

Found-RL: foundation model-enhanced reinforcement learning for autonomous driving



Yansong Qu^a, Zihao Sheng^b, Zilin Huang^b, Jiancong Chen^a, Yuhao Luo^b, Tianyi Wang^c, Yiheng Feng^a, Samuel Labi^a, Sikai Chen^{b*}

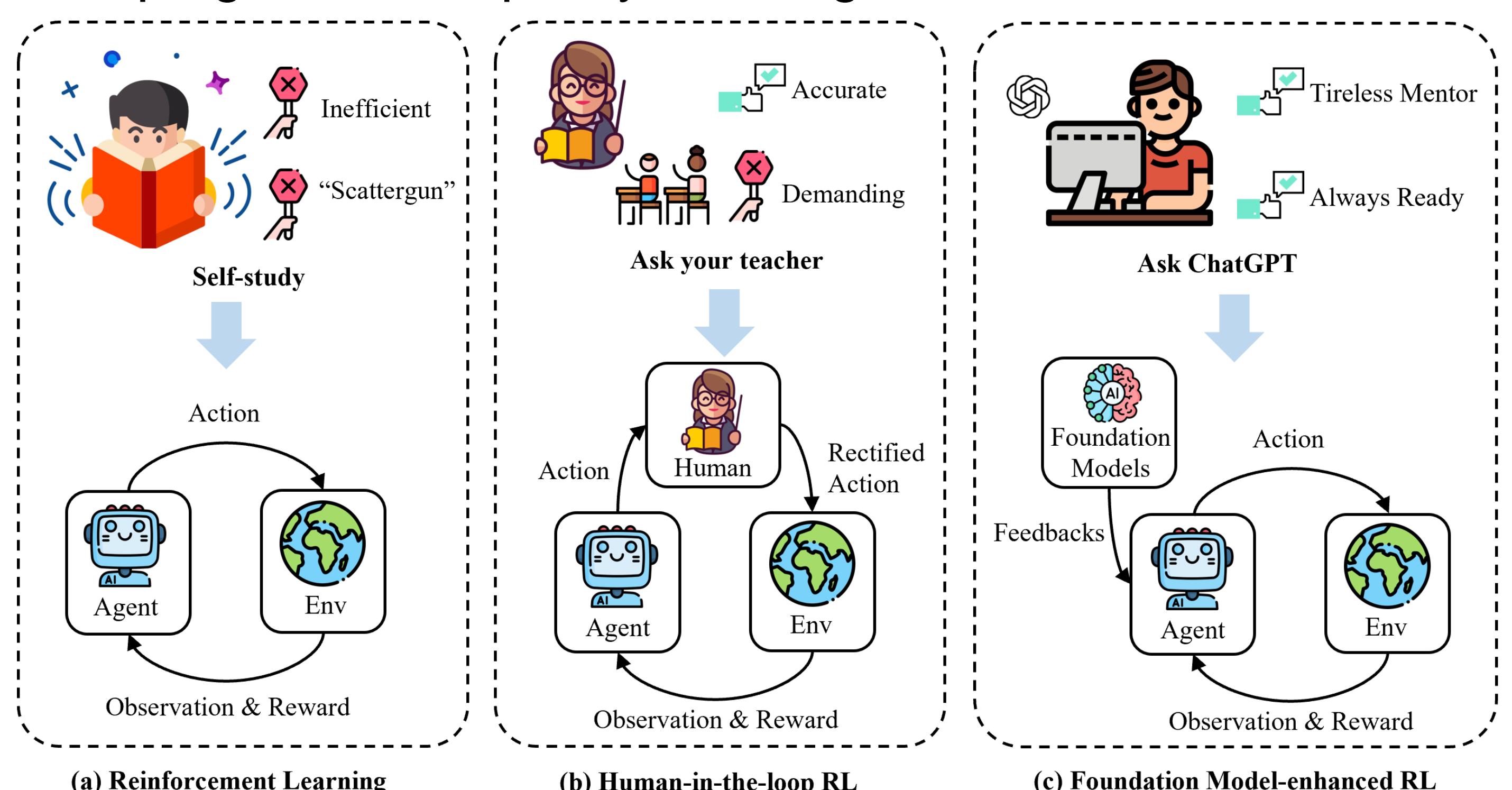
a. Purdue University

b. University of Wisconsin-Madison

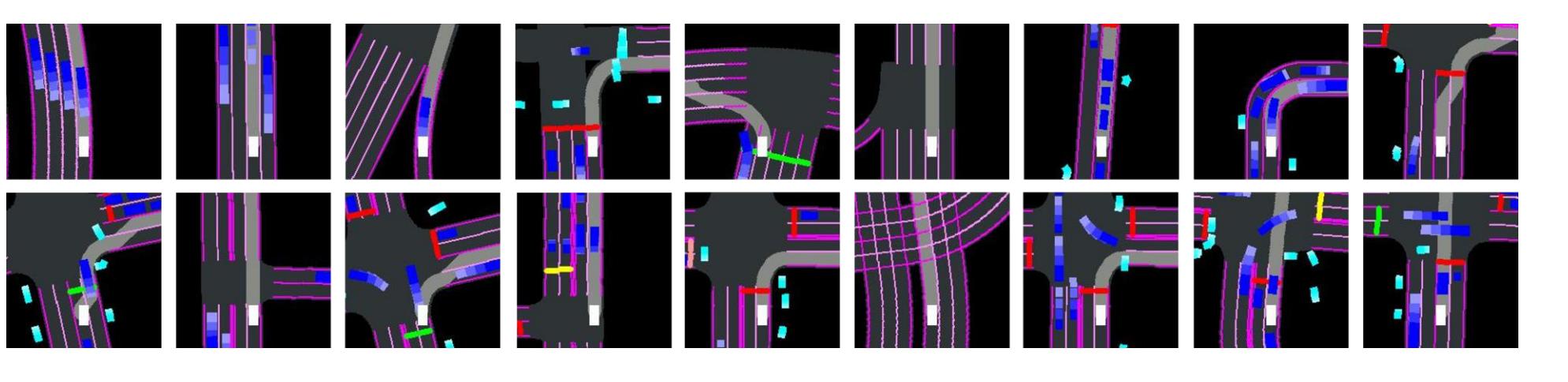
c. University of Texas at Austin

INTRODUCTION

- Motivation: RL is fast at inference but hard to train well in complex driving scenes.** End-to-end reinforcement learning is a strong paradigm for autonomous driving, yet it typically suffers from low sample efficiency and limited semantic interpretability, especially in long-tail and safety-critical scenarios.
- Gap: VLM supervision is powerful, but hard to use in closed-loop RL training.** Vision-language models can act like scalable “tireless mentors” by providing context-aware guidance, but their heavy inference cost creates major latency bottlenecks in high-frequency, multi-environment RL pipelines.
- Our solution: Found-RL enables practical foundation-model-enhanced RL.** We introduce Found-RL with an asynchronous batch inference design that decouples VLM reasoning from the simulation loop, enabling real-time/near-real-time training while supporting online action guidance (VMR/AWAG) and CLIP-based dense reward shaping for better policy learning.



EXPERIMENT SETTING

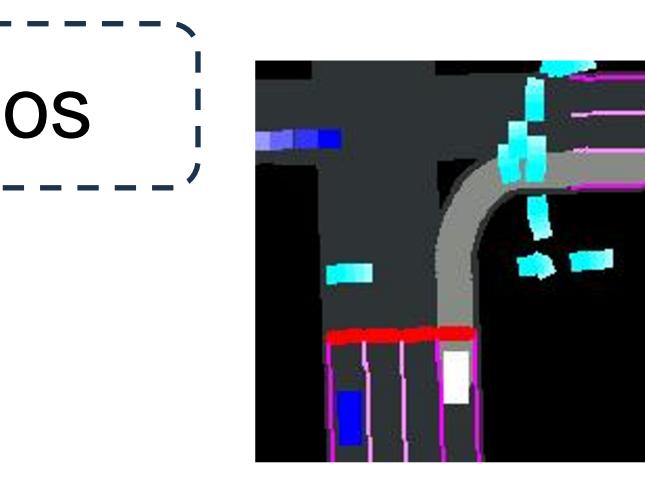


Driving scenarios.

CARLA-based comprehensive driving scenarios



(a) Observation space of VLM-based driving agents

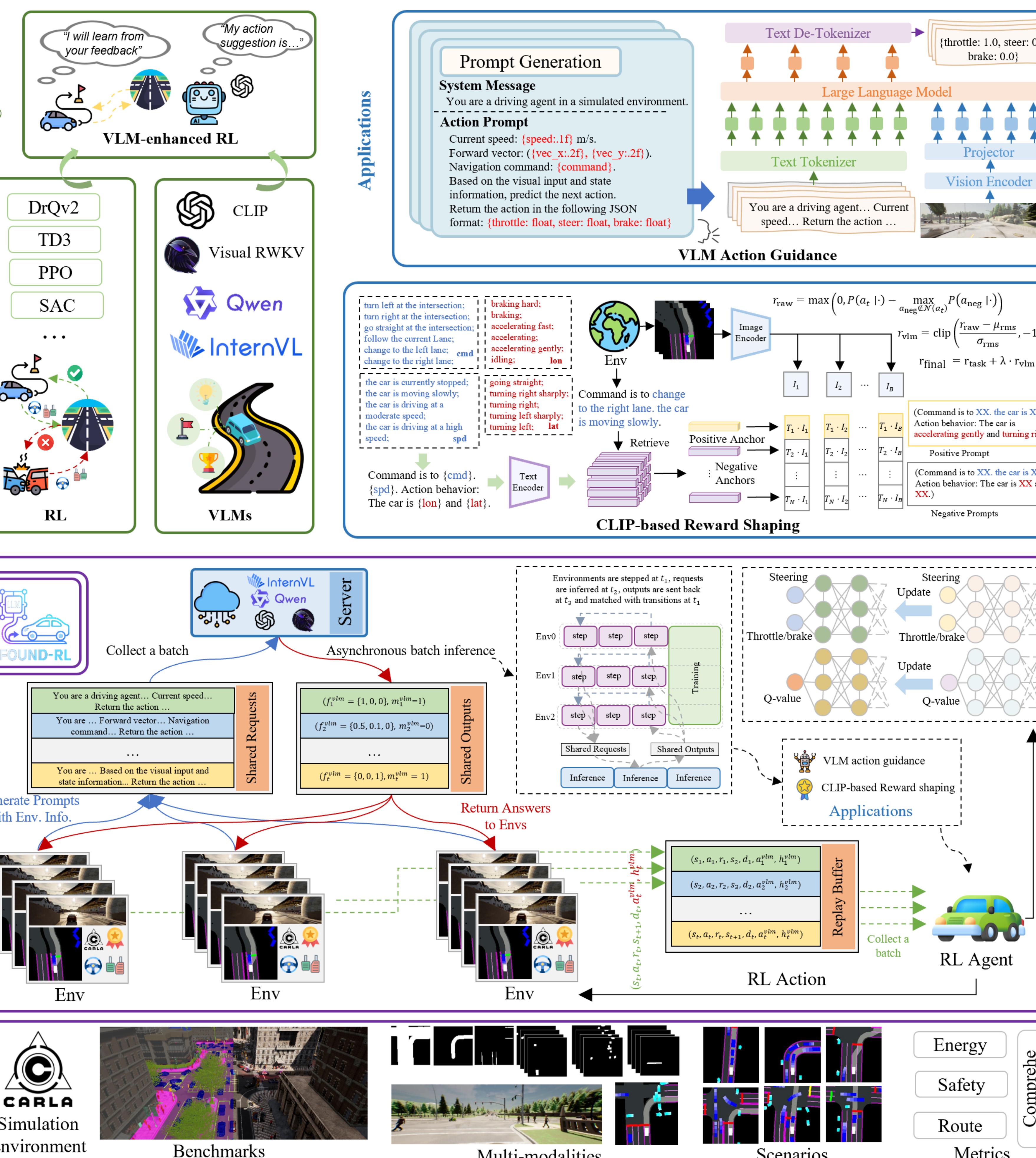


Question: You are a driving agent in a simulated environment. Current speed: 0.0 m/s. In forward vector: (4.51, 13). Navigation command: "follow the current lane". Based on the visual input and state information, predict the next action. Return the action in the following JSON format: {throttle: float, steer: float, brake: float}. Answer: {throttle: 1.0, steer: 0.0, brake: 0.0}.

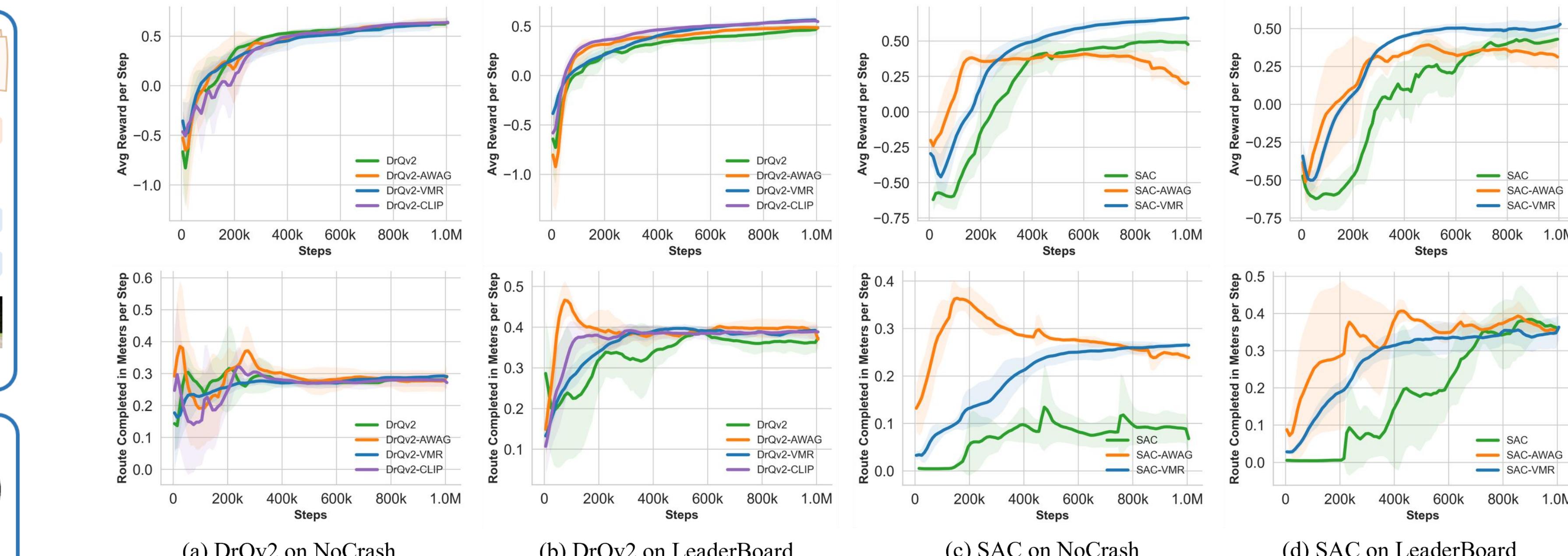
(b) Observation space of RL-based driving agents

Obs space.

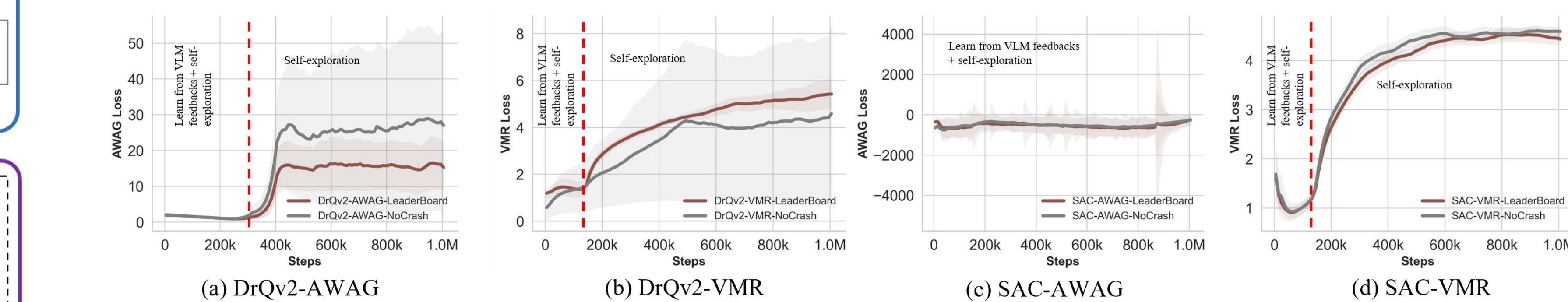
METHODOLOGY (FOUND-RL)



EXPERIMENT RESULTS



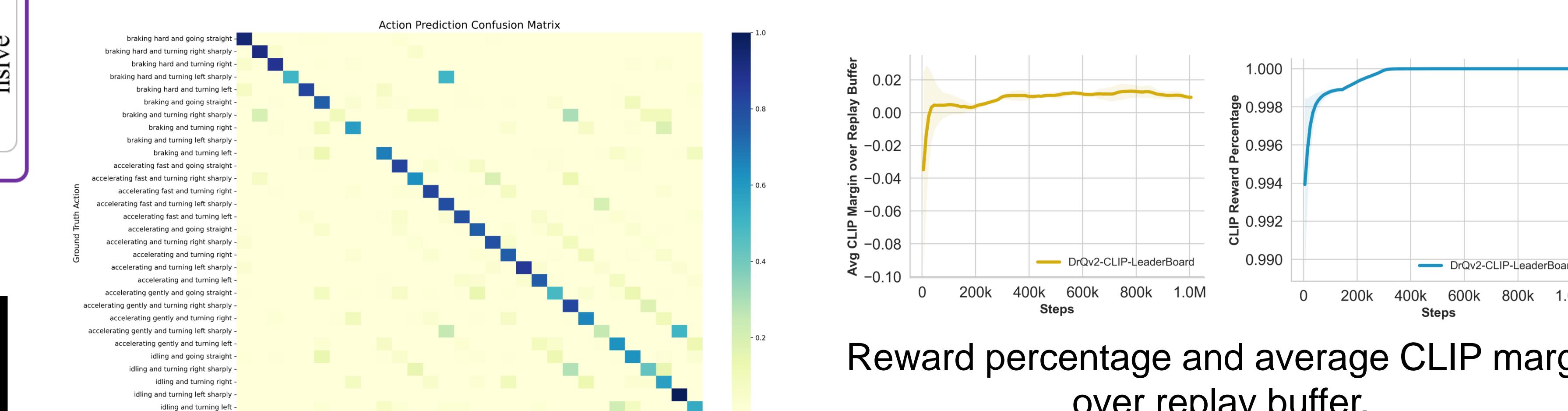
Comparison of average reward and route completion in meters per step between Found-RL and classical RLs on Leaderboard and NoCrash benchmarks.



Comparison of AWAG and VMR losses between Found-RL and classical RLs on Leaderboard and NoCrash benchmarks.

| Table I. Compare with RLs on Leaderboard benchmark | | | | | | | | | | | | |
|--|------------------|-----------------|-----------------|------------------|----------------|--------------------|--------------------|--------------------|--------------------|--------------------|-------------|-------------|
| Algorithm | Model Size | Comprehensive | | Route | | Energy | | Safety | | | | |
| | | Return ↑ | Driving Score ↑ | Infra. Penalty ↓ | Success Rate ↑ | Route Compl. ↑ | Speed ↑ | Cell 1 | Fuel Rate ↓ | | | |
| TD3 | 0.1B | BEV (192×192×3) | 2044.0 | 0.65 | 0.66 | 0.52 | 0.70 | 3.63 | 0.05 | 0.01 | 0.15 | 0.08 |
| SAC | 527.03 ± 102.66 | BEV (192×192×3) | 2702.49 | 0.76 | 0.91 | 0.65 | 0.80 | 3.50 | 0.04 | 0.01 | 0.09 | 0.05 |
| SAC-AWAG | 556.33 ± 152.86 | 0.28 ± 0.05 | 0.82 ± 0.01 | 0.08 ± 0.04 | 0.33 ± 0.06 | 5.21 ± 1.10 | 0.86 ± 0.02 | 0.02 ± 0.00 | 0.01 ± 0.01 | 0.63 ± 0.09 | 0.15 ± 0.09 | 0.05 ± 0.03 |
| SAC-VMR | 1623.34 ± 265.50 | 0.55 ± 0.10 | 0.89 ± 0.03 | 0.31 ± 0.14 | 0.58 ± 0.11 | 3.67 ± 0.27 | 0.07 ± 0.01 | 0.00 ± 0.00 | 0.20 ± 0.05 | 0.02 ± 0.02 | 0.04 ± 0.01 | 0.04 ± 0.01 |
| DrQv2 | 1507.00 ± 317.97 | 0.56 ± 0.11 | 0.88 ± 0.01 | 0.38 ± 0.13 | 0.61 ± 0.12 | 3.72 ± 0.20 | 0.09 ± 0.03 | 0.00 ± 0.00 | 0.01 ± 0.01 | 0.19 ± 0.09 | 0.02 ± 0.03 | 0.02 ± 0.02 |
| DrQv2-AWAG | 1626.55 ± 497.37 | 0.61 ± 0.10 | 0.90 ± 0.02 | 0.46 ± 0.15 | 0.66 ± 0.10 | 3.68 ± 0.27 | 0.66 ± 0.01 | 0.00 ± 0.00 | 0.01 ± 0.01 | 0.13 ± 0.03 | 0.02 ± 0.02 | 0.02 ± 0.02 |
| DrQv2-VMR | 2237.51 ± 173.69 | 0.72 ± 0.04 | 0.89 ± 0.03 | 0.60 ± 0.04 | 0.78 ± 0.05 | 3.56 ± 0.09 | 0.09 ± 0.05 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.16 ± 0.04 | 0.02 ± 0.01 | 0.02 ± 0.01 |
| DrQv2-CLIP | 2188.54 ± 294.03 | 0.77 ± 0.05 | 0.90 ± 0.02 | 0.57 ± 0.06 | 0.77 ± 0.05 | 3.60 ± 0.10 | 0.13 ± 0.03 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.14 ± 0.05 | 0.01 ± 0.00 | 0.01 ± 0.00 |

Compare found-rl between RLs and VLMs, our found-rl can achieve near-VLM performance with real-time inference speed



Personal Website