Scaling DRL for Roundabout Management via Multi-Resolution SUMO–CARLA Co-Simulation

Ali Nadar and Jérôme Härri

Communication Systems, EURECOM, 450 route des Chappes 06904 Sophia-Antipolis, France
{ali.nadar, haerri}@eurecom.fr
7th Symposium on Management of Future Motorway and Urban Traffic Systems
September 9-11, 2026, Naples, Italy

Keywords: Reinforcement learning, SUMO, CARLA, Multi-Resolution, Roundabout

1 Introduction

Roundabouts are safety-critical bottlenecks where local yield/merge decisions influence not only vehicle-level safety and passenger comfort, but also network-level performance (delay, queues, and throughput). Automated vehicles must therefore master their management not to negatively impact traffic. In our previous work (Nadar & Härri, 2025), we demonstrated that a Proximal Policy Optimization (PPO) agent can learn safe, efficient, and comfort-aware longitudinal control for a *single* entry leg in CARLA Town03 by explicitly penalizing unsafe and abrupt maneuvers. However, extending a CARLA-only setup to multi-legs, dense traffic, and multiple roundabout approaches is computationally costly, which limits training throughput and prevents realistic, network-scale evaluation. To address this, we propose a multi-resolution SUMO-CARLA co-simulation architecture: SUMO handles network-scale traffic generation and evolution, while CARLA is activated only in a predefined roundabout zone, where high-fidelity dynamics and comfort constraints are required. Inside this zone, the policy is supported by (i) a Roundabout Exit Probability (REP) model to anticipate circulating vehicles that will exit before the conflict point and reduce unnecessary conservatism, and (ii) a learned TTC predictor to guide safer and smoother entry behavior. A single shared PPO policy is trained across all entry legs using a curriculum strategy to improve robustness across traffic densities and geometries.

2 Background

The approach builds on our earlier two-stage pipeline (Nadar & Härri, 2025) depicted on Figure 1). In the offline stage, lightweight predictors are trained to enrich situational awareness. The online stage then trains a PPO controller to decide whether to yield or proceed while optimizing safety, efficiency, and comfort.

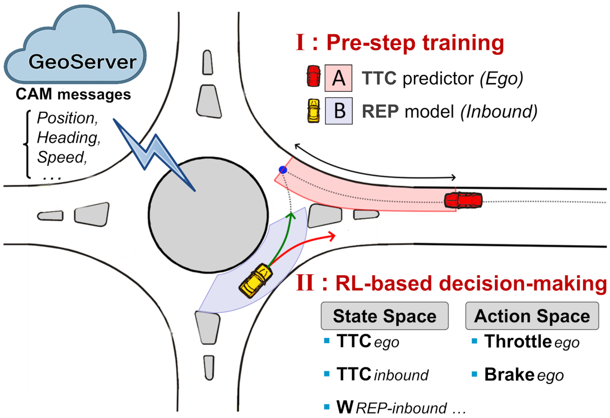
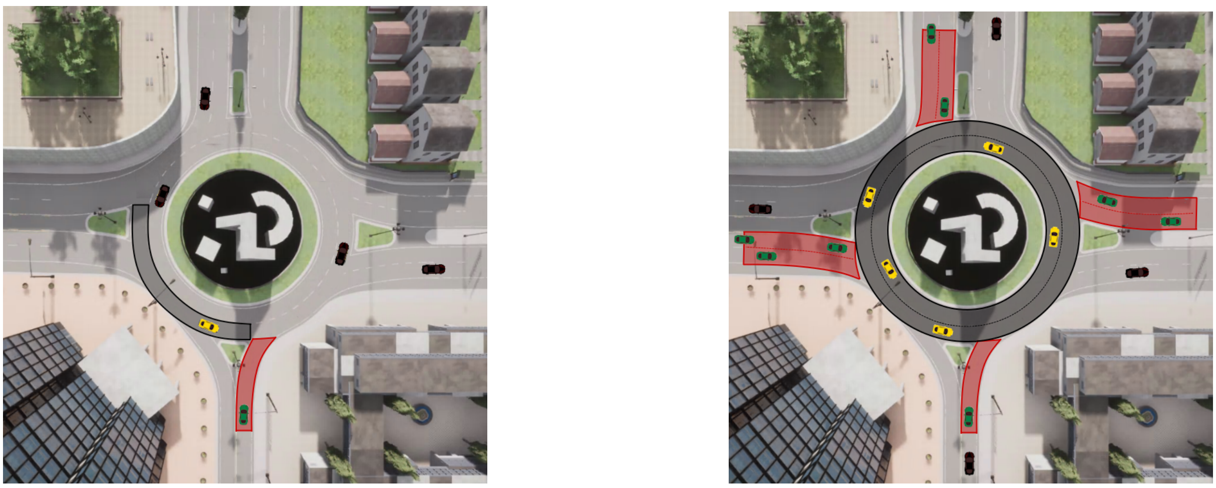


Figure 1 – *Overview on applying DRL for Roundabout Management*

Specifically, REP estimates whether a circulating vehicle is likely to exit before reaching the ego vehicle's conflict point, helping to ignore non-threatening vehicles and reduce over-conservative yielding. In addition, a TTC predictor provides a more reliable safety signal for reward shaping while respecting comfort-related constraints (bounded acceleration and jerk). While effective in a single controlled entry scenario Fig. 2(a), this CARLA-only setup does not scale well to a multi-entry, multi-traffic regime Fig. 2(b), where interactions vary across approaches, priorities, and demand levels. For instance, *any* intra-roundabout vehicles must provide high movement granularity for the in-bound ego-vehicle's perception unit to accurately estimate REP and TTC. Also, realistic roundabout vehicular in-bound and out-bound distributions are required to model realistic REP. This motivates a scalable multi-resolution training and evaluation architecture that preserves high-resolution control & perception where needed, while scaling efficiently through lower-resolution traffic modeling elsewhere..



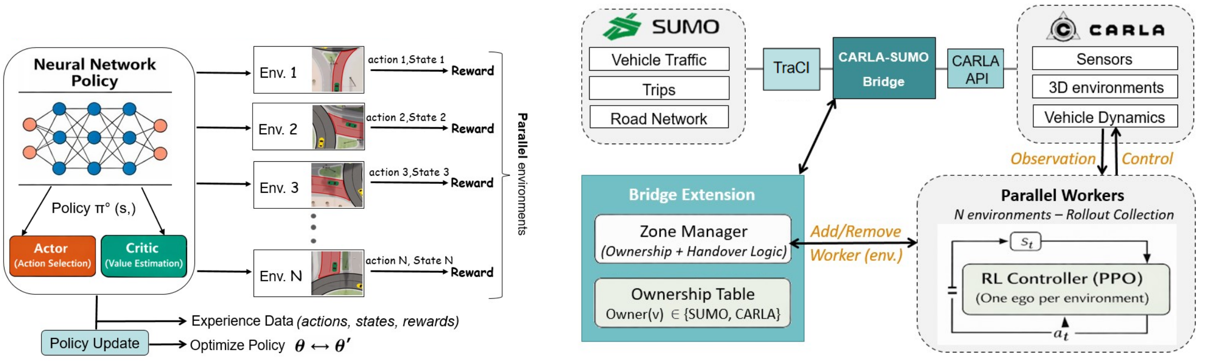
(a) *Single-topology context (single entry leg in one roundabout).*

(b) *Multi-topology context (all entry legs in all roundabouts).*

Figure 2 – *Scaling challenge: from single to multi topology RL control for roundabout management*

3 Methodology

Compared to (Nadar & Härri, 2025), we move from synthetic, hand-crafted episodes to realistic vehicle lifecycles. Vehicles are generated at the network scale and approach the roundabout under routing, car-following, and lane-changing behaviors. Any vehicle entering the observation zone temporarily becomes the RL-controlled ego vehicle, enabling large-scale experience collection under diverse micro-scale interactions. In a nutshell, different from Yilmaz-Niewerth *et al.* Yilmaz-Niewerth *et al.* (2024) *mixed* SUMO–CARLA multi-granularity approach, we propose a *split* multi-granularity architecture, where large-scale simulation areas are split into exclusive modeling zones with high or low granularity.



(a) *RL: Single-agent Parallel-Environment* (b) *Architecture: Scaling RL with contextual co-simulation*

Figure 3 – *Zone-based control handover for RL parallel training using co-sim SUMO-CARLA*

3.1 Scalable policy learning via parallel single-agent PPO

Although deployment involves dense interactions, we retain a single-agent PPO formulation to preserve learning stability. To jointly increase sample efficiency and improve representational diversity during training, we employ N parallel co-simulation workers (Fig. 3a). Each worker corresponds to one vehicle currently located inside the predefined observation zone. In practice, N dynamically reflects the number of vehicles simultaneously approaching or merging at the roundabout. When a vehicle enters the roundabout control region (starting d meters before the yield line), control authority is transferred to CARLA and its PPO controller; after exiting the control zone (i.e. after the roundabout exit), authority is returned to SUMO.

Beyond accelerating rollout collection, this split multi-granularity formulation exposes the policy to heterogeneous traffic realizations, including different entry legs, demand levels, vehicle spacings, and interaction topologies. As a result, learning becomes not only faster but also more representative of realistic multi-vehicle traffic conditions, while avoiding the instability of fully multi-agent RL. The actor–critic architecture follows our prior design: an MLP (500–300–100, tanh) mapping the same normalized state vector to continuous throttle and brake commands, with a mirrored critic network for value estimation and advantage computation.

3.2 Context-based dynamic control in SUMO–CARLA co-simulation

Our training formulation requires that multiple vehicles be controlled concurrently inside the roundabout control zone, while the rest of the road network evolves under realistic, network-scale traffic. CARLA alone is costly for dense network simulation, whereas SUMO does not provide the physics-based dynamics needed for comfort- and safety-aware RL. SUMO-CARLA co-simulation support was introduced by CARLA through its SUMO integration interface[1], and a recent work by Yilmaz-Niewerth *et al.* Yilmaz-Niewerth *et al.* (2024) demonstrated that coupling CARLA with SUMO can enable large-scale urban co-simulation by embedding high-fidelity vehicles within microscopic traffic flows. However, this approach assumes a largely static ownership of vehicles across simulators: vehicles spawned and managed by SUMO remain SUMO-controlled, and CARLA-controlled vehicles remain CARLA-controlled.

We therefore extend the SUMO–CARLA bridge with a *context-based dynamic handover mechanism* (Fig. 3b). Unlike conventional co-simulation approaches based on static vehicle ownership,

[1] https://carla.readthedocs.io/en/latest/adv_sumo

our framework performs zone-driven control reassignment through a dedicated zone manager, triggered when a vehicle enters the roundabout observation area. The manager enforces single authoritative control by maintaining an exclusive ownership table that tracks simulator responsibility and governs transfer events. During handover, state variables (position, velocity, control inputs) are synchronized to prevent inconsistencies. This split multi-granularity design activates CARLA only within critical regions while preserving network-scale efficiency under synchronous co-simulation.

4 On-going Results and Discussion

We evaluate the framework along three axes: (i) **Scalability** (real-time factor, stability, total simulated vehicles, and number of concurrent CARLA-controlled agents), (ii) **Efficiency** (throughput, delay, stops, and queue length per entry leg), and (iii) **Safety and comfort** (collisions, TTC violations, jerk distribution, and harsh-braking events). Fig. 4 illustrates the zone-based control handover. Vehicles are governed by SUMO or CARLA depending on their location. When vehicle C (green) enters the predefined approach action zone, control authority transfers from SUMO to CARLA, where the learned policy operates under high-fidelity dynamics using state variables (d, v, a) (distance to conflict point, speed, acceleration). After completing its maneuver, vehicle S1 exits the zone and control reverts to SUMO, preserving CARLA resources for active learning agents. Vehicle S3 remains SUMO-controlled until it crosses the zone boundary, at which point the zone manager dynamically assigns it to CARLA. Vehicle S2 represents an adversarial inbound vehicle; it remains SUMO-controlled but is observed by C through (d_2, v_2) and the exit probability w_{REP}, influencing the RL decision process. The zone manager guarantees exclusive ownership and synchronized handover, ensuring state consistency and simulation stability.

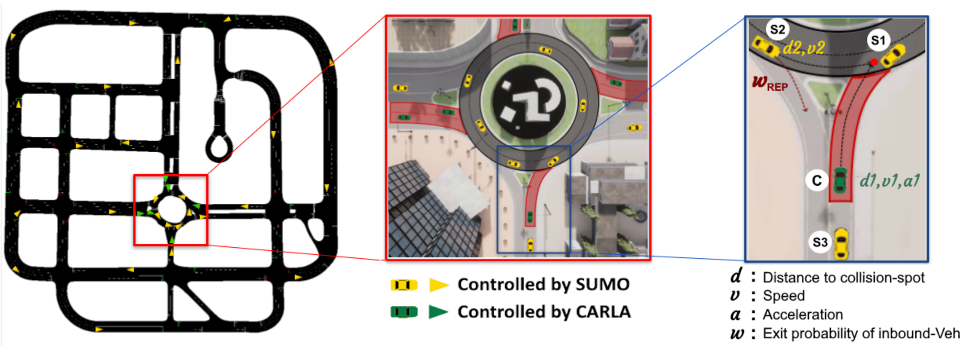


Figure 4 – *Context-based SUMO-CARLA co-simulation on Town03*

Acknowledgements

This work has benefited from a government grant managed by the "*Agence Nationale de la Recherche*" under the France 2030 program with NF-FITNESS, grant "ANR-22-PEFT-0007".

References

Nadar, Ali, & Härri, Jérôme. 2025. Optimizing Roundabout Management via Deep Reinforcement Learning with Safety and Comfort Constraints. *In: Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC)*. Gold Coast, Australia: IEEE.

Yilmaz-Niewerth, Sefa, Häbel, Riklas, & Friedrich, Bernhard. 2024. Developing a comprehensive large-scale co-simulation for replication of automated driving in urban traffic scenarios. *Transportation Research Procedia*, **78**, 522–529. 25th Euro Working Group on Transportation Meeting (EWGT 2023).