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

Zhejun Zhang[1], Alexander Liniger[1], Dengxin Dai[1,2], Fisher Yu[1], Luc Van Gool[1]

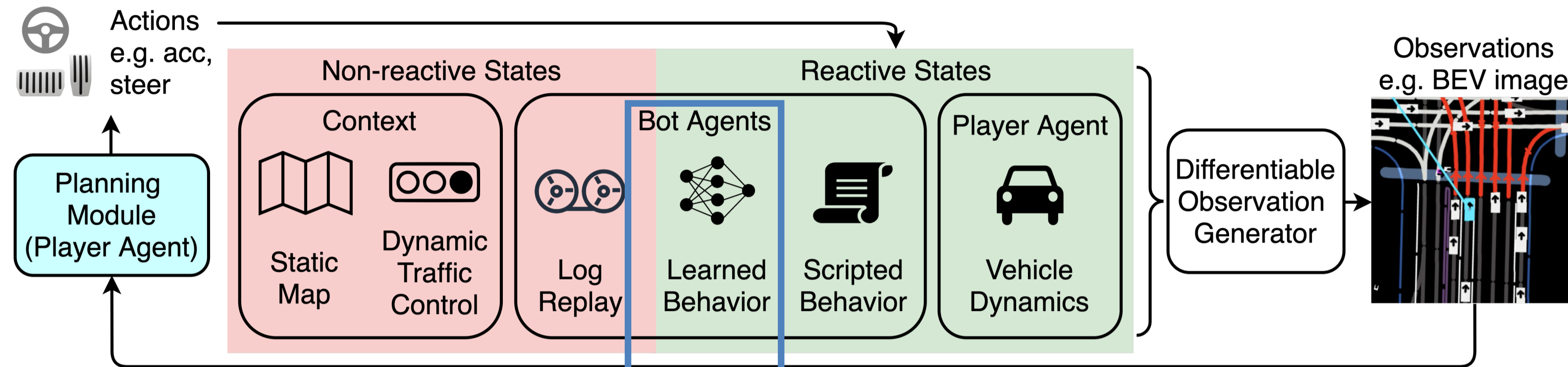[1]Computer Vision Lab, ETH Zurich. [2]MPI for Informatics. [3]PSI, KU Leuven.

World Model for Autonomous Driving Planning Modules

Problem: In previous simulations, the behavior of bot agents is **unnatural**.

Consequences: Planning algorithms developed in these simulations will **overfit** the unnatural behavior and fail to interact with real human.
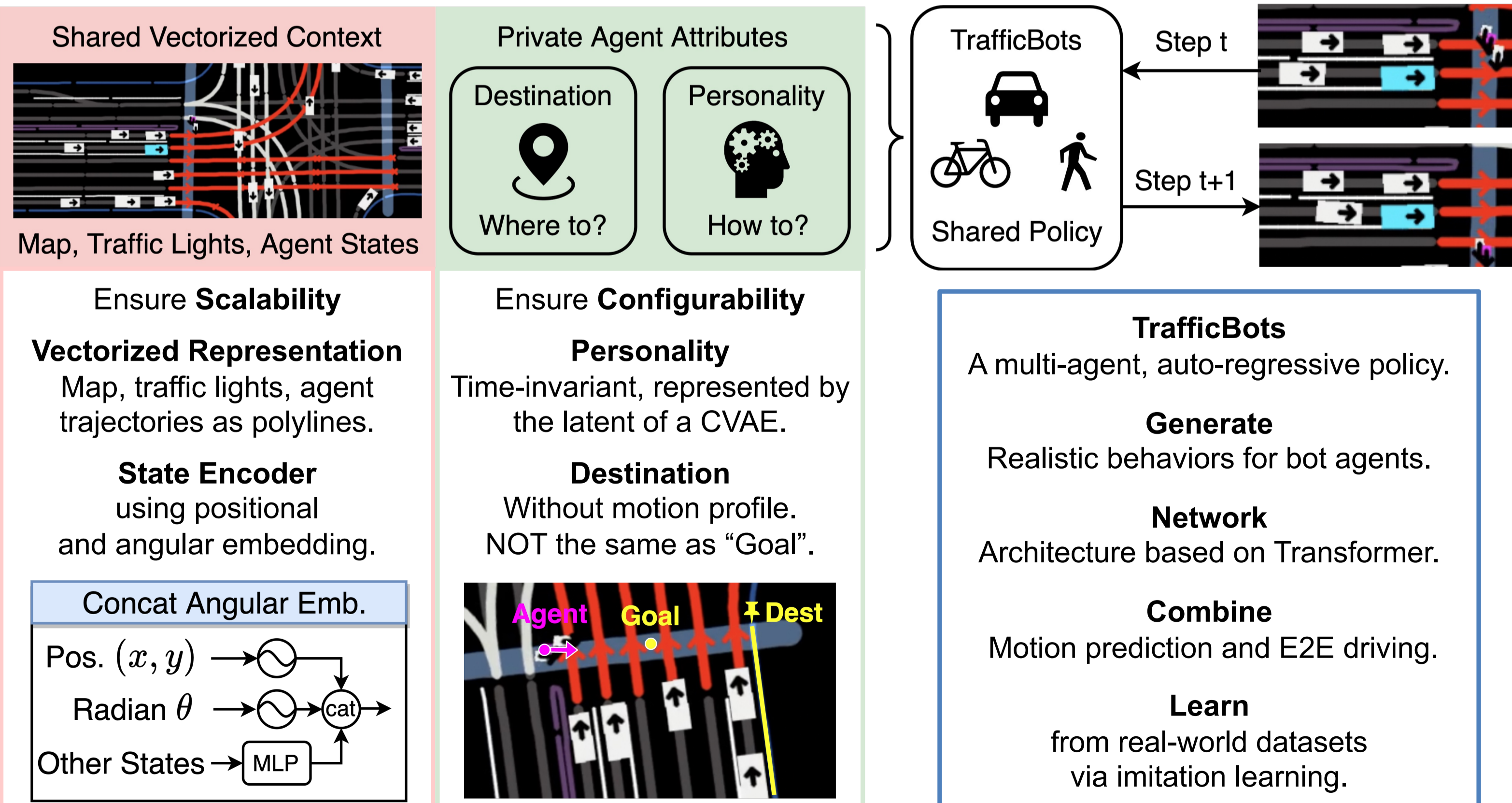
Solution: Generate **human-like** behavior for bots.

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

TrafficBots: A Multi-Agent Policy Learned from Real-World Motion Prediction Datasets

Shared Vectorized Context
Map, Traffic Lights, Agent States

Private Agent Attributes
Destination — Where to?
Personality — How to?

TrafficBots — Shared Policy
Step t → Step t+1

Ensure **Scalability**

Vectorized Representation
Map, traffic lights, agent trajectories as polylines.

State Encoder
using positional and angular embedding.

Concat Angular Emb.
Pos. (x, y) →
Radian θ →
Other States → MLP → cat

Ensure **Configurability**

Personality
Time-invariant, represented by the latent of a CVAE.

Destination
Without motion profile.
NOT the same as "Goal".

Agent — Goal — Dest

TrafficBots
A multi-agent, auto-regressive policy.

Generate
Realistic behaviors for bot agents.

Network
Architecture based on Transformer.

Combine
Motion prediction and E2E driving.

Learn
from real-world datasets via imitation learning.

Performance

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

WOMD[1] *test*	soft mAP ↑	mAP ↑	minADE ↓	minFDE ↓	miss rate ↓	overlap rate ↓
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] *valid*	a priori sim K=6 (motion prediction)					a posteriori sim K=1				
	mAP ↑	min ADE ↓	min FDE ↓	miss rate ↓	overl. rate ↓	diff. pos ↓	diff. rot ↓	veh col %, ↓	run red %, ↓	passive %, ↓
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.
TrafficBots achieve **SOTA** performance on closed-loop simulation task.

Qualitative Results

Trajectories — **Destinations**
Vehicle / Cyclist

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.