OpenDriveLab

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

Not yet...

Could Foundation Models really resolve End-to-end Autonomy?

Hongyang Li

Research Scientist, **Shanghai AI Lab** /
Assistant Professor, **University of Hong Kong**

June 18, 2024

Slides credit:
Jiazhi, Zetong, Li, Huijie, Shenyuan, etc.

Outline

- **Introduction to End-to-end Autonomous Driving (E2E AD)**
  - Setup / Definition
  - Datasets and Evaluation
  - Motivation
  - Classical Approaches Walkthrough
- **Research Panorama**
  - Past / Present / Future
  - Concurrent Work and Future
    - GenAD (CVPR 2024 Highlight)
    - Vista (in arXiv)
- **Challenges and Closing Remarks**
  - Data / Methodology / Compute / Goal

OpenDriveLab

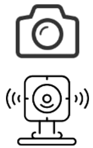
上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Part 1:
Introduction to End-to-end Autonomous Driving

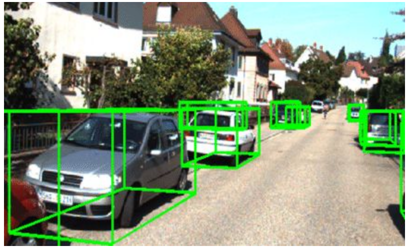
Setup / Metric / Motivation

Preliminary | Problem Setup

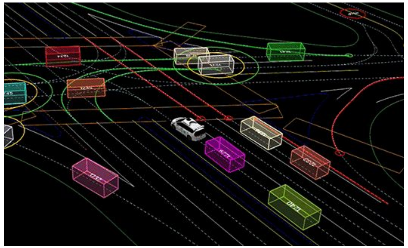
Challenge | Various weathers, illuminations, and scenarios



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



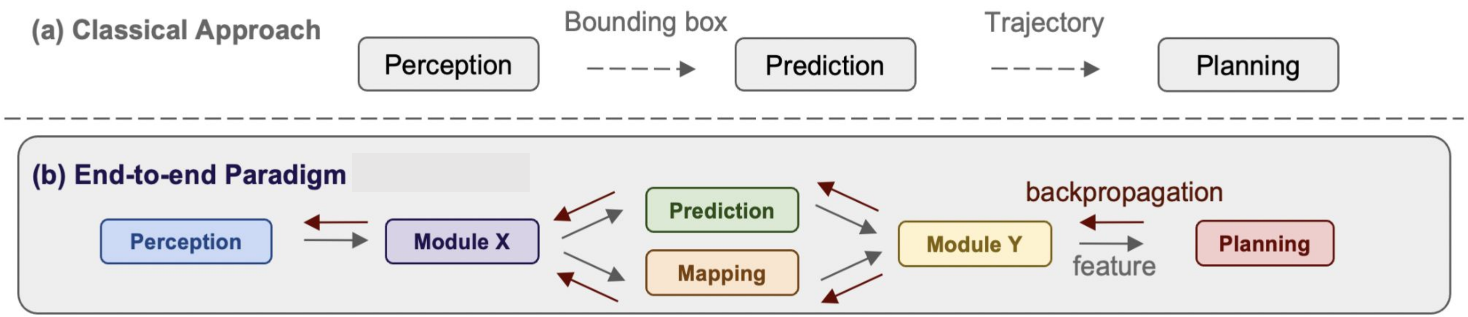
What are around?



How will they go in the future?

Where should I go?

End-to-end | Definition



End-to-end autonomous driving system - A suite of fully differentiable programs that:

- take raw sensor data as input
- produce a plan and/or low-level control actions as output

Preliminary | Datasets and Evaluation

Note:
https://github.com/autonomousvision/navsim/blob/main/docs/metrics.md

Real-world Collected

Dataset	Scale	Behavior & Interaction	Planning Task Evaluation	
			Strategy	Metrics
nuScenes	5.5 h	Realistic	Open-loop (Log-replay)	- L2 Error - Collision Rate
Waymo*	11 h			
Argoverse2*	4.2 h			
nuPlan*	120 h	ML-based	Closed-loop (Interactive)	- Average Displacement Error (ADE) - Final Displacement Error (FDE) - Collision Rate - Comfort Score - PDM Score [Note]

*Perception subset (with visual inputs)

Preliminary | Datasets and Evaluation

Note:
https://github.com/autonomousvision/navsim/blob/main/docs/metrics.md

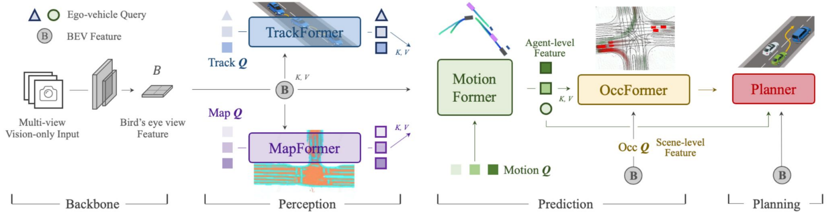
	Dataset	Scale	Behavior & Interaction	Planning Task Evaluation	
				Strategy	Metrics
Real-world Collected	nuScenes	5.5 h	Realistic	Open-loop (Log-replay)	- L2 Error - Collision Rate
	Waymo*	11 h			
	Argoverse2*	4.2 h			
	nuPlan*	120 h	ML-based	Closed-loop (Interactive)	- Average Displacement Error (ADE) - Final Displacement Error (FDE) - Collision Rate - Comfort Score - PDM Score [Note]
Synthetic generated	DriveSim	Unlimited	Handcrafted & ML-based	Closed-loop (Interactive)	- N/A
	Carla				- Driving Score = Route Completion $*\prod$ Infraction Penalty

*Perception subset (with visual inputs)

Motivation | Why end to end?

Advantages

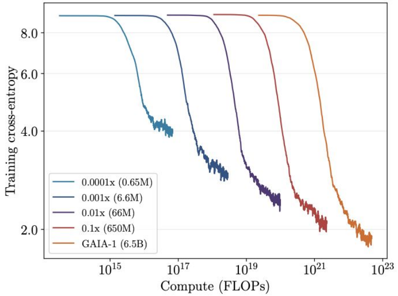
- **Global optimization:** when perception fails/inferior, planning still could work.



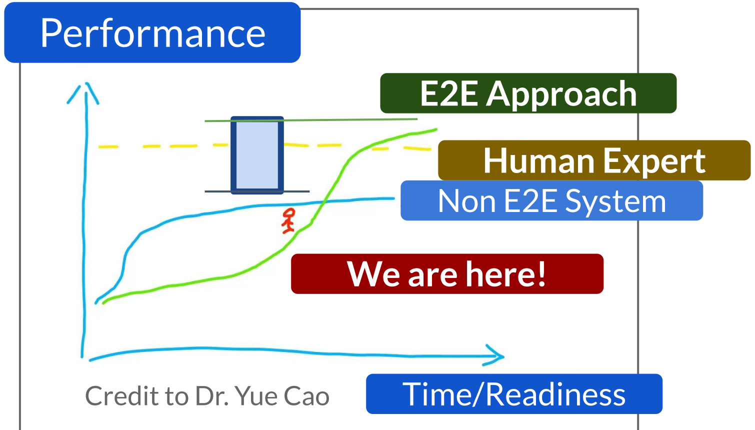
Hu et al. Planning-oriented Autonomous Driving. CVPR 2023.

- ~~"**Efficiency**" / faster due to one single net?~~ ✅ ❌

- **Scaling law:** massive amount of data + infra/compute —> **strong generalization**



Hu et al. GAIA-1: A Generative World Model for Autonomous Driving.



Credit to Dr. Yue Cao

Motivation | Why end to end?

Disadvantages

- **Lack of interpretability**, due to the e2e neural network. ✅

- ~~Unfair evaluation? E.g. open-loop L2 metric~~ ❌
  - [Ref] Li et.al, Is Ego Status All You Need for Open-Loop End-to-End Autonomous Driving? CVPR 2024

- ~~Lack of data / Simulation (sim2real) / etc..~~ ❌

Classic algorithm: TransFuser (1/2) - Motivation

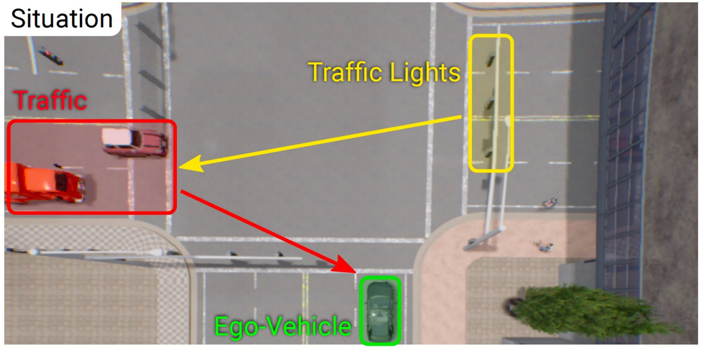
LiDAR Point Cloud

- 3D information
- Robustness for weather variations

RGB Camera

- Traffic light state
- Long-range perception

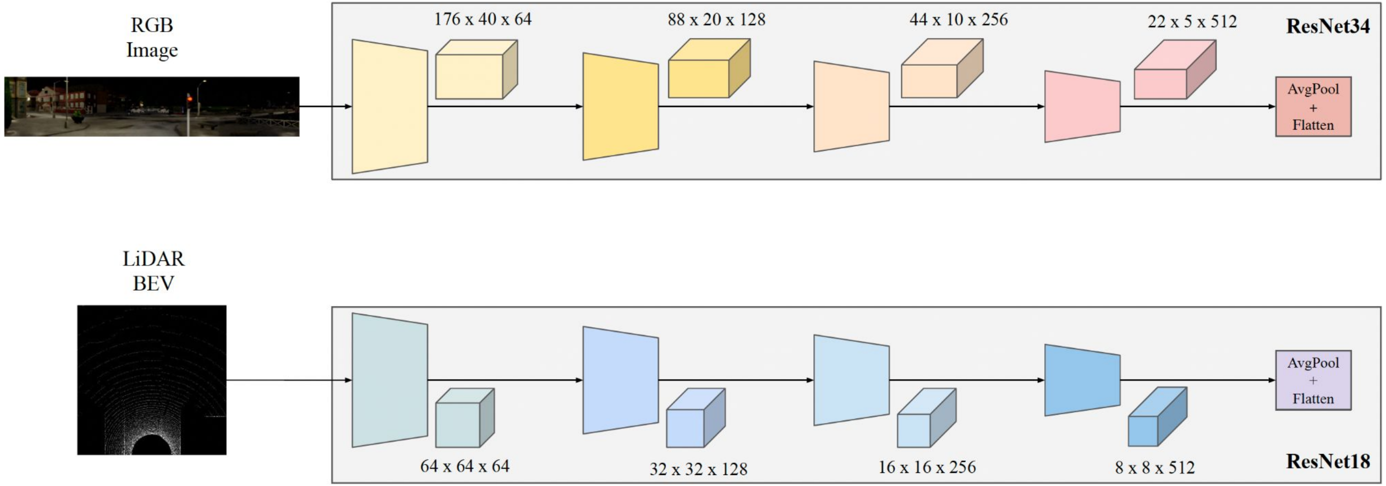
Combine the best of both worlds



Whole-scene understanding
for safe driving

Prakash et al. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving. CVPR 2021.

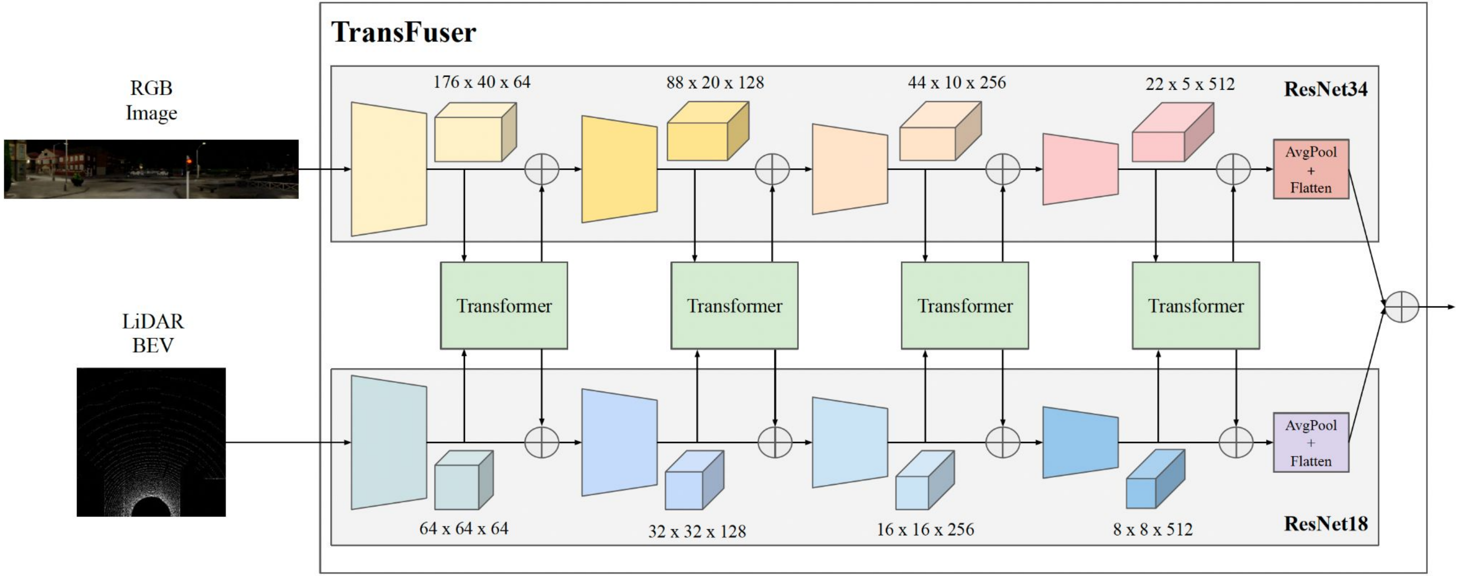
Classic algorithm: TransFuser (2/2)



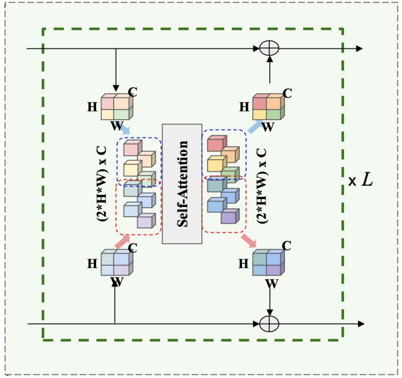
- **Dual-stream network** to extract modality-specific features

Prakash et al. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving. CVPR 2021.

Classic algorithm: TransFuser (2/2)



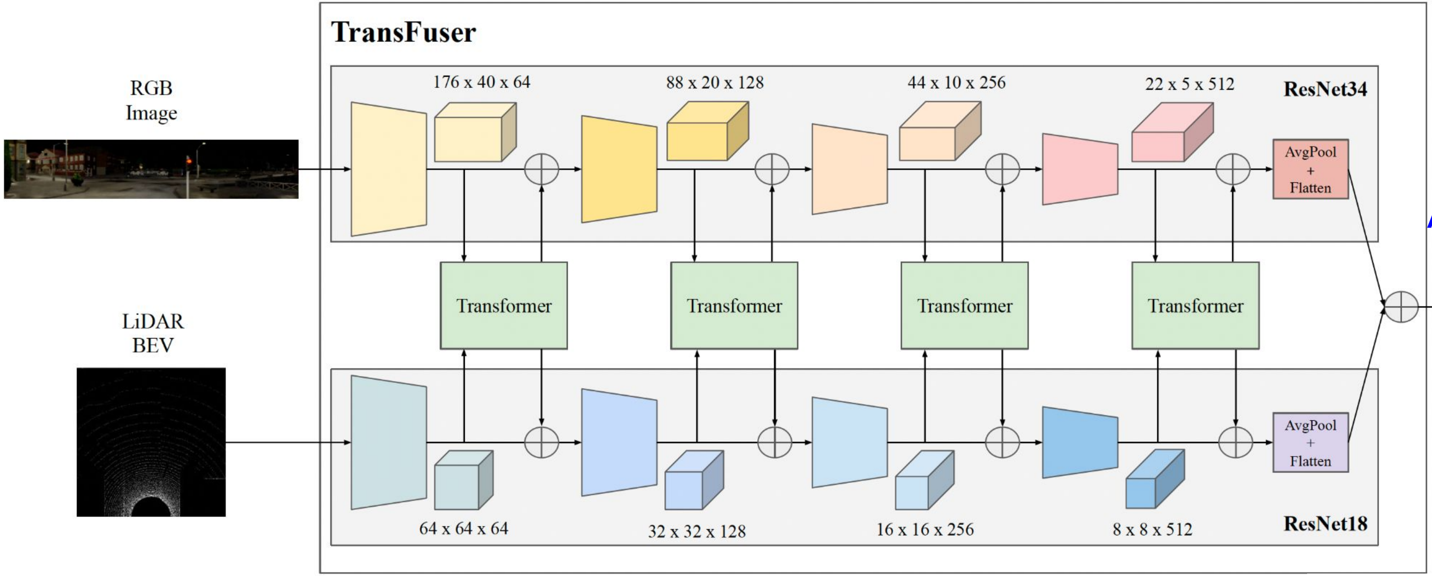
Transformer



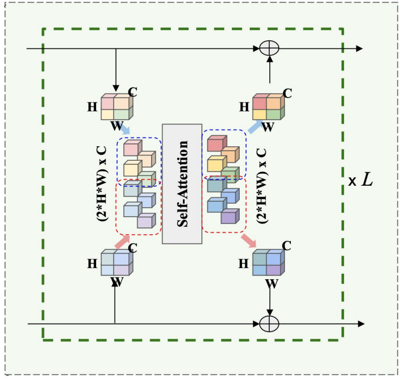
- **Dual-stream network** to extract modality-specific features
- **Transformer** to effectively fuse feature across modalities

Prakash et al. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving. CVPR 2021.

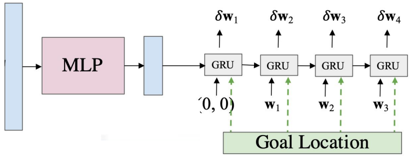
Classic algorithm: TransFuser (2/2)



Transformer



A sequence of waypoints (w1…w4)



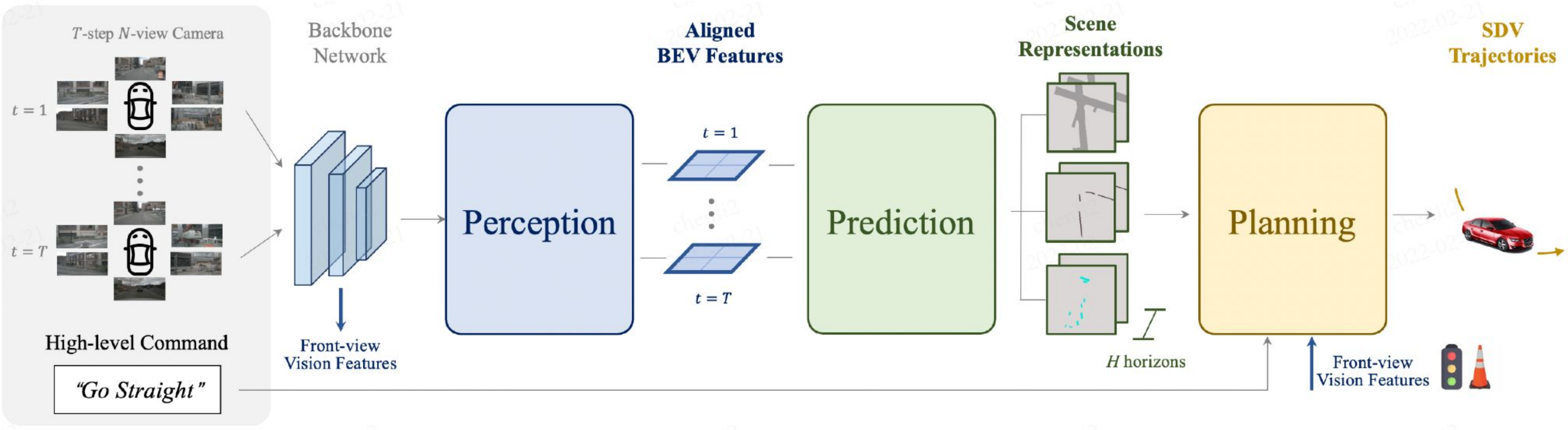
GRU Head

- **Dual-stream network** to extract modality-specific features
- **Transformer** to effectively fuse feature across modalities
- **Simple GRU head** to convert global context into waypoints

Method	Driving Score ↑
Late Fusion	22 ± 4
Geometric Fusion	27 ± 1
TransFuser (Ours)	$\mathbf{47} \pm \mathbf{6}$
Privileged Expert	*77 ± 2*

Prakash et al. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving. CVPR 2021.

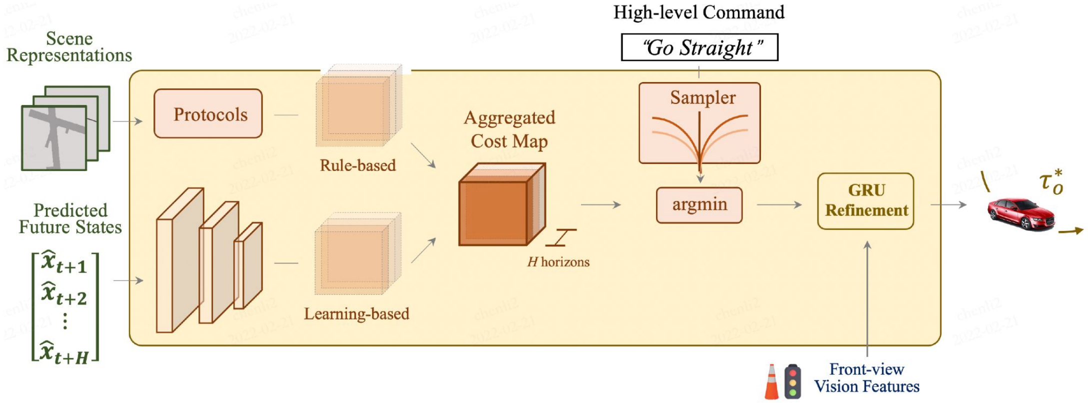
Classic algorithm: ST-P3 (1/2)



- Incorporate perception and prediction tasks to **enrich feature learning**

Hu et al. ST-P3: End-to-end Vision-based Autonomous Driving via Spatial-Temporal Feature Learning. ECCV 2022.

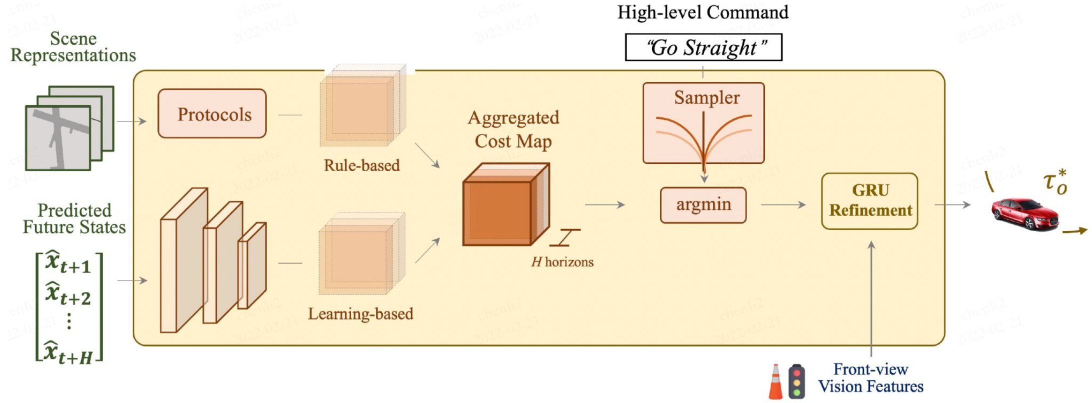
Classic algorithm: ST-P3 (2/2)

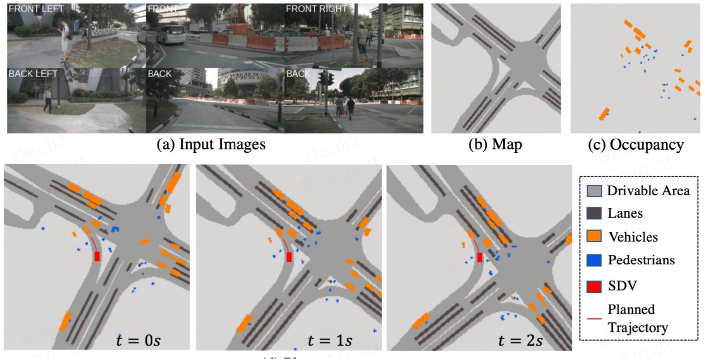


- Incorporate perception and prediction tasks to **enrich feature learning**
- Plan safe routes with **cost optimization**

Hu et al. ST-P3: End-to-end Vision-based Autonomous Driving via Spatial-Temporal Feature Learning. ECCV 2022.

Classic algorithm: ST-P3 (2/2)



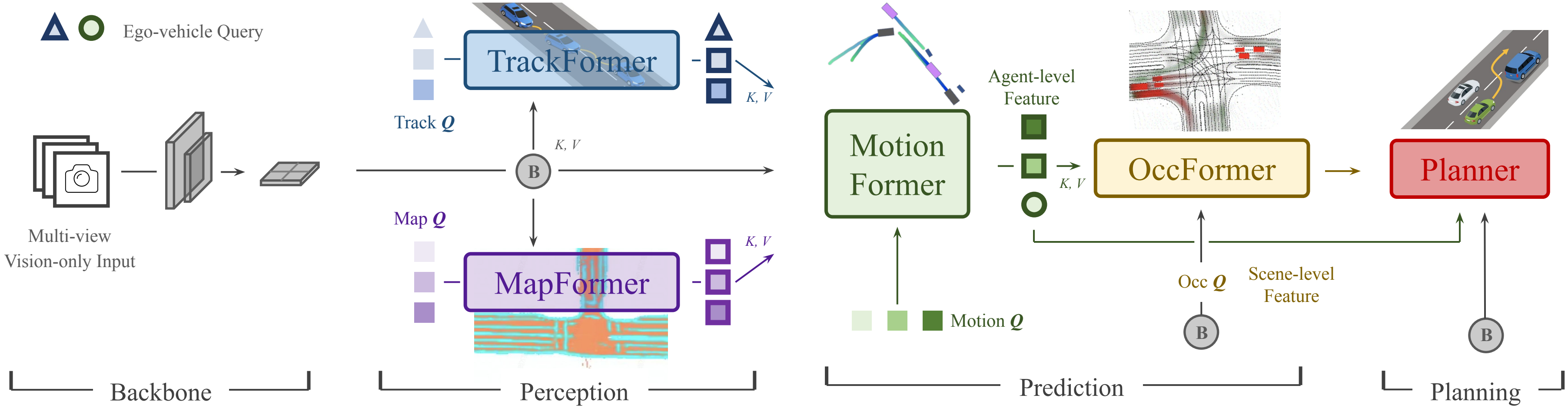


- Incorporate perception and prediction tasks to **enrich feature learning**
- Plan safe routes with **cost optimization**
- End-to-end driving with **interpretable scene representations**

Hu et al. ST-P3: End-to-end Vision-based Autonomous Driving via Spatial-Temporal Feature Learning. ECCV 2022.

Classic algorithm: UniAD

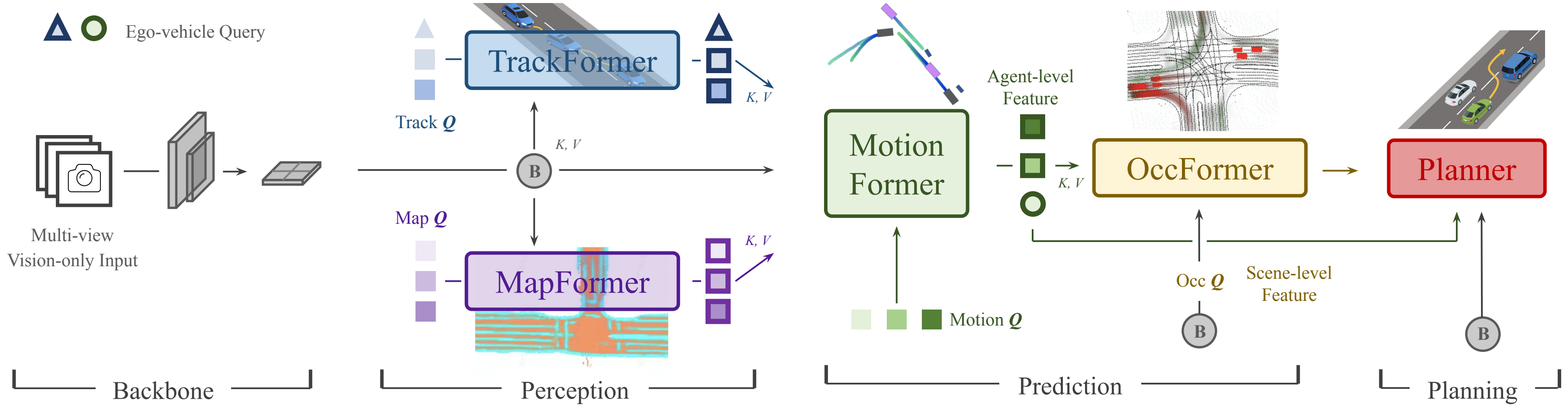
- Track agents across time
- Predict long-term trajectory
- Plan final trajectory
- Avoid collision



- Segment map elements
- Scene-level representation

Hu et al. Planning-oriented Autonomous Driving. CVPR 2023.

Classic algorithm: UniAD



- Entire pipeline connected by queries
- Tasks coordinated with queries
- Interactions modeled by attention

Unified Query

Transformer-based

First time to unify full-stack AD tasks!

Hu et al. Planning-oriented Autonomous Driving. CVPR 2023.

Core in UniAD: Planning-oriented, not a MTL framework.

Tasks benefit🚀each other and contribute to safe planning

ID	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.↓
	Modules					Tracking			Mapping		Motion Forecasting			Occupancy Prediction				Planning	
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**

Task Synergy Effect:

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

Hu et al. Planning-oriented Autonomous Driving. CVPR 2023.

Why mention these Classic algorithms?

Table 2. **Open-Loop Evaluation on nuScenes.** FeD achieves state-of-the-art open-loop evaluation performance on nuScenes [5] validation set compared with both none-LLM based methods and LLM-based GPT-Driver [58]. We evaluate FeD on two different measures of metrics for fair comparison[1].

Metrics	Method	L2 (m) ↓ 1s	2s	3s	Avg.	Collision (%) ↓ 1s	2s	3s	Avg.
ST-P3	ST-P3 [34]	1.33	2.11	2.90	2.11	0.23	0.62	1.27	0.71
	VAD [40]	**0.17**	0.34	0.60	0.37	0.07	0.10	0.24	0.14
	GPT-Driver [58]	0.20	0.40	0.70	0.44	0.04	0.12	0.36	0.17
	FeD	0.21	**0.33**	**0.49**	**0.34**	**0.00**	**0.03**	**0.15**	**0.06**

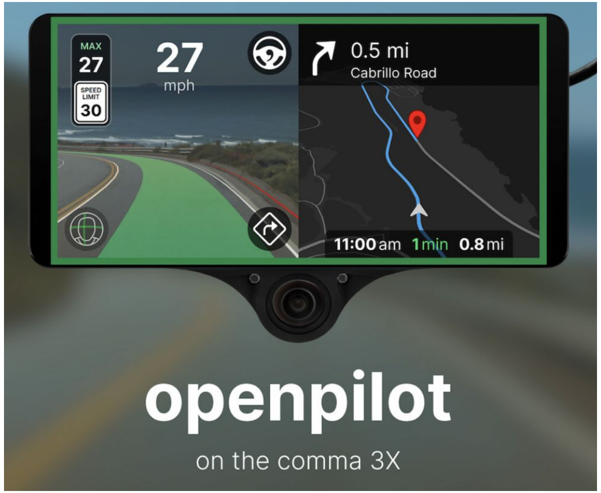
UniAD	NMP [94]	-	-	2.31	-	-	-	1.92	-
	SA-NMP [94]	-	-	2.05	-	-	-	1.59	-
	FF [33]	0.55	1.20	2.54	1.43	0.06	0.17	1.07	0.43
	EO [43]	0.67	1.36	2.78	1.60	0.04	0.09	0.88	0.33
	UniAD [35]	0.48	0.96	1.65	1.03	0.05	0.17	0.71	0.31
	GPT-Driver [58]	**0.27**	0.74	1.52	0.84	0.07	0.15	1.10	0.44
	FeD	**0.27**	**0.53**	**0.94**	**0.58**	**0.00**	**0.04**	**0.52**	**0.19**

Baselines of Today's Literature in End—to-end autonomous driving

Snapshot from Zhang et al., Feedback-Guided Autonomous Driving, CVPR 2024.

Industry Credit: Openpilot (~2016)

comma.ai



- *Openpilot is an open source driver assistance system.*
- *Openpilot performs the functions of Automated Lane Centering (ALC) and Adaptive Cruise Control (ACC) for 250+ supported car makes and models.*

A minor (yet respectful) technical report by our team:
https://arxiv.org/abs/2206.08176

Li et al. Level 2 Autonomous Driving on a Single Device: Diving into the Devils of Openpilot.

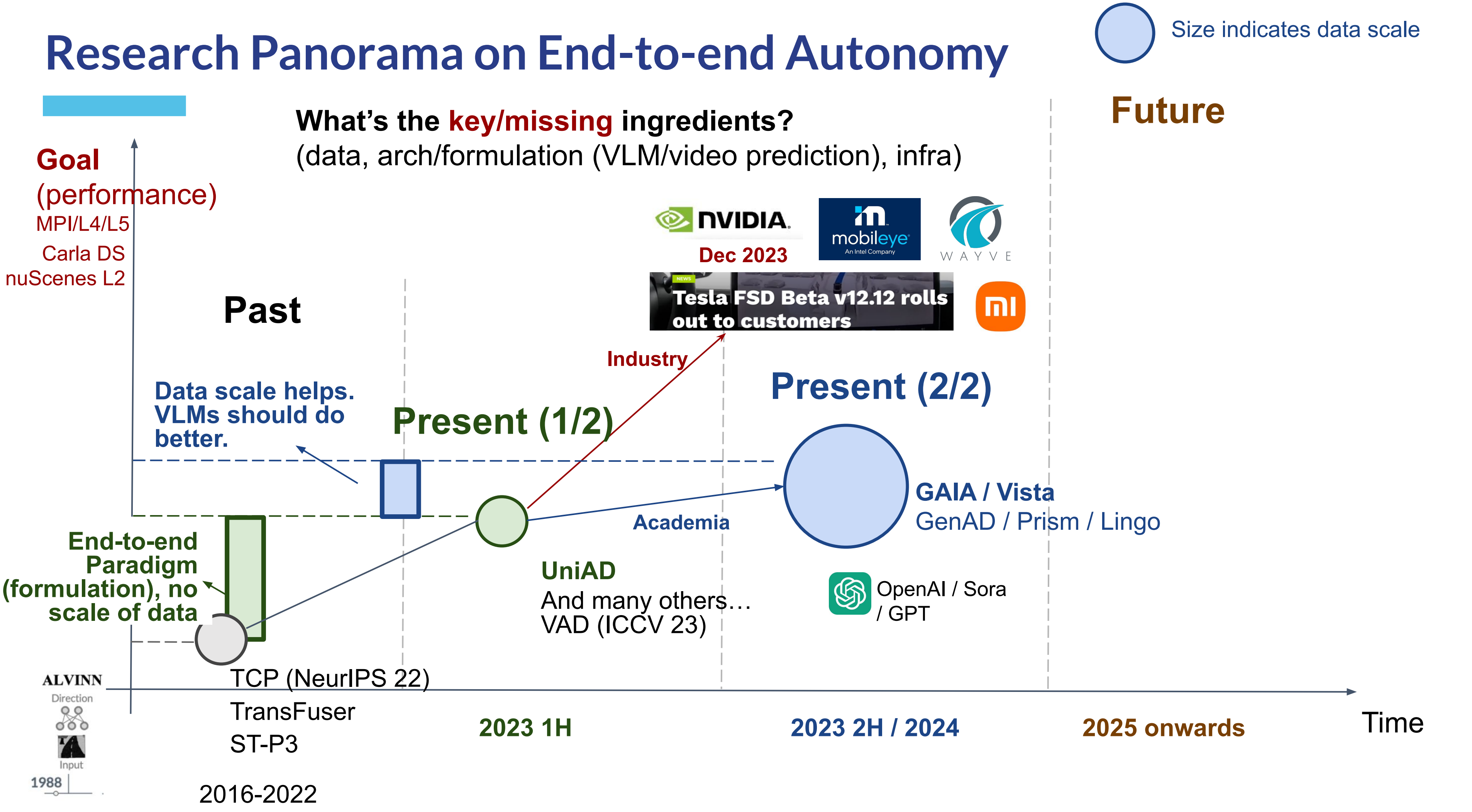
OpenDriveLab

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

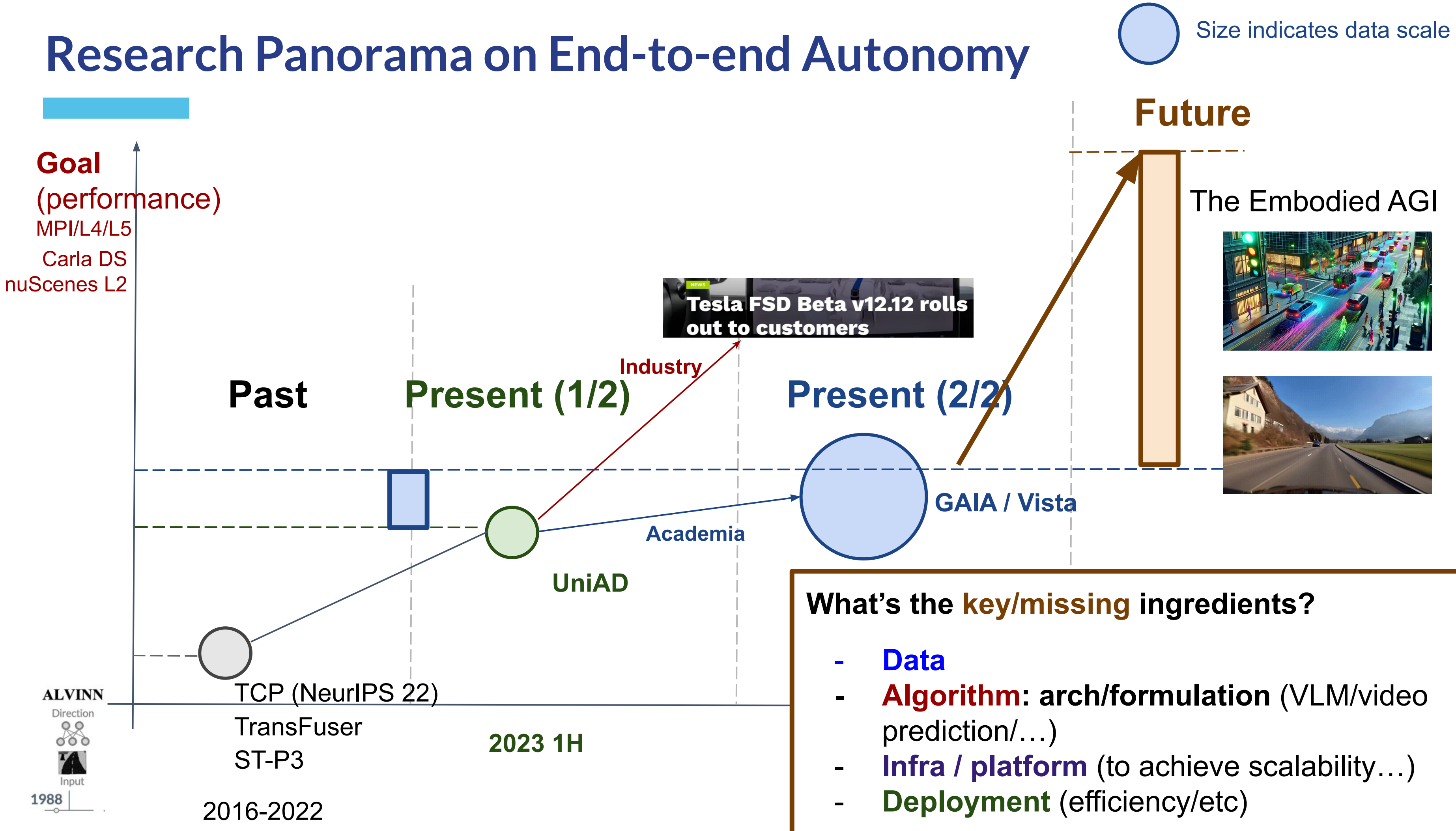
Part 2:
Research Panorama

Past / Present / Future

Research Panorama on End-to-end Autonomy



Research Panorama on End-to-end Autonomy



What's the key/missing ingredients?

- **Data**
- **Algorithm: arch/formulation** (VLM/video prediction/…)
- **Infra / platform** (to achieve scalability…)
- **Deployment** (efficiency/etc)

Our Take on Generalizable End-to-end Autonomy Systems

https://github.com/OpenDriveLab/DriveAGI

DriveAGI

GenAD/ACO **Data**

Massive Data Collection

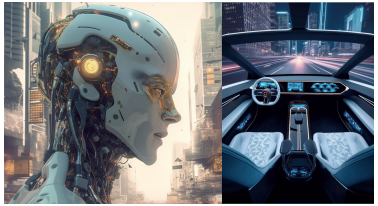
Demonstration Collection

Reward Feedback

Effective Action (policy) Learning

UniAD/DriveAdapter/MP3

Algorithm



Interactive Closed-loop Simulation

Vista/UniSim/NeuRAD

Algorithm

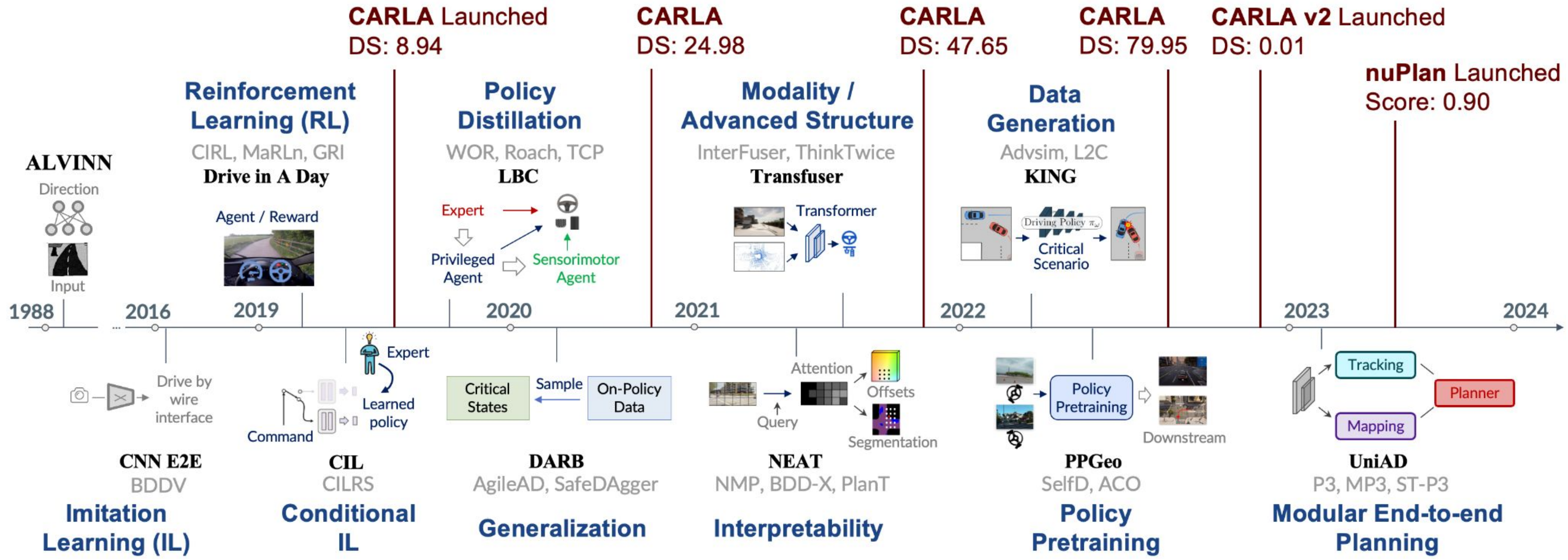
Interpretable Causal Reasoning

DriveLM/Lingo

Algorithm

Safe **Deployment**

Taking it seriously: Roadmap | End-to-end Autonomous Driving



Chen et al. End-to-end Autonomous Driving: Challenges and Frontiers
https://arxiv.org/abs/2306.16927

Concurrent Work

GenAD / Vista / GAIA / etc.

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

OpenDriveLab

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

How to scale up the autonomous driving models?

GenAD: Generalized Predictive Model for Autonomous Driving

CVPR 2024, Highlight

arxiv.2403.09630

Motivation (1/3) | What Makes for Generalized AD Model?

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

Data Distinction:

+ LLMs pretrained on **trillions of unlabeled text tokens** exhibit strong generalization in a variety of domains and applications

- However, existing AD models are established on **limited labeled data**, which hampers their generalization

LLMs

Unlabeled Text Data

✅ **Internet scale**: World knowledge.

✅ **Free of labeling**: Easy to collect and scale up.

✅ *Strong generalization*

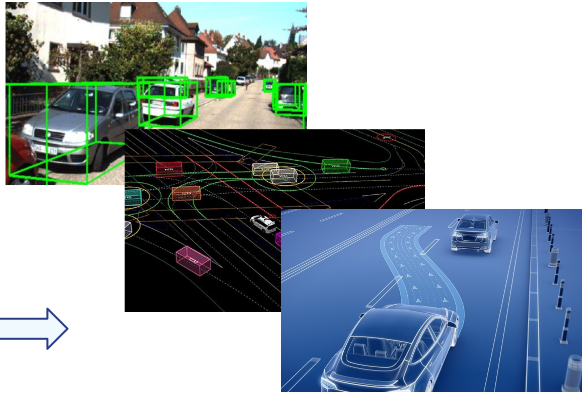
Existing AD Models

Labeled Driving Data

❌ **Small scale**: Limited domain knowledge.

❌ **Intricate labeling process**: Unscalable.

❌ *Poor generalization*



- **Bbox, HD map, etc.**

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

Motivation (2/3) | What Makes for Generalized AD Model?

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

Learning Objective:

- Supervised by **3D labels**
  ❌ Hard to scale without sufficient labeled data

No accessible labeled data

Model ⇢ ❌ ⇢ Model-XL

NUSCENES

- Supervised by **expert features**
  ✅ Scalable with developed expert models (e.g., DINOv2)
  ✅ Focusing on specific objects (e.g., centered or large ones)
  ❌ Ignoring critical details (e.g., small objects)

- Feature map visualization from DINOv2

❌ ***Undesirable** for modeling challenging driving scenes*

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

Motivation (3/3) | What Makes for Generalized AD Model?

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

Our Initiative:

Data: Massive online driving videos

Learning Objective:

- Supervised by "**pixels of future frames**" → **Video Prediction**

✅ **Scalable Data** (easy to collect from the web)
✅ No 3D labeling needed
✅ Better detail preservation
✅ Learning **world knowledge** and **how to drive** inherently

✅ ***Strong** generalization*

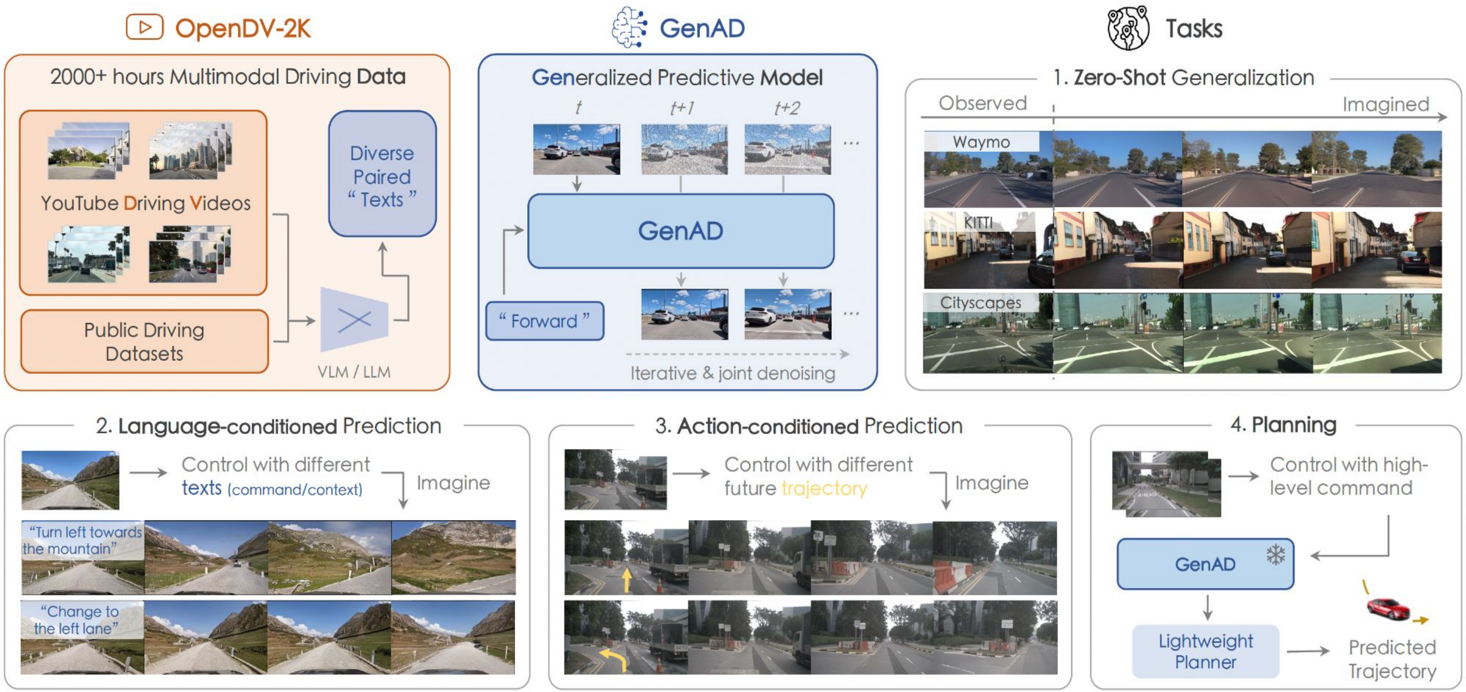
Massive YouTube videos, collected worldwide

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

GenAD | At a Glance

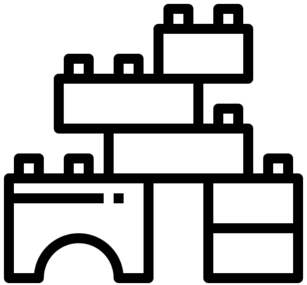
Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

Summary: A **billion-scale video prediction model** trained on **web-scale driving videos**, demonstrating **strong generalization across** a wide spectrum of **domains and tasks**.

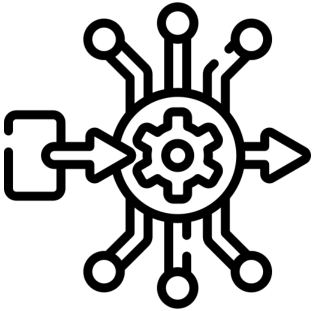


Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

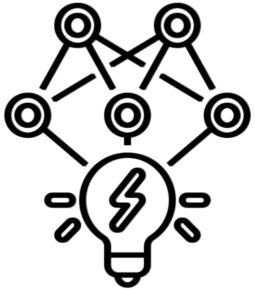
GenAD - Overview



Data



Model

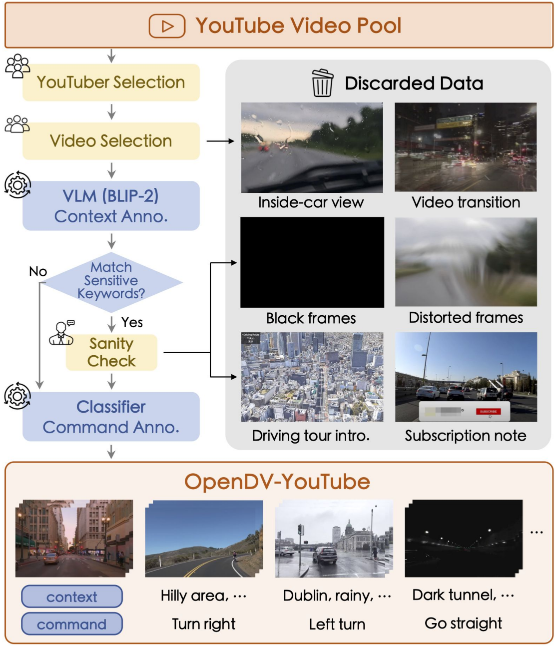


Tasks

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

GenAD | Dataset

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

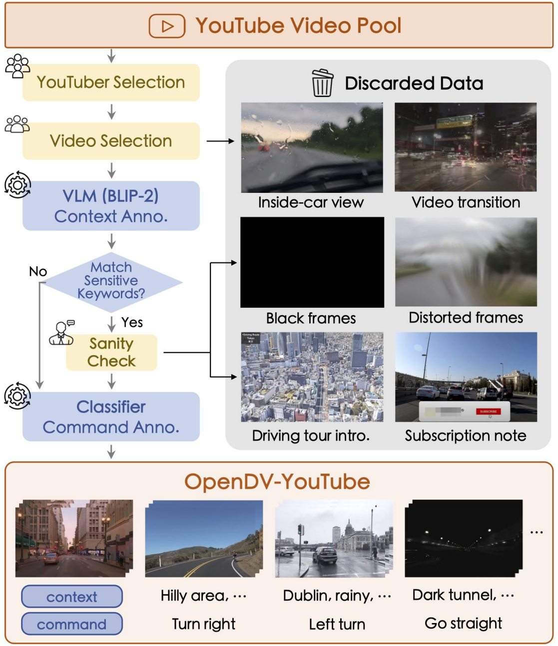


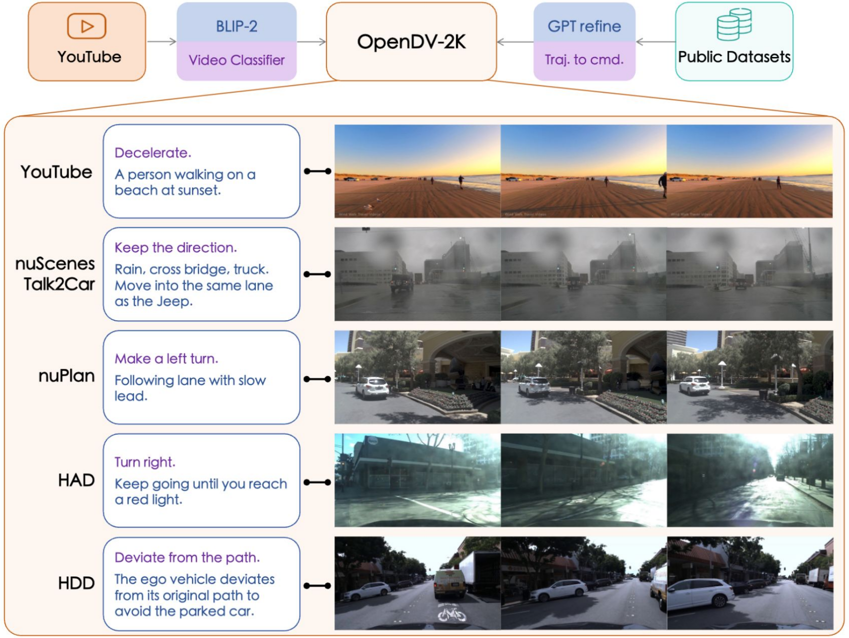
- Rigorous **data collection** and **filtering** strategy

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

GenAD | Dataset

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5





- Rigorous **data collection** and **filtering** strategy

- **Multi-modal** and **Multi-source Nature**
  - Sourced from both **online videos** and **public datasets** for diversity
  - Paired with textual **context** and **command**

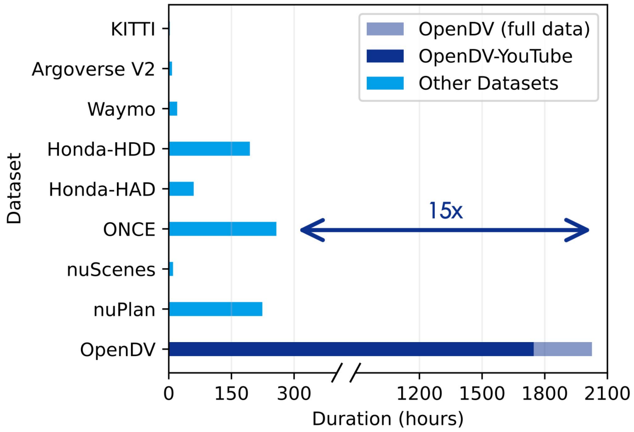
Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

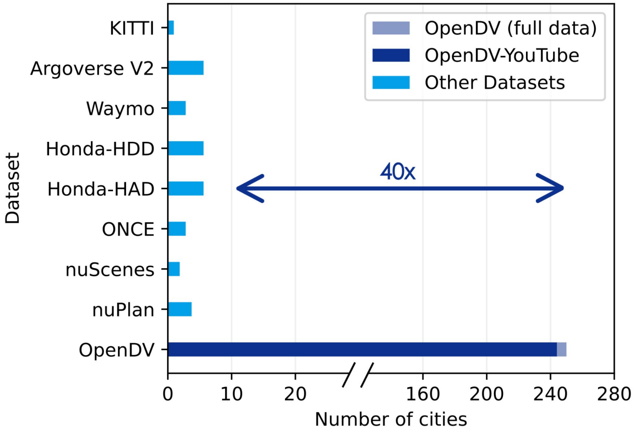
GenAD | Dataset

- ***Largest public dataset*** for autonomous driving
- **≥ 2059 hours, ≥ 244 cities**

	Dataset	Duration (hours)	Front-view Frames	Geographic Diversity Countries	Cities	Sensor Setup
✗	KITTI [30]	1.4	15k	1	1	fixed
✗	Cityscapes [21]	0.5	25k	3	50	fixed
✗	Waymo Open⋆ [97]	11	390k	1	3	fixed
✗	Argoverse 2⋆ [109]	4.2	300k	1	6	fixed
✓	nuScenes [12]	5.5	241k	2	2	fixed
✓	nuPlan⋆ [13]	120	4.0M	2	4	fixed
✓	Talk2Car [24]	4.7	-	2	2	fixed
✓	ONCE [72]	144	7M	1	-	fixed
✓	Honda-HAD [51]	32	1.2M	1	-	fixed
✓	Honda-HDD-Action [84]	104	1.1M	1	-	fixed
✓	Honda-HDD-Cause [84]	32	-	1	-	fixed
✓	OpenDV-YouTube (Ours)	1747	60.2M	$\geq 40^\dagger$	$\geq 244^\dagger$	uncalibrated
-	**OpenDV-2K** (Ours)	**2059**	**65.1M**	$\mathbf{\geq 40}^\dagger$	$\mathbf{\geq 244}^\dagger$	**uncalibrated**

OpenDV-2K (Ours) 🚀

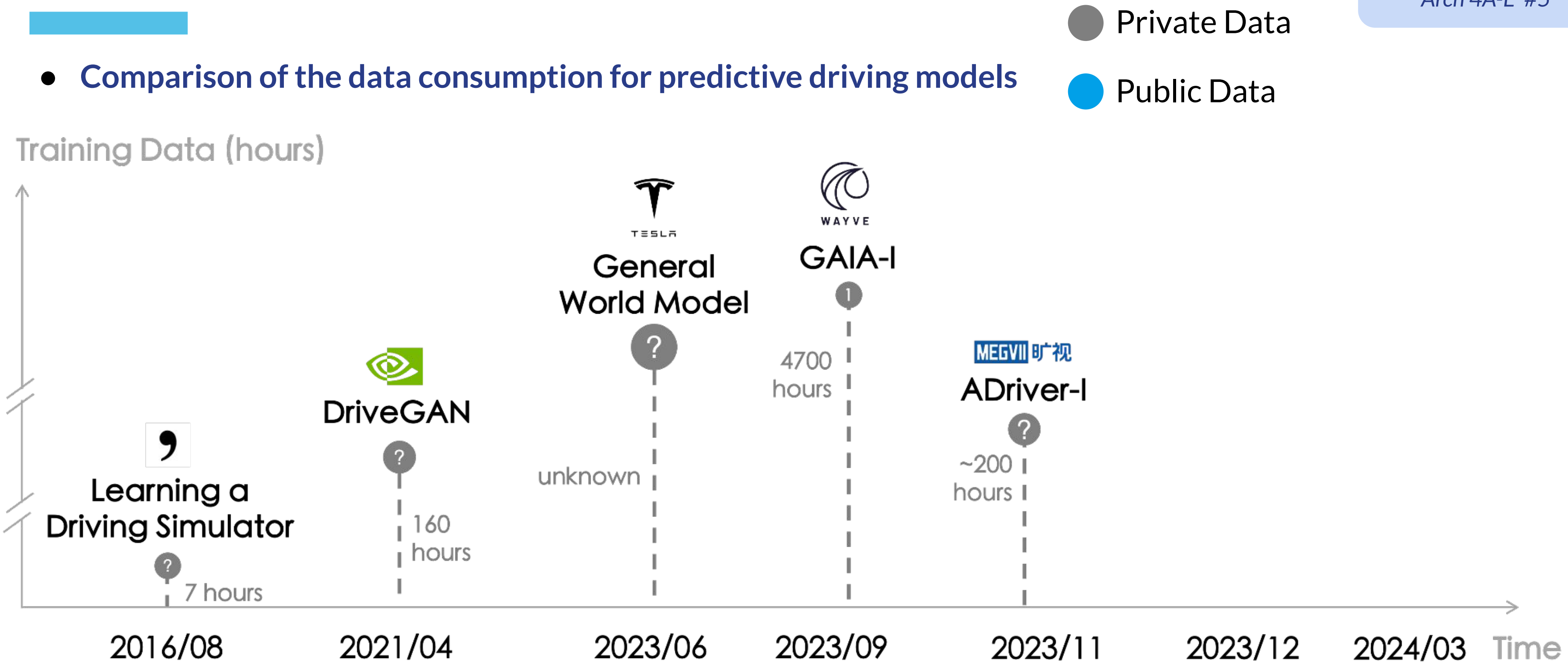




Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

GenAD | Dataset

- **Comparison of the data consumption for predictive driving models**

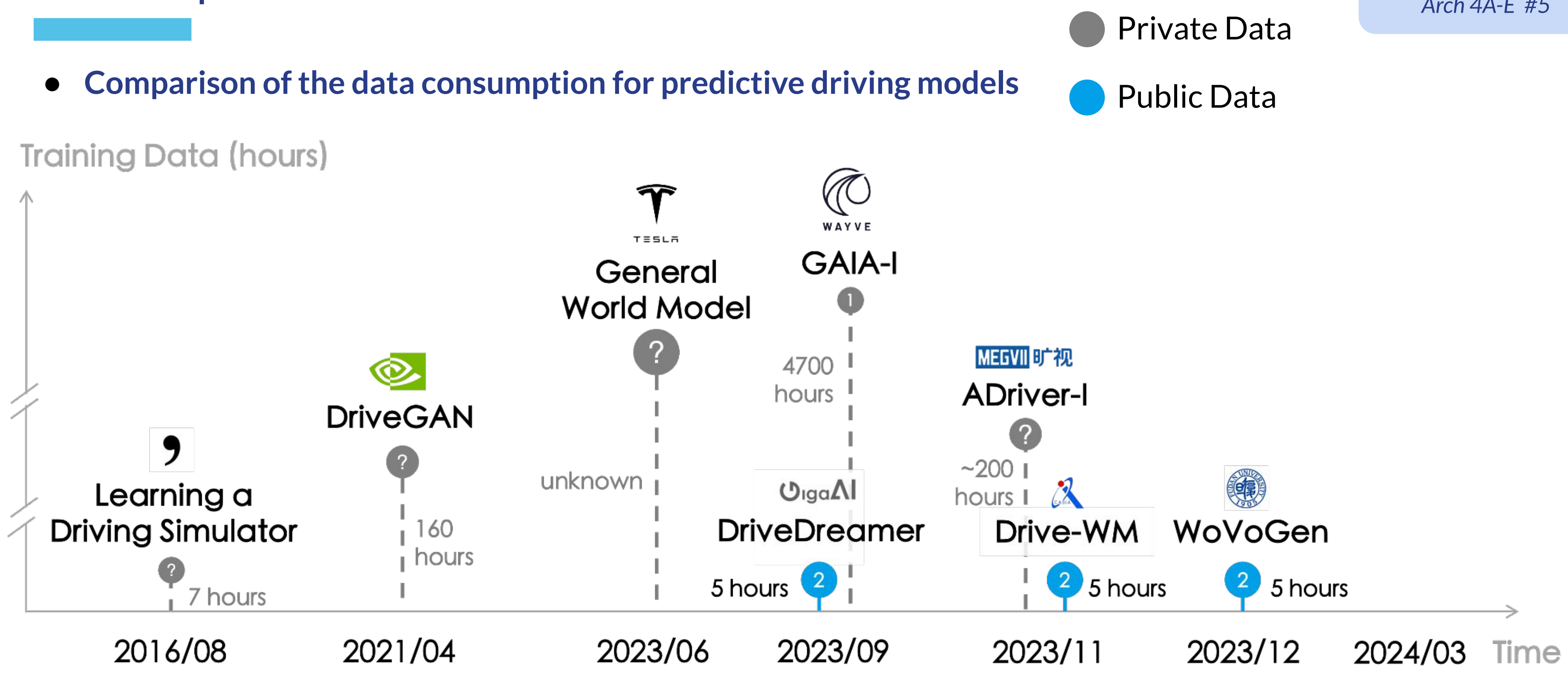


Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

GenAD | Dataset

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

- **Comparison of the data consumption for predictive driving models**

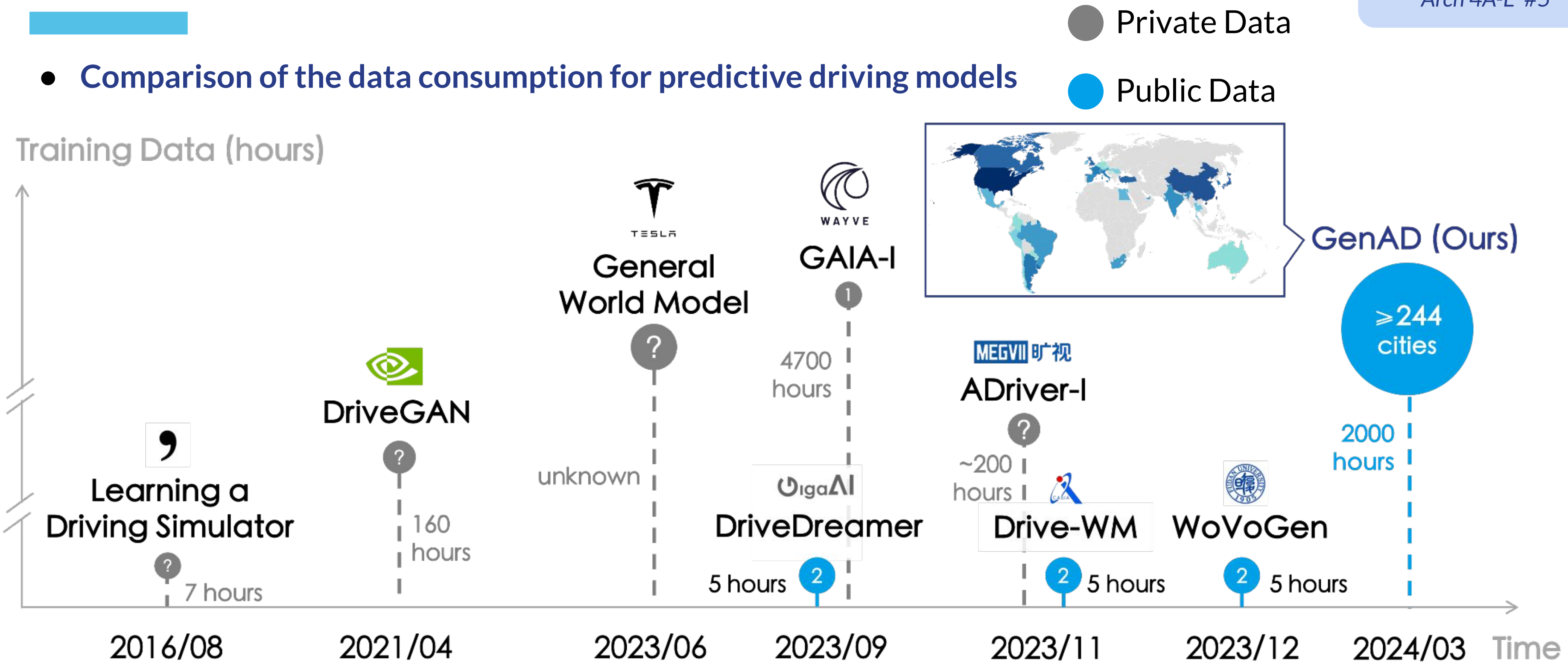


Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

GenAD | Dataset

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

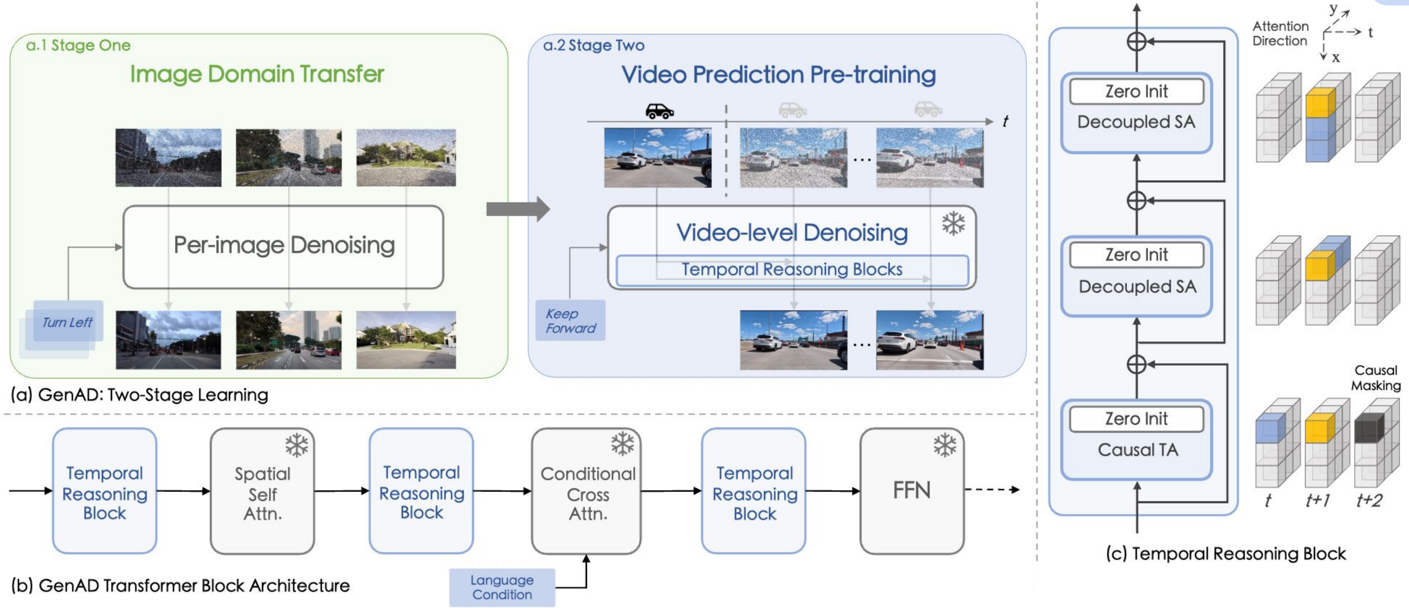
- **Comparison of the data consumption for predictive driving models**



Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

Algorithm | Video Prediction Model for Driving

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5



- **Two-stage Training**:
  - Tuning the **image generation model** (SDXL) into a highly-capable **video prediction model**
- **Model Specializations for Driving:**
  - Causal Temporal Attention: coherent and consistent future prediction
  - Decoupled Spatial Attention: efficient long-range modeling
  - Interleaved temporal blocks: sufficient spatiotemporal interaction

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

Result on Tasks (1/4) | Zero-shot Generalization (Video Prediction)

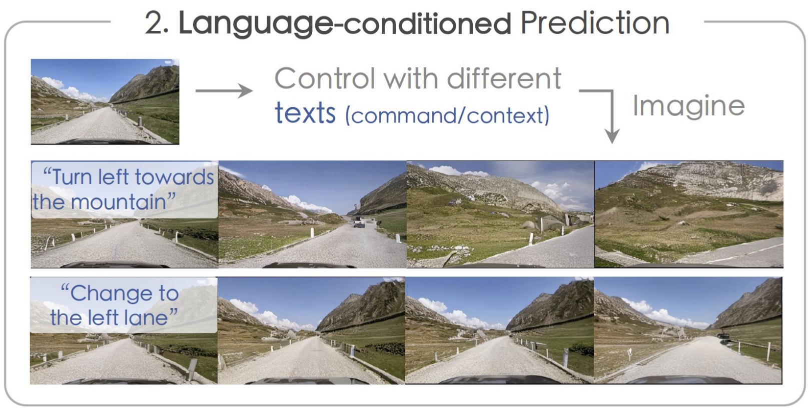
Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

- **Zero-shot video prediction** on unseen datasets including Waymo, KITTI and Cityscapes
- Outperforming competitive general video generation models

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

Result on Tasks (2/4) | Language-conditioned Prediction

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5



Controlling the future evolvement with **language**

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

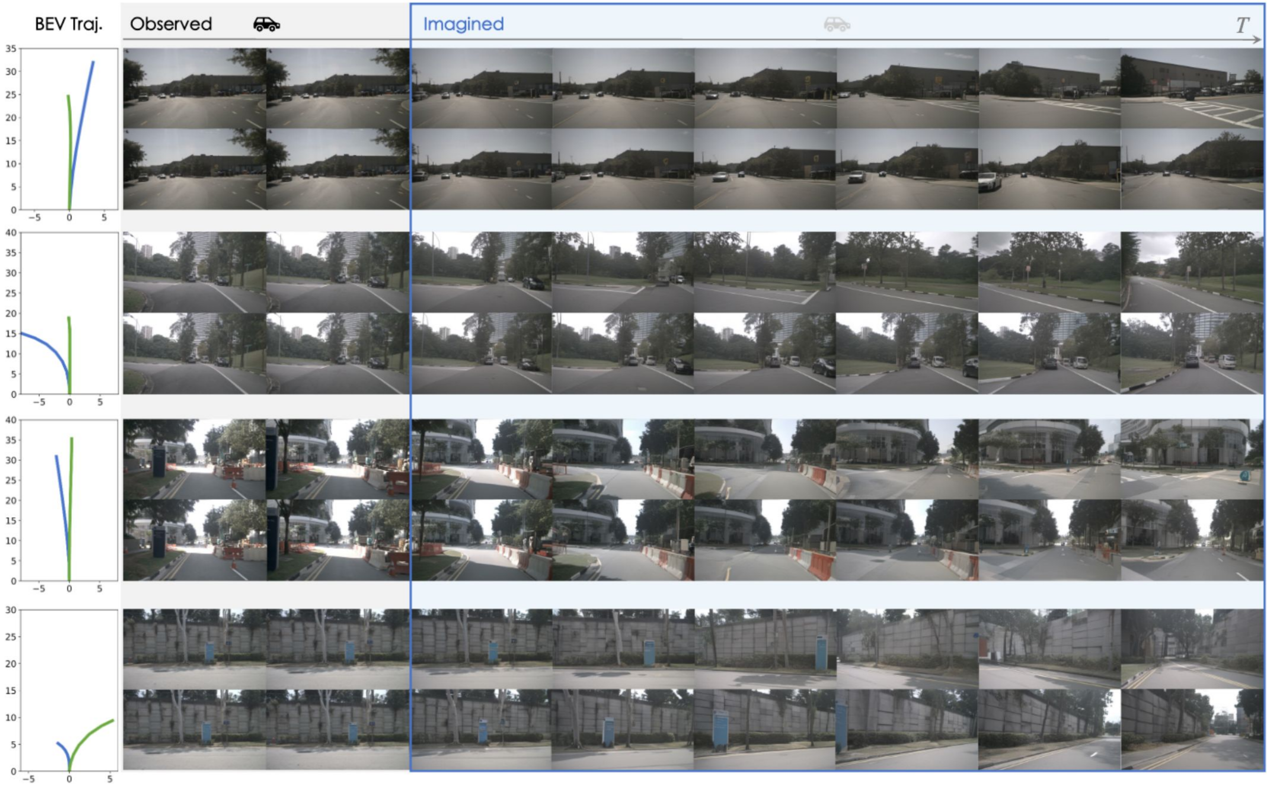
Result on Tasks (3/4) | Action-conditioned Prediction (Simulation)

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

Method	Condition	nuScenes Action Prediction Error (↓)
Ground truth	-	0.9
GenAD	text	2.54
GenAD-act	text + traj.	**2.02**

Table 4. **Task on Action-conditioned prediction**. Compared to GenAD with text conditions only, GenAD-act enables more precise future predictions that follow the action condition.

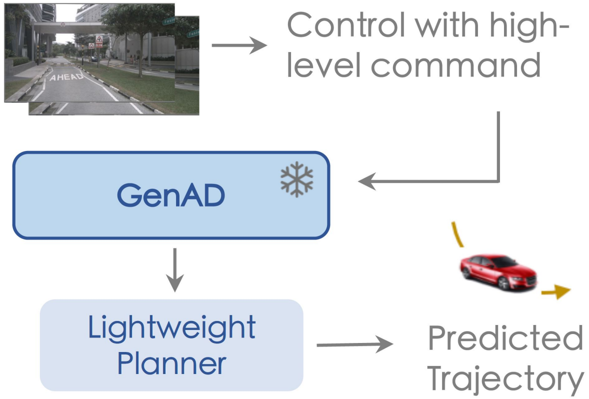
Simulating the future with
user-specified trajectory



Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

Result on Tasks (4/4) | Planning

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5



Method	# Trainable Params.	nuScenes ADE (↓)	nuScenes FDE (↓)
ST-P3* [20]	10.9M	2.11	2.90
UniAD* [22]	58.8M	1.03	1.65
GenAD (Ours)	0.8M	1.23	2.31

Table 5. **Task on Planning**. A lightweight MLP with *frozen* GenAD gets competitive planning results with 73× fewer trainable parameters and front-view image alone. *: multi-view inputs.

- Speeding up training by **3400 times** *(vs. UniAD)*
- Demonstrating the **effectiveness of** the learned spatiotemporal **representations**

Yang et al., GenAD: Generalized Predictive Model for Autonomous Driving, CVPR 2024

Summary

Poster Session
Thu, 5: 15- 6:45 p.m
Arch 4A-E #5

- **Largest Public Driving Dataset:**
  - **OpenDV-2K** provides ***2059 hours*** of ***worldwide*** driving videos.
- **Generalized Predictive Model for Autonomous Driving:**
  - **GenAD** can predict plausible futures with ***language*** conditions and generalize to ***unseen*** datasets in a ***zero-shot*** manner.
- **Broad Applications:**
  - **GenAD** can readily adapt to ***planning*** and ***simulation***.

OpenDriveLab

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

How to build a generally applicable driving world model?

Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability

Open Release

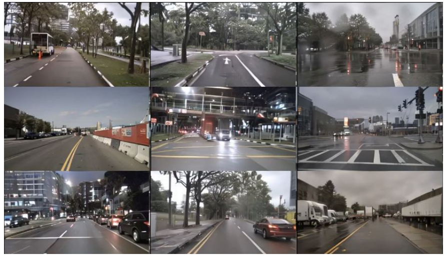
arxiv.2405.17398

Limitations of Existing Driving World Models

Open Release

- **Generalization:** limited data scale and geographical coverage

5h
within Singapore & Boston
nuScenes



- **Representation capacity:** low resolution and low frame rate

- **Control flexibility:** single modality, incompatible with planning algorithms

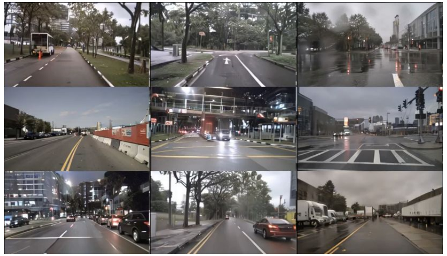
Gao et al., Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability

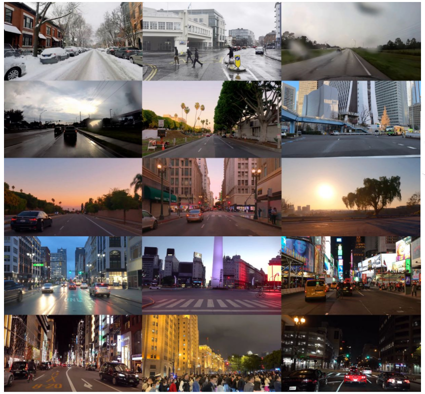
Our Investigation: A Generalizable Driving World Model

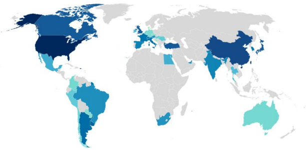
Open Release

- **Generalization:** largest driving video dataset

5h
within Singapore & Boston
nuScenes







1740h
worldwide

- **Representation capacity:** high spatiotemporal resolution

576×1024

- **Control flexibility:** multi-modal action inputs

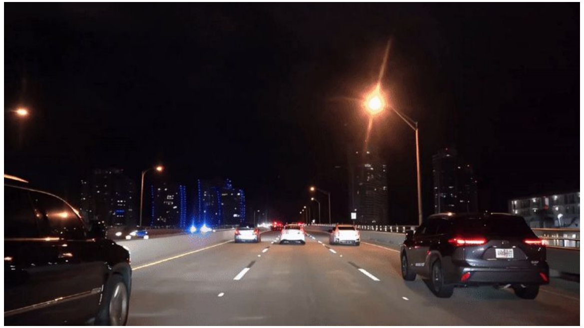
Gao et al., Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability

Capability of Vista

Open Release

- **High-fidelity future prediction**

- **Continuous long-horizon rollout (15 seconds)**



Gao et al., Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability

Capability of Vista

Open Release

- **Zero-shot action controllability**

turn left | go straight | turn right | stop

- **Provide reward without ground truth actions**

Summary

Open Release

- **Vista is a generalizable driving world model that can:**
  - *Predict high-fidelity futures in open-world scenarios.*
  - *Extend its predictions to continuous and long horizons.*
  - *Execute multi-modal actions (steering angles, speeds, commands, trajectories, goal points).*
  - *Provide rewards for different actions without accessing ground truth actions.*

OpenDriveLab

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

Part 3:
Challenges & Closing Remarks

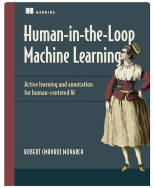
Data / Methodology / Compute / Goal

Challenges | End-to-end Autonomy

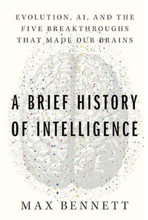
L4/L5, with driving comfort / experience considered
(Goals should be the same from two domains)

Task / Goal

Dimension	Research ("academia")	Engineering ("industry")
Data High quality. Large-scale	**High-quality / controllable Simulation** **Unlimited** - Neural rendering - 3DGS / AIGC (e.g. CVPR / Siggraph 2024)	**Scalable collection / Sanity check** - Data Flywheel **At least 10k of hours?** C.f. nuScenes 4.5h
Algorithm/Methodology Efficient and scalable	**Closed-loop Feedback / Long-horizon Planning** - World Model / - Video generation (e.g. Sora) / etc..	**Efficiency / Deployment** - Dual system (Sys1/Sys2) - Model compression / etc. - **Perception** …
Compute/Infra	~50-200 GPUs **Stable Training / fast I/O**	500+ GPUs **preferably 10k? / I've no idea**

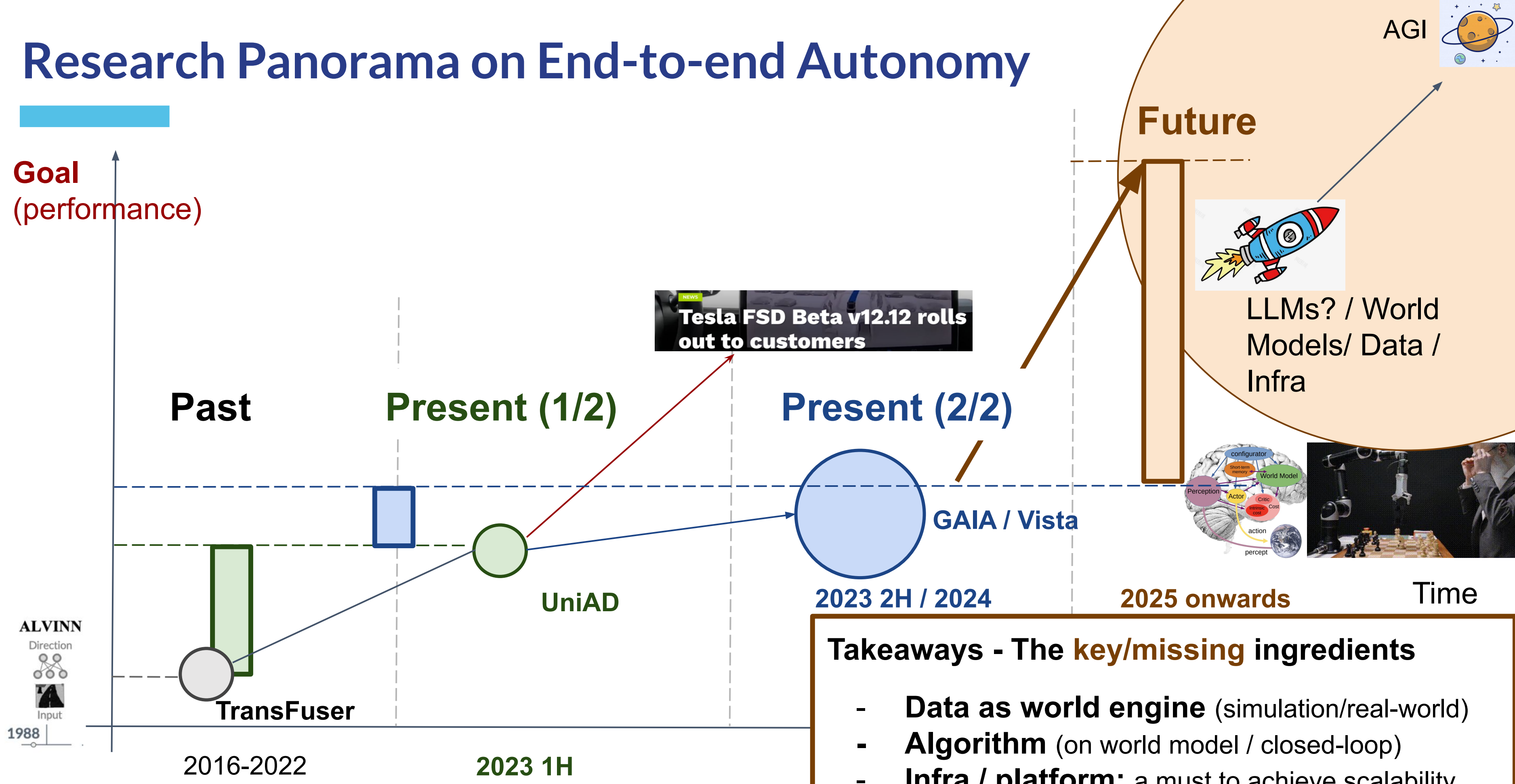


Gray area
Paradigm shift
(dichotomy?)



Details:
Chen et al. End-to-end Autonomous Driving: Challenges and Frontiers
https://arxiv.org/abs/2306.16927

Research Panorama on End-to-end Autonomy



Takeaways - The key/missing ingredients

- **Data as world engine** (simulation/real-world)
- **Algorithm** (on world model / closed-loop)
- **Infra / platform:** a must to achieve scalability
- **Deployment**: dual system / onboard-chip

Kudos to Our Fantastic Members / Collaborators

Also the slide credit

Meet our team in
Seattle @CVPR 2024!!!

Jiazhi Yang

GenAD

Shenyuan Gao

Vista

Li Chen

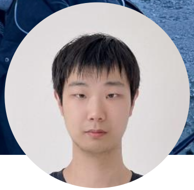
UniAD

Chonghao Sima

DriveLM

Huijie Wang

OpenLane



Zetong Yang

ViDAR

Yunsong Zhou

ELM

And many others remote…

Yihang Qiu

Tianyu Li

Kashyap Chitta

Jia Zeng

Andreas Geiger

End-of-Talk Questions?