EVALUATING ROBUSTNESS OF DEEP REINFORCEMENT LEARNING FOR AUTONOMOUS DRIVING

By **Mehmet Alp Sozuduz** Supervisors: Matthijs Spaan, Moritz Zanger

msozuduz@student.tudelft.nl

TU Delft

RESEARCH QUESTION – 1

How does the **discretization of the action space** and various Deep Q Network (DQN) **extensions** affect algorithm training and the robustness?

BACKGROUND INFORMATION – 2

Deep Q Network - deep RL algorithm, Q-value estimation, Temporal difference loss, experience replay

Action Space Discretization - optimum number of values to represent an infinite amount of actions

DQN Extensions - improvements proposed to DQN

- Double Q-Learning: separate action selection and evaluation to reduce overestimation bias [1]
- Prioritized Experience Replay: prioritize unique experience (high loss) when sampling experiences [2]

METHODOLOGY – 3

Carla - a highly realistic as well as customizable traffic simulator

DQN - Implementation exists in the literature [3] as well as cleanRl

Train for 500k steps on Delft Blue to check **training**.

Check 3 different simulation maps for **robustness**.

Action Space Discretization:

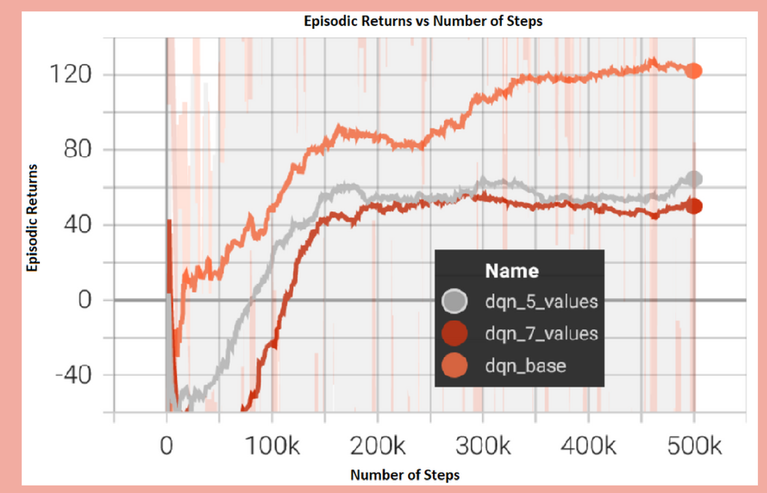
- Change steering angle and acceleration values.
- Use three agents with 3, 5, and 7 values for each property.

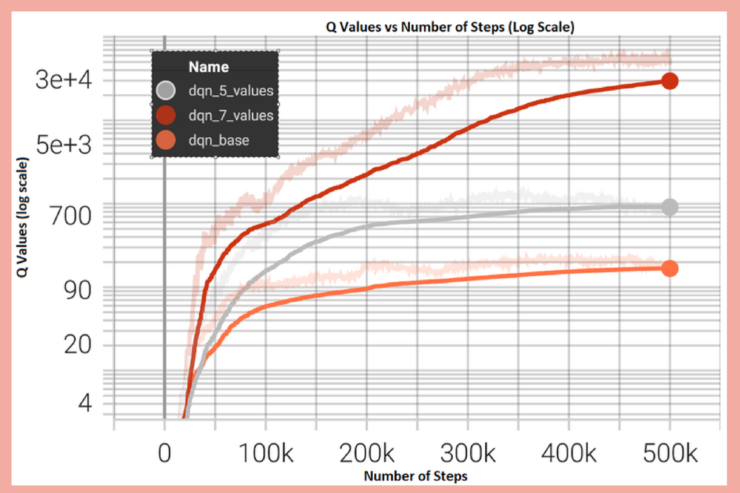
DQN Extensions

- Use three agents: default, with Double Q-Learning, with Prioritized Experience Replay

ACTION SPACE DISCRETIZATION – 4A

Training Performance





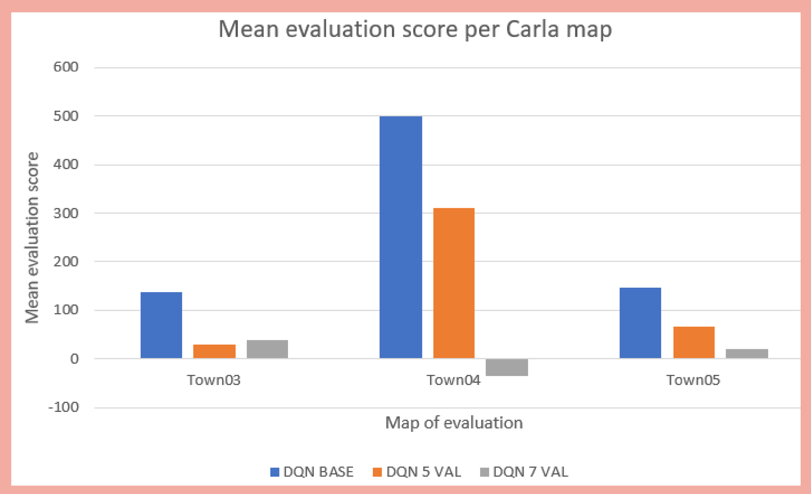
Higher Q-values as discretization increases.

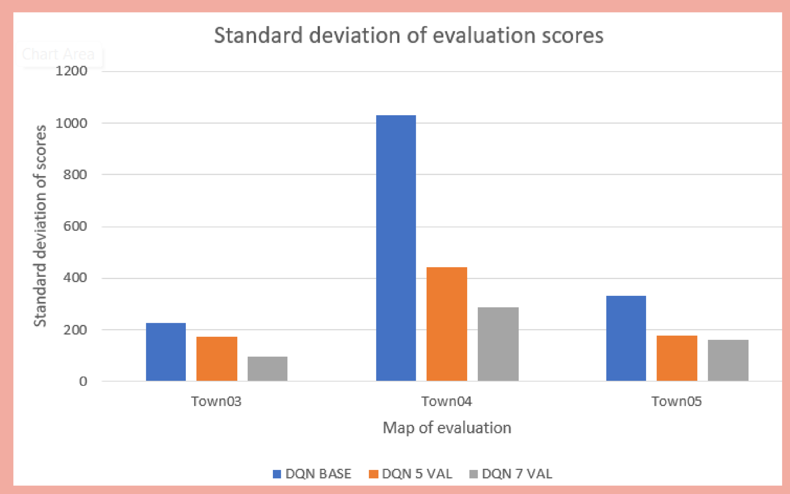
Worse training performance as discretization increases.

Robustness Testing

Worse mean evaluation scores for higher discretization

Lower standard deviation seen for higher discretization

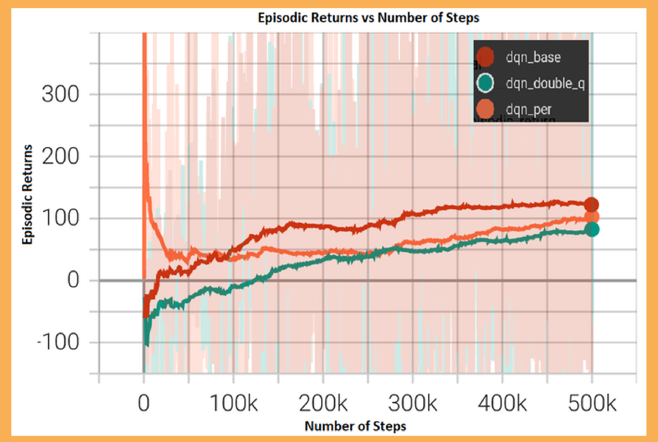




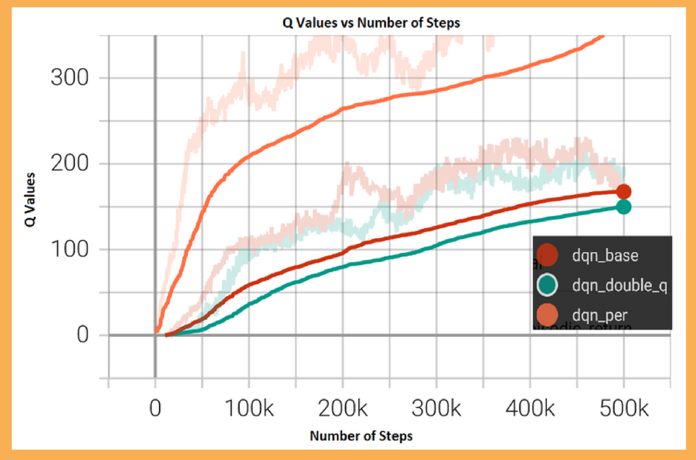
DQN EXTENSIONS – 4B

Training Performance

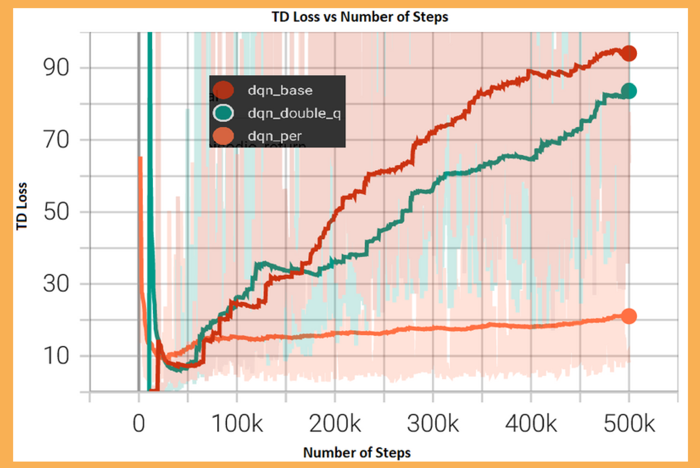
Slightly worse training performance



Double Q lower performance, Q-values, as well as TD loss



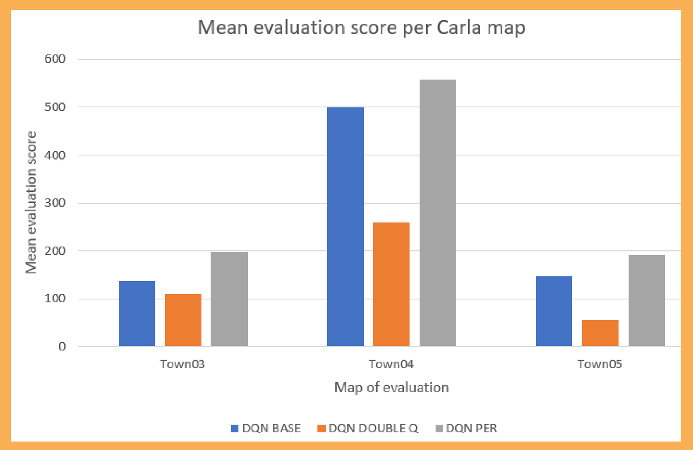
PER lower performance, high Q-values, really low TD loss

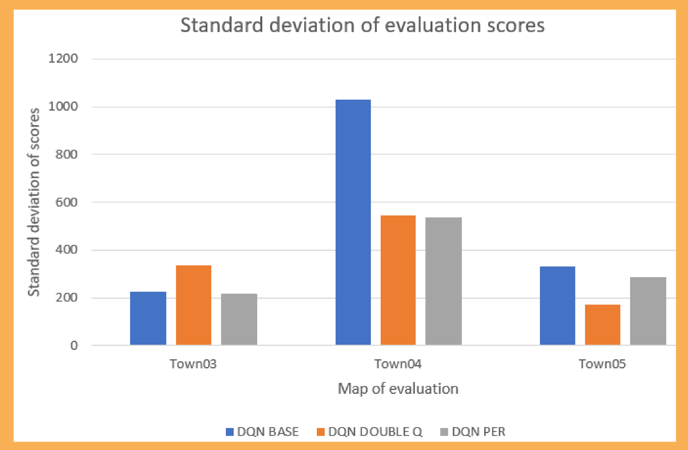


Robustness Testing

Double Q, lower mean evaluation scores.

PER, higher mean evaluation scores performance.





CONCLUSIONS – 5

Higher discretization, leads to high Q-value estimations and potentially divergence.

Double Q slightly worsens training performance as well as robustness.

PER extension, decreases training performance, but increases robustness.

LIMITATIONS– 6

Resource/time limitations lead to hardship in hyperparameter tuning.

Some graphs still show an upwards trend, more training could allow them to find a better point for convergence.

RELATED LITERATURE

[1] Hado van Hasselt, Arthur Guez, and David Silver. "Deep reinforcement learning with Double Q-learning". In: (Dec. 2015). arXiv: 1509.06461 [cs.LG].
[2] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. November 2015..
[3] Volodymyr Mnih et al. "Playing Atari with deep reinforcement learning". In: (Dec. 2013). arXiv: 1312.5602 [cs.LG].
[4] Yunhao Tang and Shipra Agrawal. "Discretizing Continuous Action Space for On-Policy Optimization". In: (Jan. 2019). arXiv: 1901.10500 [cs.LG].