RL for Autonomous Driving

Group 13: YYDS

Ang Li[1] Maggie Yao[2] Weijie Zhu[3] Spiro Li[4]

[1]Department of ECE anguoft.li@mail.utoronto.ca

[2]Department of ECE maggieyyx.yao@mail.utoronto.ca

[3]Department of ECE weijie.zhu@mail.utoronto.ca

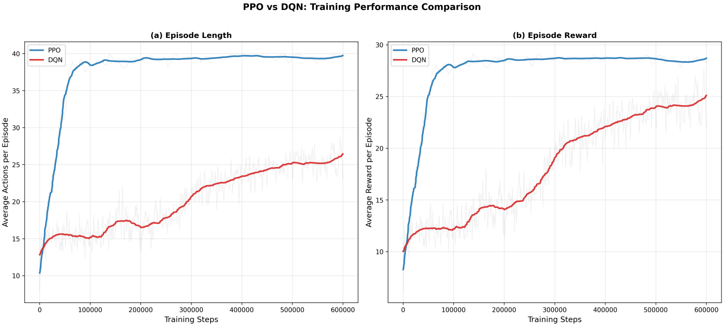
[4]Department of ECE supeng.li@mail.utoronto.ca

Reinforcement Learning – Fall 2025

Agenda

- Default Highway Environment
  - OPD — Optimal Policy Deterministic (Rule-Based Policy)
  - DQN — Deep Q-Network
  - PPO — Proximal Policy Optimization
- Finetune DQN
  - Buffer Size vs Gamma vs Learning Rate
- Modified Highway Environment
  - Vehicle Density 1.0 → 1.25
- Speed Weights Experiment with PPO
  - Speed Weight 0.4 → 0.6 → 0.8

Default Highway Environment

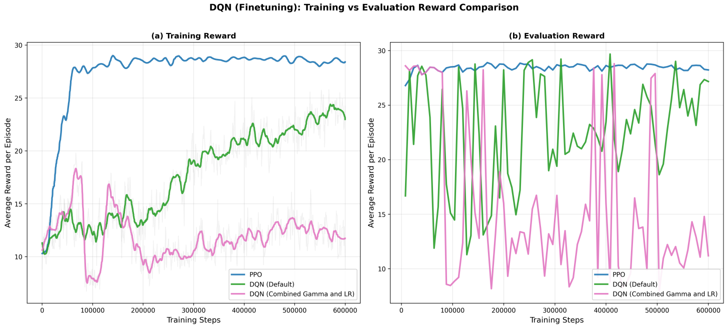


Model Variant	Collision Rate *	Avg. Reward *	Avg. Speed (m/s) *	Avg. Inference Time (ms) **	Training Time (h) ***
OPD (Rule-based)	2%	39.3	29.79	721	-
PPO	2%	28.36	20.16	0.16	15.83
DQN	6%	28.67	21.23	0.18	17.76

*: Averaged across 100 episodes. **: Per action step. ***: Measured on MacBook Air with Apple M4.

Table: OPD, DQN, and PPO Testing Metrics: Comparison Across Variants

Finetune DQN



Model Variant (Finetune)	Collision Rate *	Avg. Reward *	Avg. Speed (m/s) *	Avg. Inference Time (ms) **	Training Time (h) ***
PPO (Baseline)	2%	28.36	20.16	0.30	27.26
DQN (**Default**) †	6%	28.64	21.23	0.18	26.46
DQN (Buffer 25k)	62%	28.26	28.38	0.18	26.38
DQN (Buffer 100k)	34%	29.01	25.41	0.17	27.27
DQN (Gamma 0.99)	3%	28.82	20.65	0.19	27.42
DQN (Linear LR Schedule)	4%	28.75	20.66	0.10	27.44
DQN (Gamma and LR)	0%	29.04	20.41	0.10	27.46

†: Default hyperparameters: Replay buffer size of 50,000; Discount factor (γ) of 0.95; Constant learning rate of 5×10^{-4}.
*: Averaged across 100 episodes. **: Per action step. ***: Measured on MacBook Air with Apple M3.

Table: DQN Fine-Tuning Testing Metrics: Comparison Across Variants

Modified Highway Environment

Varied the vehicle density from 1.0 to 1.25 to test our models' scalability.

Model Variant	Collision Rate*	Avg. Reward*	Avg. Speed (m/s)*
OPD (1.0)	0.00%	39.57	29.63
OPD (1.25)	0.00%	39.05	29.18
DQN (1.0→1.0)**	6%	28.64	21.23
DQN (1.0→1.25)**	19.00%	26.91	21.61
DQN (1.25→1.25)**	2.00%	29.39	21.07
PPO (1.0→1.0)**	2%	28.36	20.16
PPO (1.0→1.25)**	16.00%	26.40	19.92
PPO (1.25→1.25)**	5.00%	28.02	20.02

**(x→y): Trained at vehicle density of x, tested at vehicle density of y.

Speed Weights Experiment

Model Variant	Collision Rate*	Avg. Reward*	Avg. Speed (m/s)*
PPO (Speed Weight: 0.4)	2.00%	28.36	20.16
PPO (Speed Weight: 0.6)	2.00%	24.92	20.16
PPO (Speed Weight: 0.8)	4.00%	21.99	20.22

* Averaged across 100 episodes.

UNIVERSITY OF TORONTO