

Multi-Agent Reinforcement Learning for Safe Autonomous Driving Under Pedestrian Behavioral Uncertainty

Prakash Aryan¹, Kaushik Raghupathruni¹, Timo Kehrner¹, and Sebastiano Panichella^{1,2}

Abstract—Simulation-based testing of self-driving cars (SDCs) typically relies on scripted pedestrian models that do not capture the heterogeneity and uncertainty of real crossing behavior, limiting the realism of safety assessments, especially for jaywalking, which is governed by latent personality traits the vehicle cannot observe. We hypothesize that jointly training pedestrians and the SDC with multi-agent reinforcement learning (MARL) yields more realistic interaction scenarios than training against fixed pedestrian policies, and that the behavior gap between predictable and unpredictable crossings can be measured directly from trajectories. We co-train an SDC and 12 pedestrians using Multi-Agent Proximal Policy Optimization (MAPPO): pedestrian locomotion follows scripted Dijkstra pathfinding while an RL policy controls high-level go/wait decisions, and jaywalking probability depends on a per-pedestrian trait sampled at episode start and hidden from the SDC. In 500-episode evaluations, the co-trained SDC reached 78% of goals with a 14% collision rate, versus 35%/33% for the best rule-based baseline. A speed differential metric shows the SDC traveled 2.65 m/s faster near jaywalkers than near crosswalk users at close range (0–3 m), indicating jaywalking encounters were not anticipated. Jaywalking was 13% of crossing events but 62% of collisions, and co-training reduced collisions by 30% relative to single-agent RL as pedestrians learned to wait when the SDC approached at speed.

I. INTRODUCTION

Pedestrian-vehicle interactions are a recurring source of urban collisions, and self-driving cars (SDCs) must account for them during training and evaluation. Existing simulation-based testing of SDCs typically relies on scripted pedestrian motion or simplified crossing rules that do not reflect the heterogeneity of real human behavior [1], [2], raising a question for SDC assessment: are we training and testing SDCs against pedestrian models realistic enough to expose latent-intent situations such as jaywalking? We hypothesize that co-training pedestrians and the SDC with multi-agent reinforcement learning (MARL), with a personality-driven jaywalking mechanism invisible to the vehicle, yields more realistic interaction scenarios than single-agent training, and that the gap between predictable and unpredictable crossings can be measured from trajectories. We instantiate this idea with Multi-Agent Proximal Policy Optimization (MAPPO) [3].

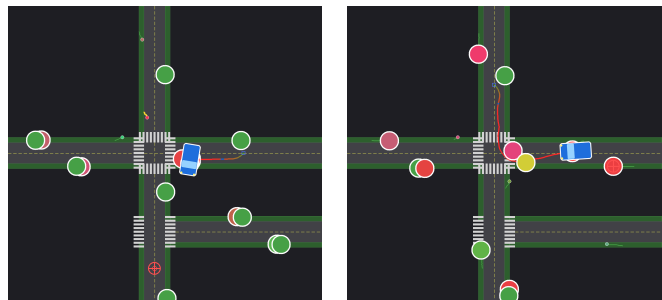
This work is supported by the SNSF SwarmOps project (No. 200021_219732).

Code: <https://github.com/prakash-aryan/marl-sdc-pedestrian-uncertainty>

¹University of Bern, Bern, Switzerland. {prakash.aryan, timo.kehrner, sebastiano.panichella}@unibe.ch, kaushik.raghupathruni@students.unibe.ch

²AI4I, The Italian Institute of Artificial Intelligence, Turin, Italy. sebastiano.panichella@ai4i.it

Prior work on pedestrian-vehicle interaction has examined risk-aware RL for jaywalker encounters [4], social force based simulation [1], [5], and survey evidence on jaywalking near automated vehicles [6]. MARL has been applied to cooperative driving at intersections [7], [8]. However, these works either fix the pedestrian model or train the SDC in isolation; none co-train pedestrians and the SDC with explicit, trait-driven behavioral uncertainty.



(a) Collision with jaywalker

(b) Successful avoidance

Fig. 1: Two outcomes in our environment. The blue rectangle is the SDC; coloured dots are pedestrians (hue encodes jaywalking tendency, green cautious to red reckless; yellow segments mark active jaywalking); grey is road, green strips are sidewalks, and the red crosshair is the goal. (a) Collision with a jaywalker. (b) The SDC steers around a jaywalker and reaches its goal. A collision is registered when the SDC and a pedestrian centre come within 1.5 m.

This paper makes three contributions. *First*, a MARL environment for pedestrian-SDC co-training in which jaywalking is driven by a latent personality trait hidden from the SDC, extending prior MARL intersection work that uses at most a handful of scripted pedestrians [7]. *Second*, a speed differential metric that quantifies, from trajectories, how the SDC responds differently to predictable crosswalk and unpredictable jaywalking encounters. *Third*, empirical evidence that co-training produces emergent cooperative waiting and reduces collisions by 30% relative to single-agent RL, even when the single-agent SDC is later paired with the MARL pedestrian policy.

II. RELATED WORK

Multi-agent RL. Cooperative MARL methods include MADDPG [9], QMIX [10], COMA [11], and IPPO [12]. MAPPO [3], built on proximal policy optimization (PPO) [13] with generalized advantage estimation (GAE) [14], uses the centralized training with decentralized

execution (CTDE) paradigm [15]: a centralized critic during training, decentralized actors at execution. We adopt MAPPO with a shared centralized critic.

Pedestrian behavior. The social force model [5] is foundational for pedestrian simulation; personality traits have been added to heterogeneous models [16], and recent work integrates sensory-motor constraints into RL-based pedestrian policies [17]. Trajectory prediction methods such as Social LSTM [18] and Trajectron++ [19] forecast paths but do not model crossing decisions. Khuzam et al. [20] studied jaywalking via Markov games, and Zhang et al. [4] proposed risk-aware RL for jaywalker interactions with the SDC trained in isolation.

Uncertainty and simulation testing. Kendall and Gal [21] distinguish aleatoric and epistemic uncertainty, Hoel et al. [22] apply uncertainty estimation to tactical driving, and Wang et al. [23] survey uncertainty quantification (UQ) methods for autonomous vehicles. Standard SDC testbeds include CARLA [24] and SUMO [25]. Birchler et al. [2] show simulation-based SDC testing does not always align with human perception of realism, motivating richer pedestrian models. GPU-accelerated RL via JAX [26] is demonstrated by JaxMARL [27] and CleanRL [28]. Table I summarizes the comparison; its final column marks that our method, like all listed systems, is validated only in simulation.

TABLE I: Comparison with Related Systems

System	Co-train	Jay-walk	UQ Metric	GPU Accel	Ped RL	Real Eval
Zhang [4]	×	✓	×	×	×	×
Khuzam [20]	×	✓	×	×	×	×
Yu [7]	✓	×	×	×	×	×
Rashid [1]	×	×	×	×	×	×
Ours	✓	✓	✓	✓	✓	×

III. SYSTEM DESIGN

Fig. 2 overviews the framework: RL policies choose high-level actions (pedestrians go/wait, the SDC accelerates/steers), a scripted layer routes pedestrians along pre-computed shortest paths (the SDC bypasses it), and a physics layer advances the bicycle model, all under CTDE training (Fig. 2a–d). The environment, training, and inference run in JAX on a single graphics processing unit (GPU) via `jax.vmap` (512 parallel environments) and `jax.lax.scan` (rollout collection).

A. Environment

The environment covers a 120×120 m urban map with a four-way intersection and a T-junction (3 road segments, 20 sidewalk segments, 6 crosswalks), so a single episode spans both intersection topologies. Simulation runs at $dt=0.1$ s (10 Hz) for 500 steps (50 s) per episode.

Pedestrians. The environment contains 12 pedestrian agents, far more than the up to 3 in the recent CARLA-based MARL intersection study of Yu et al. [7], to sustain multiple concurrent pedestrian-SDC interactions per episode.

Pedestrians navigate via Dijkstra shortest-path on a 40-node navigation graph (34 sidewalk waypoints plus 6 crosswalk midpoints). The RL policy outputs a binary go/wait decision. When the policy selects “go,” a personality-driven roll determines whether the pedestrian crosses via a designated crosswalk or jaywalks across the road:

$$P(\text{jaywalk} \mid \text{go}) = \tau_j \times 0.25 \quad (1)$$

where $\tau_j \in [0, 1]$ is the jaywalking tendency, sampled uniformly at episode start and *not observable* by the SDC. This trait-based parameterization of pedestrian heterogeneity is in the same spirit as the risk-taking, cautious, and distracted pedestrian types used in prior social-force simulations of AV-pedestrian interaction [1]. Walking speed ranges from 1.0 to 2.0 m/s.

SDC setting. The SDC uses a kinematic bicycle model [29] with wheelbase 2.5 m, maximum speed 8.33 m/s (30 km/h), and maximum steering angle 0.52 rad. A hard constraint keeps the SDC on the road, applied as a post-physics projection outside the policy and reward: any off-road position is snapped back to within a 0.5 m road margin and the speed is reduced to 30%. An episode terminates on collision (SDC-pedestrian centre distance below 1.5 m), on goal (distance below 3.0 m), or on timeout.

B. MAPPO Training

Algorithm 1 summarizes the training loop. We use clipped PPO objectives with GAE and a shared centralized critic, following the common MAPPO design [3]. The number of updates (5,000) was chosen empirically by monitoring the learning curves of all three networks until they plateaued on held-out seeds; combined with 512 parallel environments and 256-step rollouts, this corresponds to 6.55×10^8 environment steps in total. The centralized critic observes a 58-dim global state (pedestrian positions, speeds, jaywalking tendencies, and SDC state); the pedestrian and SDC actors use 20-dim and 34-dim local observations (own state, traits or goal, surface or lane information, and nearby agents). All 12 pedestrians share actor parameters, as is standard for homogeneous MARL agents [3]. Each actor and the critic is a two-layer multilayer perceptron (MLP) with ReLU activations: 128 units per layer for the pedestrian actor and 256 for the SDC actor and critic. Table II lists the full configuration.

Algorithm 1 MAPPO Co-Training

- 1: Initialize ped actor π_θ^p , SDC actor π_θ^s , shared critic V_ψ
 - 2: **for** update = 1 to 5,000 **do**
 - 3: Collect rollouts across 512 parallel envs for 256 steps
 - 4: **for** each env step **do**
 - 5: Ped obs \leftarrow local (20-dim); SDC obs \leftarrow local (34-dim)
 - 6: Ped actions $\sim \pi_\theta^p$; SDC action $\sim \pi_\theta^s$
 - 7: Step env; record (o, a, r, d)
 - 8: **end for**
 - 9: Compute $V(s)$ using critic with global state (58-dim)
 - 10: Compute GAE advantages ($\gamma=0.995$, $\lambda=0.95$)
 - 11: Update π_θ^p , π_θ^s , V_ψ with clipped PPO ($\epsilon=0.2$, 4 epochs, 8 mini-batches)
 - 12: **end for**
-

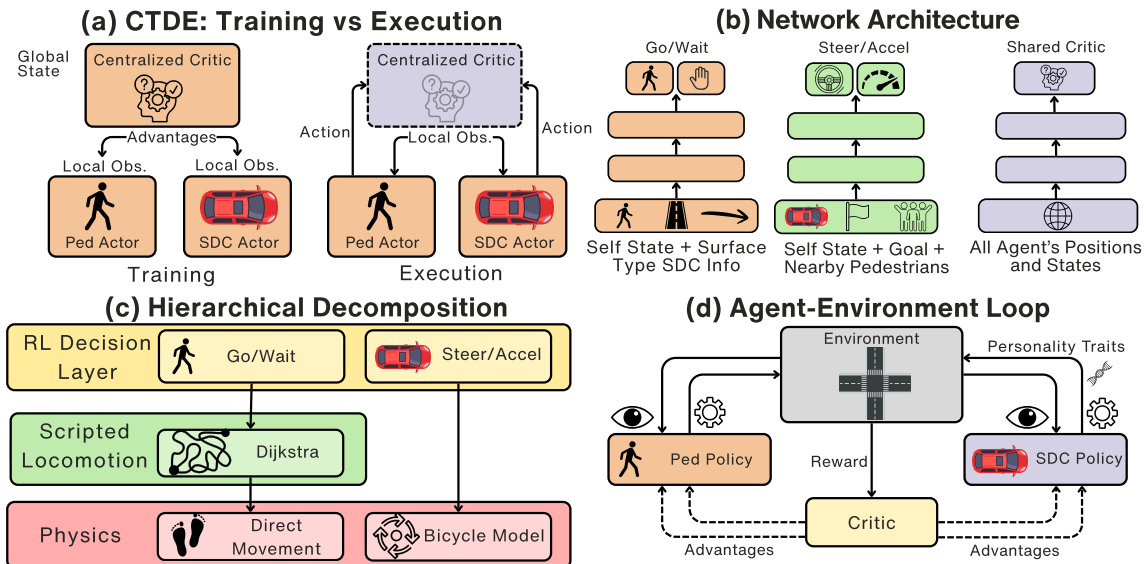


Fig. 2: System architecture. (a) CTDE: the centralized critic uses global state during training and is discarded at execution. (b) Network architecture for pedestrian actor, SDC actor, and shared critic. (c) Hierarchical decomposition: RL controls go/wait and accel/steer; scripted Dijkstra handles locomotion (SDC bypasses this layer); physics handles motion. (d) Agent-environment loop with personality traits feeding into the environment.

TABLE II: System configuration and hyperparameters. CPU: central processing unit; RAM: random-access memory; NN lib: neural-network library; LR: learning rate; FPS: frames per second; Ent: entropy coefficient.

System	Spec	Param	Value
GPU	RTX 5070 Ti 16 GB	Parallel envs	512
CPU	Ryzen 7 9700X	Rollout len	256
RAM	32 GB DDR5	PPO epochs	4
Framework	JAX 0.6.2	Minibatches	8
	+ CUDA 12.6	Updates	5,000
		LR	3×10^{-4}
NN lib	Flax 0.10.7	γ / λ	0.995 / 0.95
Optimizer	Optax 0.2.8	ϵ_{clip}	0.2
Distributions	Distrax	Ent (ped/SDC)	0.03 / 0.01
Training FPS	$\sim 558\text{k}$	Grad norm	0.5
Total steps	655M		

The critic trains on a blended reward (50% mean pedestrian, 50% SDC) for the cooperative objective. Pedestrians are rewarded for waypoint progress ($+2.0 \Delta d$) and reaching waypoints ($+5.0$), penalized for collisions (-25.0), and given a smart-waiting bonus ($+0.3$) near a fast-approaching SDC. The SDC is rewarded for goal progress and arrival ($+50.0$) and penalized for collisions (-50.0), speeding near occupied crosswalks and jaywalkers, off-lane driving, and heading misalignment, with no reward for stopping (preventing indefinite yielding).

C. Uncertainty Quantification

We distinguish two encounter types from the SDC's perspective: *predictable* (crosswalk crossings, anticipatable from crosswalk proximity and trajectory) and *uncertain* (jaywalking, governed by the latent trait τ_j invisible to the SDC). We quantify uncertainty via a **Speed Differential** metric. For each timestep where a pedestrian of type $c \in \{\text{cw}, \text{jw}\}$ is

within distance bin $[d_1, d_2]$ of the SDC, we record the SDC speed $v_{\text{sd}c}$:

$$\bar{v}_c(d_1, d_2) = \frac{1}{|T_c|} \sum_{t \in T_c} v_{\text{sd}c}(t), \quad T_c = \{t : d_t^c \in [d_1, d_2]\} \quad (2)$$

The gap $\Delta v = \bar{v}_{\text{jw}} - \bar{v}_{\text{cw}}$ measures how much faster the SDC travels near jaywalkers than crosswalk users; a positive Δv indicates an unanticipated encounter. We also report **collision attribution**, the fraction of collisions involving each pedestrian type.

IV. EXPERIMENTS

Policies are trained on randomized episodes with personality traits resampled at the start of each episode. All results use 500-episode evaluations on a fixed seed set identical across methods, so every method is compared on the same scenarios. We report goal rate and collision rate, standard safety and success metrics for RL-based autonomous driving evaluation (CARLA uses similar metrics [24]). Scenario seeds in Fig. 1 were selected by fixed programmatic criteria on encounter proximity rather than hand-picked.

A. Baseline Comparison

Fig. 3 compares the MARL SDC against four non-learning baselines and a single-agent RL SDC over 500 episodes. The four non-learning baselines are *random* (uniform acceleration/steering), *constant speed* (fixed throttle, no steering or awareness), *rule-based* (full throttle steering toward the goal), and *rule-based + braking* (rule-based plus braking when a pedestrian is ahead), reaching 11%/25%, 23%/39%, 35%/33%, and 33%/35% (goals/collisions); rule-based full throttle is best, and braking does not help because it cannot steer around pedestrians. The co-trained MARL SDC

reached 78%/14%, and the single-agent RL SDC reached 65%/20%.

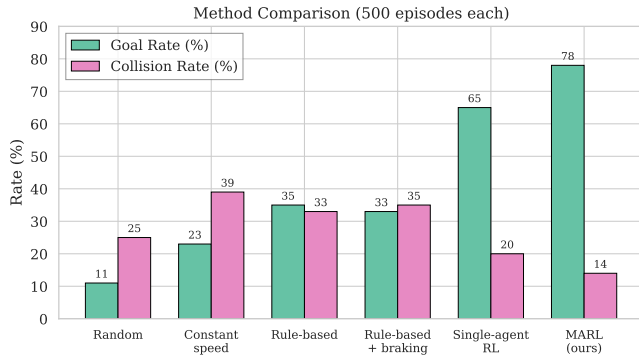


Fig. 3: Goal and collision rates across methods (500 episodes each). Single-agent RL was trained with scripted pedestrians.

B. Uncertainty Results

Fig. 4 shows the speed differential from (2). At 0–3 m, the SDC was 2.65 m/s faster near jaywalkers (8.08 vs. 5.43 m/s), and the gap persists across all distance bins (0.80–1.60 m/s at 3–12 m). Near crosswalk users the SDC had already decelerated, while near jaywalkers it maintained a higher speed, consistent with the crossing being unanticipated.

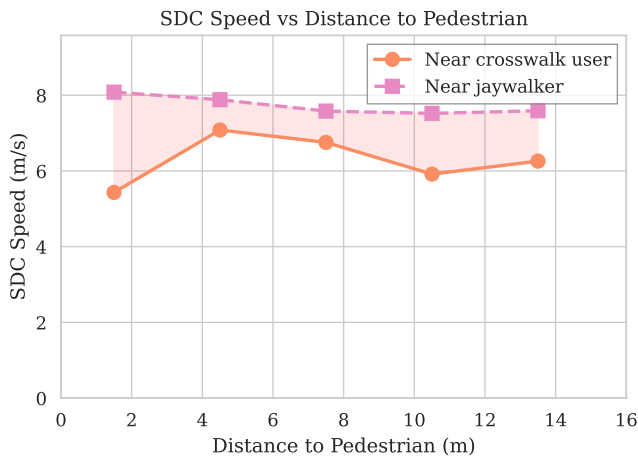


Fig. 4: SDC speed vs. distance to nearest pedestrian, separated by encounter type. The shaded region indicates the speed gap between jaywalker and crosswalk encounters.

Collision attribution. Jaywalking accounted for 13% of crossing events but was associated with 62% of collisions. Two further indicators corroborate the speed differential: the 5th-percentile minimum approach distance was 3.25 m for jaywalker encounters versus 3.49 m for crosswalk encounters, and the mean time from a jaywalker entering the road to the SDC initiating braking was 3.38 s (33.8 steps).

Personality mapping. Jaywalking rate increases monotonically with the trait: Q1 ($\tau_j < 0.25$) = 3.2%, Q2 = 9.4%, Q3 = 15.7%, Q4 ($\tau_j > 0.75$) = 21.6%, consistent with (1).

C. MARL vs. Single-Agent RL

Table III presents a 2×2 comparison. Co-training reduces collisions from 20% to 14%. The pedestrian RL policy learns to wait when the SDC is nearby and moving fast, whereas scripted pedestrians never wait. The co-trained SDC also generalizes well against scripted pedestrians (76% vs. 65% for the single-agent SDC).

TABLE III: MARL vs. Single-Agent RL (500 episodes each)

SDC → Ped Setting	Goal%	Col%
Single-agent → Always go	65	20
Single-agent → MARL peds	65	16
MARL → Always go	76	16
MARL → MARL peds	78	14

D. Ablation: Jaywalking Rate

We scaled the jaywalking probability via the 0.25 multiplier in Eq. (1), giving effective jaywalking rates of 0%, 5%, 13% (default), 30%, and 50%, with goal/collision rates of 77%/15%, 77%/15%, 76%/16%, 73%/18%, and 64%/28% respectively. Performance degrades gracefully then sharply: the collision rate rises only from 15% to 18% between 0% and 30% jaywalking, then jumps to 28% at 50%, indicating the SDC tolerates a realistic amount of unpredictable crossings but degrades non-linearly once a large fraction of pedestrians bypass the crosswalks, consistent with jaywalking acting as a tunable source of aleatoric uncertainty.

V. DISCUSSION AND CONCLUSION

The speed differential metric gives a trajectory-level measure of uncertainty that complements human-perception-based SDC assessment [2]. The realism of the pedestrian model shapes which situations an SDC faces: a scripted-pedestrian testbed underrepresents jaywalking and its surprise (62% of collisions from 13% of crossings), whereas co-trained, trait-driven pedestrians expose the SDC to more latent-intent situations and yield a metric computable for any policy under test. Co-training is a tool for generating and stress-testing realistic scenarios, not a claim that real pedestrians optimize a joint objective; likewise, the 14% collision rate is a relative-improvement result under deliberately adversarial jaywalking, not an absolute safety guarantee, and a deployment would add explicit safety layers such as control barrier functions [30].

Our evaluation is limited to a single synthetic map and a fixed seed set, so generalization to unseen layouts and traffic densities is the primary next step. Future work includes cross-map and varied-density evaluation, comparison against stronger planning- and risk-aware baselines, validation of the speed-differential metric against real-world data, a speed-based collision criterion, per-agent-type critics, human-in-the-loop scenario validation [2], and scaling to larger heterogeneous traffic.

REFERENCES

- [1] M. M. Rashid, M. Seyedi, and S. Jung, "Simulation of pedestrian interaction with autonomous vehicles via social force model," *Simulation Modelling Practice and Theory*, vol. 132, p. 102901, Apr. 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1569190X24000157>
- [2] C. Birchler, T. K. Mohammed, P. Rani, T. Nechita, T. Kehrler, and S. Panichella, "How Does Simulation-Based Testing for Self-Driving Cars Match Human Perception?" *Replication Package - "How does Simulation-based Testing for Self-driving Cars match Human Perception?"*, vol. 1, no. FSE, pp. 42:929–42:950, Jul. 2024. [Online]. Available: <https://dl.acm.org/doi/10.1145/3643768>
- [3] C. Yu, A. Velu, E. Vinitzky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The surprising effectiveness of PPO in cooperative multi-agent games," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, ser. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., Nov. 2022, pp. 24 611–24 624.
- [4] Z. Zhang, H. Li, T. Chen, N. N. Sze, W. Yang, Y. Zhang, and G. Ren, "Decision-making of autonomous vehicles in interactions with jaywalkers: A risk-aware deep reinforcement learning approach," *Accident Analysis & Prevention*, vol. 210, p. 107843, Feb. 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0001457524003889>
- [5] D. Helbing and P. Molnar, "Social Force Model for Pedestrian Dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, May 1995, arXiv:cond-mat/9805244. [Online]. Available: <http://arxiv.org/abs/cond-mat/9805244>
- [6] X. Dong, E. Guerra, and R. A. Daziano, "Will automated vehicles encourage more jaywalking? Results from a stated preference survey," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 103, pp. 217–229, May 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1369847824000858>
- [7] T. Yu, K. Wang, Z. Li, T. Yu, and K. Sakaguchi, "Multi-Agent Reinforcement Learning-based Cooperative Autonomous Driving in Smart Intersections," May 2025, arXiv:2505.04231 [cs]. [Online]. Available: <http://arxiv.org/abs/2505.04231>
- [8] R. Zhang, J. Hou, F. Walter, S. Gu, J. Guan, F. Röhrbein, Y. Du, P. Cai, G. Chen, and A. Knoll, "Multi-Agent Reinforcement Learning for Autonomous Driving: A Survey," Aug. 2024, arXiv:2408.09675 [cs]. [Online]. Available: <http://arxiv.org/abs/2408.09675>
- [9] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 6382–6393. [Online]. Available: <https://dl.acm.org/doi/10.5555/3295222.3295385>
- [10] T. Rashid, M. Samvelyan, C. S. De Witt, G. Farquhar, J. Foerster, and S. Whiteson, "Monotonic value function factorisation for deep multi-agent reinforcement learning," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 178:7234–178:7284, Jan. 2020. [Online]. Available: <https://dl.acm.org/doi/10.5555/3455716.3455894>
- [11] J. N. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson, "Counterfactual multi-agent policy gradients," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'18/IAAI'18/EAAI'18. New Orleans, Louisiana, USA: AAAI Press, Feb. 2018, pp. 2974–2982. [Online]. Available: <https://dl.acm.org/doi/10.5555/3504035.3504398>
- [12] C. S. d. Witt, T. Gupta, D. Makoviichuk, V. Makoviychuk, P. H. S. Torr, M. Sun, and S. Whiteson, "Is Independent Learning All You Need in the StarCraft Multi-Agent Challenge?" Nov. 2020, arXiv:2011.09533 [cs]. [Online]. Available: <http://arxiv.org/abs/2011.09533>
- [13] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Aug. 2017, arXiv:1707.06347 [cs]. [Online]. Available: <http://arxiv.org/abs/1707.06347>
- [14] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-Dimensional Continuous Control Using Generalized Advantage Estimation," Oct. 2018, arXiv:1506.02438 [cs]. [Online]. Available: <http://arxiv.org/abs/1506.02438>
- [15] F. A. Oliehoek and C. Amato, *A Concise Introduction to Decentralized POMDPs*, ser. Springer Briefs in Intelligent Systems. Cham: Springer International Publishing, 2016. [Online]. Available: <http://link.springer.com/10.1007/978-3-319-28929-8>
- [16] Z. Xue, Q. Dong, X. Fan, Q. Jin, H. Jian, and J. Liu, "Fuzzy Logic-Based Model That Incorporates Personality Traits for Heterogeneous Pedestrians," *Symmetry*, vol. 9, no. 10, p. 239, Oct. 2017. [Online]. Available: <https://www.mdpi.com/2073-8994/9/10/239>
- [17] Y. Wang, A. R. Srinivasan, Y. M. Lee, and G. Markkula, "Modeling Pedestrian Crossing Behavior: A Reinforcement Learning Approach with Sensory Motor Constraints," Sep. 2024, arXiv:2409.14522 [cs]. [Online]. Available: <http://arxiv.org/abs/2409.14522>
- [18] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," 2016, pp. 961–971. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/Alahi_Social_LSTM_Human_CVPR_2016_paper.html
- [19] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data," Jan. 2021, arXiv:2001.03093 [cs]. [Online]. Available: <http://arxiv.org/abs/2001.03093>
- [20] E. A. Khuzam, G. Lanzaro, and T. Sayed, "Impact of jaywalking on pedestrian interaction behavior: A multiagent Markov Game-based analysis," *Accident Analysis & Prevention*, vol. 220, p. 108141, Sep. 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0001457525002271>
- [21] A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" Oct. 2017, arXiv:1703.04977 [cs]. [Online]. Available: <http://arxiv.org/abs/1703.04977>
- [22] C.-J. Hoel, K. Wolff, and L. Laine, "Tactical Decision-Making in Autonomous Driving by Reinforcement Learning with Uncertainty Estimation," Apr. 2020, arXiv:2004.10439 [cs]. [Online]. Available: <http://arxiv.org/abs/2004.10439>
- [23] K. Wang, C. Shen, X. Li, and J. Lu, "Uncertainty Quantification for Safe and Reliable Autonomous Vehicles: A Review of Methods and Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 3, pp. 2880–2896, Mar. 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/10879299>
- [24] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," Nov. 2017, arXiv:1711.03938 [cs]. [Online]. Available: <http://arxiv.org/abs/1711.03938>
- [25] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wiebner, "Microscopic Traffic Simulation using SUMO," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui, HI, USA: IEEE Press, Nov. 2018, pp. 2575–2582. [Online]. Available: <https://doi.org/10.1109/ITSC.2018.8569938>
- [26] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, Y. Katariya, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: composable transformations of Python+NumPy programs," 2018. [Online]. Available: <http://github.com/google/jax>
- [27] A. Rutherford, B. Ellis, M. Gallici, J. Cook, A. Lupu, G. Ingvarsson, T. Willi, R. Hammond, A. Khan, C. S. d. Witt, A. Souly, S. Bandyopadhyay, M. Samvelyan, M. Jiang, R. T. Lange, S. Whiteson, B. Lacerda, N. Hawes, T. Rocktaschel, C. Lu, and J. N. Foerster, "JaxMARL: Multi-Agent RL Environments and Algorithms in JAX," Nov. 2024, arXiv:2311.10090 [cs]. [Online]. Available: <http://arxiv.org/abs/2311.10090>
- [28] S. Huang, R. F. J. Dossa, C. Ye, J. Braga, D. Chakraborty, K. Mehta, and J. G. M. Araújo, "CleanRL: High-quality Single-file Implementations of Deep Reinforcement Learning Algorithms," *Journal of Machine Learning Research*, vol. 23, no. 274, pp. 1–18, 2022. [Online]. Available: <http://jmlr.org/papers/v23/21-1342.html>
- [29] P. Polack, F. Alché, B. d'Andréa Novel, and A. de La Fortelle, "The kinematic bicycle model: A consistent model for planning feasible trajectories for autonomous vehicles?" in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2017, pp. 812–818.
- [30] P. Aryan and S. Panichella, "Physics-informed machine learning for precision unmanned aerial vehicle control: Adaptive transformers with safety guarantees," *Engineering Applications of Artificial Intelligence*, vol. 172, p. 114379, 2026.