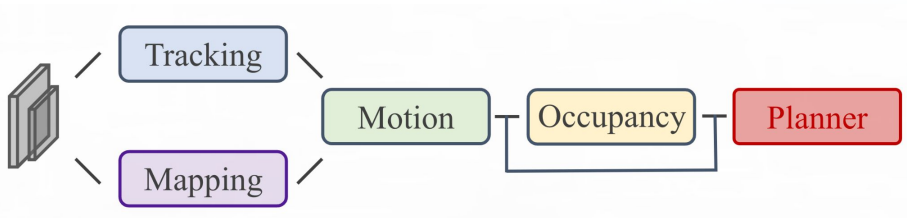
Interact with environment

Find free space

Planning

- ...

Incorporate all tasks in a ***hierarchical*** manner

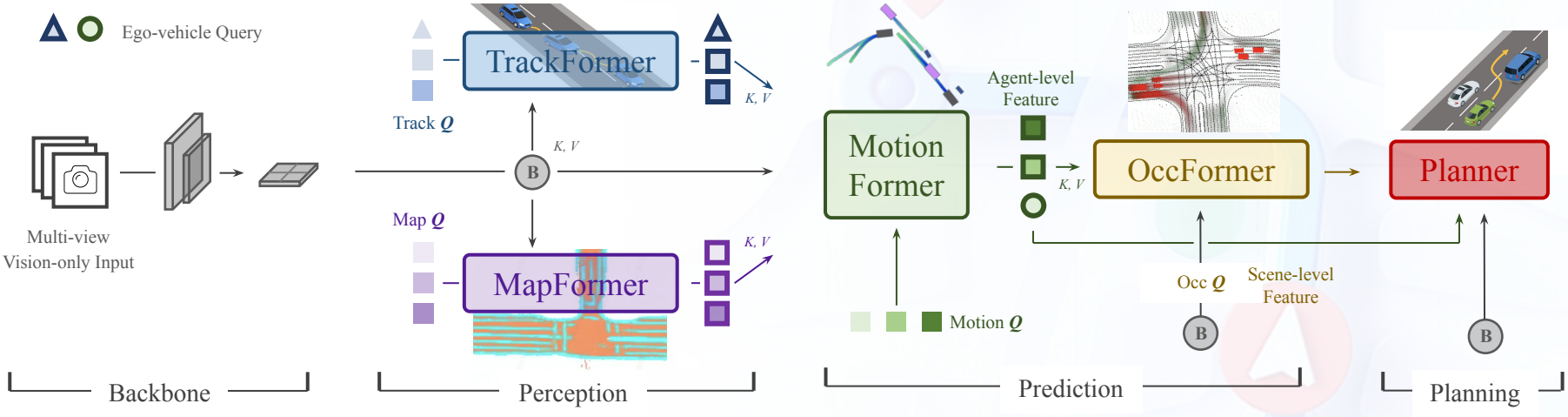


- Five **safety-critical tasks**: Model the static and dynamic information
- Task hierarchy: Tasks are ***well-organized*** to optimize information flow to the planner.

UniAD - How to Construct?

Pipeline

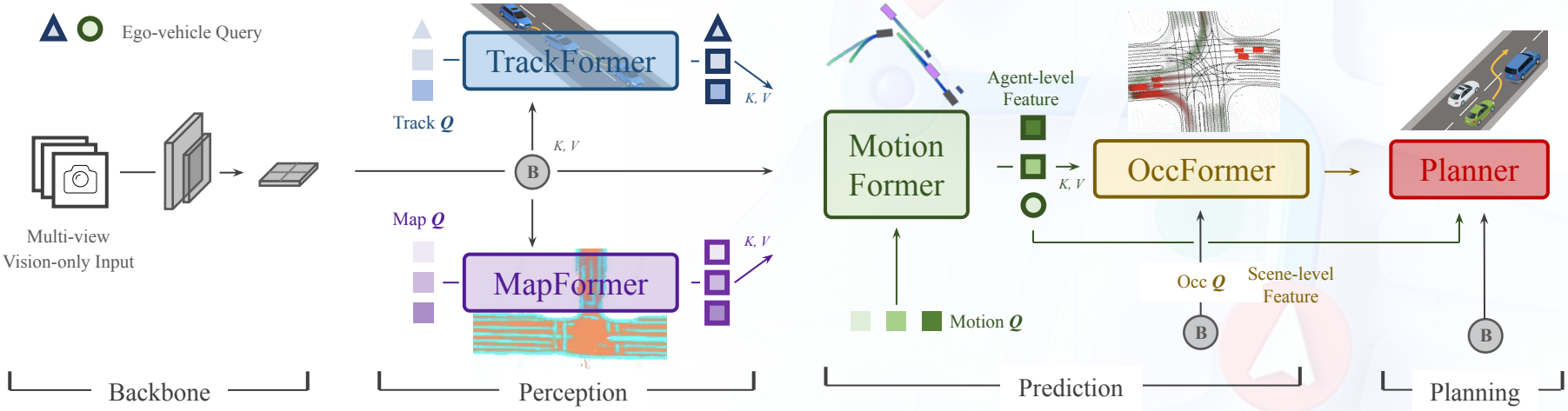
- Track agents across time
- Predict long-term trajectory
- Plan final trajectory
- Avoid collision



- Segment map elements
- Incorporate agent-wise knowledge

UniAD - How to Construct?

Pipeline



- Connect pipeline
- Coordinate tasks
- Transmit knowledge
- Model interactions

Unified Query

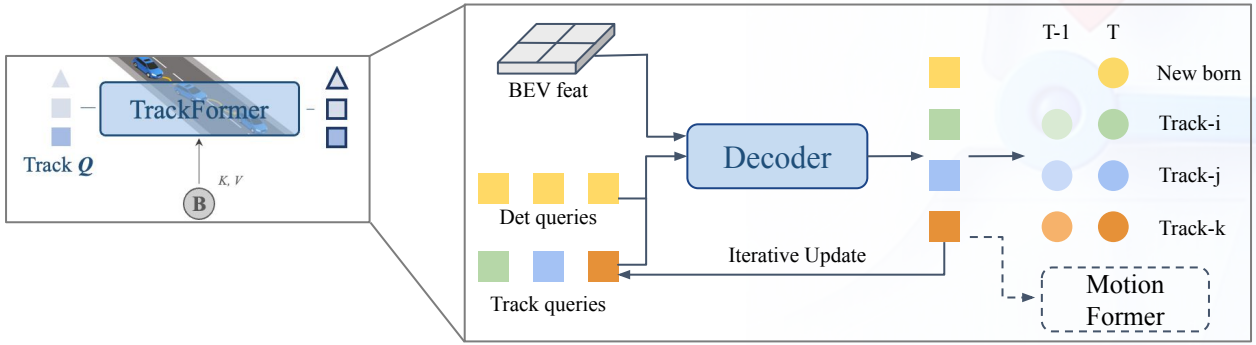
Transformer-based

First time to unify full-stack AD tasks!

UniAD - How to Construct?

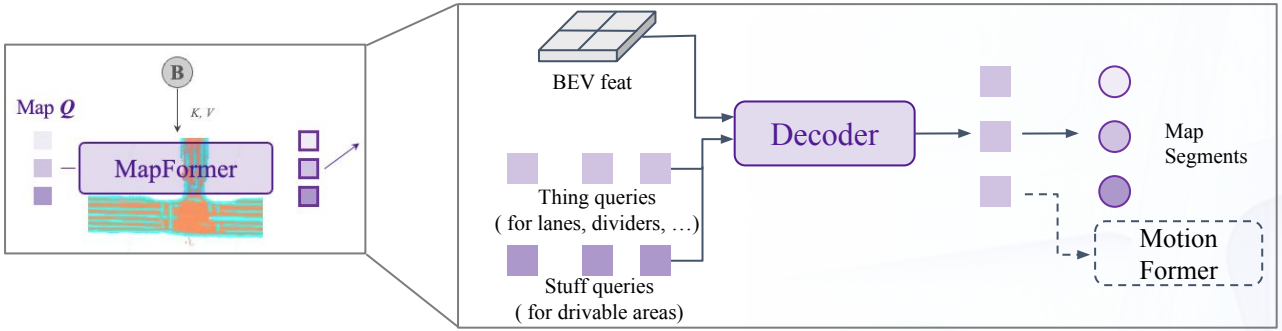
Perception — Prediction — Planning

TrackFormer - MOTR (ECCV 2022)



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

MapFormer - Panoptic SegFormer (CVPR 2022)

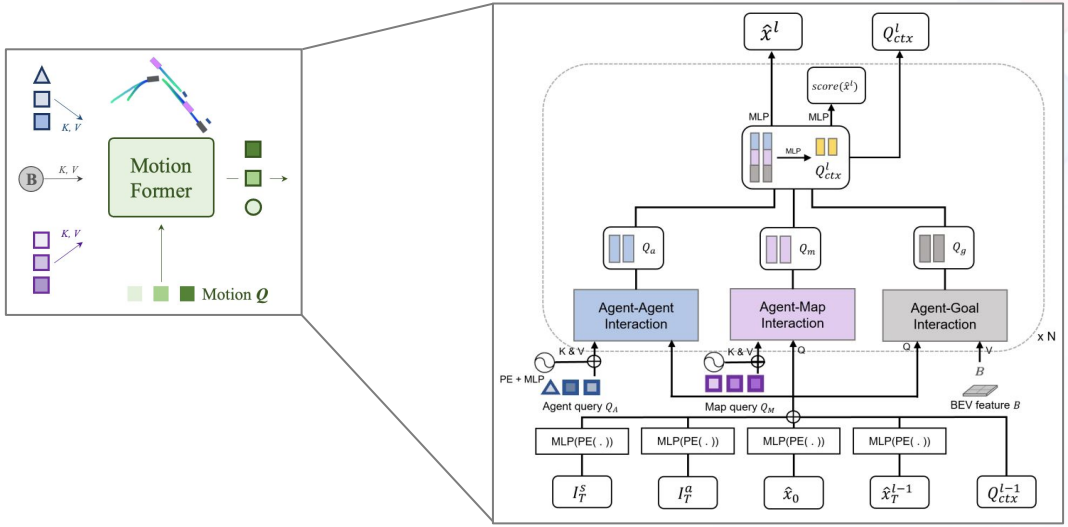


- Each query represents a map element

UniAD - How to Construct?

Perception | Prediction | Planning

MotionFormer (Proposed in UniAD)



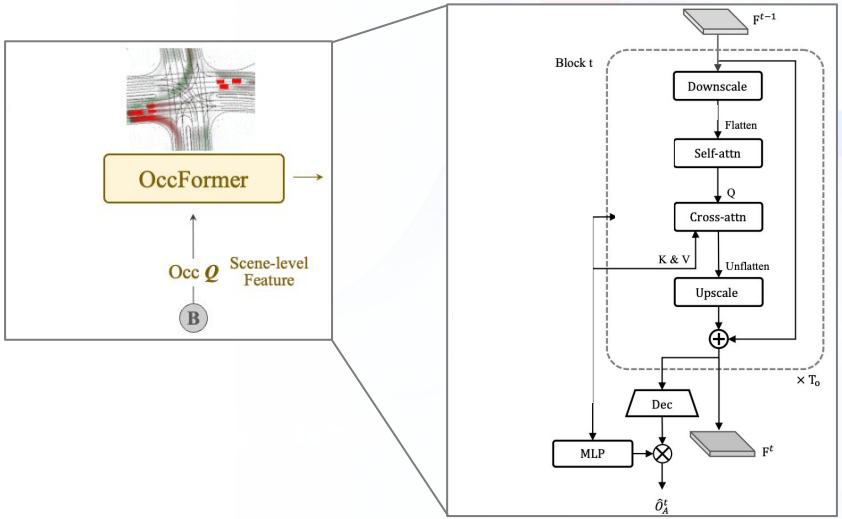
- Diverse **relation modelings** via attentions: Agent-agent, agent-map, agent-goal

- **Non-linear optimization**: Adjust ground-truth trajectory based on upstream predictions

UniAD - How to Construct?

Perception | Prediction | Planning

OccFormer (Proposed in UniAD)

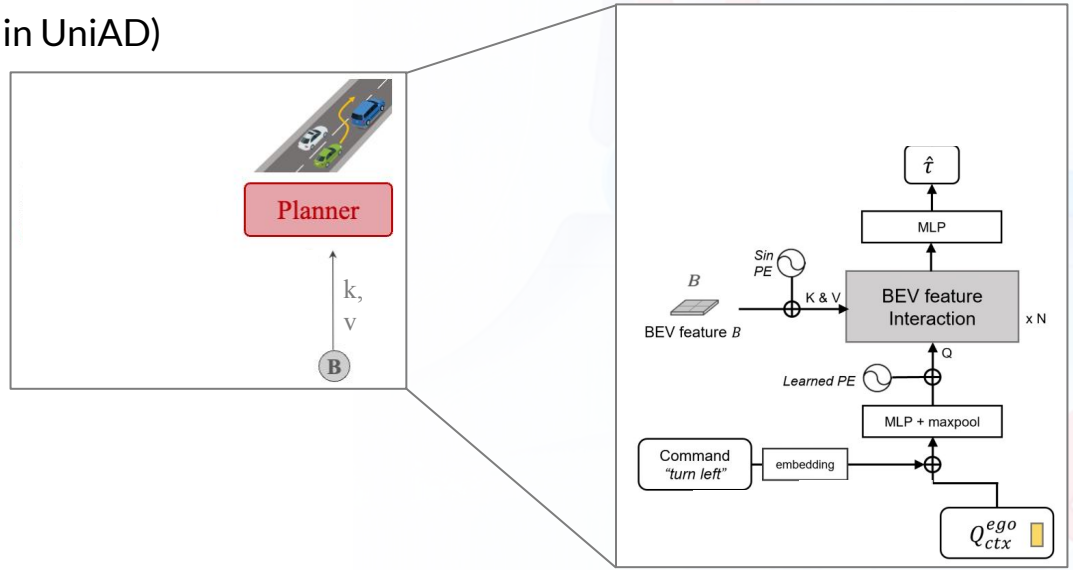


- **Encode agent-wise knowledge** into the scene representation
- Predict **occupancy as attention mask** to restrict the interactions between the agents and their corresponding BEV features.

UniAD - How to Construct?

Perception — Prediction — **Planning**

Planner (Proposed in UniAD)

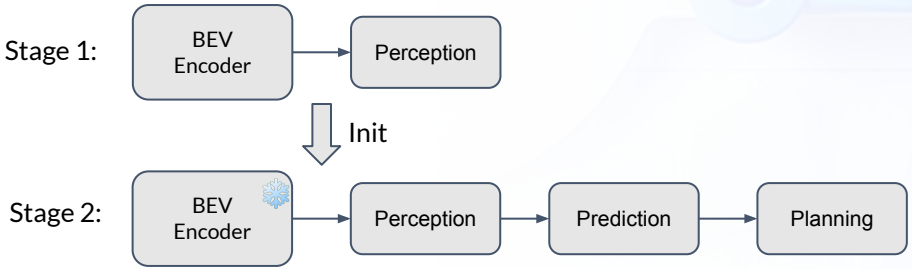


- **Ego-vehicle query:** consistently models the ego-vehicle
- **Collision optimization:** Steer the predicted trajectories clear of predicted occupancy.

The Recipe - How to Train?

Two-phase 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
- Converging faster

Planning-oriented Autonomous Driving

Experiments

UniAD - Ablation Results

Tasks benefit 🚀 each other and contribute to safe planning

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

Even outperforms LiDAR-based counterparts on planning

Planning

†: LiDAR-based

Camera-based

Method	L2(m)↓				Col. Rate(%)↓			
	1s	2s	3s	Avg.	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**

UniAD - Results

SOTA performance on all investigated tasks

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

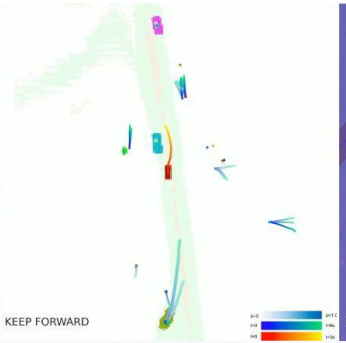
Method	Lanes↑	Drivable↑	Divider↑	Crossing↑
VPN [63]	18.0	76.0	-	-
LSS [66]	18.3	73.9	-	-
BEVFormer [48]	23.9	**77.5**	-	-
BEVerse† [92]	-	-	**30.6**	**17.2**
UniAD	**31.3**	69.1	25.7	13.8

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



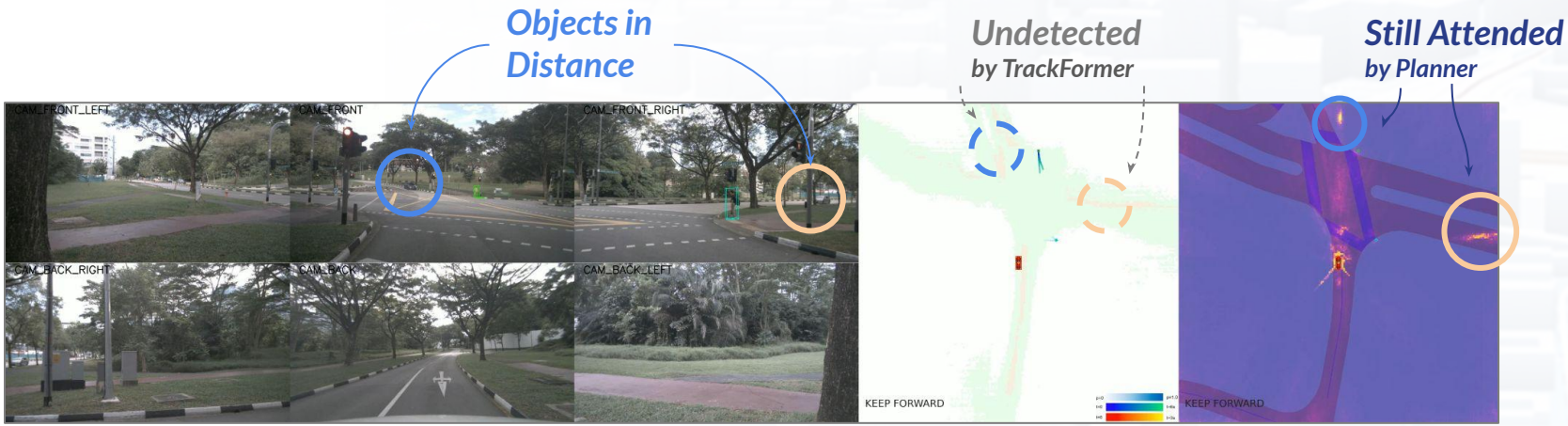
Attention on ***Forward***

Change Lane!

Attention on ***Backward***

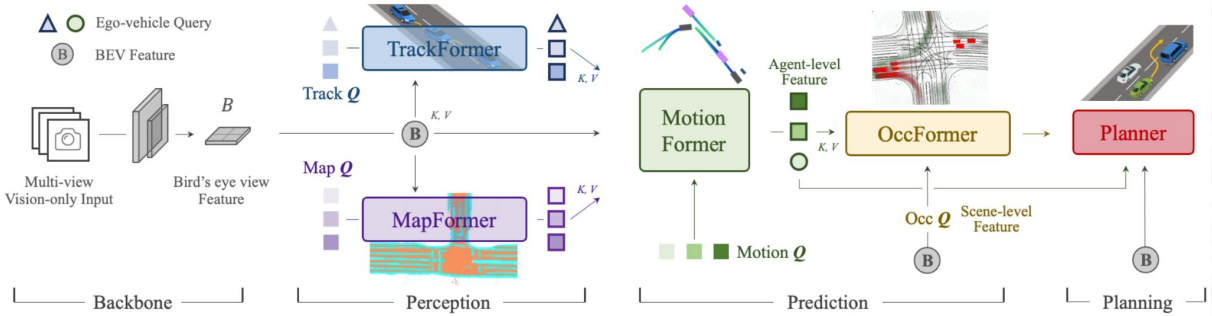
UniAD - Recover from Upstream Errors

Planner could still attend to ‘undetected’ regions/objects



One-page Summary

- **Planning-oriented Philosophy**: An end-to-end autonomous driving (AD) framework in pursuit of safe planning, equipped with a wide span of AD tasks.
- **Unified Query** design: *Queries* as interfaces to connect all tasks, and transmit upstream knowledge to planner.
- **State-of-the-art (SOTA) Performance** with vision-only input.
- **First Step towards Autonomous Driving Foundation Models**



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

What's next? beyond UniAD

Embracing Foundation Models for Autonomous Driving

UniAD v2

Data & Training Strategy

- Multiple datasets with labels for various tasks?

Shippable Algorithm

- More modules integration, extensible to applications (e.g. V2X)

Closed-loop System

- Closed-loop training and testing in simulator & real world

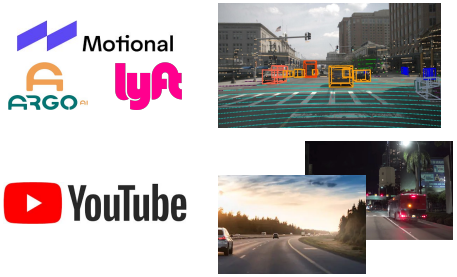
Check out the latest **Survey Paper**!

https://github.com/OpenDriveLab/End-to-end-Autonomous-Driving

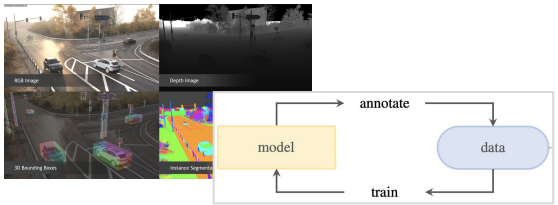
Beyond UniAD: DriveAGI

Data-centric Pipeline

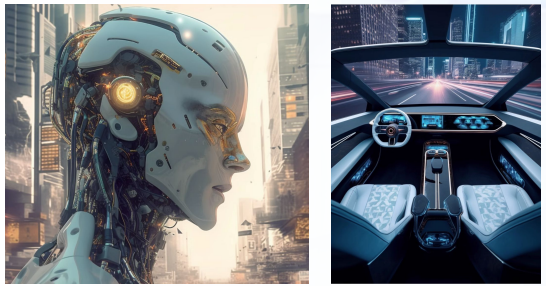
Data Collection



Data Generation



Pre-training DriveCore



Universal Foundation Model
for autonomous driving

How to formulate?
What's the objective goal?

Applications

Autonomous Driving

Broader Impact

Partial photo by courtesy of online resources.

OpenDriveLab

Poster: THU-AM-131

THANKS

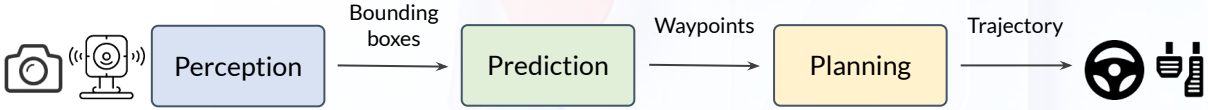
https://opendrivelab.com

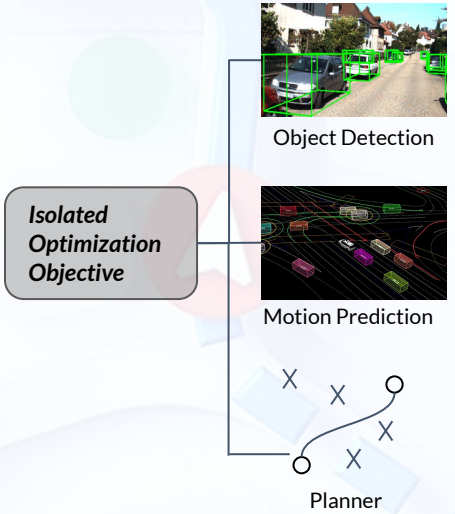
上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

WUHAN UNIVERSITY 1893 武汉大学

sensetime

Background - Autonomous Driving (AD) Systems

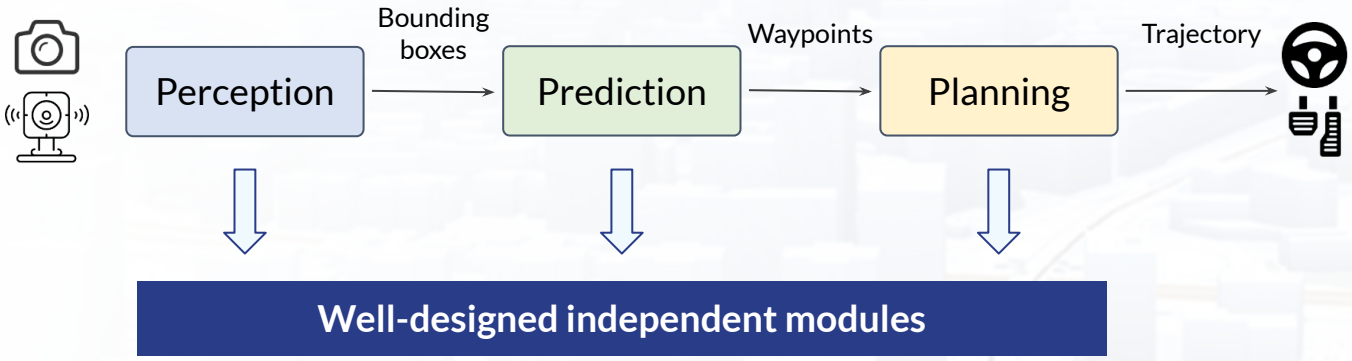


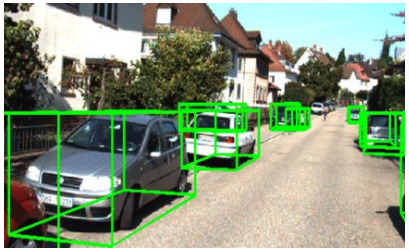


DARPA 2007 Urban Challenge
Photo credit to
https://www.darpa.mil/about-us/timeline/darpa-urban-challenge

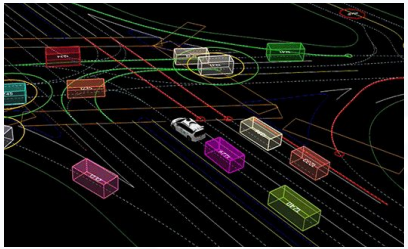
Background - Autonomous Driving (AD) Systems

Photos credit to DARPA 2007 Urban Challenge, Waymo, Cruise, and other online resources.

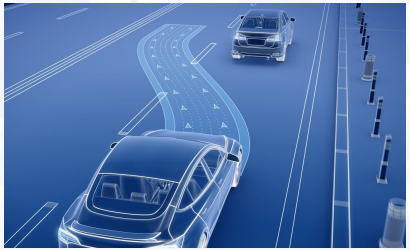




What are around?



How will they go in the future?

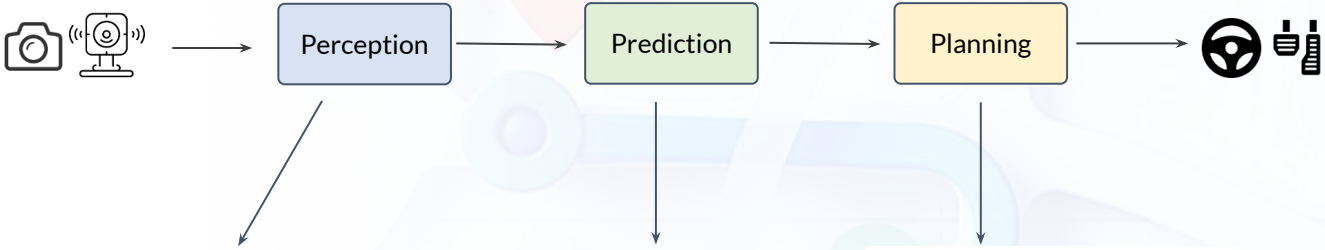


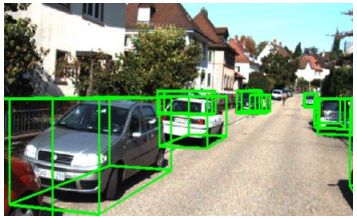
Where should I go?

Background - Autonomous Driving (AD) Systems

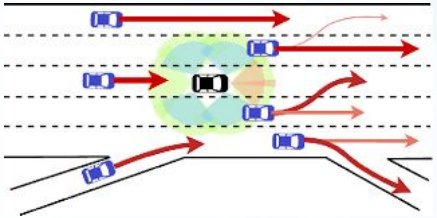
DARPA 2007 Urban Challenge

Photo credit to
https://www.darpa.mil/about-us/timeline/darpa-urban-challenge

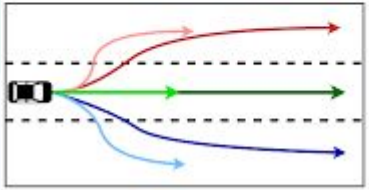




What is situation around?



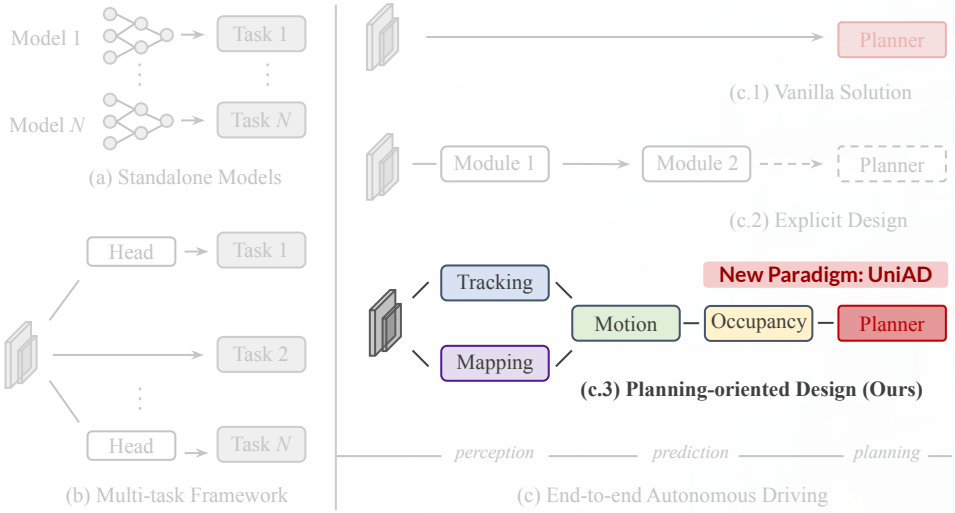
What situation will be in the future?



In which direction should I go?

Motivation- Towards Reliable Planning

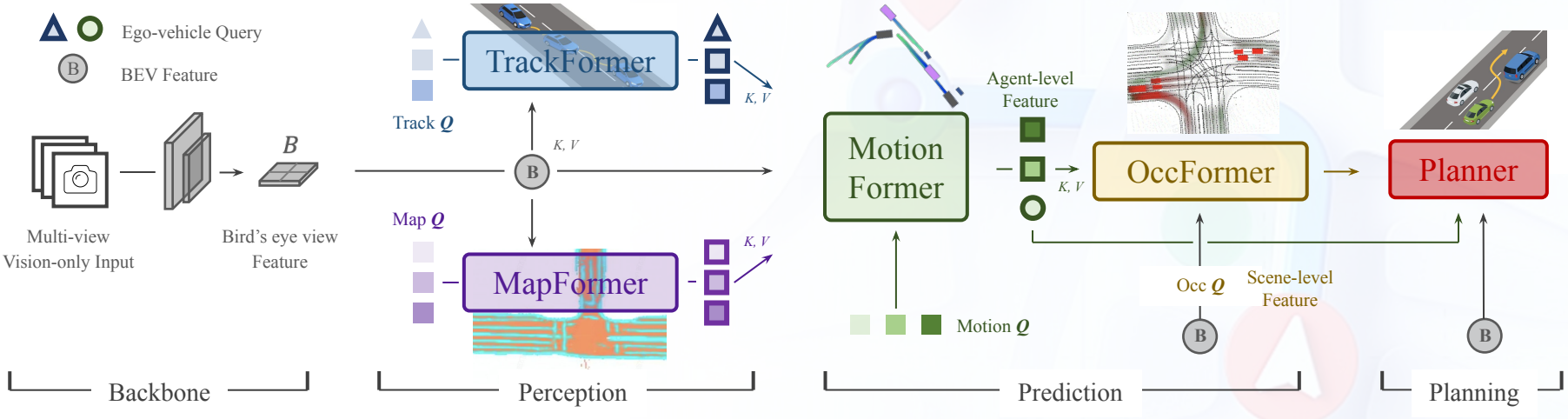
Ours: Planning-oriented Autonomous Driving



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 optimize** all (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 - How to Construct? Pipeline

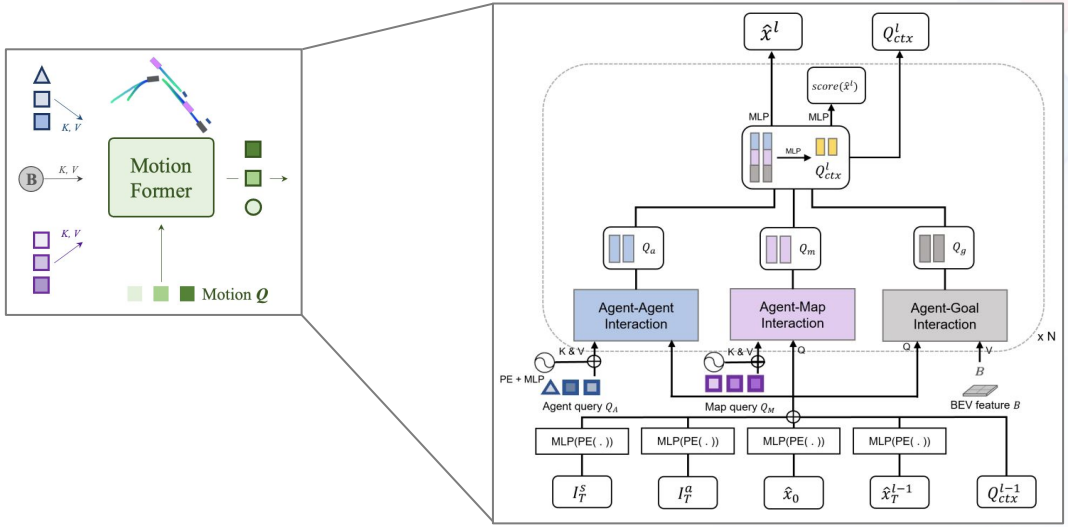


Transformer-based Structure

UniAD - How to Construct?

Perception | Prediction | Planning

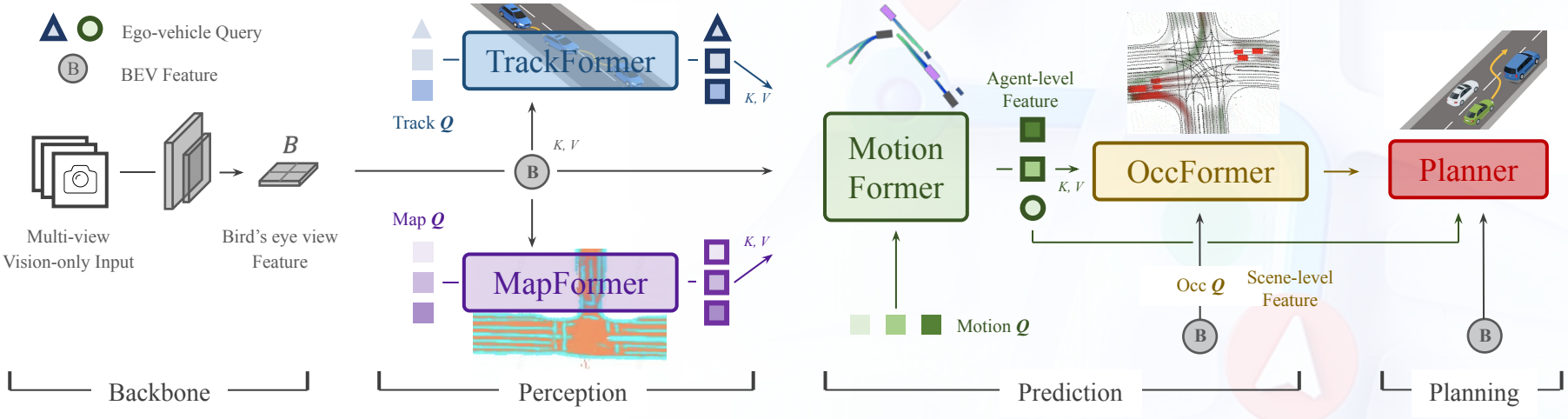
MotionFormer (Proposed in UniAD)



- Diverse **relation modelings** via attentions: Agent-agent, agent-map, agent-goal

- **Non-linear optimization**: Adjust ground-truth trajectory based on upstream predictions

UniAD - How to Construct? Pipeline



- 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**

UniAD - Visualizations

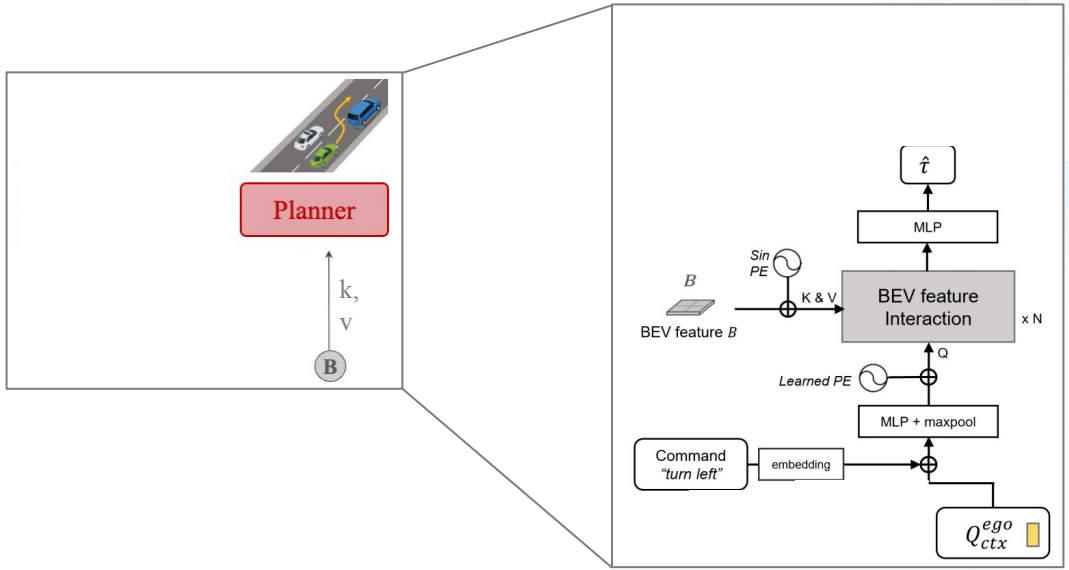
Planner attends to crucial areas in complex scenes

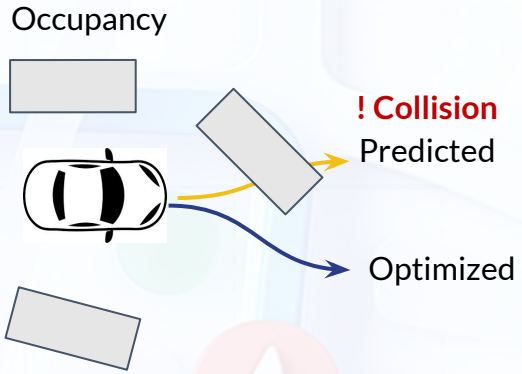
UniAD - Recover from Upstream Errors

Planner could still attend to ‘undetected’ regions/objects

UniAD - How to Construct?

Perception — Prediction — **Planning**





- **Ego-vehicle query:** consistently models the ego-vehicle
- **Collision optimization:** Steer the predicted trajectories clear of predicted occupancy.

UniAD v2

Data & Training strategy

- Multiple datasets with labels for various tasks?

Algorithm

- More modules integration, extend to V2X

Closed-loop

- Closed-loop training and testing in simulator & real world

Scale-up

- Scale up data & model toward a foundation model!

Beyond UniAD: DriveAGI

Data-centric Pipeline

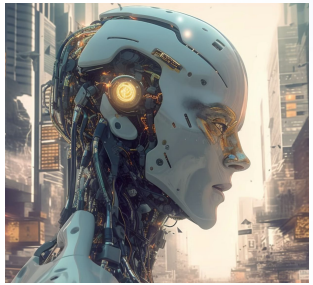
Data Collection

- Unlabeled data
- In-domain data
- Out-of-domain data

Data Generation

- AIGC
- Data-driven simulator
- Data engine

Pre-training DriveCore



Universal Foundation Model
for autonomous driving

How to formulate?
What's the objective goal?

Applications

Autonomous Driving

- Perception
- Prediction
- Decision (Planning)
- Embodied (Language-driven)

Broader Impact

- Robotics
- Embodiment
- V2X application
- ...

Beyond UniAD: DriveAGI

Data-centric Pipeline

Data Processing

- Public datasets (w/ labels)
- Online data (w/o labels)

Good annotation | Large domain gap | Unknown configuration | Large scale

Data Generation

- AIGC
- Data-driven Simulation
- Data engine (eg, SAM)

Long-tail/corner cases | Large scale | Pseudo labels | Data veracity

Pre-training DriveCore

What is a (universal) foundation model for autonomous driving?

Multi-modality | Large-scale data | Generalization on tasks | Decision Intelligence | Large model | Publicly available

Key challenges:

- **General object/scene representation**
- Domain generalization / OOD cases
- General pre-training tasks
- Incorporate LLMs “effectively”

Applications

AD Tasks

- Perception
- Prediction
- Decision (Planning)
- Embodied (Language-driven)

Broader Impact

- Robotics
- V2X application
- ...