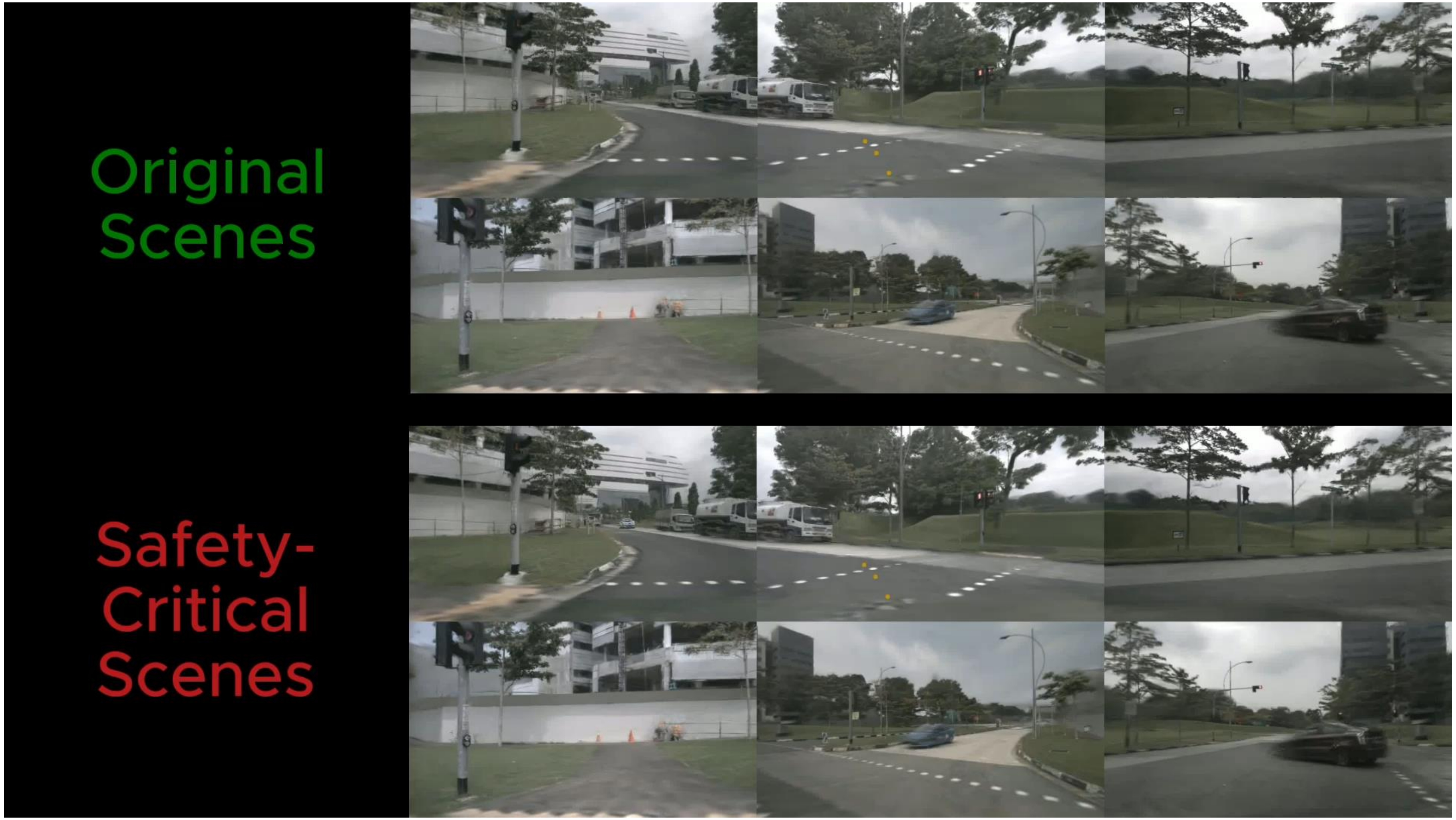
Model-Based Policy Adaptation for Closed-Loop End-to-End Autonomous Driving

Haohong Lin[1], Yunzhi Zhang[2], Wenhao Ding[3], Jiajun Wu[2], Ding Zhao[1]

[1]CMU, [2]Stanford, [3]NVIDIA

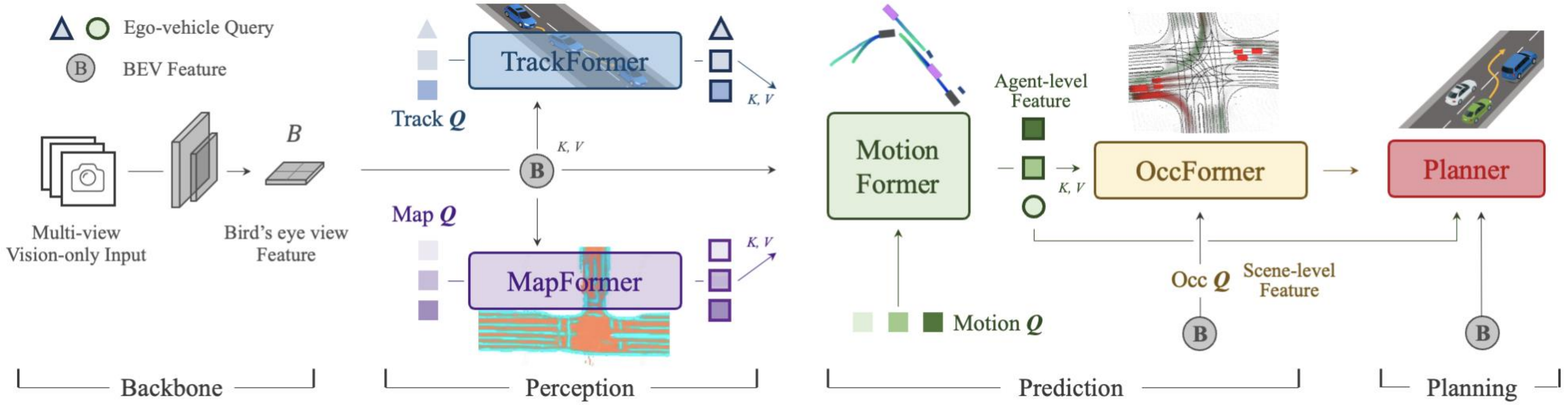
NeurIPS 2025

Teaser: Closed-Loop Evaluation of E2E Driving



Background: End-to-End Autonomous Driving

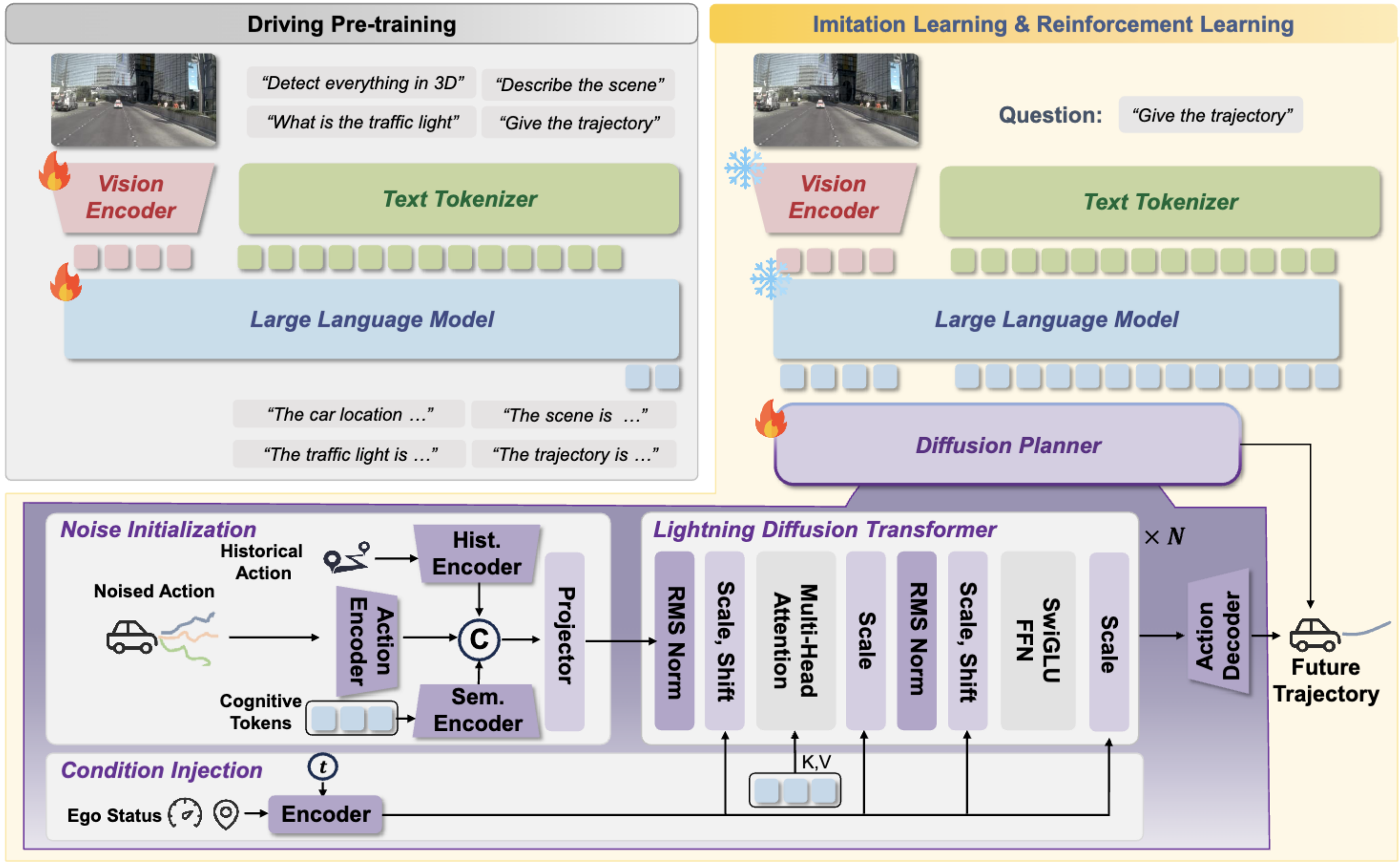
- Modularized, end-to-end joint training pipeline



Hu, Yihan, et al. "Planning-oriented autonomous driving." *CVPR 2023.*

Background: Foundation Models in Self-Driving

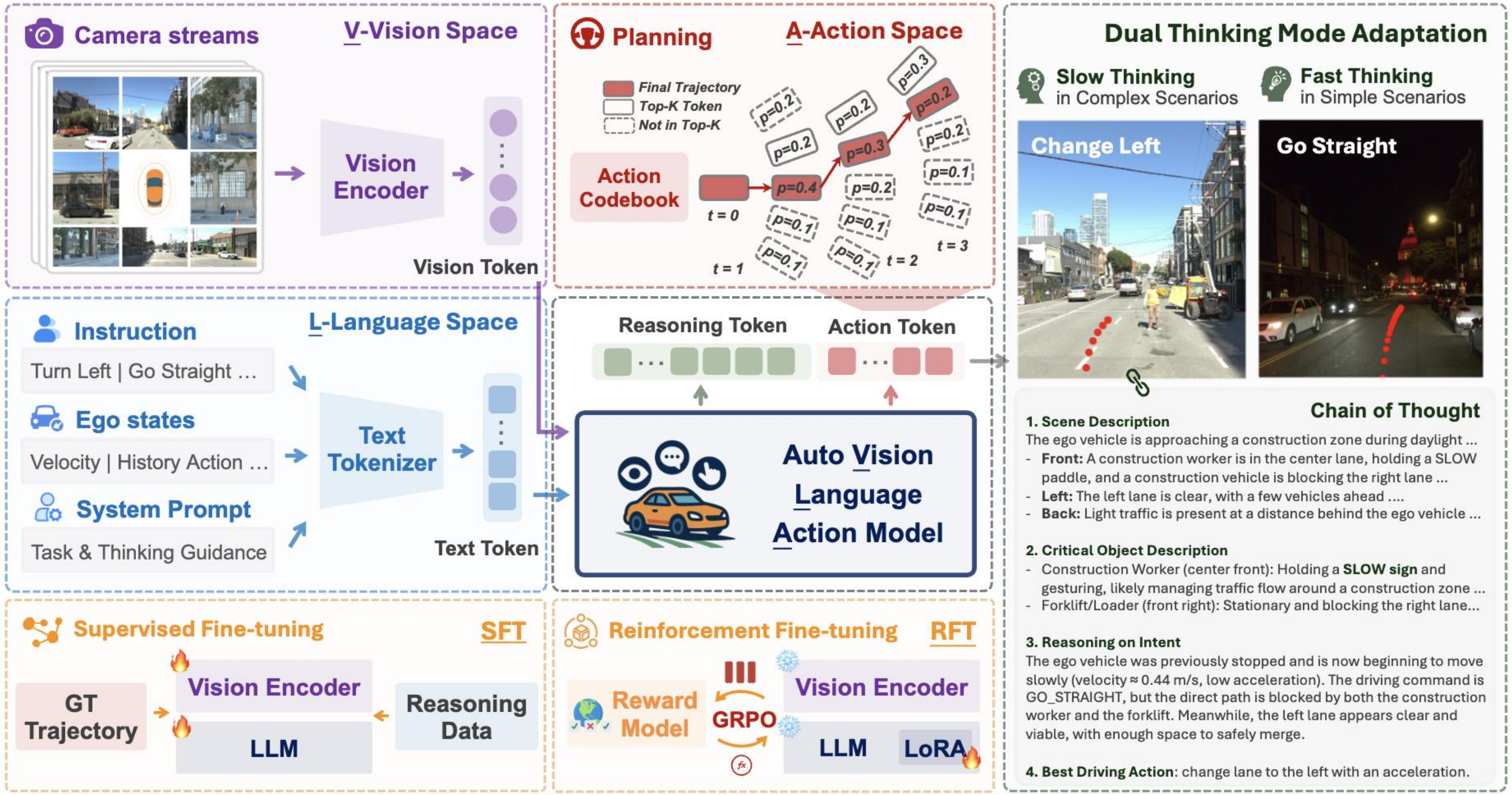
- CoT Data curation
- VLM backbone pretraining
- Diffusion head for trajectory decoding
- RL Finetuning



Li, Yongkang, et al. "ReCogDrive: A Reinforced Cognitive Framework for End-to-End Autonomous Driving." *ArXiv 2025*

Background: Foundation Models in Self-Driving

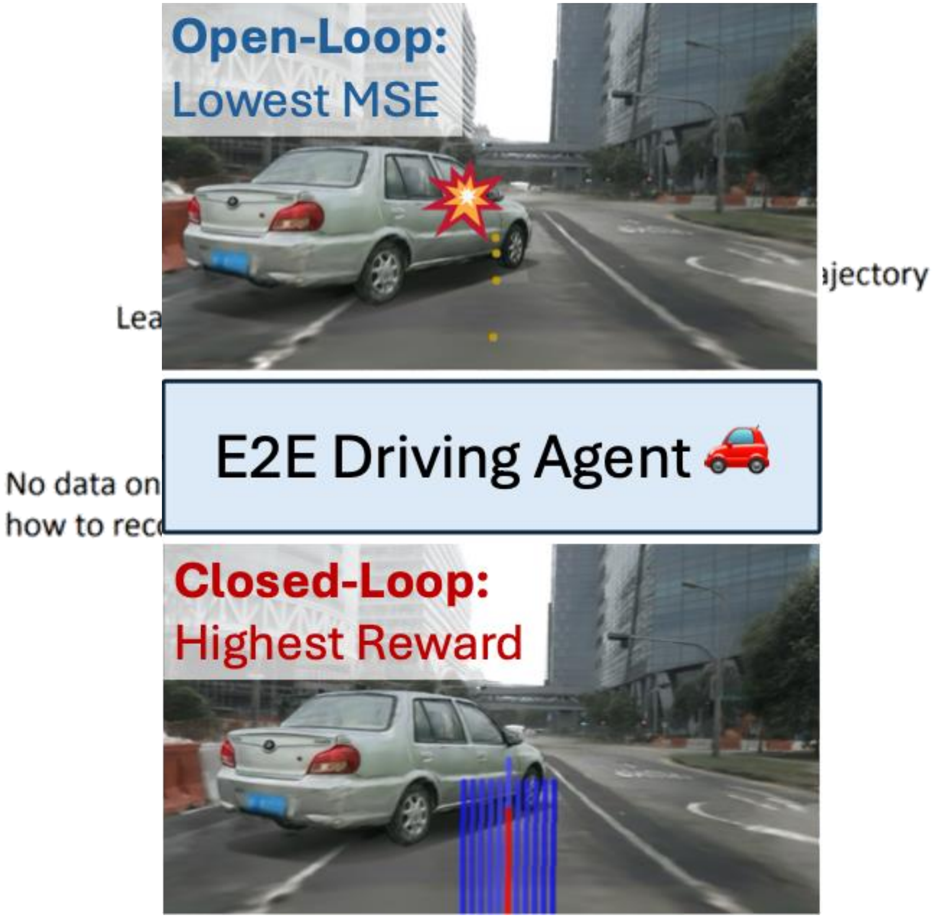
- CoT Data curation
- VLM backbone pretraining
- RL Finetuning



Zhou, Zewei, et al. "AutoVLA: A Vision-Language-Action Model for End-to-End Autonomous Driving with Adaptive Reasoning and Reinforcement Fine-Tuning." *ArXiv 2025*

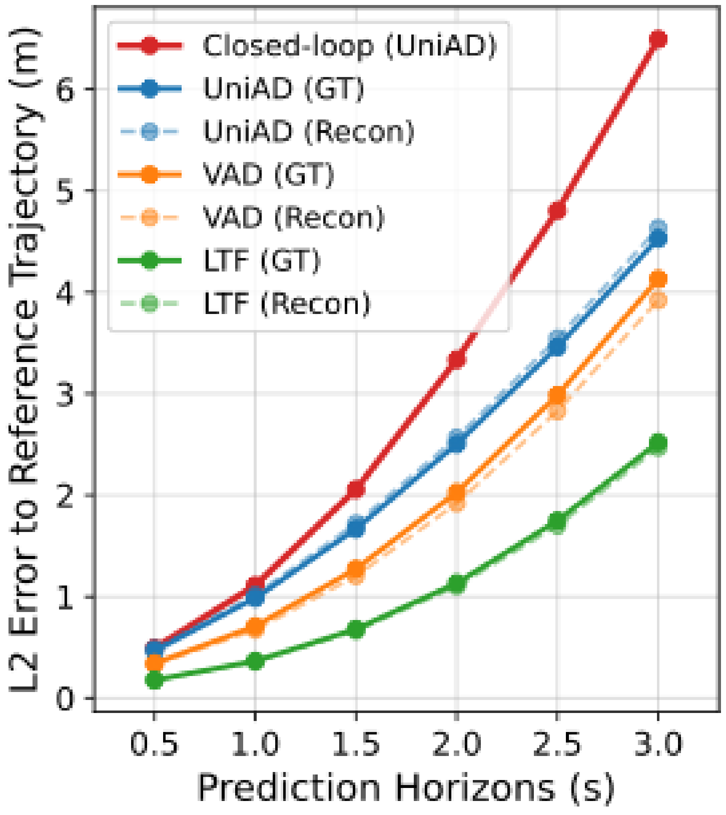
Motivation

- Missing Evaluation in Closed-Loop Rollout
  - Open-Loop Evaluation looks good
  - Compounding error leads to some failure mode
- Challenges in the Safety-Critical Behaviors
  - Imitation Learning: lower empirical risks (L2 error)
  - Online policy: reward maximization
  - **Objective Mismatch!**



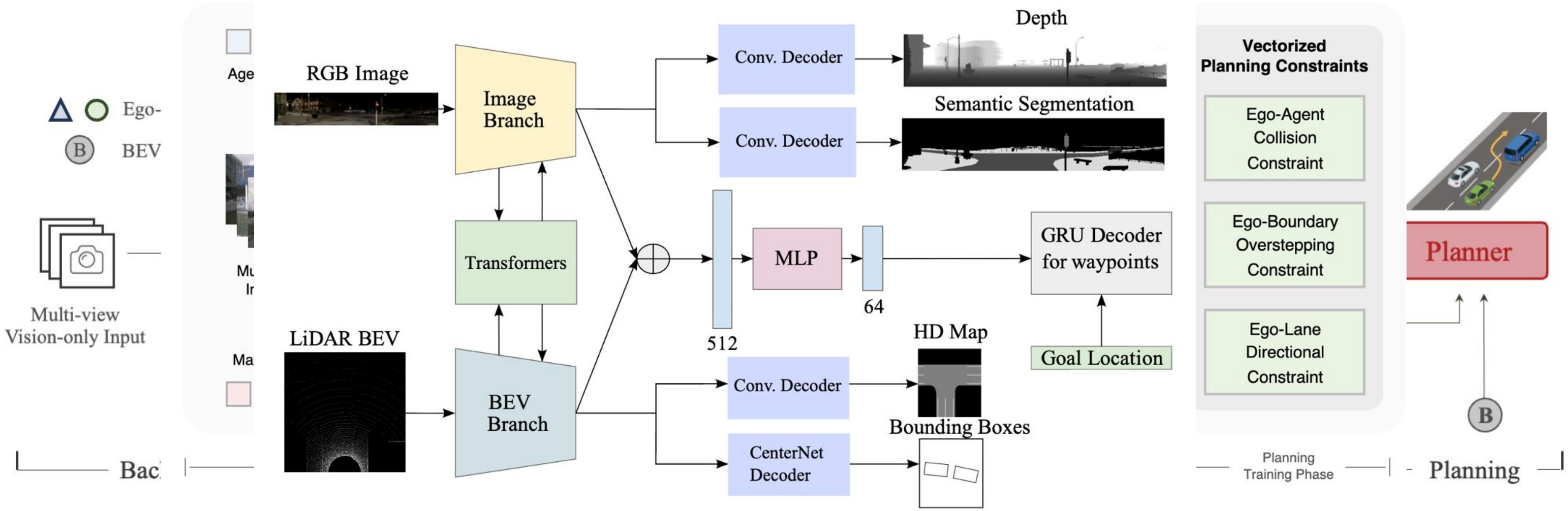
Motivation: Where does the Gap Come from?

- **Hypothesis of Error Sources:**
  - Sensor Sim. Error
  - Compounding Error
- **Preliminary experiments:** teacher-forcing rollout
- Observation:
  - Sensor simulation does not cause too much error...
  - **Compounding error** is more significant!



Related Works

- End-to-End Autonomous Driving

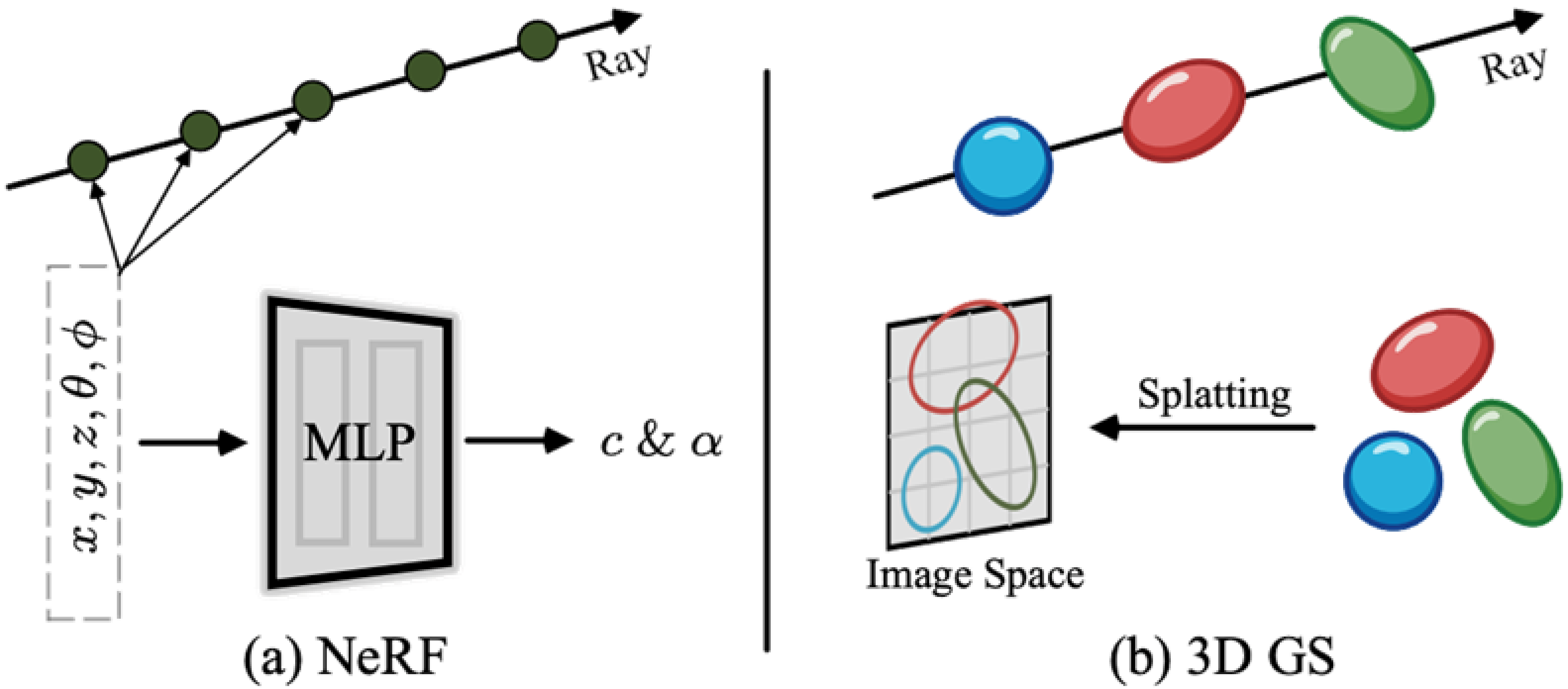


Jiang, Bo, et al. "Vad: Vectorized scene representation for efficient autonomous driving." *CVPR 2023*

Chitta, Kashyap, et al. "Transfuser: Imitation with transformer-based sensor fusion for autonomous driving." *TPAMI 2022*

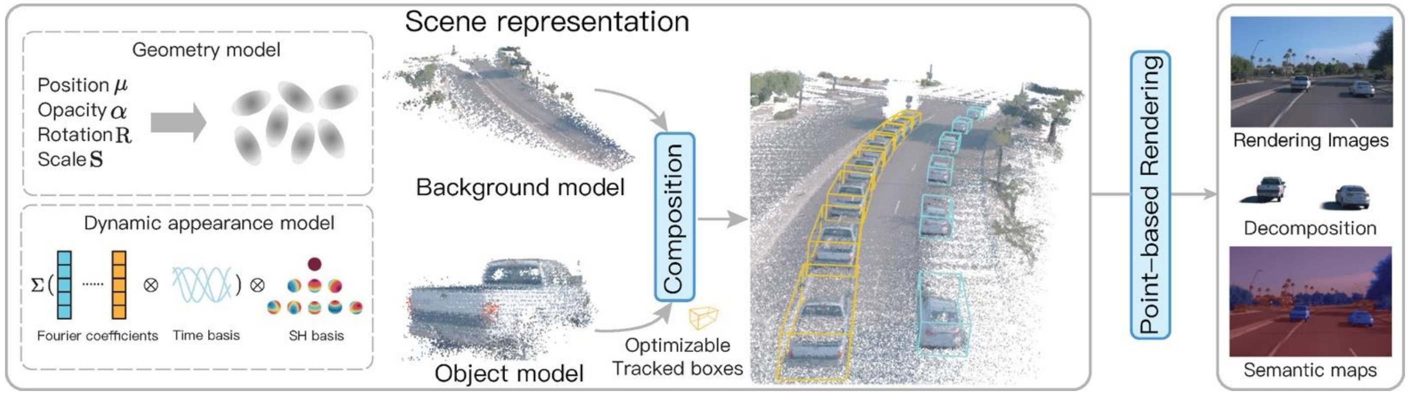
Preliminary: 3D Gaussian Splatting

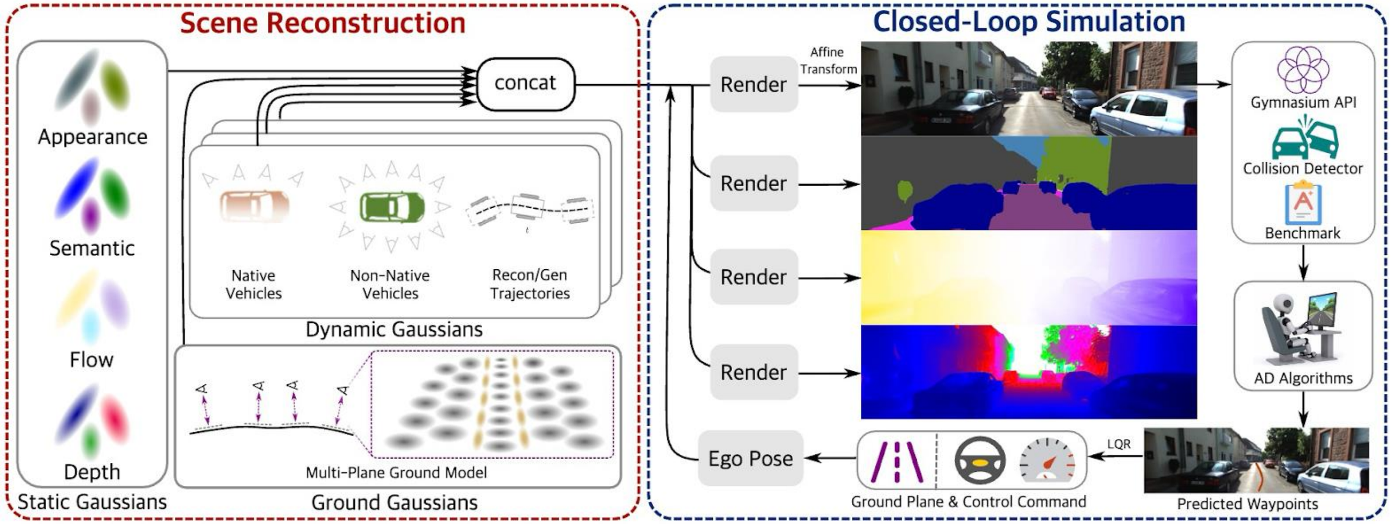
- **Pros:** 3DGS is faster at inference time compared to NeRF! Also it is parameter efficient
- **Cons:** (Both 3DGS and NeRF) cannot generalize to unseen scenes without any camera views



Preliminary: 3D Gaussian Splatting for Urban Scenes

- StreetGaussians
  - Take LiDAR Inputs
  - Decompose background / objects
- HUGSIM
  - Fine-grained structure: road surface, static, dynamic objects, etc.
  - No LiDAR dependencies

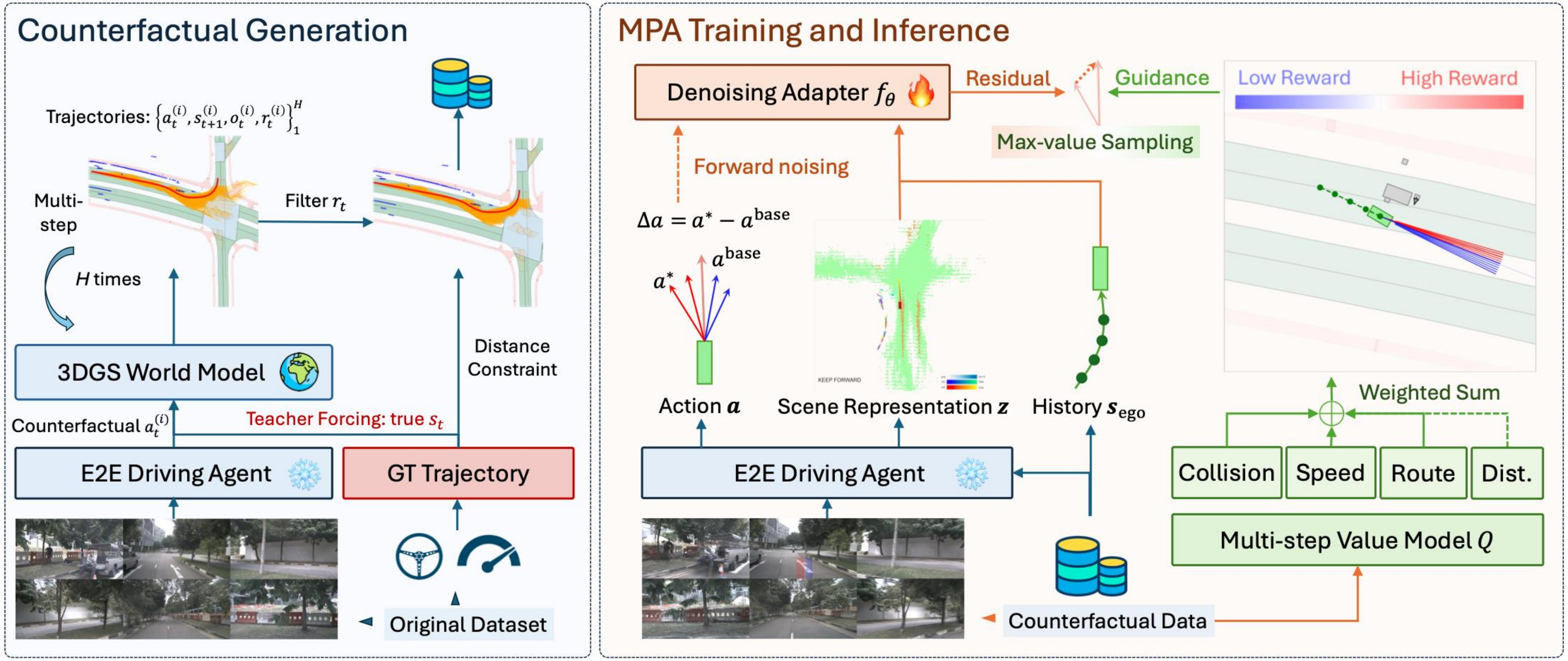




[1] Yan, Yunzhi, et al. "Street gaussians: Modeling dynamic urban scenes with gaussian splatting." *ECCV 2024*.
[2] Zhou, Hongyu, et al. "Hugsim: A real-time, photo-realistic and closed-loop simulator for autonomous driving." *arXiv 2024.*

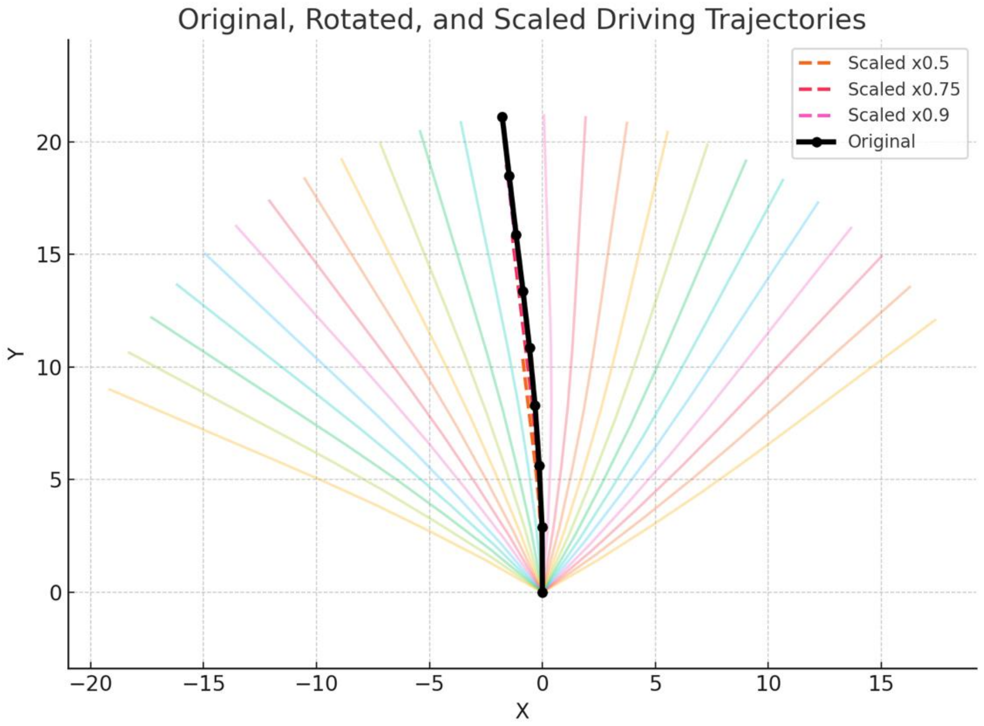
Overview of the Proposed Method

- Counterfactual Data Generation | Policy Adapter | Value Function



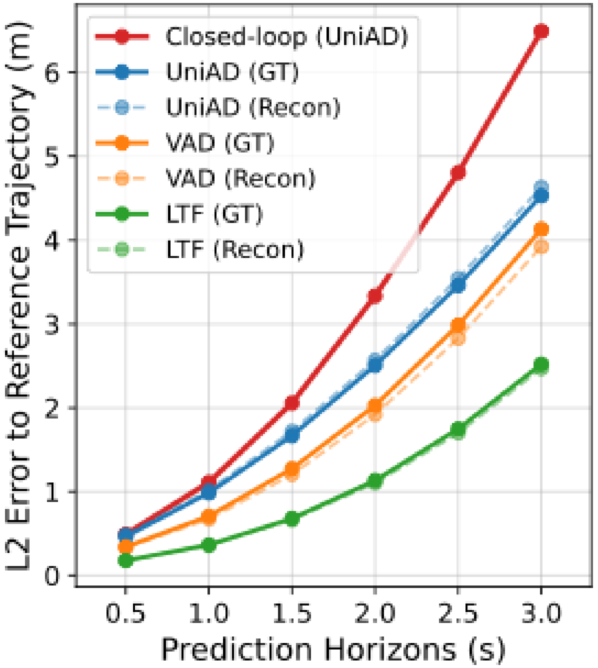
Counterfactual Data Synthesis

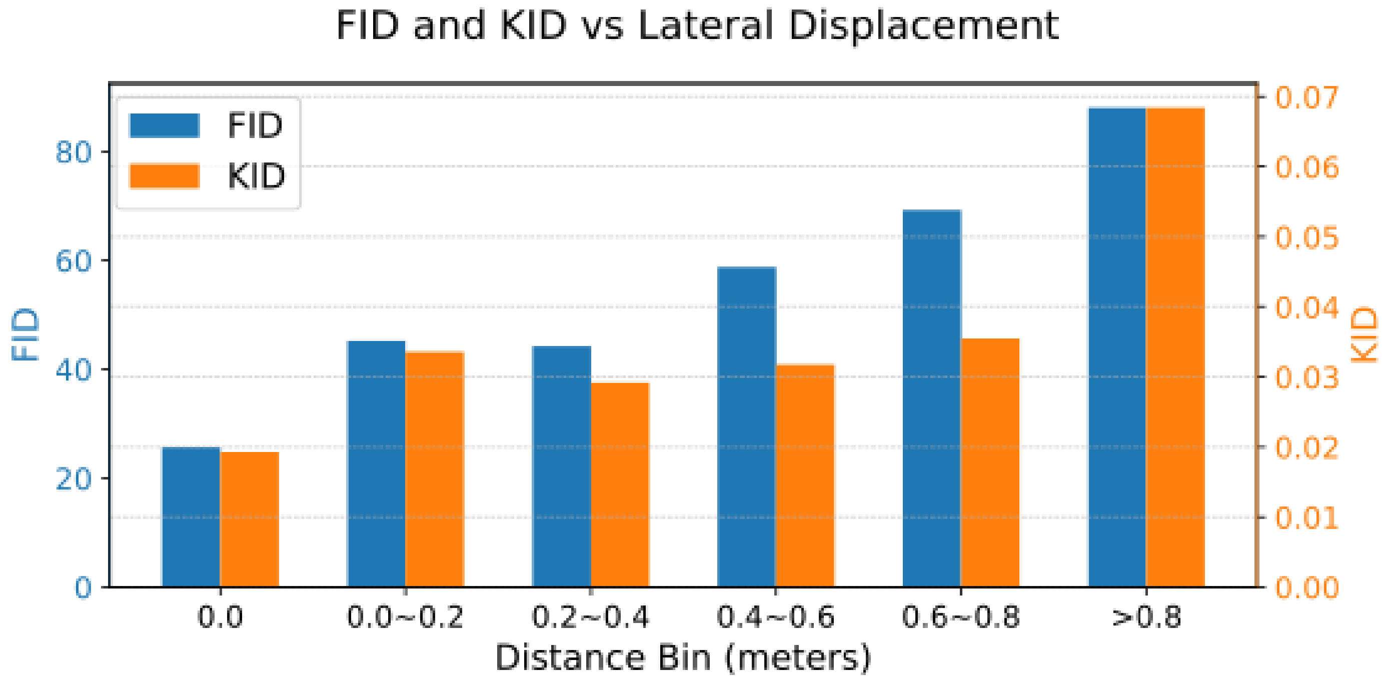
- Randomly transform the predicted trajectories with:
  - Warping
  - Rotation
  - Noising
- Rollout and accumulate the feedback



Properties of 3DGS?

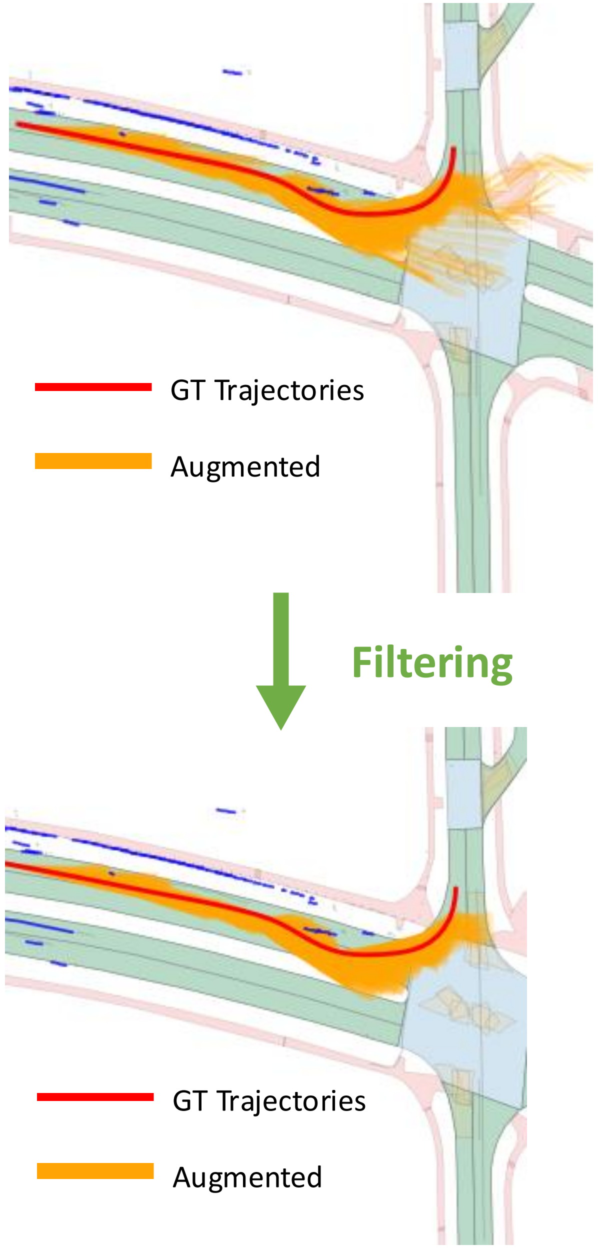
- What is the quality of the rendered results?
  - a) Impact to the policy; b) Inception Distance w.r.t. Lateral Displacement





Counterfactual Data Synthesis

- How to guarantee the realism?
- Constraining the distance
  - Between the current poses with the demonstration data
  - If exceeds, reject these samples



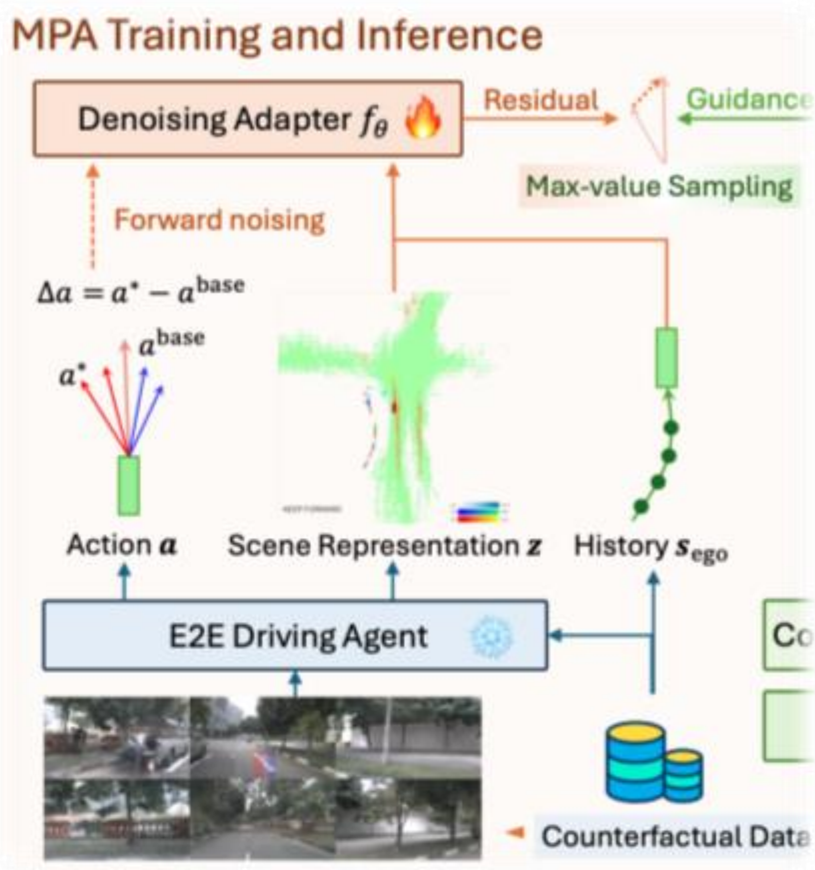
DDIM Sampler for Policy Adaptation

- Process of DDIM Sampler

$$\mathbb{E}_{\Delta a^{(0)},k,\epsilon} \min_i \left\| f_\theta(\Delta a^{(k)}, k, z, \boldsymbol{s}_{\text{ego}}, a^{\text{base}})[i] - \Delta a^{(0)} \right\|_2^2,$$

where $\Delta a^{(k)} = \sqrt{\bar{\alpha}_k}\Delta a^{(0)} + \sqrt{1-\bar{\alpha}_k}\epsilon$, with $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

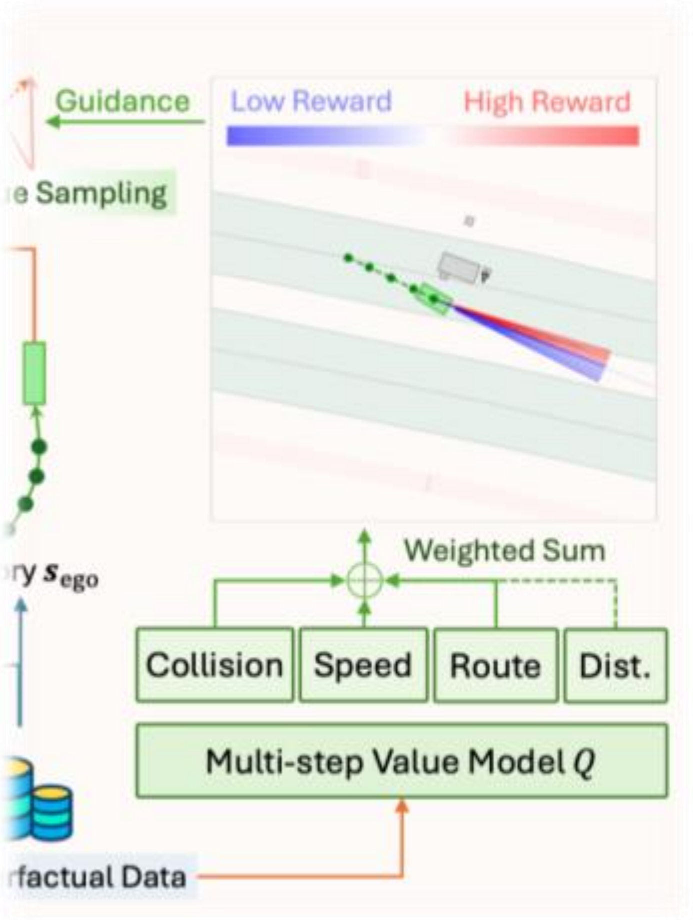
$$a^{\text{adapt}}[i] = a^{\text{base}} + \Delta a^{(0)}[i], \ \forall i \in [N].$$



Multi-Principled Q-Value Heads

- Reward Shaping
  - (Longitude) Route progression
  - (Lateral) Driveable area compliance
  - (Safety) Collision penalty
  - (Safety) Off-road penalty
  - (Comfort) Overspeed penalty

- Mult-Head Truncated Q Value
  - $\min_{\theta} \mathrm{E}[|Q_\theta(s_t, a_t) - \sum_k r(s_{t+k}, a_{t+k})|^2]$



Experiment Results

- Setting: nuScenes + HUGSIM closed-loop evaluation
  - Training on ~290 normal scenes with counterfactual generation (Singapore)
  - Testing on 70 unseen, normal scenes (Boston)
  - Testing on 10 seen, safety-critical scenes (Singapore)

Singapore (Normal)

Boston (Normal)

Experiment Results

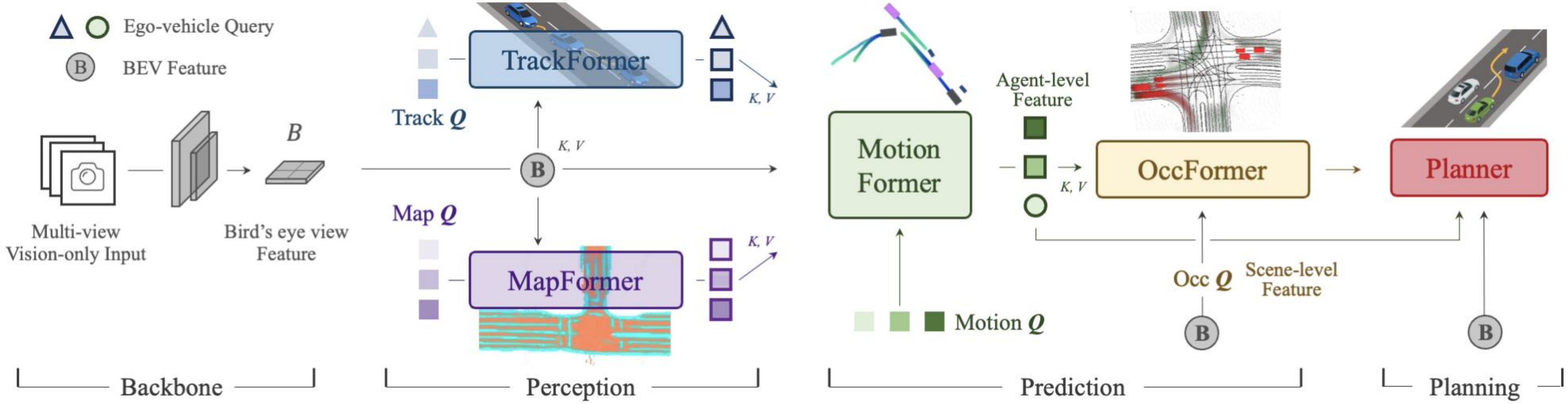
- Metrics [1, 2]
  - Route Completion (RC)
  - Collision related: Non-Collision (NC), Time-to-Collision (TTC)
  - Driving style: Comfort (COM), Driveable Area Compliance (DAC)
  - HDScore

$$\mathrm{HDScore} = \mathrm{RC} \times \frac{1}{T}\sum_{t=0}^{T}\left\{\prod_{m\in\{\mathrm{NC},\,\mathrm{DAC}\}} \mathrm{score}_m \times \frac{\sum_{m\in\{\mathrm{TTC},\,\mathrm{COM}\}} \mathrm{weight}_m \times \mathrm{score}_m}{\sum_{m\in\{\mathrm{TTC},\mathrm{COM}\}} \mathrm{weight}_m}\right\}_t.$$

[1] Zhou, Hongyu, et al. "Hugsim: A real-time, photo-realistic and closed-loop simulator for autonomous driving." *arXiv 2024.*
[2] Dauner, Daniel, et al. "Navsim: Data-driven non-reactive autonomous vehicle simulation and benchmarking." NeurIPS 2024

Experiment Results

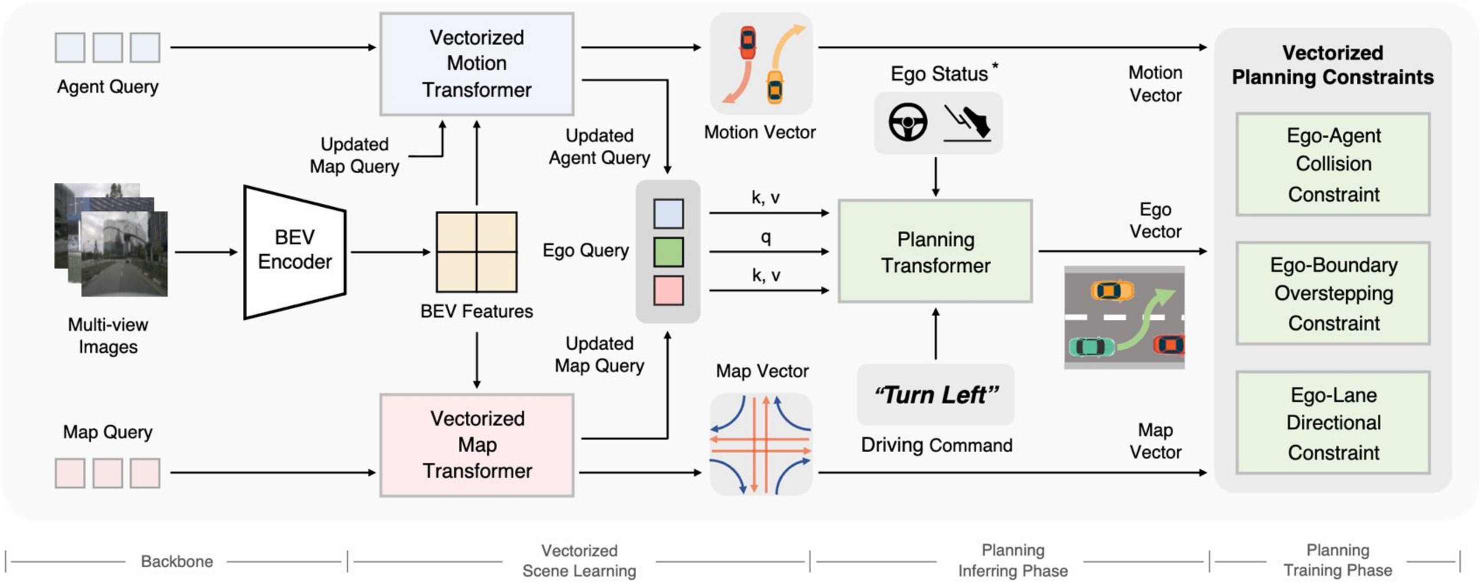
- Baselines in Comparison
  - **Pretrained Policies:** UniAD / VAD / LTF
  - **Trained w/ Counterfactual data:** AD-MLP / BC-Safe / Diffusion



Hu, Yihan, et al. "Planning-oriented autonomous driving." *CVPR 2023.*

Experiment Results

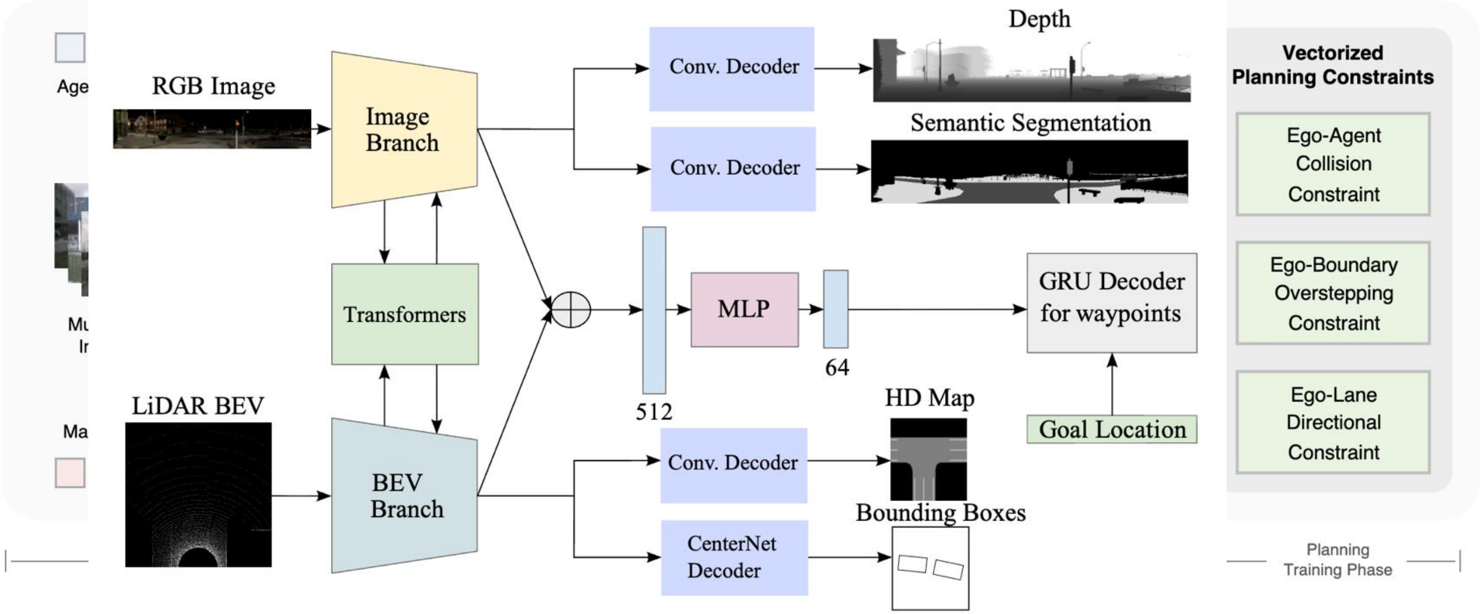
- Baselines in Comparison
  - **Pretrained Policies:** UniAD / VAD / LTF
  - **Trained w/ Counterfactual data:** AD-MLP / BC-Safe / Diffusion



Jiang, Bo, et al. "Vad: Vectorized scene representation for efficient autonomous driving." *CVPR 2023*

Experiment Results

- Baselines in Comparison
  - **Pretrained Policies:** UniAD / VAD / LTF
  - **Trained w/ Counterfactual data:** AD-MLP / BC-Safe / Diffusion



Chitta, Kashyap, et al. "Transfuser: Imitation with transformer-based sensor fusion for autonomous driving." *TPAMI 2022*

Main Results

- Better closed-loop driving performance for in-domain scenarios!

Model	Ego Status	Camera	Curation	RC	NC	DAC	TTC	COM	HDScore
UniAD	✓	✓	✗	39.4	56.9	75.1	52.1	**98.7**	19.4
VAD	✓	✓	✗	50.1	68.4	87.2	66.1	90.2	31.9
LTF	✓	✓	✗	65.2	71.3	92.1	67.6	98.4	46.7
AD-MLP	✓	✗	✓	13.4	**80.2**	86.2	**79.4**	90.1	6.5
BC-Safe	✓	✓	✓	57.0	59.8	87.9	55.2	89.4	33.6
Diffusion	✓	✓	✓	71.8	67.4	88.1	64.5	91.5	45.1
MPA (UniAD)	✓	✓	✓	93.6	76.4	92.8	72.8	91.8	66.4
MPA (VAD)	✓	✓	✓	**94.9**	75.4	**93.6**	72.5	92.8	**67.0**
MPA (LTF)	✓	✓	✓	93.1	70.8	90.9	67.9	94.9	60.0

Main Results

- Better performance in the OOD Scenarios

	Unseen Nominal Scenes						Safety-Critical Scenes					
Model	**RC**	**NC**	**DAC**	**TTC**	**COM**	**HDScore**	**RC**	**NC**	**DAC**	**TTC**	**COM**	**HDScore**
UniAD	39.3	56.6	74.0	52.6	**98.2**	22.2	11.4	76.2	82.1	57.8	95.9	4.5
VAD	45.4	64.8	86.2	62.0	95.9	29.3	25.4	77.0	88.3	73.2	88.4	16.0
LTF	63.3	64.8	86.5	62.8	**98.2**	41.9	35.1	80.9	96.8	<u>78.1</u>	**100.0**	24.2
AD-MLP	7.6	**71.6**	82.2	**69.8**	92.3	3.3	4.9	**93.5**	96.2	**93.4**	85.9	4.3
BC-Safe	59.2	59.8	81.2	56.3	95.9	34.6	20.2	80.1	91.7	67.3	86.7	13.5
Diffusion	57.9	62.1	83.5	58.3	96.2	35.1	20.9	<u>84.3</u>	92.3	72.4	86.3	13.1
MPA (UniAD)	**93.7**	69.5	<u>92.9</u>	66.6	97.6	<u>60.9</u>	<u>95.1</u>	76.8	<u>98.9</u>	74.2	97.7	<u>70.4</u>
MPA (VAD)	90.9	<u>71.0</u>	**94.4**	<u>68.8</u>	97.7	**61.2**	**96.6**	79.8	**99.0**	77.3	97.7	**74.7**
MPA (LTF)	<u>91.8</u>	68.3	91.0	66.5	96.9	57.0	87.3	72.0	94.0	66.9	<u>97.8</u>	56.3

Ablation Studies

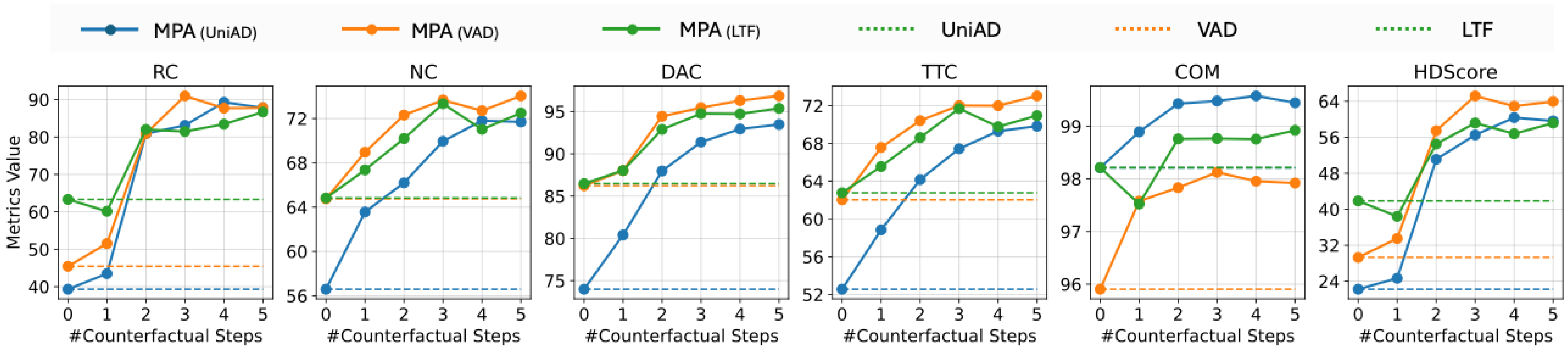
- Quantitative Studies of the Value Heads:

ID	Q_{route}	Q_{dist}	$Q_{\text{collision}}$	Q_{speed}	Adapter	RC	NC	DAC	TTC	COM	HDScore
1		✓	✓	✓		6.9	**81.2**	95.1	81.0	**100**	5.1
2	✓		✓	✓		83.9	57.0	81.0	53.6	99.4	43.2
3	✓	✓		✓		89.2	70.8	**95.6**	68.6	99.4	60.8
4	✓	✓	✓			90.4	68.9	91.8	65.4	99.4	56.6
5	✓	✓	✓	✓		91.1	71.5	94.1	69.4	99.4	**60.9**
6	✓	✓	✓	✓	✓	**93.7**	69.5	92.9	66.6	97.6	**60.9**

ID	Q_{route}	Q_{dist}	$Q_{\text{collision}}$	Q_{speed}	Adapter	RC	NC	DAC	TTC	COM	HDScore
1		✓	✓	✓		4.6	**86.0**	98.3	**79.3**	90.1	3.6
2	✓		✓	✓		65.1	65.6	85.7	53.8	86.5	39.5
3	✓	✓		✓		57.7	82.4	**99.0**	69.6	84.6	39.2
4	✓	✓	✓			79.3	82.9	98.5	68.0	93.9	50.1
5	✓	✓	✓	✓		75.6	81.2	98.8	78.6	**99.7**	55.3
6	✓	✓	✓	✓	✓	**95.1**	76.8	98.9	74.2	97.7	**70.4**

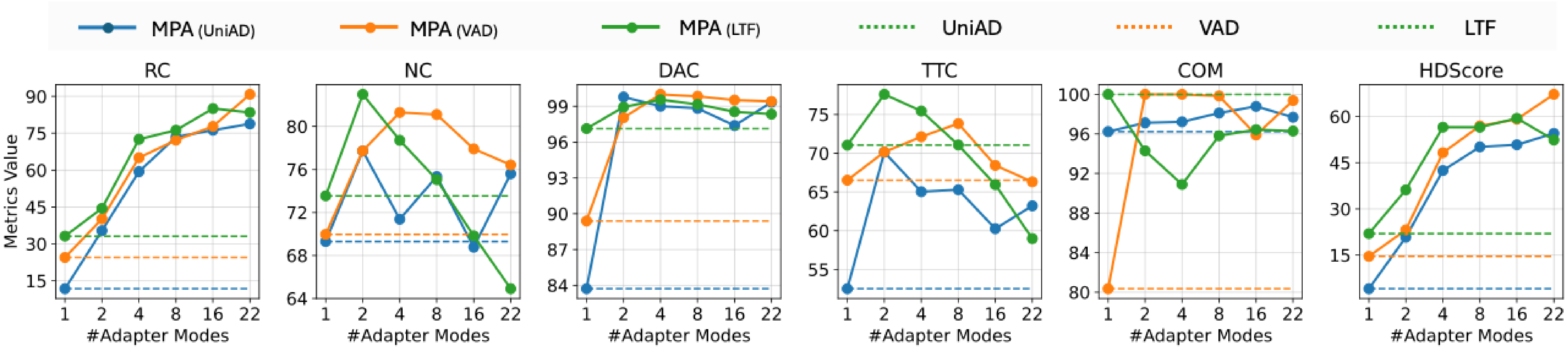
Ablation Studies --- Counterfactual Rollout Steps

- Impact of Counterfactual Rollout Steps
  - Longer rollout steps give better future awareness in planning



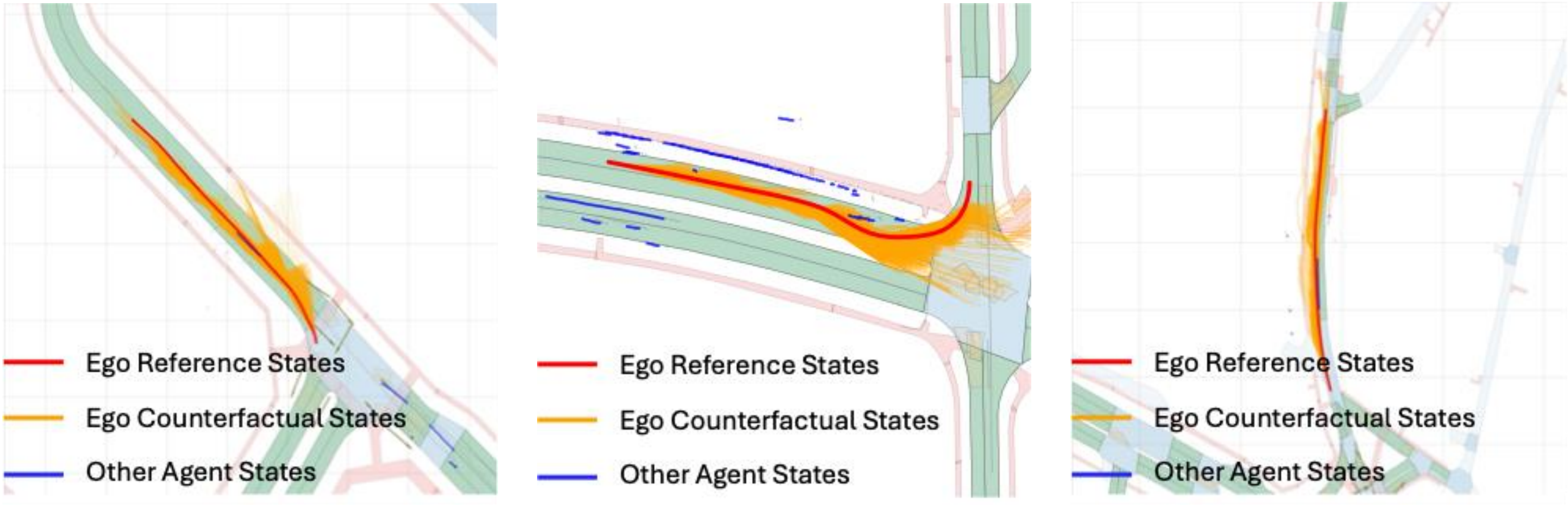
Ablation Studies --- Capacity of Policy Adapter

- Impact of #adapter modes of the diffusion head
  - More modes bring better diversity in action proposal!



Qualitative Results on the Counterfactual Dataset

- **Counterfactual dataset** has better coverage in driving behavior!



Qualitative Results

Qualitative Results (Safety-Critical Scenes)

Takeaways

- Mitigate the performance gap between open- and closed-loop evaluation:
  - Counterfactual Data Generation
  - Diffusion policy adapter for diverse action proposal
  - Inference-time Scaling to search the optimal actions
- Next step:
  - Synthesize safety-critical scenes at scale
  - Sim-to-real / real-to-sim gap?

NEURAL INFORMATION PROCESSING SYSTEMS

Thanks for Listening!

Project Page

Haohong Lin | CMU

Website: https://hhlin.info

Email: haohongl@andrew.cmu.edu

Poster session: Wed, Dec 3rd, 4:30pm-7:30pm (PST)