EVALUATING THE ROBUSTNESS OF DQN AND QR-DQN IN TRAFFIC SIMULATION:

ANALYZING THE EFFECT OF QUANTILE MANIPULATION AND ENVIRONMENTAL VARIABILITY

AUTHOR Cristian Flaviu Toadere
c.f.toadere@student.tudelft.nl
SUPERVISORS Dr. Mustafa Celikok, Dr. Frans Oliehoek
EXAMINER Dr. Annibale Panichella

TU Delft

I. INTRODUCTION

- Recent advances in autonomous driving require reliable and robust machine learning algorithms.
- Deep Q-Network (DQN) [1] works well with discrete action spaces, but suffers from overestimation bias and out-of-distribution performance.
- Quantile Regression Deep Q-Network (QR-DQN) [2] improves on DQN by estimating quantiles of the value distribution for better return prediction.
- The effect of utilising QR-DQN's quantile range when predicting actions has not been sufficiently studied.

II. RESEARCH QUESTIONS

- *Does utilizing only **parts** of **QR-DQN's quantiles** determine the model to employ a **conservative approach** that **improves** its **performance**?*
- *How does the **robustness** of **DQN** and **QR-DQN compare** when evaluated across progressively varying **traffic environments** that differ from the training setting?*

III. METHODOLOGY

- **Models:** taken from Stable Baselines 3 library, having consistent hyperparameters and 5 seeds for each.
- **Proposed model:** modifies standard QR-DQN by using only lower quantiles to guide conservative decision-making.
  - Added 'quantile_fraction' parameter to control the quantile range used when predicting actions (values 0.1 and 0.4);
  - Referred to as Risk-Averse QR-DQN (RA QR-DQN).
- **Environment**: HighwayEnv's highway scenario with configurable lane count, traffic density, and vehicle behavior.
- **Experiment**: trained each model in the same fashion and tested them in five varied environments for 1000 epsiodes.
- **Metrics**: average reward and collision rate.

IV. MODEL TRAINING

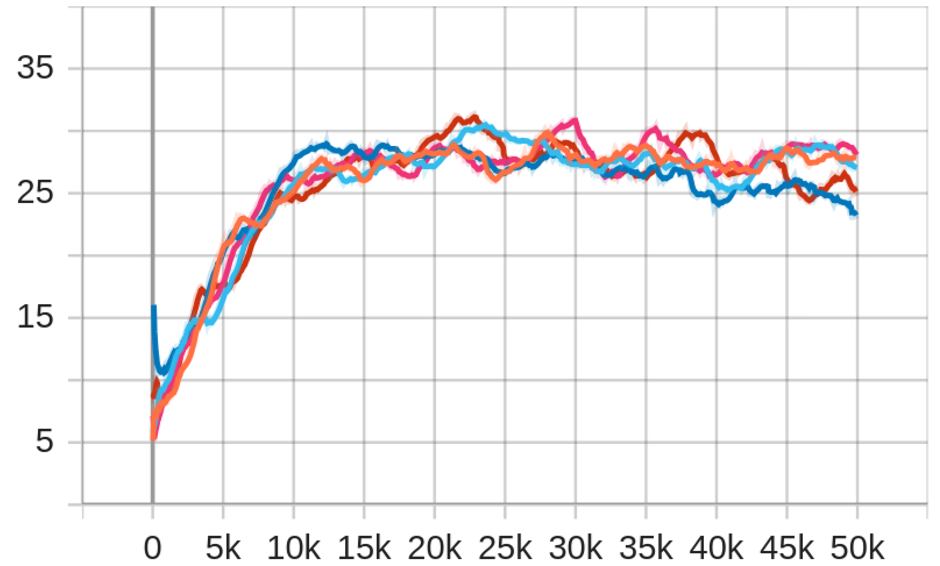

Fig. 1: DQN training graph for 5 seeds

DQN (Fig. 1)
- slower learning
- lower final reward (avg. ~26.3)
- signs of instability in some seeds.

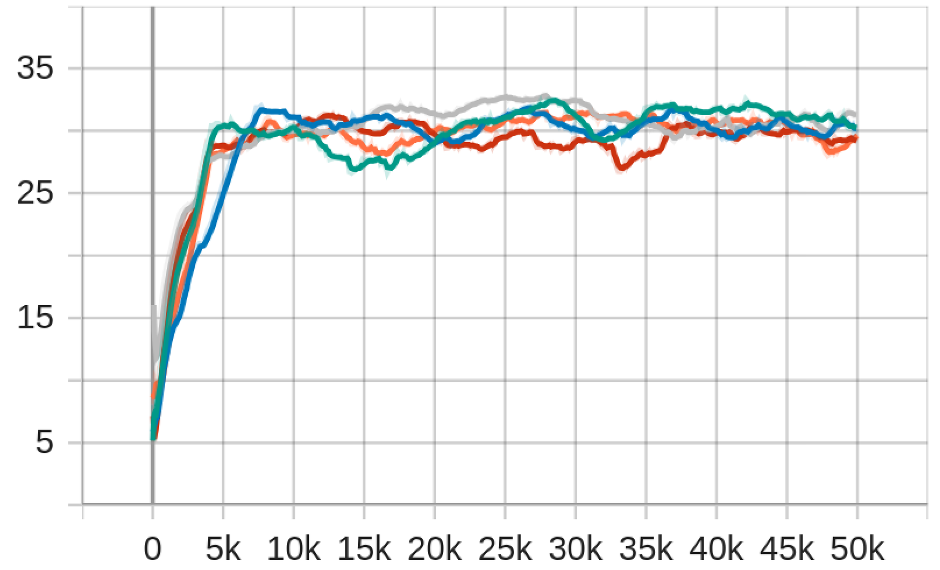

Fig. 2: QR-DQN training graph for 5 seeds

QR-DQN (Fig. 2)
- learned faster
- reached higher reward average (~30.1)
- shows more stable performance to DQN

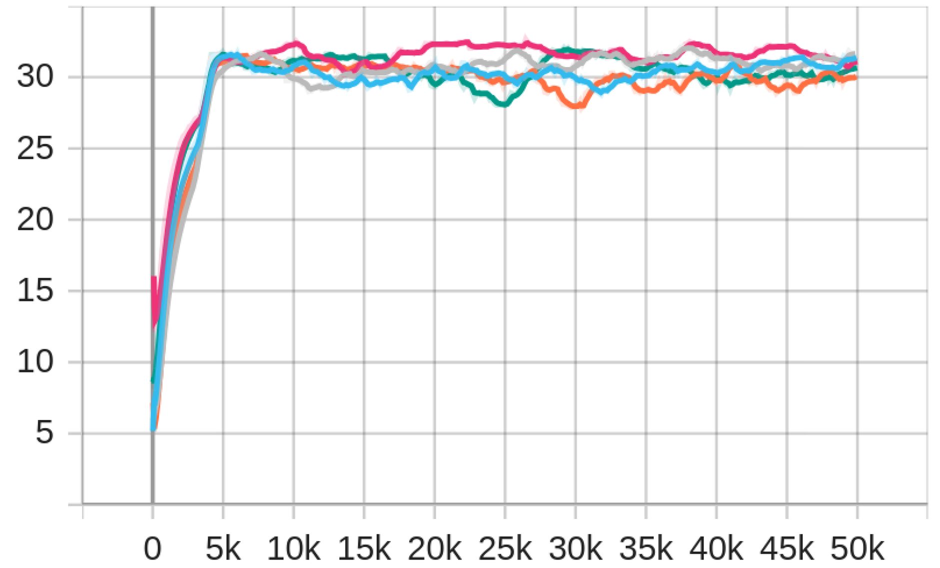

Fig. 3: RA QR-DQN 0.4 training graph for 5 seeds

RA QR-DQN 0.4 (Fig. 3)
- outperforms standard QR-DQN
- average reward (~31.0)
- even more stable in learning

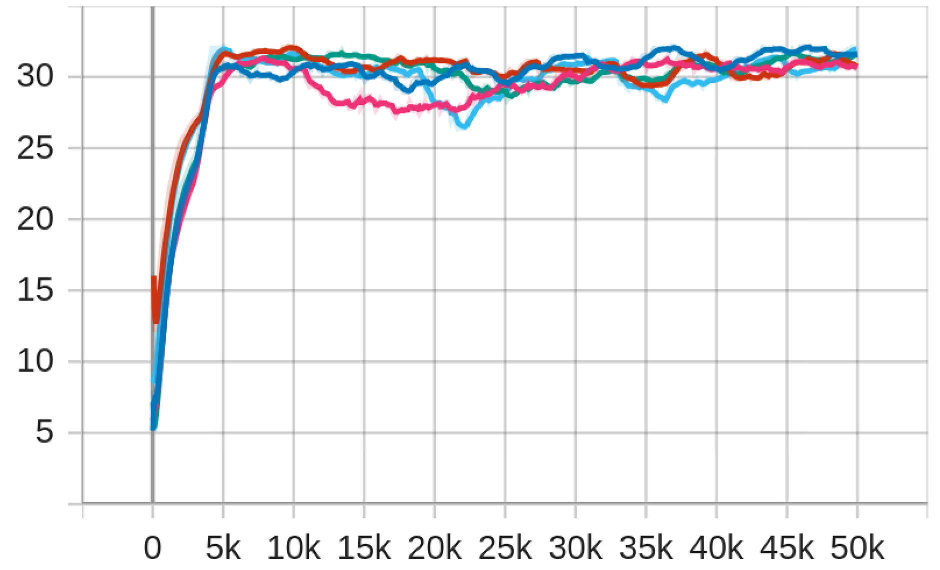

Fig. 4: RA QR-DQN 0.1 training graph for 5 seeds

RA QR-DQN 0.1 (Fig. 4)
- best performance
- average reward (~31.3)
- fewer quantile usage encouraged stable and more efficient learning

V. RESULTS

Table 1: Mean return ± standard error, over 5 seeds in each test environment

Model	3 lanes	4 lanes	6 lanes	aggressive	traffic
DQN	30.392 ± 1.010	29.959 ± 0.953	29.339 ± 0.850	29.050 ± 0.971	17.099 ± 0.603
QR-DQN	31.162 ± 0.616	32.637 ± 0.343	32.268 ± 0.392	28.320 ± 0.711	15.647 ± 0.971
RA QR-DQN 0.4	31.400 ± 0.259	31.664 ± 0.207	31.097 ± 0.435	29.606 ± 0.392	15.493 ± 0.791
RA QR-DQN 0.1	31.696 ± 0.165	32.370 ± 0.134	32.057 ± 0.114	28.819 ± 0.420	16.131 ± 0.297

Table 2: Mean collision rate (%) ± standard error, over 5 seeds in each test environment

Model	3 lanes	4 lanes	6 lanes	aggressive	traffic
DQN	20.36 ± 5.914	20.78 ± 5.351	20.8 ± 4.727	24.48 ± 5.272	86.8 ± 1.499
QR-DQN	15.42 ± 3.159	5.22 ± 1.501	5.6 ± 1.689	25.04 ± 3.222	86.8 ± 2.946
RA QR-DQN 0.4	5.72 ± 1.673	4.44 ± 1.902	6.62 ± 3.088	13.04 ± 2.199	94.58 ± 1.579
RA QR-DQN 0.1	8.68 ± 1.399	2.74 ± 0.603	2.6 ± 0.252	20.14 ± 2.379	83.88 ± 0.833

VI. CONCLUSION

- **DQN** presents its limitations, **dropping in performance** in different environments **due to overestimation bias**.
- **QR-DQN**'s **quantile** utlization showed **better adaptability** to new environments, achieving **better results** in most cases.
- **RA QR-DQN** further **reduces collision** rates by using a **risk-sensitive approach** in quantile selection which employs a conservative behaviour, with **minor reward trade-off**.

Limitations and Future Work

- **Results** are **limited** to the highway scenario, they **cannot be generalised** for other scenarios.
- Model **performance constrained** by current configuration and **training** scope, with **limited** RA QR-DQN **models considered**.
- **Further study** into **dynamic quantile range** selection for **more adaptive**, context-aware agents.
- **Optimize** HighwayEnv **reward function** for **better balance** between exploration and safety.

REFERENCES

[1] B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, et al. Deep Reinforcement Learning for Autonomous Driving: A Survey. IEEE Transactions on Intelligent Transportation Systems, 23(6):4909–4926, June 2022.
[2] Hado van Hasselt, Arthur Guez, and David Silver. Deep Reinforcement Learning with Double Q-Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 30(1), March 2016. Number: 1.