Autonomous Lane Following in CARLA Using Deep Q-Networks

Charles Gu, Matthew Xiao, Jarvis Zou
Department of Computer Science
Northeastern University
gu.qiu@northeastern.edu

***Abstract*—Autonomous driving presents a challenging sequential decision-making problem with long-term safety and stability requirements. In this work, we study a simplified autonomous lane-following task in the CARLA simulator using a Deep Q-Network (DQN) agent operating on a compact continuous state representation and a discrete steering action space. We investigate both online reinforcement learning and a hybrid offline–online approach that incorporates expert trajectories from the PDM_Lite_Carla_LB2 dataset via replay buffer prefill. While our baseline DQN learns stable lane-following behavior, we observe that naïvely integrating offline expert data leads to significant performance degradation due to distribution mismatch and bootstrapping error. Our results highlight both the effectiveness and the limitations of value-based reinforcement learning for autonomous driving control.**

I. Introduction and Motivation

Driving is a fundamental capability in modern society, yet it is a skill that demands significant time, training, and access to resources. The process of obtaining a driving license is often lengthy and stressful, and serves as a gatekeeper to essential freedoms such as access to employment, healthcare, education, and social participation. While public transportation partially mitigates these limitations, its coverage is incomplete, and long-term reliance on hired drivers remains economically impractical. Autonomous driving therefore represents a transformative technological opportunity to reduce the burden of human driving, improve safety, and increase societal productivity.

Reinforcement learning (RL) offers a compelling framework for autonomous driving because it directly addresses sequential decision-making under uncertainty. Driving tasks naturally involve long-term planning, delayed consequences, and complex system dynamics that are difficult to model explicitly. Simulators such as CARLA enable safe, scalable experimentation and provide a controlled environment in which RL agents can be trained and evaluated. In this work, we focus on a simplified but foundational autonomous driving subtask—lane following—and investigate whether a Deep Q-Network (DQN) can learn stable lane-keeping behavior using a compact state representation and discrete steering actions. We further examine whether expert driving trajectories from the PDM_Lite_Carla_LB2 dataset can be incorporated into online learning through replay-buffer prefill, and analyze the resulting benefits and pitfalls.

II. Related Work and Background

Deep reinforcement learning was popularized by the introduction of the Deep Q-Network (DQN) [1], which extended classical tabular Q-learning to continuous state spaces through neural function approximation. By combining experience replay, target networks, and ε-greedy exploration, DQN demonstrated stable learning in domains previously considered intractable for value-based methods. These stabilization techniques were crucial in addressing the instability caused by the interaction of function approximation, bootstrapping, and off-policy learning.

In autonomous driving research, RL has been applied both to end-to-end vision-based control and to low-dimensional control-oriented tasks. The CARLA simulator [2] has become a standard benchmark for evaluating autonomous driving systems, and the CARLA Leaderboard highlights the difficulty of achieving robust generalization across diverse environments. Lane following constitutes a core competency underlying most driving pipelines and serves as a useful testbed for evaluating low-level control strategies.

Offline reinforcement learning has recently gained attention as a means to leverage large expert datasets without online exploration. However, prior work has shown that naïvely applying Q-learning to fixed datasets often leads to divergence or performance collapse due to distributional mismatch and extrapolation error [3], [4]. The PDM_Lite_Carla_LB2 dataset provides high-quality expert trajectories for autonomous driving, but its integration into standard DQN remains nontrivial.

III. Methodology

We formulate the lane-following task as a Markov Decision Process (MDP) defined by the tuple

$$(\mathcal{S}, \mathcal{A}, P, R, \gamma).$$

All experiments are conducted in CARLA version 0.9.16 using synchronous simulation with a fixed timestep of $\Delta t = 0.05$ seconds. The environment consists of a straight-road lane-following scenario in Town03. Each episode terminates when the agent departs the lane, exceeds a safe heading deviation, or reaches a maximum horizon of $T_{\max} = 300$ steps.

The state space $\mathcal{S}$ is continuous and represented as a four-dimensional real-valued vector,

$$\mathcal{S} \subset \mathbb{R}^4, \quad s_t = \begin{bmatrix} d_{\text{lat}} \\ \theta_{\text{err}} \\ v_t \\ t_{\text{norm}} \end{bmatrix}.$$

Here, d_{lat} denotes the signed lateral offset from the lane center, θ_{err} denotes the angular difference between the vehicle's heading and the lane tangent, v_t is the forward velocity, and $t_{\text{norm}} = \frac{t}{T_{\max}}$ is the normalized timestep.

The action space $\mathcal{A}$ is discrete and defined as

$$\mathcal{A} = \{a_0, a_1, a_2\},$$

corresponding to steering left, steering straight, and steering right. Although steering is inherently continuous, discretization enables tractable maximization over actions.

The reward function at timestep t is defined as

$$r_t = 0.1 - 1.0 \cdot \min\left(\frac{|d_{\text{lat}}|}{2.0}, 1\right) - 0.5 \cdot \min\left(\frac{|\theta_{\text{err}}|}{\pi/6}, 1\right) + 0.3 \cdot \min\left(\frac{v_t}{8}, 1\right) - 20 \cdot \mathbf{1}_{\text{terminal}}.$$

The DQN approximates the action-value function

$$Q_\theta : \mathcal{S} \times \mathcal{A} \to \mathbb{R},$$

using a multilayer perceptron with two hidden layers of 128 units and ReLU activations. Actions are selected using an ε-greedy behavior policy. Learning minimizes the mean squared temporal-difference error

$$L(\theta) = \mathbb{E}\left[(Q_\theta(s_t, a_t) - y_t)^2\right],$$

where

$$y_t = \begin{cases} r_t, & \text{if terminal,} \\ r_t + \gamma \max_{a'} Q_{\theta^-}(s_{t+1}, a'), & \text{otherwise.} \end{cases}$$

IV. Experiments and Results

A. Baseline DQN vs Random Policy

Figure 1 illustrates that the baseline DQN significantly outperforms a random policy. The DQN agent rapidly improves its return during training and consistently completes evaluation episodes without lane departure, confirming that the task requires nontrivial sequential decision-making and that the learned policy captures meaningful lane-following behavior.

B. Baseline DQN vs Large-Capacity DQN

As shown in Figure 2, increasing the capacity of the Q-network does not improve performance. Although the larger network occasionally produces smoother steering trajectories, it exhibits higher variance and more frequent catastrophic failures. This behavior is consistent with the known instability of off-policy temporal-difference learning when combined with expressive function approximation.

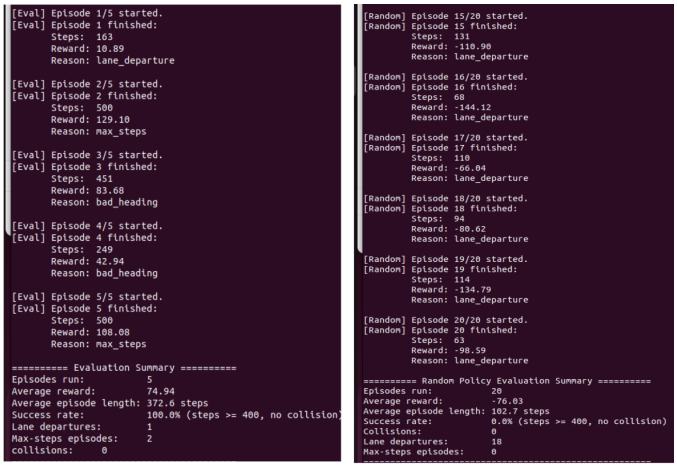

Fig. 1. Comparison between the baseline DQN agent and a random policy. The DQN agent consistently achieves higher episodic returns and maintains stable lane-following behavior, while the random policy fails early due to frequent lane departures.

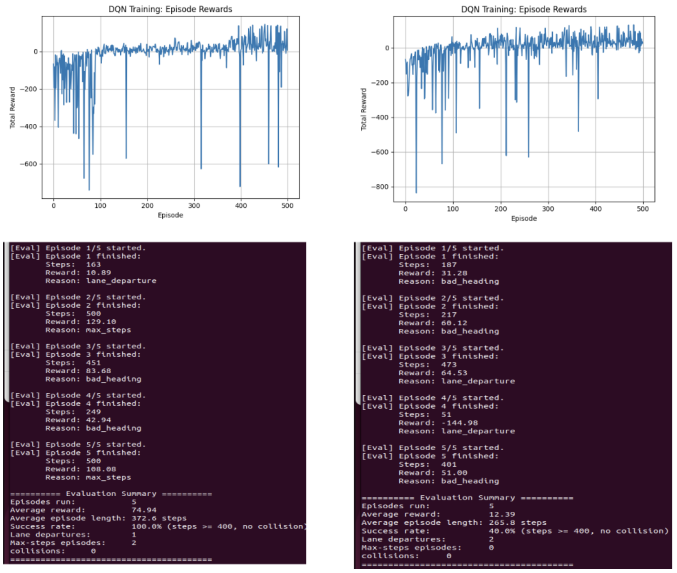

Fig. 2. Performance comparison between the baseline DQN and a larger-capacity DQN. Increasing network size leads to higher variance in learning curves and degraded evaluation performance, indicating increased instability in off-policy value learning.

C. Baseline DQN vs PDM-Prefilled DQN

Figure 3 demonstrates that replay-buffer prefill with expert trajectories degrades quantitative performance under the task reward function. While the prefilled agent exhibits smoother and more human-like steering, it fails to complete evaluation episodes reliably. This performance drop is attributed to distributional mismatch between offline expert data and the online environment, leading to bootstrapping error and Q-value misestimation.

V. Discussion and Conclusions

Our experiments demonstrate that a lightweight DQN agent can successfully learn stable lane-following behavior

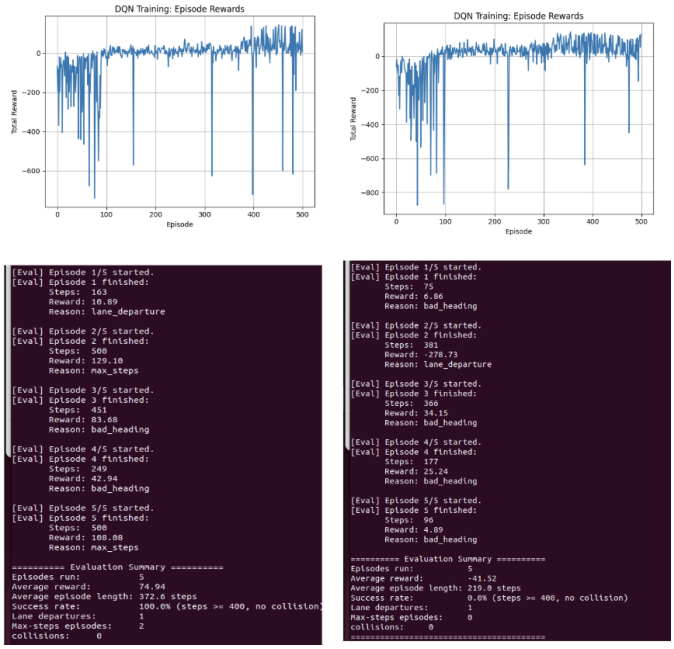

Fig. 3. Evaluation results comparing the baseline DQN with a replay-buffer-prefilled DQN using expert trajectories from the PDM_Lite_Carla_LB2 dataset. The prefilled agent demonstrates smoother qualitative behavior but achieves significantly lower returns due to offline-to-online distribution mismatch.

in CARLA using a compact continuous state representation and discrete steering actions. Classical DQN stabilization techniques—experience replay, target networks, and ε-greedy exploration—are essential for preventing divergence. Increasing network capacity degrades performance, consistent with the instability induced by the deadly triad.

While replay-buffer prefill with expert trajectories produces qualitatively smoother driving behavior, it significantly degrades quantitative performance under the defined reward function due to distribution mismatch and off-policy bootstrapping error. These findings highlight the limitations of naïvely combining offline expert data with online value-based reinforcement learning.

Future work will explore principled offline RL algorithms, vision-based state representations, and Vision-Language-Action models for scalable autonomous driving.

REFERENCES

[1] V. Mnih, K. Kavukcuoglu, D. Silver *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[2] A. Dosovitskiy, G. Ros *et al.*, "Carla: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[3] A. Kumar, A. Zhou *et al.*, "Conservative q-learning for offline reinforcement learning," *Advances in Neural Information Processing Systems*, 2020.

[4] S. Fujimoto, D. Meger, and D. Precup, "Off-policy deep reinforcement learning without exploration," *arXiv preprint arXiv:1812.02900*, 2019.