Comparative Analysis of Exploration Algorithms in Deep Reinforcement Learning for Autonomous Driving

By Efe Sözen | Supervisors: Moritz Zanger, Matthijs Spaan | E.Sozen@student.tudelft.nl

Research Question - 1

- How does epsilon-greedy, random network distillation, bootstrapped DQN affect **training** and the **robustness** of final policies under various testing conditions in autonomous driving?

Background Information - 2

- Deep Q-Networks [2]
- Exploration vs Exploitation
- Exploration
  - Epsilon-Greedy
  - Random Network Distillation (RND): target network, predictor network, intrinsic reward [1]
  - Bootstrapped DQN (BDQN): value heads, bootstrapping, Thompson sampling [3]
- Environments: CARLA, CarRacing

Methodology - 3

- Implementations
  - E-Greedy: epsilon decayed over time
  - RND: DQN instead of PPO, episodic instead of non-episodic, observation normalization
  - BDQN: Masking distribution, Thompson sampling are not used
- Training on CarRacing, CARLA
- Evaluating robustness on different CARLA maps

Figure 4: Image provided by gym-carla [4]

Results - 4

Training (Fig. 1):

- Similar time needed to train
- BDQN had higher episodic returns
- RND performed worse than expected - better on CarRacing

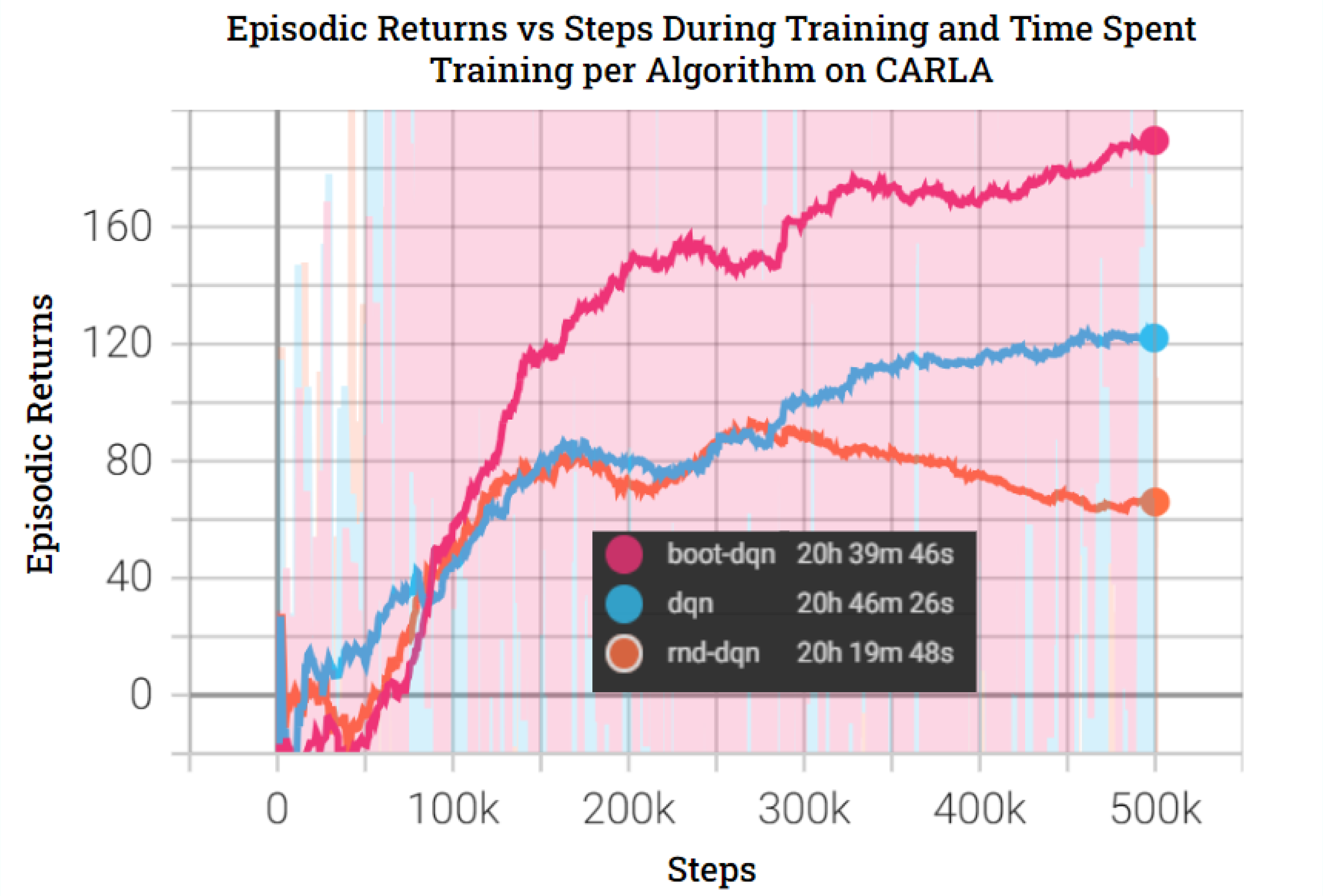

Figure 1: Episodic Returns vs Steps During Training

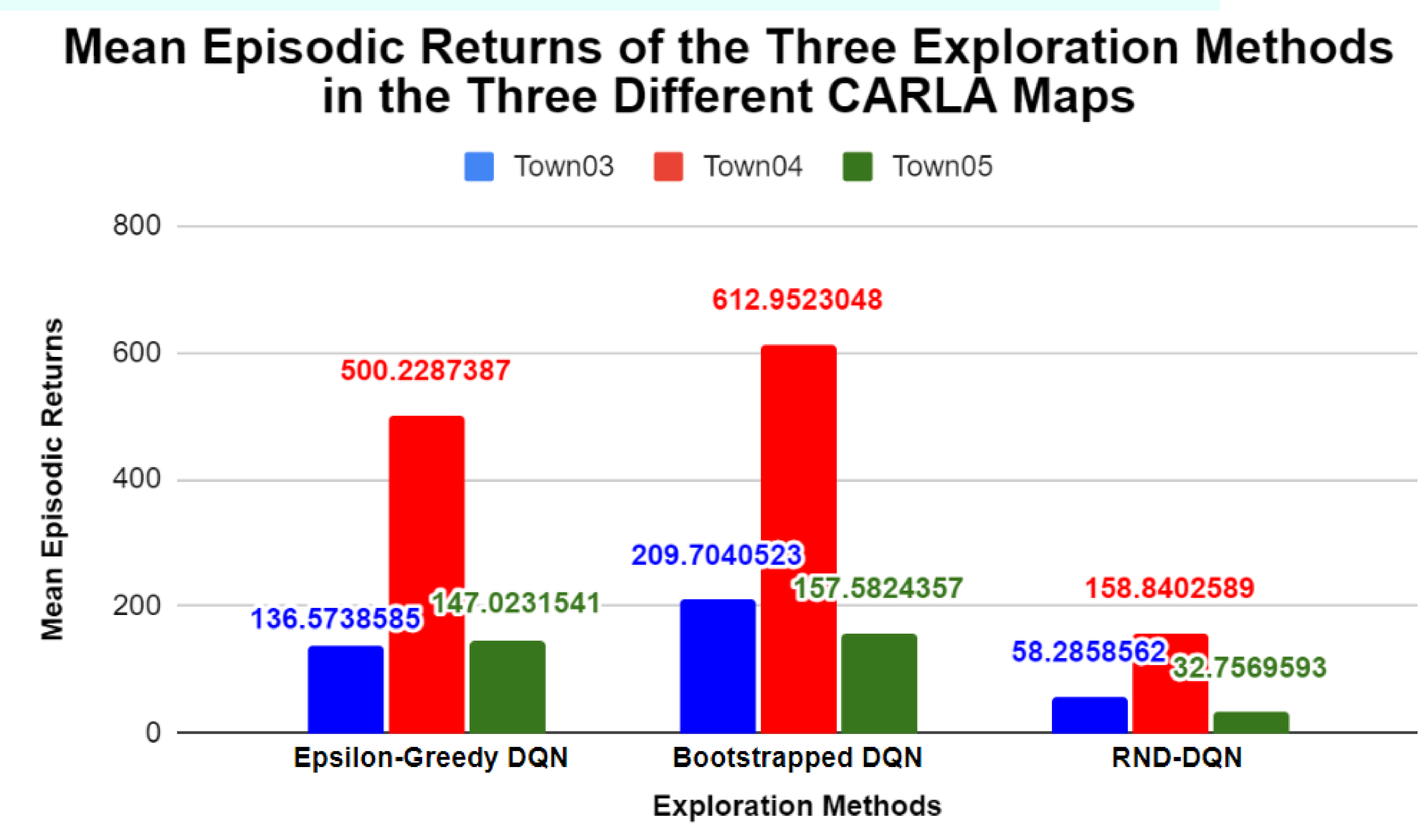

Figure 2: Mean Episodic Returns

Evaluation:

- BDQN with highest returns on all maps (Fig. 2)
- RND with lowest returns (Fig. 2) and highest standard deviation (Fig. 3) on episodic returns
- Town04 highest returns (Fig. 2) and highest standard deviation (Fig. 3)
- RND with inferior ability to learn
- BDQN outperformed E-Greedy by: 55%, 22% and 7% on Town03, Town04 and Town05 respectively (Fig. 2)

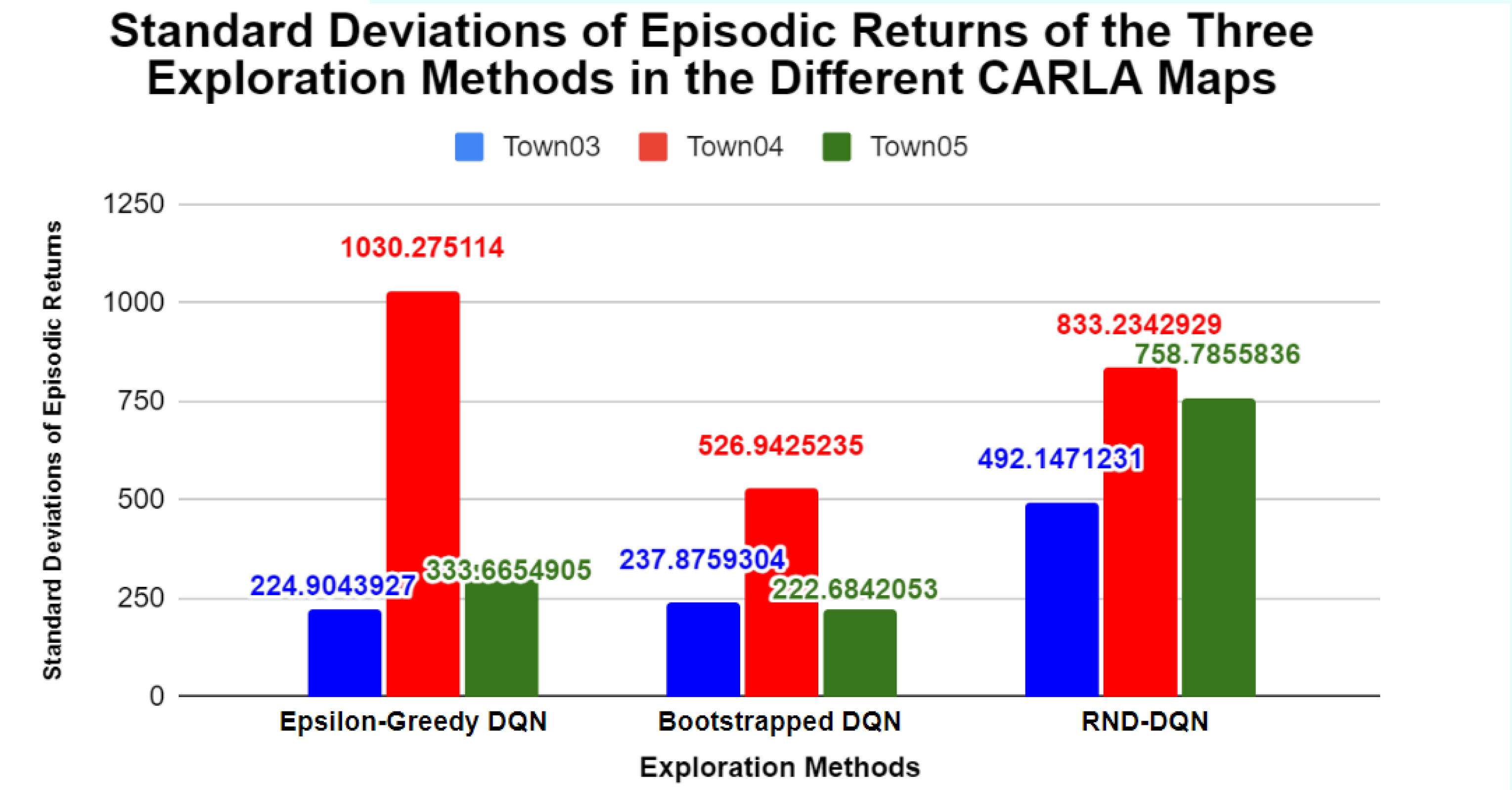

Figure 3: Standard Deviation of Episodic Returns

Limitations - 5

- Hyperparameter tuning
- Training on different maps
- Training with different input (Lidar, camera)
- Evaluating on even more maps
- Allowing to train with more steps
- Implementation differences

Conclusion - 6

- BDQN clearly outperformed E-Greedy
- RND had issues learning on CARLA
  - Implementation differences
  - Limitations of experiment
- Train with more steps and optimize hyperparameters
- NoisyNets or Diversity-Driven Exploration

References

[1] Yuri Burda et al. Exploration by Random Network Distillation. 2018. arXiv: 1810.12894 [cs.LG].
[2] Volodymyr Mnih et al. "Playing Atari with deep reinforcement learning". In: (Dec. 2013). arXiv: 1312.5602 [cs.LG].
[3] Ian Osband et al. "Deep Exploration via Bootstrapped DQN". In: (Feb. 2016). arXiv: 1602.04621 [cs.LG].
[4] Jianyu Chen. gym-carla. https://github.com/cjy1992/gym-carla, 2020.