



TrafficBots: Towards World Models for Autonomous Driving Simulation and Motion Prediction

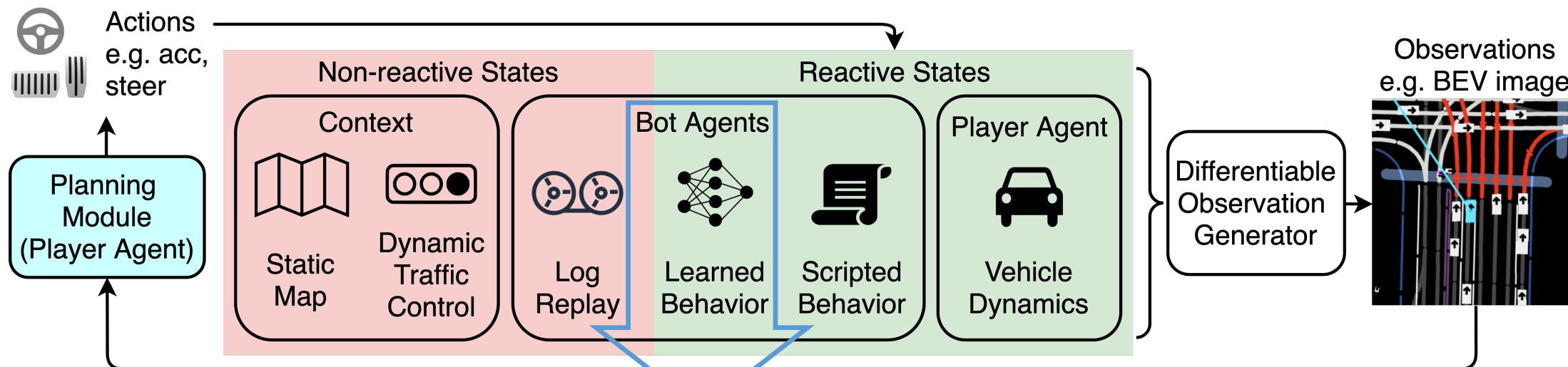
Zhejun Zhang¹, Alexander Liniger¹, Dengxin Dai^{1,2},
Fisher Yu¹, Luc Van Gool¹

¹Computer Vision Lab, ETH Zurich. ²MPI for Informatics.

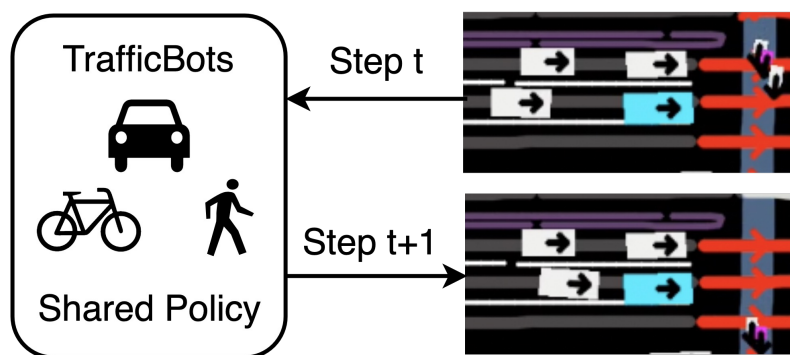
³PSI, KU Leuven



World Model for Autonomous Driving Planning

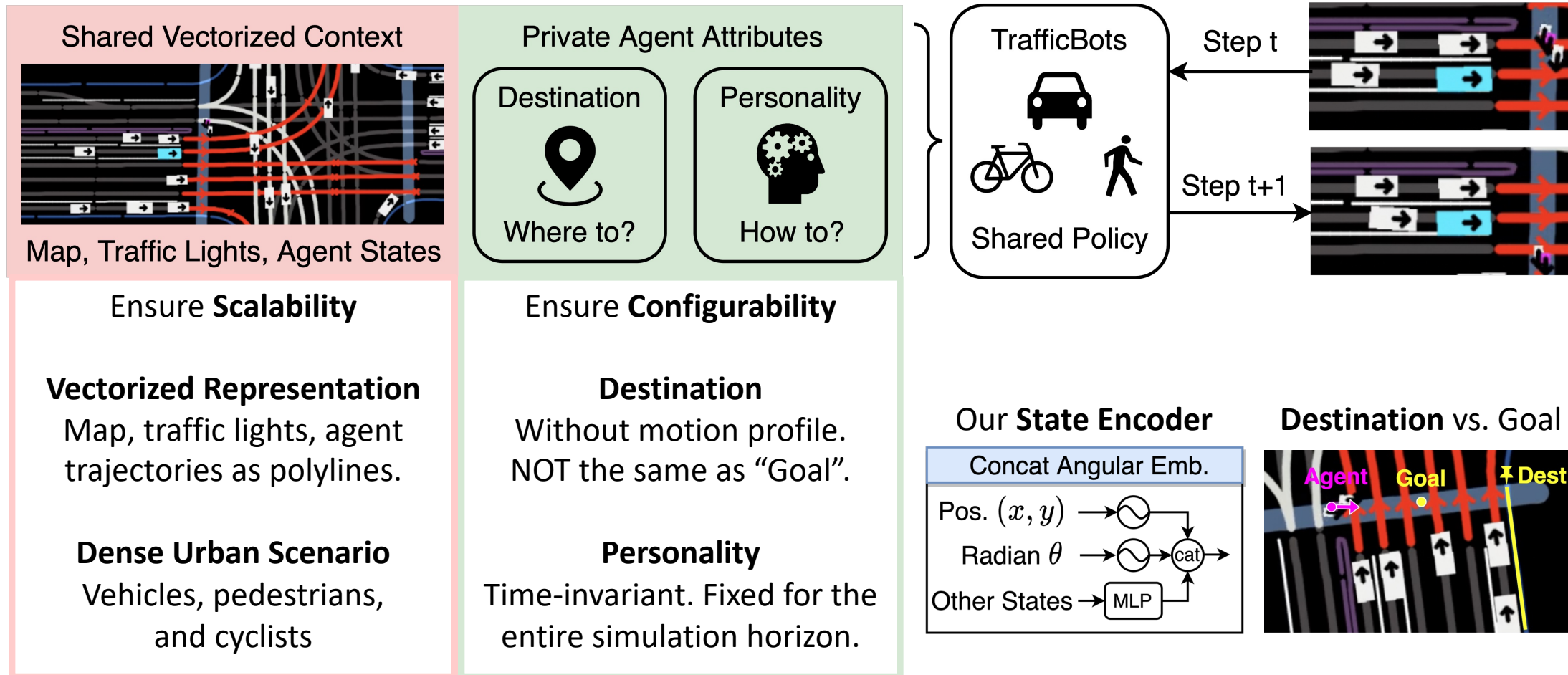


- Goal: Create a **data-driven differentiable** simulation for
1. Training and evaluating AD planning algorithms.
 2. Enhancing the behavior-fidelity of full-stack simulators.

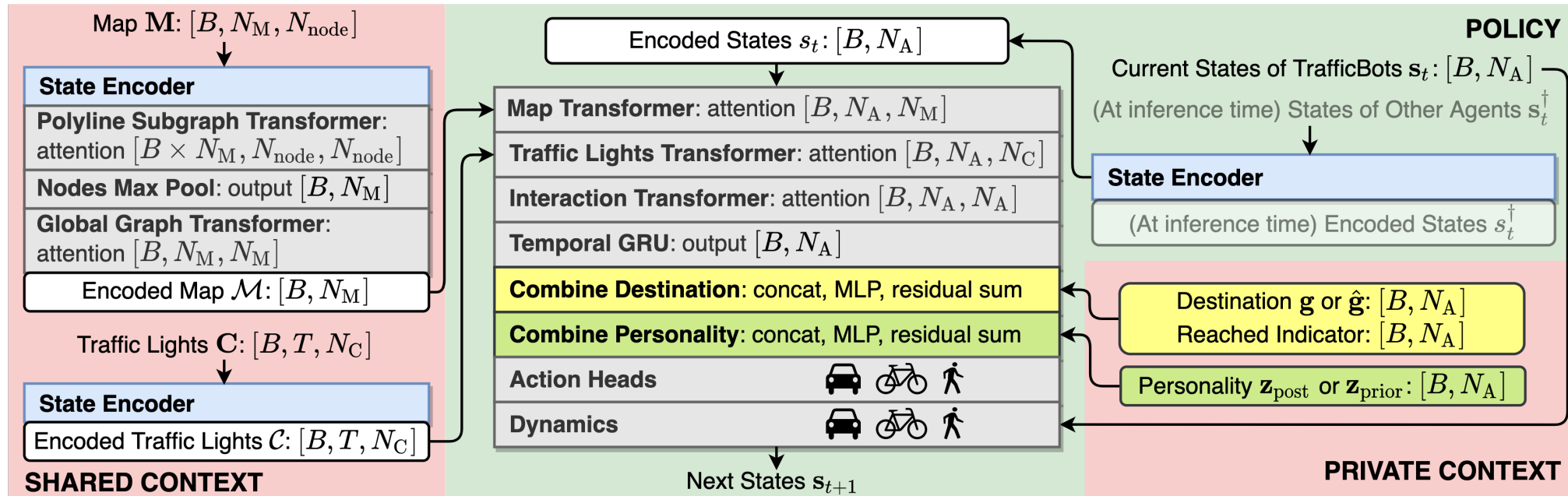


Generate **human-like** behavior for bot agents.

TrafficBots: A Multi-Agent Auto-Regressive Policy



TrafficBots: Network Architecture



- **Transformer** with self-attention and cross-attention.
- **State Encoder** using positional and angular embedding.
- **Personality** represented by the latent of conditional VAE.
- **Destination** predicted via multi-class classification over polylines.
- **Dynamics** simplified to unicycle model.

Performance

Table 1: Performance on the Waymo (marginal) motion prediction leaderboard

WOMD ^[1] <i>test</i>	soft mAP \uparrow	mAP \uparrow	minADE \downarrow	minFDE \downarrow	miss rate \downarrow	overlap rate \downarrow
SceneTransformer ^[2]	N/A	0.279	0.612	1.212	0.156	0.147
Waymo LSTM ^[1]	0.182	0.176	1.007	2.355	0.375	0.190
TrafficBots (a priori)	0.219	0.212	1.313	3.102	0.344	0.145

Table 2: Ablation. All models are trained for 48 hours.

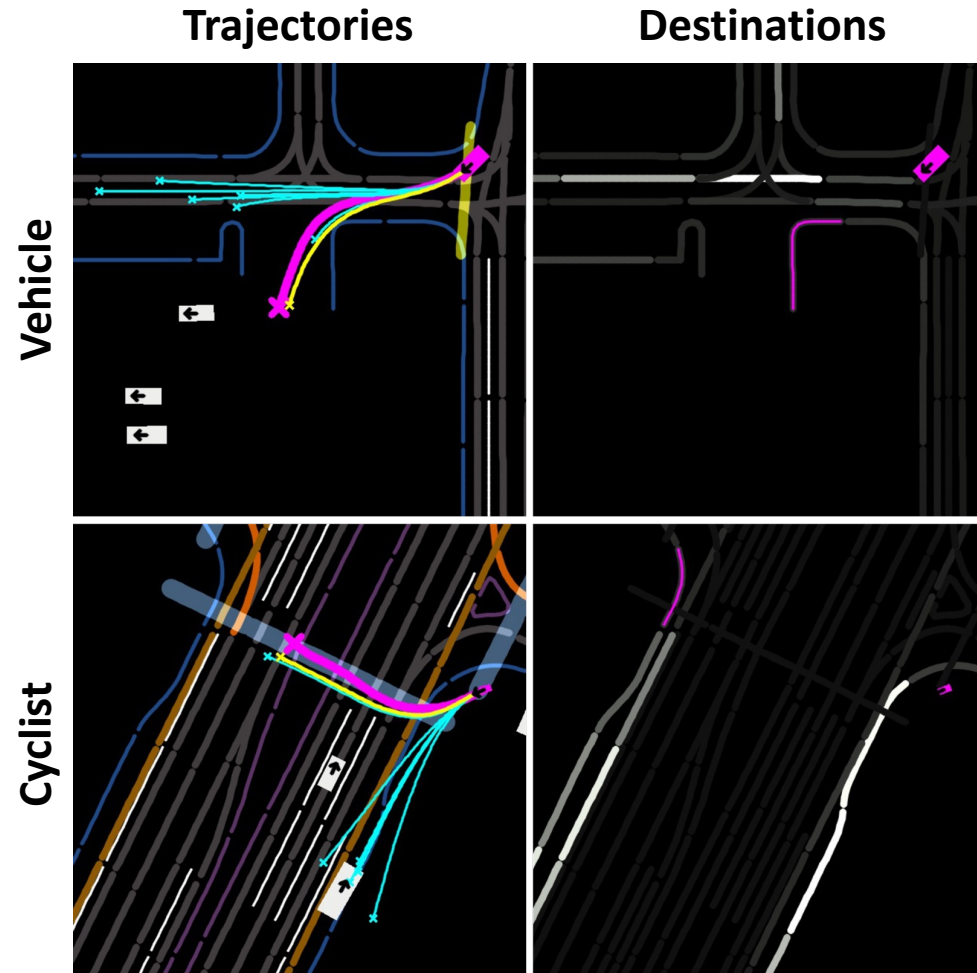
WOMD ^[1] <i>valid</i>	a priori sim K=6 (motion prediction)					a posteriori sim K=1				
	mAP \uparrow	min ADE \downarrow	min FDE \downarrow	miss rate \downarrow	overl. rate \downarrow	diff. pos \downarrow	diff. rot \downarrow	veh col %, \downarrow	run red %, \downarrow	passive %, \downarrow
TrafficBots	0.18	1.49	3.66	0.39	0.15	0.80	2.84	11.5	1.31	19.1
w/o angular emb.	0.12	1.74	4.48	0.48	0.18	0.74	3.05	14.7	1.47	19.4
w/o personality	0.06	1.66	4.09	0.48	0.15	1.29	3.63	13.6	1.50	19.2
w/o dest. w/ goal	0.17	1.47	3.44	0.40	0.16	0.78	2.68	12.3	1.35	20.2
SimNet ^[3]	0.01	2.76	7.77	0.76	0.21	2.27	7.37	21.9	1.59	19.6

Task 1: **A Priori** Simulation, i.e. motion prediction, multi-modal.

Task 2: **A Posteriori** Simulation, i.e. scenario reconstruction, single-modal.

TrafficBots achieve **baseline** performance on open-loop motion prediction task, and **SOTA** performance on closed-loop simulation task.

Prediction and Simulation Results



- In **Magenta**:
 - Agent of interest
 - GT Future trajectory
 - GT destination polyline
- In **Cyan**:
 - **A priori** trajectories.
 - GT future and destination are not given.
- In **Yellow**:
 - **A posteriori** trajectory.
 - GT future and destination are given.
- All agents are simulated simultaneously. A selected one is visualized.

Summary



- High-Fidelity Simulation for AD
 - **Behavior-Realism**: for E2E or planning algorithms.
 - Photo-Realism: for E2E or perception algorithms.
- TrafficBots, A Multi-Agent Policy:
 - **Realistic behavior** learned from real-world datasets.
 - **Game AI** for bot-agents in AD simulation.
 - **Faster** than real-time simulation.
 - **Parallelizable** on single GPU.
 - **All** traffic participants in dense urban scenarios.
 - **Configurable and interpretable** via personality and destination.