Nocturne
a scalable driving benchmark for bringing multi-agent learning one step closer to the real world

Eugene Vinitsky[*12], Nathan Lichtlé[*13], Xiaomeng Yang[*2], Brandon Amos[2], Jakob Foerster[4]

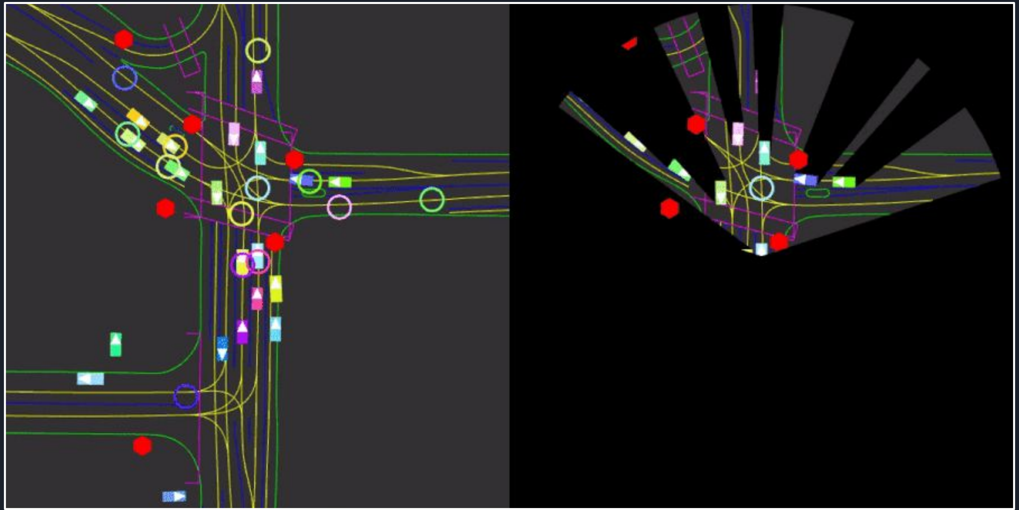
[*] Contributed equally
[1] UC Berkeley
[2] Meta AI
[3] École des Ponts ParisTech
[4] University of Oxford

What is Nocturne?

A fast, partially observable driving environment

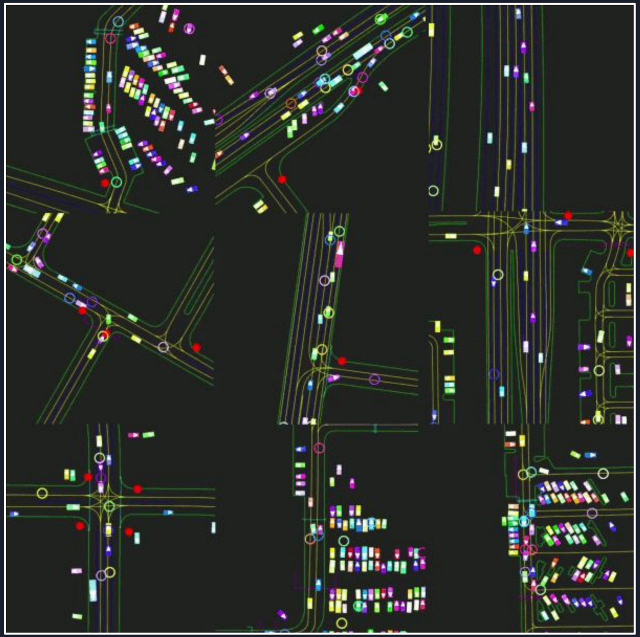
A benchmark for human-level driving

➔ Get each agent to its goal without collisions.
➔ Behave similarly to human behavior.



How is Nocturne a distinct multi-agent benchmark?

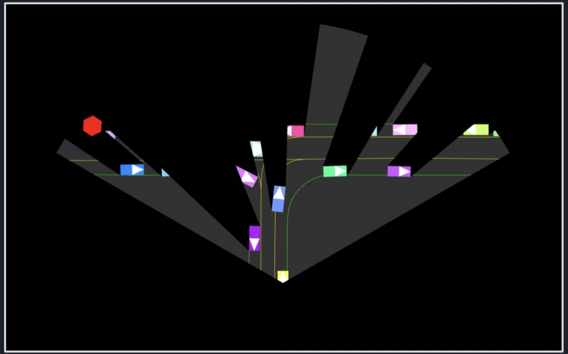
- ➔ Many-agent
- ➔ Mixed-motive
- ➔ Partial observability
- ➔ High-dimensional observations
- ➔ Continuous actions
- ➔ Extensive evaluation on over 130k data-driven scenes from Waymo Motion



Design of the Benchmark - Structures

Efficient visibility checks (computational bottleneck) through data structures:

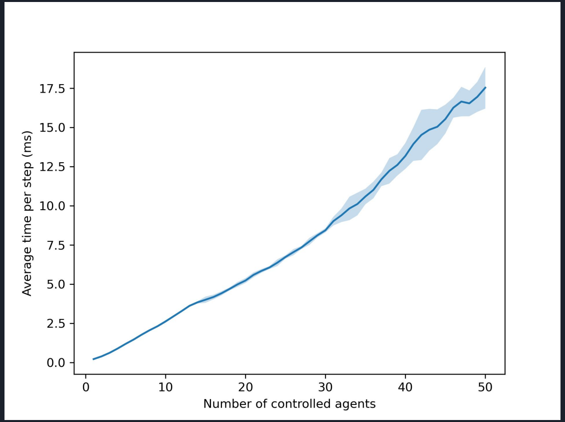
- Bounding volume hierarchies for vehicles and road edges
- 2D range-trees for road points
- Brute force ray-casting on candidates



➔ **Single-agent: ~3400 steps per second**

➔ **Multi-agent: ~3800 frames of experience per second**

➔ Linear scaling with number of agents



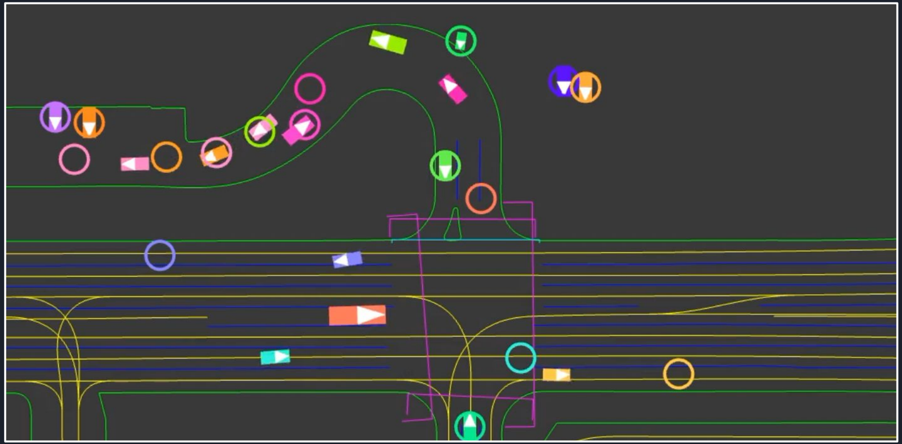
Design of the Benchmark - Environment

Each vehicle has a target goal (optionally speed and heading targets)

Termination conditions:

- Agent reaches target goal
- Agent collides with another vehicle or a road edge

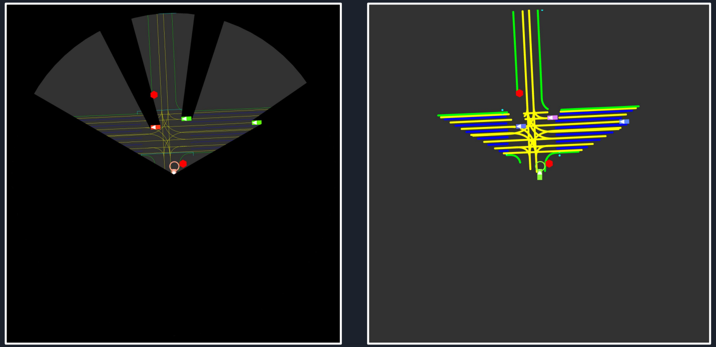
Vehicle dynamics use bicycle model



Design of the Benchmark - POMDP

Vectorized state space.
(around 13000 egocentric features):

- Agent state (speed, heading, etc.)
- Target goal (position, speed, heading)
- Visible vehicles
- Visible road points (in VectorNet representation)



Visible obstructed region
120º cone, 80m view distance

Action space. Accelerate, steer, rotate head (view cone).

Reward function. We treat our agents as goal-driven:

$$r_t = 0.2 \times \left(1 - \frac{||x_t - x_g||_2}{||x_0 - x_g||_2}\right) + 0.2 \times \left(1 - \frac{||v_t - v_g||_2}{40}\right) + 0.2 \times \left(1 - \frac{f(h_t, h_g)}{2\pi}\right)$$

Approach target position   Approach target speed   Approach target heading

Experiments

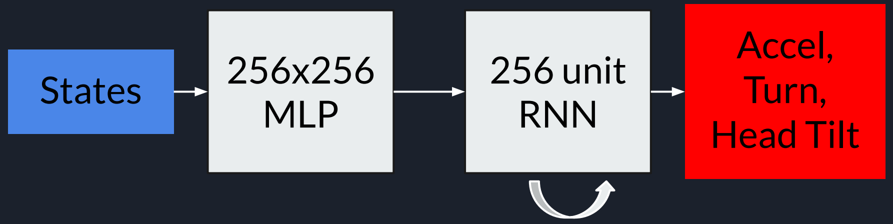
Reinforcement Learning

- Algorithm: **Asynchronous Proximal Policy Optimization** (APPO)
- Using SampleFactory* library
- Achieves 60'000 SPS on 1 GPU, 10 CPUs

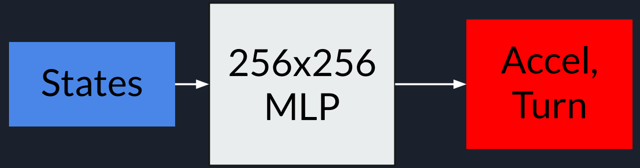
Imitation Learning

- Algorithm: **Behavioral Cloning** (BC)
- Invert bicycle model to get actions
- Stack 5 states for memory
- Achieves 20'000 SPS on 1 GPU, 10 CPUs

Architecture



Architecture



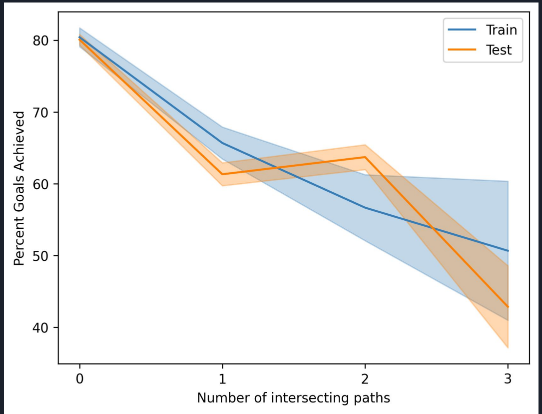
*Petrenko, Aleksei, et al. "Sample factory: Egocentric 3d control from pixels at 100000 fps with asynchronous reinforcement learning." *International Conference on Machine Learning*. PMLR, 2020.

Experiments results

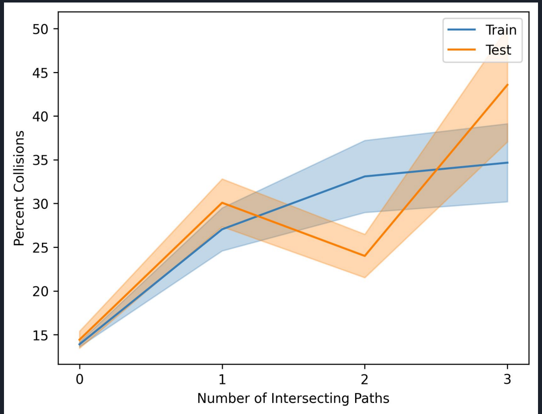
Table 2: Overview of metrics across methods for an 8 second rollout.

Algorithm	Collision Rate (%)	Goal Rate (%)	ADE (m)	FDE (m)
Expert Playback	4.9	100	0	0
APPO	20.3 ± 0.8	71.7 ± 0.7	3.1 ± 0.2	6.1 ± 0.3
BC	$38.2 \pm .1$	25.3 ± 0.1	5.6 ± 0.1	9.2 ± 0.1

Goal rate



Collision rate



Thank you!

Upcoming: Nocturne v2

- ❑ Traffic lights
- ❑ Pedestrians & cyclists
- ❑ GPU backend
- ❑ Scene completion
- ❑ Support for additional datasets

Try out Nocturne at: **github.com/facebookresearch/nocturne**