

Reinforcement Learning: DQN

- As before, please keep the names of the layer consistent with what is requested in model.py. Otherwise the test functions will not work
- You will need to fill in the model.py, DQN.py, replay_buffer.py, and env_wrapper.py.

To avoid pain with installation and model training, please use Colab for this project. DO NOT use Windows for this project. If you use Mac, there might be some potential test case failures you need to address with extra effort.

Set up

```
In [ ]: # !git clone https://github.com/kbaseba/239AS_2.git
```

```
Cloning into '239AS_2'...
remote: Enumerating objects: 360, done.
remote: Total 360 (delta 0), reused 0 (delta 0), pack-reused 360 (from 2)
Receiving objects: 100% (360/360), 865.41 MiB | 15.48 MiB/s, done.
Resolving deltas: 100% (114/114), done.
Updating files: 100% (163/163), done.
```

```
In [ ]: # mount it if you use Colab
# from google.colab import drive
# drive.mount('/content/drive')
```

Mounted at /content/drive

```
In [ ]: # TODO: change the dir to your folder
# %cd /content/239AS_2/Project_4/
```

/content/239AS_2/Project_4

```
In [1]: %load_ext autoreload
%autoreload 2
```

```
In [2]: !pip install numpy torch wandb swig matplotlib termcolor
!pip install gymnasium
!pip install gymnasium[box2d]
```

Requirement already satisfied: numpy in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (2.2.6)

Requirement already satisfied: torch in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (2.7.0)

Requirement already satisfied: wandb in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (0.20.0)

Requirement already satisfied: swig in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (4.3.1)

Requirement already satisfied: matplotlib in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (3.10.3)

Requirement already satisfied: termcolor in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (3.1.0)

Requirement already satisfied: filelock in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (3.18.0)

Requirement already satisfied: typing-extensions>=4.10.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (4.14.0)

Requirement already satisfied: sympy>=1.13.3 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (1.14.0)

Requirement already satisfied: networkx in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (3.5)

Requirement already satisfied: jinja2 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (3.1.6)

Requirement already satisfied: fsspec in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (2025.5.1)

Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.6.77)

Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.6.77)

Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.6.80)

Requirement already satisfied: nvidia-cudnn-cu12==9.5.1.17 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (9.5.1.17)

Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.6.4.1)

Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (11.3.0.4)

Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (10.3.7.77)

Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (11.7.1.2)

Requirement already satisfied: nvidia-cusparselt-cu12==12.5.4.2 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.5.4.2)

Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.5.4.2)

Requirement already satisfied: nvidia-cusparselt-cu12==0.6.3 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (0.6.3)

Requirement already satisfied: nvidia-nccl-cu12==2.26.2 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (2.26.2)

Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in /home/tilboon/Do

cuments/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (12.6.85)
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (1.11.1.6)
Requirement already satisfied: triton==3.3.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from torch) (3.3.0)
Requirement already satisfied: setuptools>=40.8.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from triton==3.3.0->torch) (78.1.1)
Requirement already satisfied: click!=8.0.0,>=7.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (8.2.1)
Requirement already satisfied: gitpython!=3.1.29,>=1.0.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (3.1.44)
Requirement already satisfied: packaging in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (25.0)
Requirement already satisfied: platformdirs in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (4.3.8)
Requirement already satisfied: protobuf!=4.21.0,!5.28.0,<7,>=3.19.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (6.31.1)
Requirement already satisfied: psutil>=5.0.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (5.9.0)
Requirement already satisfied: pydantic<3 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (2.11.5)
Requirement already satisfied: pyyaml in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (6.0.2)
Requirement already satisfied: requests<3,>=2.0.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (2.32.3)
Requirement already satisfied: sentry-sdk>=2.0.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (2.29.1)
Requirement already satisfied: setproctitle in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from wandb) (1.3.6)
Requirement already satisfied: annotated-types>=0.6.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from pydantic<3->wandb) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.2 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from pydantic<3->wandb) (2.33.2)
Requirement already satisfied: typing-inspection>=0.4.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from pydantic<3->wandb) (0.4.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from requests<3,>=2.0.0->wandb) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from requests<3,>=2.0.0->wandb) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from requests<3,>=2.0.0->wandb) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from requests<3,>=2.0.0->wandb) (2025.4.26)

Requirement already satisfied: contourpy>=1.0.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from matplotlib) (1.3.2)

Requirement already satisfied: cyclor>=0.10 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from matplotlib) (0.12.1)

Requirement already satisfied: fonttools>=4.22.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from matplotlib) (4.58.1)

Requirement already satisfied: kiwisolver>=1.3.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from matplotlib) (1.4.8)

Requirement already satisfied: pillow>=8 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from matplotlib) (11.2.1)

Requirement already satisfied: pyparsing>=2.3.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from matplotlib) (3.2.3)

Requirement already satisfied: python-dateutil>=2.7 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from matplotlib) (2.9.0.post0)

Requirement already satisfied: gitdb<5,>=4.0.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gitpython!=3.1.29,>=1.0.0->wandb) (4.0.12)

Requirement already satisfied: smmap<6,>=3.0.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gitdb<5,>=4.0.1->gitpython!=3.1.29,>=1.0.0->wandb) (5.0.2)

Requirement already satisfied: six>=1.5 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from python-dateutil>=2.7->matplotlib) (1.17.0)

Requirement already satisfied: mpmath<1.4,>=1.1.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from sympy>=1.13.3->torch) (1.3.0)

Requirement already satisfied: MarkupSafe>=2.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from jinja2->torch) (3.0.2)

Requirement already satisfied: gymnasium in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (1.1.1)

Requirement already satisfied: numpy>=1.21.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium) (2.2.6)

Requirement already satisfied: cloudpickle>=1.2.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium) (3.1.1)

Requirement already satisfied: typing-extensions>=4.3.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium) (4.14.0)

Requirement already satisfied: farama-notifications>=0.0.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium) (0.0.4)

Requirement already satisfied: gymnasium[box2d] in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (1.1.1)

Requirement already satisfied: numpy>=1.21.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (2.2.6)

Requirement already satisfied: cloudpickle>=1.2.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (3.1.1)

Requirement already satisfied: typing-extensions>=4.3.0 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (4.14.0)

Requirement already satisfied: farama-notifications>=0.0.1 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (0.0.4)

Requirement already satisfied: box2d-py==2.3.5 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (2.3.5)

Requirement already satisfied: pygame>=2.1.3 in /home/tilboon/Documents/239AS_2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (2.1.3)

```
S_2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (2.6.1)
Requirement already satisfied: swig==4.* in /home/tilboon/Documents/239AS_
2/.conda/lib/python3.11/site-packages (from gymnasium[box2d]) (4.3.1)
```

```
In [2]: import rl_test
        from utils import *
```

Introduction to the Enviroment (5 points)

We will be training a DQN agent to play the game of CarRacing. The agent will be trained to play the game using the pixels of the game as an input. The reward structure is as follows for each frame:

- -0.1 for each frame
- +1000/N where N is the number of tiles visited by the car in the episode

The overall goal of this game is to design a agent that is able to play the game with a average test score of above 600. In discrete mode the actions can take 5 actions,

- 0: Do Nothing
- 1: Turn Left
- 2: Turn Right
- 3: Accelerate
- 4: Brake

First let us visualize the game and understand the environment.

```
In [3]: import gymnasium as gym
        import numpy as np
        env = gym.make('CarRacing-v3', continuous=False, render_mode='rgb_array')
        env.seed(42)
```

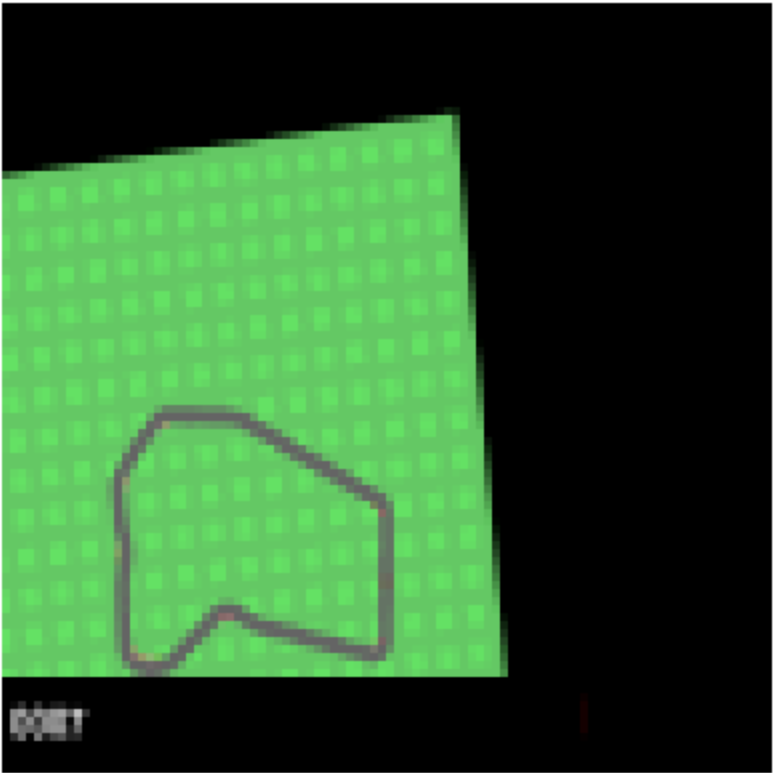
```
In [13]: from IPython.display import HTML

        frames = []
        s, _ = env.reset()

        while True:
            a = env.action_space.sample()
            s, r, terminated, truncated, _ = env.step(a)
            frames.append(s)
            if terminated or truncated:
                break

        anim = animate(frames)
        HTML(anim.to_jshtml())
```

Out[13]:



-

⏮

⏪

⏩

⏭

⏴

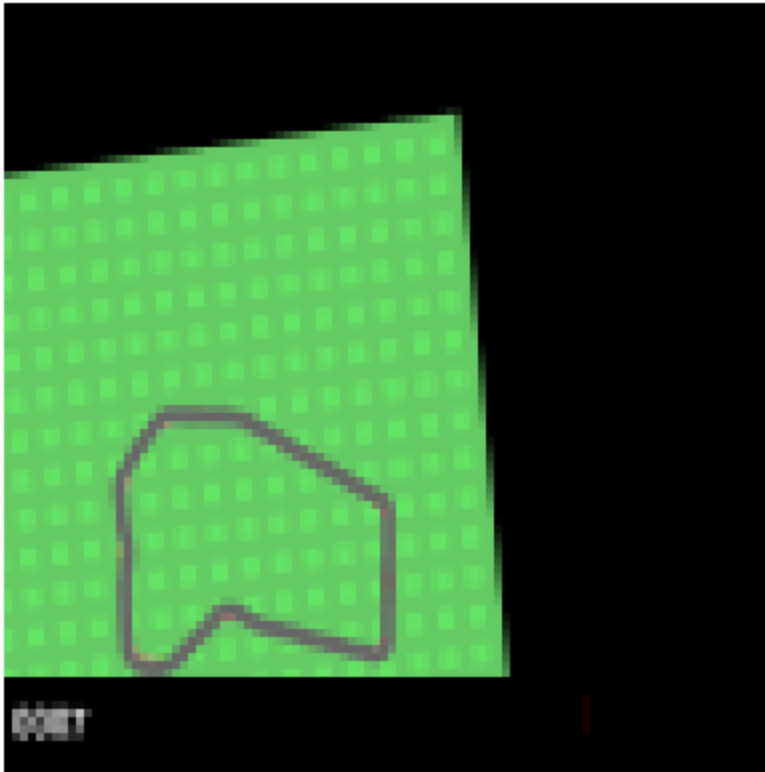
⏵

+

Once

Loop

Reflect



So a couple things we can note:

- at the beginning of the game, we have 50 frames of the game slowly zooming into the car, we should ignore this period, ie no-op during this period.
- there is a black bar at the bottom of the screen, we should crop this out of the observation.

In addition, another thing to note is that the current frame doesn't give much information about the velocity and acceleration of the car, and that the car does not move much for each frame.

Environment Wrapper (5 points)

As a result, you will need to complete `EnvWrapper` in `env_wrapper.py`. You can find more information in the docstring for the wrapper, however the main idea is that it is a wrapper to the environment that does the following:

- skips the first 50 frames of the game
- crops out the black bar and reshapes the observation to a 84x84 image, as well as turning the resulting image to grayscale
- performs the actions for `skip_frames` frames
- stacks the last `num_frames` frames together to give the agent some information about the velocity and acceleration of the car.

```
In [4]: from env_wrapper import EnvWrapper  
rl_test.test_wrapper(EnvWrapper)
```

Passed reset
Passed step

DQN (25 points)

CNN Model (5 points)

Now we are ready to build the model. Our architecture of the CNN model is the one proposed by Mnih et al in "Human-level control through deep reinforcement learning". Specifically this consists of the following layers:

- A convolutional layer with 32 filters of size 8x8 with stride 4 and relu activation
- A convolutional layer with 64 filters of size 4x4 with stride 2 and relu activation
- A convolutional layer with 64 filters of size 3x3 with stride 1 and relu activation
- A fully connected layer with 512 units and relu activation
- A fully connected layer with the number of outputs of the environment

Please implement this model `Nature_Paper_Conv` in `model.py` as well as the helper `MLP` class.

```
In [5]: import model  
rl_test.test_model_DQN(model.Nature_Paper_Conv)
```

Passed

Now we are ready to implement the DQN algorithm.

Algorithm 1 Deep Q-learning with Experience Replay

```
Initialize replay memory  $\mathcal{D}$  to capacity  $N$ 
Initialize action-value function  $Q$  with random weights
for episode = 1,  $M$  do
    Initialise sequence  $s_1 = \{x_1\}$  and preprocessed sequenced  $\phi_1 = \phi(s_1)$ 
    for  $t = 1, T$  do
        With probability  $\epsilon$  select a random action  $a_t$ 
        otherwise select  $a_t = \max_a Q^*(\phi(s_t), a; \theta)$ 
        Execute action  $a_t$  in emulator and observe reward  $r_t$  and image  $x_{t+1}$ 
        Set  $s_{t+1} = s_t, a_t, x_{t+1}$  and preprocess  $\phi_{t+1} = \phi(s_{t+1})$ 
        Store transition  $(\phi_t, a_t, r_t, \phi_{t+1})$  in  $\mathcal{D}$ 
        Sample random minibatch of transitions  $(\phi_j, a_j, r_j, \phi_{j+1})$  from  $\mathcal{D}$ 
        Set  $y_j = \begin{cases} r_j & \text{for terminal } \phi_{j+1} \\ r_j + \gamma \max_{a'} Q(\phi_{j+1}, a'; \theta) & \text{for non-terminal } \phi_{j+1} \end{cases}$ 
        Perform a gradient descent step on  $(y_j - Q(\phi_j, a_j; \theta))^2$  according to equation 3
    end for
end for
```

Replay Buffer (5 points)

First start by implementing the DQN replay buffer `ReplayBufferDQN` in `replay_buffer.py`. This buffer will store the transitions of the agent and sample them for training.

```
In [6]: from replay_buffer import ReplayBufferDQN

rl_test.test_DQN_replay_buffer(ReplayBufferDQN)
```

Passed

DQN (15 points)

Now implement the `_optimize_model` and `_sample_action` functions in `DQN` in `DQN.py`. The `_optimize_model` function will sample a batch of transitions from the replay buffer and update the model. The `_sample_action` function will sample an action from the model given the current state. Train the model over 200 episodes, validating every 50 episodes for 30 episodes, before testing the model for 50 episodes at the end.

```
In [7]: import wandb
```

```
In [8]: wandb.login()
```

```
wandb: Currently logged in as: kbaseba (kbaseba-ucla) to https://api.wandb.ai. Use `wandb login --relogin` to force relogin
```

```
Out[8]: True
```

```
In [9]: wandb.init(project="racing-car-dqn")
```

Tracking run with wandb version 0.20.0

Run data is saved locally in

/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250608_115956-4dulynil

Syncing run **golden-sun-7** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/4dulynil>

Out[9]: Display W&B run

```
In [11]: import DQN
import utils
import torch

# if running with Colab T4 with wandb, it takes 60-70 mins
# if running with Colab T4 w/o wandb, it takes ~30 mins
# your test mean reward ideally should be larger than 400
trainerDQN = DQN.DQN(EnvWrapper(env),
                      model.Nature_Paper_Conv,
                      lr = 0.00025,
                      gamma = 0.95,
                      buffer_size = 100000,
                      batch_size = 32,
                      loss_fn = "mse_loss",
                      use_wandb = True,
                      device = 'cuda',
                      seed = 42,
                      epsilon_scheduler = utils.exponential_decay(1, 700, 0.1),
                      save_path = utils.get_save_path("DQN", "./runs/"))

trainerDQN.train(200, 50, 30, 50, 50)
```

saving to ./runs/DQN/run0

Finishing previous runs because reinit is set to 'default'.

wandb: **ERROR** The nbformat package was not found. It is required to save notebook history.

View run **smart-firefly-1** at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/t9wk404r>

View project at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

Find logs at: ./wandb/run-20250603_185854-t9wk404r/logs

Tracking run with wandb version 0.20.0

Run data is saved locally in

/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250603_185917-ajn0n650

Syncing run **neat-microwave-2** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/ajn0n650>

Episode: 0: Time: 6.133239984512329 Total Reward: -12.26618705035983 Avg_Loss: 0.0
Episode: 1: Time: 8.21942663192749 Total Reward: -62.948717948718894 Avg_Loss: 0.9044264260750667
Episode: 2: Time: 8.034026622772217 Total Reward: -32.95620437956208 Avg_Loss: 0.855911413692877
Episode: 3: Time: 8.44169282913208 Total Reward: -73.57142857142873 Avg_Loss: 0.825164575834109
Episode: 4: Time: 8.884541988372803 Total Reward: -20.925925925926414 Avg_Loss: 0.7258721724731707
Episode: 5: Time: 8.549166917800903 Total Reward: -38.66197183098643 Avg_Loss: 0.9092964216628495
Episode: 6: Time: 8.060179948806763 Total Reward: -18.66412213740521 Avg_Loss: 0.7732668172787217
Episode: 7: Time: 8.099495649337769 Total Reward: 14.375000000001073 Avg_Loss: 1.035137609085616
Episode: 8: Time: 8.225048780441284 Total Reward: 28.188405797103286 Avg_Loss: 1.4093751558924423
Episode: 9: Time: 8.339138507843018 Total Reward: 75.4545454545497 Avg_Loss: 1.1264866085237815
Episode: 10: Time: 8.327865362167358 Total Reward: -55.784313725490904 Avg_Loss: 1.2182710207059604
Episode: 11: Time: 7.740700006484985 Total Reward: -42.28555956678336 Avg_Loss: 1.3704961291720381
Episode: 12: Time: 8.910240888595581 Total Reward: 62.09969788520029 Avg_Loss: 4.109065845358272
Episode: 13: Time: 8.5490562915802 Total Reward: 17.627986348121983 Avg_Loss: 3.806811966315037
Episode: 14: Time: 5.6029887199401855 Total Reward: -93.01355932203455 Avg_Loss: 4.880658785008496
Episode: 15: Time: 8.656633853912354 Total Reward: 91.49517684887832 Avg_Loss: 11.253737637475759
Episode: 16: Time: 8.690074920654297 Total Reward: -28.3333333333374 Avg_Loss: 4.611919908839114
Episode: 17: Time: 8.704070568084717 Total Reward: 121.94915254237577 Avg_Loss: 6.624895627132985
Episode: 18: Time: 8.76585054397583 Total Reward: 179.14330218069014 Avg_Loss: 4.442848708887561
Episode: 19: Time: 8.740228176116943 Total Reward: 31.21359223300972 Avg_Loss: 4.493706668625359
Episode: 20: Time: 8.527551174163818 Total Reward: 35.281690140847566 Avg_Loss: 3.7453338874488318
Episode: 21: Time: 8.382328987121582 Total Reward: 13.108108108109029 Avg_Loss: 2.711287133333062
Episode: 22: Time: 8.85209846496582 Total Reward: 561.0283687943178 Avg_Loss: 4.066573404213962
Episode: 23: Time: 8.712191104888916 Total Reward: 388.870967741929 Avg_Loss: 4.510514091042912
Episode: 24: Time: 8.741256475448608 Total Reward: 267.1794871794898 Avg_Loss: 5.981761251677986
Episode: 25: Time: 8.629642248153687 Total Reward: 256.6483516483525 Avg_Loss: 5.791294638098789
Episode: 26: Time: 8.749993562698364 Total Reward: 261.64335664334857 Avg_Loss: 6.160182711457004
Episode: 27: Time: 8.63331151008606 Total Reward: 310.5944055944049 Avg_Loss: 4.377944937022794

Episode: 28: Time: 9.600750207901001 Total Reward: 365.81504702193683 Avg_Loss: 5.0955391145553905
Episode: 29: Time: 9.376365661621094 Total Reward: 320.4929577464706 Avg_Loss: 4.9559988785190745
Episode: 30: Time: 9.507494688034058 Total Reward: 442.6344086021476 Avg_Loss: 6.523079610672317
Episode: 31: Time: 9.56268048286438 Total Reward: 117.50000000000418 Avg_Loss: 5.0505260674893355
Episode: 32: Time: 9.467845916748047 Total Reward: 323.25095057034355 Avg_Loss: 5.788560429541003
Episode: 33: Time: 9.7420973777771 Total Reward: 277.549019607846 Avg_Loss: 7.841803481598862
Episode: 34: Time: 9.438688039779663 Total Reward: 6.265822784810526 Avg_Loss: 7.497489858074348
Episode: 35: Time: 9.794541597366333 Total Reward: 307.5157232704401 Avg_Loss: 6.86144556768802
Episode: 36: Time: 9.838263511657715 Total Reward: 210.64784053156586 Avg_Loss: 6.938436169083379
Episode: 37: Time: 9.857799291610718 Total Reward: 132.414330218071 Avg_Loss: 6.415113450599318
Episode: 38: Time: 9.360185146331787 Total Reward: 259.4776119402933 Avg_Loss: 6.521061640827596
Episode: 39: Time: 9.657518148422241 Total Reward: 186.64556962025537 Avg_Loss: 7.583496148345851
Episode: 40: Time: 9.496632099151611 Total Reward: 6.818181818095 Avg_Loss: 7.929517399864037
Episode: 41: Time: 9.50952434539795 Total Reward: 454.6183206106809 Avg_Loss: 6.39722378514394
Episode: 42: Time: 9.641336917877197 Total Reward: 441.1842105263091 Avg_Loss: 5.7463160863443585
Episode: 43: Time: 9.529329776763916 Total Reward: 443.18181818096 Avg_Loss: 6.424317680987992
Episode: 44: Time: 9.58205795288086 Total Reward: 271.1016949152571 Avg_Loss: 7.109447162692287
Episode: 45: Time: 9.956891775131226 Total Reward: 69.38356164383993 Avg_Loss: 6.300570757449174
Episode: 46: Time: 9.56554889678955 Total Reward: 65.71428571429033 Avg_Loss: 6.371259860631798
Episode: 47: Time: 9.843086957931519 Total Reward: 419.3769968051065 Avg_Loss: 6.0480705915378925
Episode: 48: Time: 9.608461141586304 Total Reward: 282.55102040816325 Avg_Loss: 5.385454183866997
Episode: 49: Time: 9.557258367538452 Total Reward: 214.52380952381168 Avg_Loss: 6.065896774540429
Validation Mean Reward: 26.815853054890464 Validation Std Reward: 199.75409948699942
Episode: 50: Time: 8.60549259185791 Total Reward: 381.3513513513471 Avg_Loss: 5.764939173429954
Episode: 51: Time: 8.73774790763855 Total Reward: 128.2415902140675 Avg_Loss: 6.107253385191204
Episode: 52: Time: 8.626689910888672 Total Reward: 221.32653061224903 Avg_Loss: 7.174219978456738
Episode: 53: Time: 8.439385652542114 Total Reward: 306.4336917562713 Avg_Loss: 5.665037500758131
Episode: 54: Time: 8.489814519882202 Total Reward: 386.6176470588215 Avg_Loss: 5.723625704020011

Episode: 55: Time: 8.457855701446533 Total Reward: 394.2086330935222 Avg_Loss: 5.593486036573138
Episode: 56: Time: 8.39532732963562 Total Reward: 232.93522267206708 Avg_Loss: 5.550559617391153
Episode: 57: Time: 8.499794006347656 Total Reward: 332.5618374558277 Avg_Loss: 6.0964519131083446
Episode: 58: Time: 8.21054983139038 Total Reward: 74.67509025271121 Avg_Loss: 5.685884224266565
Episode: 59: Time: 8.140634536743164 Total Reward: 41.71875000000306 Avg_Loss: 5.766211925434465
Episode: 60: Time: 8.471680402755737 Total Reward: 190.7142857142894 Avg_Loss: 5.993246082999125
Episode: 61: Time: 8.71571159362793 Total Reward: 234.032258064518 Avg_Loss: 5.961606960837581
Episode: 62: Time: 8.590460538864136 Total Reward: 364.999999999996 Avg_Loss: 6.154986846346815
Episode: 63: Time: 8.458800077438354 Total Reward: 152.34982332155573 Avg_Loss: 5.981433028934383
Episode: 64: Time: 8.475598096847534 Total Reward: 230.5033557047015 Avg_Loss: 5.99139567583549
Episode: 65: Time: 8.747697830200195 Total Reward: 310.06329113923914 Avg_Loss: 5.7521675594714505
Episode: 66: Time: 8.908056020736694 Total Reward: 255.42735042735356 Avg_Loss: 5.509420610776468
Episode: 67: Time: 8.483033418655396 Total Reward: 118.5593220339004 Avg_Loss: 6.093845582809768
Episode: 68: Time: 8.4671471118927 Total Reward: 139.87544483986085 Avg_Loss: 6.142524041047617
Episode: 69: Time: 8.25883173942566 Total Reward: 358.0612244897911 Avg_Loss: 6.04441098806237
Episode: 70: Time: 8.473734855651855 Total Reward: 422.1102661596868 Avg_Loss: 5.681056739903298
Episode: 71: Time: 8.438021421432495 Total Reward: 318.0434782608686 Avg_Loss: 5.853102281313984
Episode: 72: Time: 8.647363424301147 Total Reward: 68.00940438871416 Avg_Loss: 6.315724428962259
Episode: 73: Time: 8.693833589553833 Total Reward: 236.18971061093538 Avg_Loss: 6.885392281187682
Episode: 74: Time: 8.43391728401184 Total Reward: 276.5277777777795 Avg_Loss: 6.391134386303044
Episode: 75: Time: 8.72618579864502 Total Reward: 209.49826989619777 Avg_Loss: 5.4345506299443604
Episode: 76: Time: 8.844269752502441 Total Reward: 270.6249999999926 Avg_Loss: 6.405602575350208
Episode: 77: Time: 8.570299625396729 Total Reward: 82.04918032787255 Avg_Loss: 5.672231001513345
Episode: 78: Time: 8.648847579956055 Total Reward: 329.13793103447654 Avg_Loss: 5.62028311480995
Episode: 79: Time: 8.735196113586426 Total Reward: 201.6666666666709 Avg_Loss: 6.056042985755856
Episode: 80: Time: 8.735011100769043 Total Reward: 206.53846153846482 Avg_Loss: 6.670787956534314
Episode: 81: Time: 8.748886823654175 Total Reward: 87.69230769231005 Avg_Loss: 6.96011440894183
Episode: 82: Time: 8.521762609481812 Total Reward: 260.63380281690377 Avg_Loss: 6.534499041172636

Episode: 83: Time: 8.802938222885132 Total Reward: 354.99999999999426 Avg_Loss: 6.265703735732231
Episode: 84: Time: 9.09769344329834 Total Reward: 296.691394658754 Avg_Loss: 6.3792630283772445
Episode: 85: Time: 8.772404670715332 Total Reward: 246.91176470588715 Avg_Loss: 6.824356597511708
Episode: 86: Time: 8.855376243591309 Total Reward: 305.6514657980422 Avg_Loss: 6.551552474999628
Episode: 87: Time: 8.791648626327515 Total Reward: 276.92982456140174 Avg_Loss: 5.756341840038781
Episode: 88: Time: 8.714039325714111 Total Reward: 193.9610389610425 Avg_Loss: 5.62295334479388
Episode: 89: Time: 8.73051905632019 Total Reward: 427.5563909774339 Avg_Loss: 5.751106779114539
Episode: 90: Time: 8.630814552307129 Total Reward: 46.06583072100583 Avg_Loss: 7.226604164147577
Episode: 91: Time: 8.414000988006592 Total Reward: 315.8527131782954 Avg_Loss: 6.247645106636176
Episode: 92: Time: 8.423139810562134 Total Reward: 300.6043956043938 Avg_Loss: 5.6599397729424865
Episode: 93: Time: 8.745530843734741 Total Reward: 257.73972602739855 Avg_Loss: 6.016664015144861
Episode: 94: Time: 8.826033115386963 Total Reward: 227.36842105263545 Avg_Loss: 7.272338969366891
Episode: 95: Time: 8.872728824615479 Total Reward: 383.72340425531763 Avg_Loss: 6.079880816595895
Episode: 96: Time: 9.15888261795044 Total Reward: 275.3703703703732 Avg_Loss: 6.11970666777186
Episode: 97: Time: 8.72473430633545 Total Reward: 207.81690140845348 Avg_Loss: 6.520103151056947
Episode: 98: Time: 9.107728242874146 Total Reward: 164.8187311178284 Avg_Loss: 5.771718839136493
Episode: 99: Time: 8.846008539199829 Total Reward: 304.9999999999992 Avg_Loss: 5.6634034364163375
Validation Mean Reward: -94.99999999999999 Validation Std Reward: 3.750876069797784e-14
Episode: 100: Time: 9.23388385772705 Total Reward: 399.9152542372824 Avg_Loss: 5.495403982010208
Episode: 101: Time: 8.987744331359863 Total Reward: 434.61672473866855 Avg_Loss: 6.438014780773836
Episode: 102: Time: 8.848567485809326 Total Reward: 389.64163822524654 Avg_Loss: 5.865145631698
Episode: 103: Time: 8.694725036621094 Total Reward: 361.0810810810778 Avg_Loss: 5.811214960923715
Episode: 104: Time: 8.781830310821533 Total Reward: 497.71523178807195 Avg_Loss: 5.756218878661885
Episode: 105: Time: 8.800401449203491 Total Reward: 223.02120141342812 Avg_Loss: 5.812086381832091
Episode: 106: Time: 9.186482429504395 Total Reward: 237.3943661971872 Avg_Loss: 6.720707616886171
Episode: 107: Time: 8.744510650634766 Total Reward: 346.17647058823195 Avg_Loss: 7.111493985692994
Episode: 108: Time: 8.76833987236023 Total Reward: 232.7027027027048 Avg_Loss: 7.632528891082571
Episode: 109: Time: 8.879911422729492 Total Reward: 335.5555555555556 Avg_Loss: 6.134643898290746

Episode: 110: Time: 9.093061447143555 Total Reward: 351.0641399416907 Avg_Loss: 6.097547522112101
Episode: 111: Time: 8.824126958847046 Total Reward: 372.532467532464 Avg_Loss: 5.8935531477968235
Episode: 112: Time: 8.824001789093018 Total Reward: 429.1379310344746 Avg_Loss: 5.585125052628397
Episode: 113: Time: 9.022949934005737 Total Reward: 614.090909090896 Avg_Loss: 5.639891529283604
Episode: 114: Time: 9.186776161193848 Total Reward: 246.5384615384594 Avg_Loss: 6.505168010206783
Episode: 115: Time: 8.986930847167969 Total Reward: 178.6156351791576 Avg_Loss: 5.978715783407708
Episode: 116: Time: 8.837463617324829 Total Reward: 399.54545454544837 Avg_Loss: 5.130526297733564
Episode: 117: Time: 9.066534519195557 Total Reward: 383.8273615635105 Avg_Loss: 5.983850382957138
Episode: 118: Time: 8.71725344657898 Total Reward: 339.9315068493129 Avg_Loss: 6.040466426801281
Episode: 119: Time: 8.542380571365356 Total Reward: 272.8571428571459 Avg_Loss: 5.996386554060864
Episode: 120: Time: 8.663614749908447 Total Reward: 303.0263157894716 Avg_Loss: 5.51860421745717
Episode: 121: Time: 8.95458173751831 Total Reward: 337.3529411764688 Avg_Loss: 6.419669531974472
Episode: 122: Time: 9.011008024215698 Total Reward: 395.44585987260365 Avg_Loss: 5.529630923471531
Episode: 123: Time: 9.059798002243042 Total Reward: 399.9494949494921 Avg_Loss: 5.55393497583245
Episode: 124: Time: 8.97329330444336 Total Reward: 254.05660377358896 Avg_Loss: 6.185687293525503
Episode: 125: Time: 8.88105845451355 Total Reward: 362.9124579124533 Avg_Loss: 6.382902432890499
Episode: 126: Time: 8.974822044372559 Total Reward: 85.42813455657964 Avg_Loss: 5.827355312700031
Episode: 127: Time: 9.043782711029053 Total Reward: 404.99999999999335 Avg_Loss: 5.433334853468823
Episode: 128: Time: 8.885504722595215 Total Reward: 294.43894389439083 Avg_Loss: 5.69522147419072
Episode: 129: Time: 8.784563064575195 Total Reward: 519.0350877192882 Avg_Loss: 5.505801822457995
Episode: 130: Time: 8.797478914260864 Total Reward: 350.22968197879345 Avg_Loss: 5.4870838347603295
Episode: 131: Time: 8.824974775314331 Total Reward: 121.39344262295147 Avg_Loss: 6.393650137075857
Episode: 132: Time: 9.025090456008911 Total Reward: 201.17834394904904 Avg_Loss: 5.600372358530509
Episode: 133: Time: 8.847079753875732 Total Reward: 463.3038869257896 Avg_Loss: 5.369048794778455
Episode: 134: Time: 8.914988994598389 Total Reward: 365.207612456741 Avg_Loss: 6.279380510835087
Episode: 135: Time: 8.66708517074585 Total Reward: 451.8164794007389 Avg_Loss: 5.631665067512448
Episode: 136: Time: 8.547762155532837 Total Reward: 282.9527559055024 Avg_Loss: 5.800314729955016
Episode: 137: Time: 7.851160287857056 Total Reward: -15.318725099601961 Avg_Loss: 5.8691440319814605

Episode: 138: Time: 9.014394044876099 Total Reward: 316.585365853655 Avg_Loss: 5.454292745149436
Episode: 139: Time: 8.92708158493042 Total Reward: 181.07361963190544 Avg_Loss: 6.26820403087039
Episode: 140: Time: 9.400326013565063 Total Reward: 296.6666666666647 Avg_Loss: 5.228830572937717
Episode: 141: Time: 8.982694387435913 Total Reward: 487.75862068964534 Avg_Loss: 5.5891978510287625
Episode: 142: Time: 9.114570379257202 Total Reward: 451.66666666666003 Avg_Loss: 5.666927215431919
Episode: 143: Time: 9.304147720336914 Total Reward: 343.80597014925087 Avg_Loss: 5.259942142903304
Episode: 144: Time: 8.967745780944824 Total Reward: 535.5084745762624 Avg_Loss: 5.771103710687461
Episode: 145: Time: 8.678895711898804 Total Reward: 269.7540983606455 Avg_Loss: 5.573106793796315
Episode: 146: Time: 9.143201112747192 Total Reward: 266.7647058823514 Avg_Loss: 6.110257084630117
Episode: 147: Time: 8.977455854415894 Total Reward: 453.0769230769168 Avg_Loss: 5.6753365272233465
Episode: 148: Time: 8.96343445777893 Total Reward: 321.6666666666657 Avg_Loss: 5.698031340326581
Episode: 149: Time: 8.91055417060852 Total Reward: 371.43109540635913 Avg_Loss: 6.076687597427048
Validation Mean Reward: 559.1889593315589 Validation Std Reward: 168.09601292013545
Episode: 150: Time: 9.74891996383667 Total Reward: 474.96587030716023 Avg_Loss: 5.399547977106912
Episode: 151: Time: 9.408512830734253 Total Reward: 583.0303030302925 Avg_Loss: 5.739971555581613
Episode: 152: Time: 9.578917980194092 Total Reward: 400.6268221574315 Avg_Loss: 6.300333744337578
Episode: 153: Time: 9.098928689956665 Total Reward: 340.89743589743523 Avg_Loss: 5.78580046701832
Episode: 154: Time: 9.274168252944946 Total Reward: 170.5172413793144 Avg_Loss: 5.894829004752536
Episode: 155: Time: 9.355210542678833 Total Reward: 240.61643835616832 Avg_Loss: 5.366021521952974
Episode: 156: Time: 9.250321865081787 Total Reward: 487.08955223879695 Avg_Loss: 6.242042586583049
Episode: 157: Time: 9.536597490310669 Total Reward: 523.4615384615292 Avg_Loss: 5.84925790193702
Episode: 158: Time: 9.176278114318848 Total Reward: 536.2056737588568 Avg_Loss: 5.709083238569629
Episode: 159: Time: 9.131852388381958 Total Reward: 463.8235294117604 Avg_Loss: 6.024519376394127
Episode: 160: Time: 9.382244348526001 Total Reward: 427.49134948096105 Avg_Loss: 6.282056907645795
Episode: 161: Time: 9.44643497467041 Total Reward: 198.91891891892226 Avg_Loss: 6.331373120556359
Episode: 162: Time: 9.365617990493774 Total Reward: 323.9189189189187 Avg_Loss: 6.872203869979923
Episode: 163: Time: 9.285917520523071 Total Reward: 381.6666666666603 Avg_Loss: 5.938843585863835
Episode: 164: Time: 9.267295122146606 Total Reward: 347.90657439446466 Avg_Loss: 6.810720825395665

Episode: 165: Time: 9.418452978134155 Total Reward: 363.46153846153686 Avg_Loss: 5.962919083963923
Episode: 166: Time: 9.433175325393677 Total Reward: 456.40186915887136 Avg_Loss: 5.8133620484536435
Episode: 167: Time: 9.432826042175293 Total Reward: 526.1604095563036 Avg_Loss: 5.4641768191041065
Episode: 168: Time: 9.377393245697021 Total Reward: 483.59531772574405 Avg_Loss: 5.8337165958741135
Episode: 169: Time: 9.260392904281616 Total Reward: 363.90410958903766 Avg_Loss: 6.328909038495617
Episode: 170: Time: 9.110474824905396 Total Reward: 506.6260162601555 Avg_Loss: 6.174251726695469
Episode: 171: Time: 9.624365329742432 Total Reward: 415.38575667655596 Avg_Loss: 6.052297151389242
Episode: 172: Time: 9.425992965698242 Total Reward: 454.99999999999125 Avg_Loss: 6.087077918172884
Episode: 173: Time: 8.892867803573608 Total Reward: 288.7638376383669 Avg_Loss: 6.220018607227742
Episode: 174: Time: 9.094950914382935 Total Reward: 547.8571428571324 Avg_Loss: 6.8448557483048
Episode: 175: Time: 9.266983032226562 Total Reward: 549.194756554299 Avg_Loss: 5.922192030594129
Episode: 176: Time: 9.274991035461426 Total Reward: 228.84341637011144 Avg_Loss: 6.27623997515991
Episode: 177: Time: 9.250582933425903 Total Reward: 598.7269372693652 Avg_Loss: 5.736074218229086
Episode: 178: Time: 9.451613426208496 Total Reward: 85.60200668896768 Avg_Loss: 6.16952095913286
Episode: 179: Time: 9.47279977798462 Total Reward: 502.35973597359094 Avg_Loss: 6.197666877458076
Episode: 180: Time: 9.59806489944458 Total Reward: 274.90595611285374 Avg_Loss: 6.164513395613983
Episode: 181: Time: 9.144271612167358 Total Reward: 627.6027397260159 Avg_Loss: 6.278574197232222
Episode: 182: Time: 9.238120555877686 Total Reward: 485.756013745697 Avg_Loss: 6.123357141719145
Episode: 183: Time: 9.756761074066162 Total Reward: 333.16091954022954 Avg_Loss: 6.510135272971723
Episode: 184: Time: 9.192201614379883 Total Reward: 608.7037037036957 Avg_Loss: 6.029204368591309
Episode: 185: Time: 9.261204957962036 Total Reward: 643.0073800737939 Avg_Loss: 6.1038742916924615
Episode: 186: Time: 9.786289691925049 Total Reward: 274.4267515923545 Avg_Loss: 7.135785656816819
Episode: 187: Time: 9.738494157791138 Total Reward: 417.1212121212061 Avg_Loss: 6.503933804375785
Episode: 188: Time: 9.409322023391724 Total Reward: 429.3055555555491 Avg_Loss: 6.025558203208346
Episode: 189: Time: 9.303148746490479 Total Reward: 408.52112676055685 Avg_Loss: 5.53696325646729
Episode: 190: Time: 9.182215690612793 Total Reward: 439.2960288808616 Avg_Loss: 6.4564922366823465
Episode: 191: Time: 9.40623927116394 Total Reward: 360.47945205479 Avg_Loss: 5.903223672834765
Episode: 192: Time: 9.426127195358276 Total Reward: 272.9245283018884 Avg_Loss: 6.271274244084077

Episode: 193: Time: 9.23553991317749 Total Reward: 529.060150375933 Avg_Loss: 6.307258181211327
 Episode: 194: Time: 9.657405853271484 Total Reward: 446.7956656346656 Avg_Loss: 5.9197905384191944
 Episode: 195: Time: 9.41226601600647 Total Reward: 439.4827586206824 Avg_Loss: 6.738179166777795
 Episode: 196: Time: 9.250213861465454 Total Reward: 386.75182481751216 Avg_Loss: 6.322154710272781
 Episode: 197: Time: 9.59204912185669 Total Reward: 389.330484330479 Avg_Loss: 6.575044793241164
 Episode: 198: Time: 9.279990673065186 Total Reward: 425.5992509363243 Avg_Loss: 6.366872571095699
 Episode: 199: Time: 9.471853017807007 Total Reward: 242.57961783439643 Avg_Loss: 6.635416620919685
 Validation Mean Reward: 623.8010792168969 Validation Std Reward: 201.97060741012447
 Test Mean Reward: 605.0771519145571 Test Std Reward: 181.9873134653002

Please include a plot of the training and validation rewards over the episodes in the report. Answer the question in no more than 2 sentences: Does the loss matter in DQN? Why or why not?

Your answer:

Certainly, loss has significance in DQN as it helps the Q-network converge towards the optimal action-value function by reducing the error associated with predicted and target Q-values. Nevertheless, performance isn't always strong when loss is low, as that low loss may not indicate how well the learned policy performs within the environment. loss != (always) = good policy - since overfitting or poor exploration can lead to low loss but suboptimal actions.

```
In [17]: from IPython.display import Image, display
import os

img_path = os.path.join('.', 'images', 'dqn_stats.png')
display(Image(filename=img_path))
```



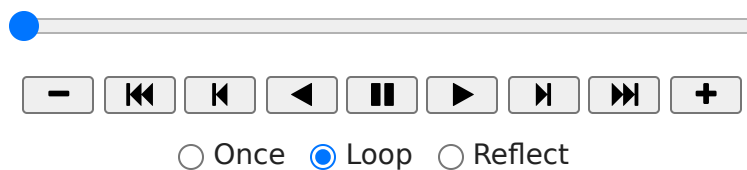
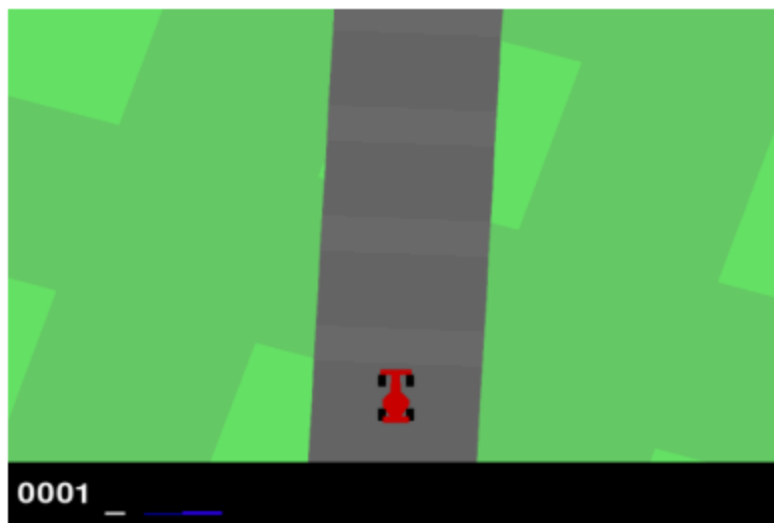
Animation

We can also draw a animation of the car in one game, the code is provided below. The animation is just for your reference to get a feeling of how good your model is. No need to submit in your submission. We will grade based on your plots of training and validation above.

```
In [14]: eval_env = gym.make('CarRacing-v3', continuous=True, render_mode='rgb_array')
eval_env = EnvWrapper(eval_env)

total_rewards, frames = trainerDQN.play_episode(0, True, 42)
anim = animate(frames)
HTML(anim.to_jshtml())
```

Out[14]:





Double DQN (10 points)

In the original paper, where the algorithm is shown above, the estimated target Q value was computed using the current Q network's weights. However, this can lead to overestimation of the Q values. To mitigate this, we can use the target network to compute the target Q value. This is known as Double DQN.

Hard updating Target Network (5 points)

Original implementations for this involved hard updates, where the model weights were copied to the target network every C steps. This is known as hard updating. This was what was used in the Nature Paper by Mnih et al 2015 "Human-level control through deep reinforcement learning"

Please implement this by implementing the `_optimize_model` and `_update_model` classes in `HardUpdateDQN` in `DQN.py`.

```
In [15]: import torch  
import utils  
import DQN  
import model  
from env_wrapper import EnvWrapper  
  
trainerHardUpdatedQDN = DQN.HardUpdatedDQN(EnvWrapper(env),  
model.Nature_Paper_Conv,  
update_freq = 100,  
lr = 0.00025,  
gamma = 0.95,  
buffer_size = 100000,  
batch_size = 32,  
loss_fn = "mse_loss",  
use_wandb = True,  
device = 'cuda',
```

```

seed = 42,
epsilon_scheduler = utils.exponential_decay(1, 1000, 0.1),
save_path = utils.get_save_path("DoubleDQN_HardUpdates/", ".")

trainerHardUpdateDQN.train(200, 50, 30, 50, 50)

```

saving to ./runs/DoubleDQN_HardUpdates/run0
 Finishing previous runs because reinit is set to 'default'.

Run history:



Run summary:

loss	6.63542
mean_reward	623.80108
mean_test_reward	605.07715
std_reward	201.97061
std_test_reward	181.98731
total_reward	242.57962

View run **neat-microwave-2** at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/ajn0n650>

View project at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

Find logs at: ./wandb/run-20250603_185917-ajn0n650/logs

Tracking run with wandb version 0.20.0

Run data is saved locally in

/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250603_200700-u0lwj2ye

Syncing run **charmed-salad-3** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/u0lwj2ye>

Episode: 0: Time: 6.266873121261597 Total Reward: -75.32786885245906 Avg_Loss: 0.0
Episode: 1: Time: 7.877405643463135 Total Reward: -27.65993265993273 Avg_Loss: 0.5757119180456658
Episode: 2: Time: 8.303340673446655 Total Reward: -49.38596491228109 Avg_Loss: 0.6556620778468978
Episode: 3: Time: 8.493849039077759 Total Reward: -25.379746835443456 Avg_Loss: 0.669369897891243
Episode: 4: Time: 8.35913872718811 Total Reward: -30.593220338983627 Avg_Loss: 0.7232532752912586
Episode: 5: Time: 8.36700439453125 Total Reward: -60.18987341772223 Avg_Loss: 0.6708114022279487
Episode: 6: Time: 8.604320049285889 Total Reward: -48.55191256830665 Avg_Loss: 0.6209662687502989
Episode: 7: Time: 8.404366254806519 Total Reward: -40.945945945946264 Avg_Loss: 0.5619077489030462
Episode: 8: Time: 8.311173677444458 Total Reward: -36.60583941605904 Avg_Loss: 0.5841519497334957
Episode: 9: Time: 8.698513507843018 Total Reward: 106.29870129870064 Avg_Loss: 0.6069765026767214
Episode: 10: Time: 8.373558044433594 Total Reward: -30.25179856115177 Avg_Loss: 0.7633207832627437
Episode: 11: Time: 8.466575384140015 Total Reward: -46.298701298701985 Avg_Loss: 0.7232574843058065
Episode: 12: Time: 8.293853759765625 Total Reward: -24.577464788733096 Avg_Loss: 0.7492708108004402
Episode: 13: Time: 8.515872478485107 Total Reward: -16.146953405018504 Avg_Loss: 0.7753119613943982
Episode: 14: Time: 8.757607460021973 Total Reward: 143.09523809524197 Avg_Loss: 0.8553179518139663
Episode: 15: Time: 8.775694131851196 Total Reward: 128.68421052631842 Avg_Loss: 1.0059000729512768
Episode: 16: Time: 8.349969148635864 Total Reward: 33.47222222222406 Avg_Loss: 1.2181579729839533
Episode: 17: Time: 8.585930347442627 Total Reward: -30.018050541516622 Avg_Loss: 1.1749991385125313
Episode: 18: Time: 8.480395793914795 Total Reward: 153.12030075188375 Avg_Loss: 1.3743790159581089
Episode: 19: Time: 8.719565629959106 Total Reward: -16.146953405018692 Avg_Loss: 1.4068439019327403
Episode: 20: Time: 8.536231279373169 Total Reward: -23.571428571429312 Avg_Loss: 1.5648912208921768
Episode: 21: Time: 8.528996229171753 Total Reward: -32.716262975779195 Avg_Loss: 1.694191945701086
Episode: 22: Time: 8.65802550315857 Total Reward: 279.9999999999893 Avg_Loss: 1.8147937647935724
Episode: 23: Time: 8.67793607711792 Total Reward: 201.29629629630028 Avg_Loss: 2.287161427886546
Episode: 24: Time: 8.77269458770752 Total Reward: 114.5238095238133 Avg_Loss: 2.563684720952972
Episode: 25: Time: 8.833858013153076 Total Reward: 99.4444444444487 Avg_Loss: 2.510592081216203
Episode: 26: Time: 8.685228824615479 Total Reward: -38.33333333333414 Avg_Loss: 2.3019592178468944
Episode: 27: Time: 8.704230308532715 Total Reward: 41.51877133106064 Avg_Loss: 1.9497923707010365

Episode: 28: Time: 8.330749034881592 Total Reward: 60.47703180212432 Avg_Loss: 2.4202153859268716
Episode: 29: Time: 8.986355543136597 Total Reward: 286.62544169611596 Avg_Loss: 2.2160942384174893
Episode: 30: Time: 8.18997836112976 Total Reward: -38.521594684386116 Avg_Loss: 2.2618502893117296
Episode: 31: Time: 8.220182418823242 Total Reward: -0.6603773584911372 Avg_Loss: 2.453497886657715
Episode: 32: Time: 8.847923517227173 Total Reward: 157.30769230769445 Avg_Loss: 2.3867683430679705
Episode: 33: Time: 8.66874885559082 Total Reward: 327.14532871971954 Avg_Loss: 2.663296186498233
Episode: 34: Time: 8.927073955535889 Total Reward: 33.12499999999906 Avg_Loss: 2.976613777274845
Episode: 35: Time: 8.83239459991455 Total Reward: 214.76430976431428 Avg_Loss: 3.030421138060193
Episode: 36: Time: 8.79420280456543 Total Reward: 91.13138686131711 Avg_Loss: 3.2349655923723173
Episode: 37: Time: 8.763880729675293 Total Reward: 183.35051546391958 Avg_Loss: 2.977044665638138
Episode: 38: Time: 8.459109544754028 Total Reward: 203.18181818182052 Avg_Loss: 2.9591680324378133
Episode: 39: Time: 8.668036937713623 Total Reward: 226.07023411371725 Avg_Loss: 3.1741584163753926
Episode: 40: Time: 8.843774318695068 Total Reward: 478.37883959043427 Avg_Loss: 3.214887613759321
Episode: 41: Time: 8.592675685882568 Total Reward: 207.63157894737213 Avg_Loss: 3.284646310475694
Episode: 42: Time: 8.550989627838135 Total Reward: 367.8975265017678 Avg_Loss: 3.5160669198557106
Episode: 43: Time: 8.566348314285278 Total Reward: 303.5765124555109 Avg_Loss: 3.638552453838477
Episode: 44: Time: 8.721383810043335 Total Reward: 357.2292993630568 Avg_Loss: 3.504936333964853
Episode: 45: Time: 8.817419052124023 Total Reward: 314.0909090909065 Avg_Loss: 3.8150374926939734
Episode: 46: Time: 8.700479984283447 Total Reward: 399.4237918215538 Avg_Loss: 4.262487466595754
Episode: 47: Time: 8.907272100448608 Total Reward: 375.9302325581332 Avg_Loss: 4.5757001262752945
Episode: 48: Time: 8.633474111557007 Total Reward: 306.9607843137267 Avg_Loss: 4.445991744013393
Episode: 49: Time: 8.868477582931519 Total Reward: -15.000000000000243 Avg_Loss: 4.550339279555473
Validation Mean Reward: -94.999999999999 Validation Std Reward: 3.3927950614478624e-14
Episode: 50: Time: 8.795664310455322 Total Reward: 235.85501858736336 Avg_Loss: 3.77336766940205
Episode: 51: Time: 8.779613494873047 Total Reward: 275.25316455695656 Avg_Loss: 4.194430611213716
Episode: 52: Time: 8.528160572052002 Total Reward: 468.99999999999494 Avg_Loss: 4.062984642361393
Episode: 53: Time: 9.083875179290771 Total Reward: 493.65248226949615 Avg_Loss: 4.050758403890273
Episode: 54: Time: 9.004987955093384 Total Reward: 190.28528528528798 Avg_Loss: 4.740397757842761

Episode: 55: Time: 8.814074039459229 Total Reward: 75.96774193548774 Avg_Loss: 4.611291925446326
Episode: 56: Time: 8.535109996795654 Total Reward: 185.15564202334767 Avg_Loss: 4.445487646996474
Episode: 57: Time: 8.850864887237549 Total Reward: 333.5714285714256 Avg_Loss: 5.015415890877988
Episode: 58: Time: 9.252506017684937 Total Reward: 230.00000000000313 Avg_Loss: 4.4381205394488425
Episode: 59: Time: 8.917612791061401 Total Reward: 298.393393393392 Avg_Loss: 4.433242987183964
Episode: 60: Time: 8.668699741363525 Total Reward: 359.8494983277565 Avg_Loss: 4.489874931443639
Episode: 61: Time: 8.80156946182251 Total Reward: 308.38983050846974 Avg_Loss: 4.613950657243488
Episode: 62: Time: 9.086069107055664 Total Reward: 356.7133956386287 Avg_Loss: 4.3322186900788
Episode: 63: Time: 8.900357484817505 Total Reward: 115.00000000000321 Avg_Loss: 4.539278195685699
Episode: 64: Time: 8.660009860992432 Total Reward: 404.9999999999949 Avg_Loss: 4.32696511014169
Episode: 65: Time: 8.683341264724731 Total Reward: 276.13402061855913 Avg_Loss: 4.945884339949664
Episode: 66: Time: 8.64181399345398 Total Reward: 177.72727272727664 Avg_Loss: 5.1018973353530175
Episode: 67: Time: 8.237791538238525 Total Reward: 332.08333333332627 Avg_Loss: 4.505151315396573
Episode: 68: Time: 8.558026790618896 Total Reward: 376.7607973421861 Avg_Loss: 4.820704914441629
Episode: 69: Time: 8.69057321548462 Total Reward: 371.6666666666612 Avg_Loss: 4.800717422942154
Episode: 70: Time: 8.474892139434814 Total Reward: 191.19528619528919 Avg_Loss: 4.555085998122432
Episode: 71: Time: 8.698236465454102 Total Reward: 185.59701492537695 Avg_Loss: 4.830152309241415
Episode: 72: Time: 8.508115768432617 Total Reward: 95.62500000000185 Avg_Loss: 4.9989546797856566
Episode: 73: Time: 8.605000257492065 Total Reward: 286.5789473684204 Avg_Loss: 4.8216129766792815
Episode: 74: Time: 8.658449649810791 Total Reward: 109.620462046205 Avg_Loss: 4.867231259826853
Episode: 75: Time: 8.379562377929688 Total Reward: 512.287449392704 Avg_Loss: 4.684149982047682
Episode: 76: Time: 8.32052206993103 Total Reward: 539.9809885931481 Avg_Loss: 4.731247884385726
Episode: 77: Time: 8.481495380401611 Total Reward: 199.30379746835635 Avg_Loss: 4.391253661207792
Episode: 78: Time: 8.703047037124634 Total Reward: 169.33121019108663 Avg_Loss: 4.810426331367813
Episode: 79: Time: 8.57757019996643 Total Reward: 253.27586206896876 Avg_Loss: 4.735198689609015
Episode: 80: Time: 8.90942668914795 Total Reward: 469.8854961831971 Avg_Loss: 4.753960826817681
Episode: 81: Time: 8.806813955307007 Total Reward: 368.41463414633836 Avg_Loss: 4.979806552414133
Episode: 82: Time: 8.80982255935669 Total Reward: 358.33333333332604 Avg_Loss: 4.614816772837599

Episode: 83: Time: 8.699279069900513 Total Reward: 246.21621621621617 Avg_Loss: 4.977071939897137
Episode: 84: Time: 8.81799864768982 Total Reward: 285.1369863013699 Avg_Loss: 5.137629239499068
Episode: 85: Time: 9.06259822845459 Total Reward: 302.43589743589894 Avg_Loss: 4.6886935294175345
Episode: 86: Time: 8.741300821304321 Total Reward: 433.30188679244674 Avg_Loss: 4.75967075594333
Episode: 87: Time: 8.876007080078125 Total Reward: 238.33333333333582 Avg_Loss: 4.707275180756545
Episode: 88: Time: 8.679618120193481 Total Reward: 318.1274131274124 Avg_Loss: 4.998854909624372
Episode: 89: Time: 8.56941843032837 Total Reward: 431.6903914590685 Avg_Loss: 5.265257913024485
Episode: 90: Time: 8.587276220321655 Total Reward: 294.5131086142313 Avg_Loss: 5.3138946084415215
Episode: 91: Time: 9.062479734420776 Total Reward: 349.4444444444401 Avg_Loss: 4.849037592150584
Episode: 92: Time: 2.7922046184539795 Total Reward: -69.42815884476568 Avg_Loss: 5.459416762376443
Episode: 93: Time: 8.994069337844849 Total Reward: 289.08304498269956 Avg_Loss: 5.381299365468386
Episode: 94: Time: 8.704679012298584 Total Reward: 323.7725631768913 Avg_Loss: 7.52458951353025
Episode: 95: Time: 8.686879873275757 Total Reward: 448.47826086956013 Avg_Loss: 5.152436749774869
Episode: 96: Time: 8.99607801437378 Total Reward: 280.7961783439511 Avg_Loss: 4.799121176996151
Episode: 97: Time: 9.041385412216187 Total Reward: 110.98006644518294 Avg_Loss: 4.897289478478312
Episode: 98: Time: 9.07898211479187 Total Reward: 249.93670886076328 Avg_Loss: 5.160477481970267
Episode: 99: Time: 8.811305046081543 Total Reward: 324.79522184299805 Avg_Loss: 5.305092581680843
Validation Mean Reward: 393.01034276666144 Validation Std Reward: 217.71491216566633
Episode: 100: Time: 9.301905393600464 Total Reward: 324.1419141914139 Avg_Loss: 5.579878955328164
Episode: 101: Time: 8.352180242538452 Total Reward: -26.654676258993515 Avg_Loss: 5.498592734336853
Episode: 102: Time: 8.80241870880127 Total Reward: 336.0954063604195 Avg_Loss: 4.97672310246139
Episode: 103: Time: 8.971456289291382 Total Reward: 69.17910447761663 Avg_Loss: 4.7959440731200855
Episode: 104: Time: 9.046841144561768 Total Reward: 406.5974440894505 Avg_Loss: 4.913562899878045
Episode: 105: Time: 8.816270589828491 Total Reward: 371.4179104477587 Avg_Loss: 5.19844705066761
Episode: 106: Time: 8.629805088043213 Total Reward: 335.65693430656455 Avg_Loss: 6.422007110940308
Episode: 107: Time: 8.607624292373657 Total Reward: 397.75362318840234 Avg_Loss: 5.111141394166386
Episode: 108: Time: 8.487945795059204 Total Reward: 112.40740740741057 Avg_Loss: 5.426862852913993
Episode: 109: Time: 8.649800062179565 Total Reward: 348.2234432234392 Avg_Loss: 5.361711390879976

Episode: 110: Time: 8.960063457489014 Total Reward: 316.1842105263152 Avg_Loss: 4.792537425746437
Episode: 111: Time: 8.667465925216675 Total Reward: 178.68421052631965 Avg_Loss: 5.728241658511282
Episode: 112: Time: 8.964413404464722 Total Reward: 373.96551724137464 Avg_Loss: 5.474235053823776
Episode: 113: Time: 8.857991695404053 Total Reward: 266.11111111110597 Avg_Loss: 5.696340840403773
Episode: 114: Time: 8.671756267547607 Total Reward: 350.6140350877171 Avg_Loss: 5.635581993255295
Episode: 115: Time: 8.958330869674683 Total Reward: 266.48648648648964 Avg_Loss: 5.608472778516657
Episode: 116: Time: 8.71580457687378 Total Reward: 58.846153846157826 Avg_Loss: 5.326121588214105
Episode: 117: Time: 8.861892461776733 Total Reward: 472.94425087107504 Avg_Loss: 4.927772820997639
Episode: 118: Time: 8.97496223449707 Total Reward: 285.95238095238216 Avg_Loss: 5.18305869613375
Episode: 119: Time: 8.070298433303833 Total Reward: 34.824561403511225 Avg_Loss: 5.118649596927547
Episode: 120: Time: 8.90090274810791 Total Reward: 365.06389776357673 Avg_Loss: 5.507736582715972
Episode: 121: Time: 8.756322145462036 Total Reward: 354.9999999999919 Avg_Loss: 5.287329765928893
Episode: 122: Time: 8.971147060394287 Total Reward: 408.3557046979815 Avg_Loss: 5.3689607241574455
Episode: 123: Time: 9.073268175125122 Total Reward: 313.49673202614355 Avg_Loss: 4.933979367508607
Episode: 124: Time: 8.740240573883057 Total Reward: 290.7677902621708 Avg_Loss: 5.818628150875829
Episode: 125: Time: 8.82532525062561 Total Reward: 342.28813559321145 Avg_Loss: 5.716743645547819
Episode: 126: Time: 8.95838737487793 Total Reward: 259.9488054607546 Avg_Loss: 5.1740575447803785
Episode: 127: Time: 8.832682371139526 Total Reward: 268.29588014981454 Avg_Loss: 5.038969169143869
Episode: 128: Time: 8.866991758346558 Total Reward: 327.48062015503996 Avg_Loss: 5.487644368860902
Episode: 129: Time: 8.960815906524658 Total Reward: 337.2580645161239 Avg_Loss: 5.407619752803771
Episode: 130: Time: 8.739174127578735 Total Reward: 465.43956043955234 Avg_Loss: 5.39601886572958
Episode: 131: Time: 8.494217872619629 Total Reward: 201.2962962962972 Avg_Loss: 4.966861033639988
Episode: 132: Time: 8.809686660766602 Total Reward: 341.89320388348983 Avg_Loss: 5.021890603193716
Episode: 133: Time: 8.684836864471436 Total Reward: 589.2105263157788 Avg_Loss: 5.2610951341500805
Episode: 134: Time: 8.9497971534729 Total Reward: 423.9393939393904 Avg_Loss: 5.268076979312577
Episode: 135: Time: 9.131422996520996 Total Reward: 398.4640522875736 Avg_Loss: 5.392292007177818
Episode: 136: Time: 9.100772857666016 Total Reward: 278.7704918032714 Avg_Loss: 5.580633123381799
Episode: 137: Time: 8.937803745269775 Total Reward: 494.65517241378063 Avg_Loss: 5.77357807179459

Episode: 138: Time: 9.112977743148804 Total Reward: 343.0664652567947 Avg_Loss: 6.186795872800491
Episode: 139: Time: 8.796247243881226 Total Reward: 516.9402985074547 Avg_Loss: 5.980726909236748
Episode: 140: Time: 9.090883731842041 Total Reward: 335.5084745762684 Avg_Loss: 5.579471842581484
Episode: 141: Time: 9.000151872634888 Total Reward: 461.3139931740543 Avg_Loss: 5.399814761486374
Episode: 142: Time: 8.923707008361816 Total Reward: 309.6052631578864 Avg_Loss: 5.489007609231131
Episode: 143: Time: 8.796001672744751 Total Reward: 518.2404181184577 Avg_Loss: 5.4803707439358496
Episode: 144: Time: 9.113590240478516 Total Reward: 166.7647058823561 Avg_Loss: 5.022120198281873
Episode: 145: Time: 8.807538270950317 Total Reward: 325.0000000000017 Avg_Loss: 5.696399968211391
Episode: 146: Time: 8.953400373458862 Total Reward: 240.57046979866087 Avg_Loss: 5.538095957090874
Episode: 147: Time: 9.371840238571167 Total Reward: 269.16184971097863 Avg_Loss: 5.738351973665862
Episode: 148: Time: 8.972156524658203 Total Reward: 404.9999999999267 Avg_Loss: 5.705823469562691
Episode: 149: Time: 9.070889472961426 Total Reward: 376.6981132075348 Avg_Loss: 7.343291616239467
Validation Mean Reward: 541.6028979697245 Validation Std Reward: 143.13115484230534
Episode: 150: Time: 9.312124729156494 Total Reward: 477.41379310343814 Avg_Loss: 6.2417543510429
Episode: 151: Time: 9.260217666625977 Total Reward: 416.7845117845062 Avg_Loss: 6.089014342853001
Episode: 152: Time: 9.269031286239624 Total Reward: 342.4999999999977 Avg_Loss: 5.264343573265717
Episode: 153: Time: 9.193868398666382 Total Reward: 445.7166123778413 Avg_Loss: 5.674996855880032
Episode: 154: Time: 8.91148591041565 Total Reward: 249.26229508197002 Avg_Loss: 5.2319815218949515
Episode: 155: Time: 8.861860513687134 Total Reward: 612.2243346007473 Avg_Loss: 5.148778899377134
Episode: 156: Time: 8.898478746414185 Total Reward: 563.9147286821656 Avg_Loss: 5.520360475828667
Episode: 157: Time: 9.061407566070557 Total Reward: 594.5306859205695 Avg_Loss: 5.9584978368102
Episode: 158: Time: 8.907414197921753 Total Reward: 283.4722222222253 Avg_Loss: 5.996825795213716
Episode: 159: Time: 9.0666024684906 Total Reward: 360.1282051282052 Avg_Loss: 6.626289459837585
Episode: 160: Time: 8.742763996124268 Total Reward: 630.2252252252143 Avg_Loss: 6.426816237072985
Episode: 161: Time: 8.882120609283447 Total Reward: 336.9066147859886 Avg_Loss: 6.362463070564911
Episode: 162: Time: 9.11575722694397 Total Reward: 297.026578073083 Avg_Loss: 5.999397999098321
Episode: 163: Time: 9.348807096481323 Total Reward: 205.8849557522143 Avg_Loss: 5.577792072496495
Episode: 164: Time: 9.173773050308228 Total Reward: 353.90510948904387 Avg_Loss: 6.673233438940609

Episode: 165: Time: 9.167500972747803 Total Reward: 503.59154929576823 Avg_Loss: 6.457473061665767
Episode: 166: Time: 9.357877969741821 Total Reward: 391.06811145510187 Avg_Loss: 5.7700759052228525
Episode: 167: Time: 9.113815069198608 Total Reward: 556.7241379310209 Avg_Loss: 5.232491083505775
Episode: 168: Time: 9.155786514282227 Total Reward: 426.8855218855114 Avg_Loss: 5.679922026746413
Episode: 169: Time: 9.422545194625854 Total Reward: 359.81927710843286 Avg_Loss: 5.681995139402502
Episode: 170: Time: 9.188294410705566 Total Reward: 540.3790613718307 Avg_Loss: 6.259949086093101
Episode: 171: Time: 9.086500644683838 Total Reward: 248.97163120567822 Avg_Loss: 5.9890271505387895
Episode: 172: Time: 9.018906354904175 Total Reward: 549.7368421052543 Avg_Loss: 5.95718705854496
Episode: 173: Time: 8.88513445854187 Total Reward: 534.7577854671212 Avg_Loss: 6.023873747897749
Episode: 174: Time: 8.997731447219849 Total Reward: 571.6666666666624 Avg_Loss: 5.970991446190522
Episode: 175: Time: 9.174432754516602 Total Reward: 509.6511627906874 Avg_Loss: 5.8785796415905995
Episode: 176: Time: 9.353077411651611 Total Reward: 316.5755627009658 Avg_Loss: 6.196056654473312
Episode: 177: Time: 9.110191106796265 Total Reward: 585.85106382978 Avg_Loss: 5.939222324796083
Episode: 178: Time: 9.137876272201538 Total Reward: 495.2777777777688 Avg_Loss: 5.44801377498803
Episode: 179: Time: 9.09185266494751 Total Reward: 272.1328671328708 Avg_Loss: 5.794509790023835
Episode: 180: Time: 9.28247857093811 Total Reward: 339.21052631578425 Avg_Loss: 5.88858454968749
Episode: 181: Time: 9.357555389404297 Total Reward: 336.8936877076314 Avg_Loss: 6.126571315677226
Episode: 182: Time: 9.073419332504272 Total Reward: 623.1467181467056 Avg_Loss: 6.261994555717757
Episode: 183: Time: 9.252177476882935 Total Reward: 477.8155339805762 Avg_Loss: 6.241817652678289
Episode: 184: Time: 9.274469375610352 Total Reward: 411.28930817609404 Avg_Loss: 5.986658345751402
Episode: 185: Time: 9.390977144241333 Total Reward: 205.2915451895049 Avg_Loss: 5.793823542715121
Episode: 186: Time: 9.03321123123169 Total Reward: 584.9999999999869 Avg_Loss: 5.42228479585728
Episode: 187: Time: 9.124595403671265 Total Reward: 618.2616487455099 Avg_Loss: 5.830481381977306
Episode: 188: Time: 9.273937702178955 Total Reward: 529.1830065359371 Avg_Loss: 5.809072942292991
Episode: 189: Time: 9.089905738830566 Total Reward: 619.8148148148014 Avg_Loss: 5.742373068793481
Episode: 190: Time: 8.872382402420044 Total Reward: 540.2941176470497 Avg_Loss: 6.3485914328519035
Episode: 191: Time: 9.39250135421753 Total Reward: 543.7959866220654 Avg_Loss: 6.840180551805416
Episode: 192: Time: 9.129821062088013 Total Reward: 248.7500000000038 Avg_Loss: 6.545473282076731

Episode: 193: Time: 9.196569442749023 Total Reward: 344.560439560428 Avg_Loss: 6.120430489548114
 Episode: 194: Time: 9.339927673339844 Total Reward: 512.9734219268989 Avg_Loss: 6.530639248735764
 Episode: 195: Time: 9.354246854782104 Total Reward: 378.52941176469756 Avg_Loss: 5.968082449516328
 Episode: 196: Time: 8.863152980804443 Total Reward: 457.816901408443 Avg_Loss: 5.998983409224438
 Episode: 197: Time: 9.094125747680664 Total Reward: 562.3208722741333 Avg_Loss: 5.484310730164792
 Episode: 198: Time: 9.246551752090454 Total Reward: 436.5315315315284 Avg_Loss: 6.7771320518325355
 Episode: 199: Time: 9.032127141952515 Total Reward: 524.7183098591468 Avg_Loss: 6.306603697167725
 Validation Mean Reward: 513.33562338453 Validation Std Reward: 170.58048857484997
 Test Mean Reward: 446.53865528587494 Test Std Reward: 175.3391396087704

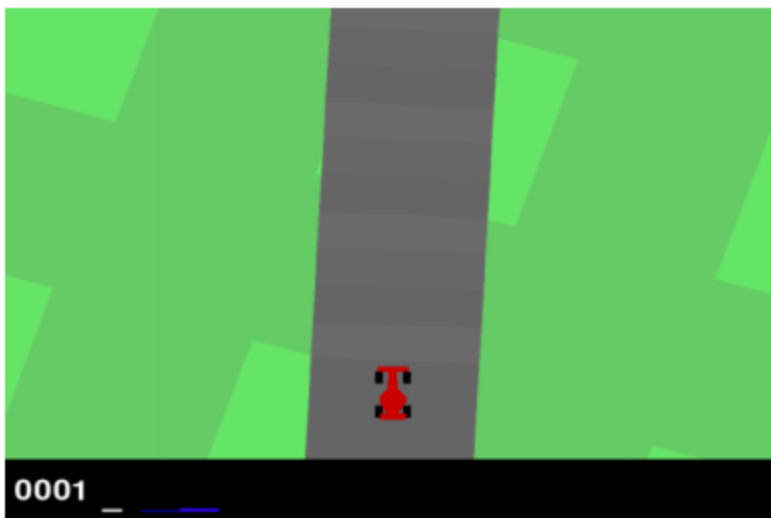
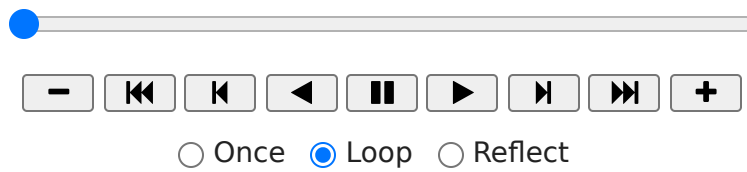
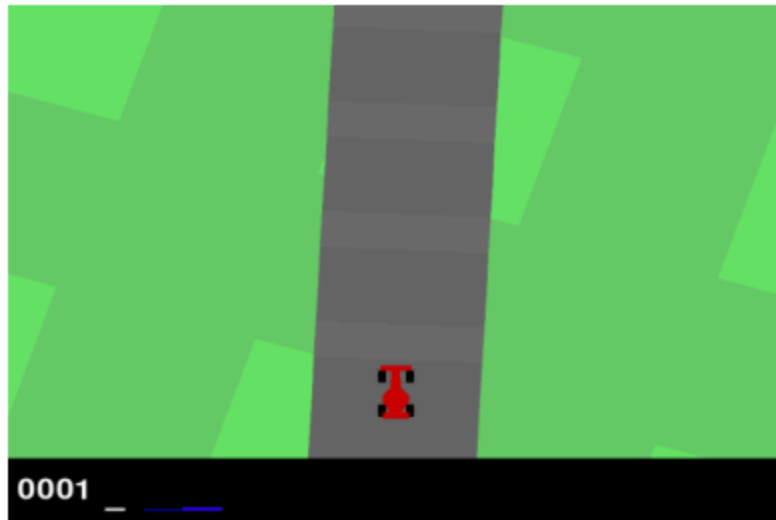
```
In [3]: img_path = os.path.join('.', 'images', 'hard_dqn_stats.png')
display(Image(filename=img_path))
```



Animation

```
In [16]: total_rewards, frames = trainerHardUpdateDQN.play_episode(0, True, 42)
anim = animate(frames)
HTML(anim.to_jshtml())
```

Out[16]:



Soft Updates (5 points)

A more recent improvement is to use soft updates, also known as Polyak averaging, where the target network is updated with a small fraction of the current model weights every step. In other words:

$$\theta_{target} = \tau \theta_{model} + (1 - \tau) \theta_{target}$$

for some $\tau \ll 1$ Please implement this by implementing the `_update_model` class in `SoftUpdatedQN` in `DQN.py`.

```
In [10]: import torch
import utils
import DQN
import model
from env_wrapper import EnvWrapper

traineSoftUpdatedQN = DQN.SoftUpdatedQN(EnvWrapper(env),
                                         model.Nature_Paper_Conv,
                                         tau = 0.01,
                                         update_freq = 1,
                                         lr = 0.00025,
                                         gamma = 0.95,
                                         buffer_size = 100000, # You can take smaller buffer size i
                                         batch_size = 32,
                                         loss_fn = "mse_loss",
                                         use_wandb = True,
                                         device = 'cuda',
                                         seed = 42,
                                         epsilon_scheduler = utils.exponential_decay(1, 1000, 0.1),
                                         save_path = utils.get_save_path("DoubleDQN_SoftUpdates", "./r

traineSoftUpdatedQN.train(200, 50, 30, 50, 50)
```

saving to ./runs/DoubleDQN_SoftUpdates/run0

Finishing previous runs because reinit is set to 'default'.

wandb: ERROR The nbformat package was not found. It is required to save notebook history.

View run **efficient-brook-5** at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/40df1uhb>

View project at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

Find logs at: ./wandb/run-20250604_173438-40df1uhb/logs

Tracking run with wandb version 0.20.0

Run data is saved locally in

/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250604_173507-c98qmc25

Syncing run **vivid-brook-6** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/c98qmc25>

Episode: 0: Time: 6.359175682067871 Total Reward: -33.848920863309544 Avg_Loss: 0.0
Episode: 1: Time: 8.767962455749512 Total Reward: -46.92307692307715 Avg_Loss: 0.7833606781331217
Episode: 2: Time: 9.387411832809448 Total Reward: -29.306569343066098 Avg_Loss: 0.7772658258488699
Episode: 3: Time: 9.651671171188354 Total Reward: -48.57142857142889 Avg_Loss: 0.8115668516574788
Episode: 4: Time: 10.046218395233154 Total Reward: -20.92592592592637 Avg_Loss: 0.7559677272909829
Episode: 5: Time: 10.012659072875977 Total Reward: -77.39436619718307 Avg_Loss: 0.7192318934728118
Episode: 6: Time: 9.452333211898804 Total Reward: -14.84732824427546 Avg_Loss: 0.7523773938980924
Episode: 7: Time: 9.38041639328003 Total Reward: -63.750000000000554 Avg_Loss: 0.7029306743638355
Episode: 8: Time: 9.65608549118042 Total Reward: -62.39130434782662 Avg_Loss: 0.6861154307149538
Episode: 9: Time: 9.332027673721313 Total Reward: -38.18181818181847 Avg_Loss: 0.7263028417565242
Episode: 10: Time: 9.829419136047363 Total Reward: -29.640522875817894 Avg_Loss: 0.776068575617646
Episode: 11: Time: 9.685279369354248 Total Reward: -48.068592057762416 Avg_Loss: 0.8432840440390992
Episode: 12: Time: 10.040468692779541 Total Reward: -22.4924471299094 Avg_Loss: 0.9710026543681362
Episode: 13: Time: 9.927687406539917 Total Reward: 85.88737201365586 Avg_Loss: 1.210566443680715
Episode: 14: Time: 9.614501953125 Total Reward: -33.983050847457775 Avg_Loss: 1.3660759304751868
Episode: 15: Time: 9.682658195495605 Total Reward: -43.5530546623803 Avg_Loss: 1.4533559619879521
Episode: 16: Time: 9.701286315917969 Total Reward: 101.49122807017751 Avg_Loss: 1.7521506163502942
Episode: 17: Time: 9.865921020507812 Total Reward: 135.508474576274 Avg_Loss: 2.1271981648286853
Episode: 18: Time: 9.575285911560059 Total Reward: -23.34890965732152 Avg_Loss: 1.9405558973300356
Episode: 19: Time: 10.124025821685791 Total Reward: 125.06472491909467 Avg_Loss: 2.338648290443821
Episode: 20: Time: 9.978367328643799 Total Reward: 130.35211267605672 Avg_Loss: 2.6769668257036128
Episode: 21: Time: 10.004230737686157 Total Reward: 151.62162162162605 Avg_Loss: 2.740618932647865
Episode: 22: Time: 9.869470357894897 Total Reward: 273.7943262411383 Avg_Loss: 2.9141142794064114
Episode: 23: Time: 9.90914249420166 Total Reward: 26.8637992831531 Avg_Loss: 2.7222721784555612
Episode: 24: Time: 9.94980502128601 Total Reward: 247.9487179487213 Avg_Loss: 2.791672555964534
Episode: 25: Time: 9.785473346710205 Total Reward: 282.28937728937746 Avg_Loss: 2.998690017381636
Episode: 26: Time: 9.756397008895874 Total Reward: 230.17482517482952 Avg_Loss: 3.0527297573931076
Episode: 27: Time: 9.902008295059204 Total Reward: 223.18181818182228 Avg_Loss: 2.9708636235289214

Episode: 28: Time: 9.911519527435303 Total Reward: 130.7053291536085 Avg_Loss: 3.0766998942659685
Episode: 29: Time: 9.761918067932129 Total Reward: 197.25352112676495 Avg_Loss: 2.8807840247114167
Episode: 30: Time: 9.700820446014404 Total Reward: 105.71684587814076 Avg_Loss: 3.26323181641202
Episode: 31: Time: 9.85072135925293 Total Reward: 36.25000000000048 Avg_Loss: 2.7935593373635235
Episode: 32: Time: 10.052670955657959 Total Reward: 76.1026615969624 Avg_Loss: 3.0351208678814543
Episode: 33: Time: 9.913760423660278 Total Reward: 166.4379084967365 Avg_Loss: 3.1833140324143803
Episode: 34: Time: 9.723971366882324 Total Reward: 126.5189873417767 Avg_Loss: 3.150880851415025
Episode: 35: Time: 9.876695156097412 Total Reward: 216.3207547169856 Avg_Loss: 3.1383642013333426
Episode: 36: Time: 9.983841896057129 Total Reward: 7.9900332225909665 Avg_Loss: 3.1697209101765096
Episode: 37: Time: 9.822298765182495 Total Reward: 266.3707165109062 Avg_Loss: 3.000276417792344
Episode: 38: Time: 9.486727952957153 Total Reward: 199.77611940298684 Avg_Loss: 3.0773827646459853
Episode: 39: Time: 9.858022212982178 Total Reward: 189.81012658228312 Avg_Loss: 2.9627301299772344
Episode: 40: Time: 9.776391983032227 Total Reward: 341.36363636363194 Avg_Loss: 3.207382644424919
Episode: 41: Time: 9.52996301651001 Total Reward: 179.80916030534854 Avg_Loss: 3.5288001564370486
Episode: 42: Time: 9.72357726097107 Total Reward: 204.34210526316204 Avg_Loss: 3.2292861986060104
Episode: 43: Time: 9.635736227035522 Total Reward: 199.54545454545814 Avg_Loss: 3.2421289602247607
Episode: 44: Time: 9.677833318710327 Total Reward: 230.42372881356243 Avg_Loss: 3.589244406764247
Episode: 45: Time: 5.585062026977539 Total Reward: -47.935616438356305 Avg_Loss: 3.3950726021738613
Episode: 46: Time: 9.655150651931763 Total Reward: 326.4285714285688 Avg_Loss: 6.927791636030213
Episode: 47: Time: 9.80740475654602 Total Reward: 135.03194888179343 Avg_Loss: 3.7365470273154124
Episode: 48: Time: 9.693898916244507 Total Reward: 115.88435374150055 Avg_Loss: 4.642089680463326
Episode: 49: Time: 9.641628503799438 Total Reward: 211.12244897959587 Avg_Loss: 3.6910098600788275
Validation Mean Reward: 136.51330664357846 Validation Std Reward: 192.55672092175595
Episode: 50: Time: 8.781256437301636 Total Reward: 155.0000000000016 Avg_Loss: 4.444873988127508
Episode: 51: Time: 8.932474613189697 Total Reward: 174.11314984709963 Avg_Loss: 4.464563330431946
Episode: 52: Time: 8.899184226989746 Total Reward: 231.53061224489994 Avg_Loss: 3.9341697662818333
Episode: 53: Time: 8.596878051757812 Total Reward: 98.54838709677597 Avg_Loss: 4.036811327233034
Episode: 54: Time: 8.623188972473145 Total Reward: -32.500000000000504 Avg_Loss: 4.3225767357008795

Episode: 55: Time: 8.595264673233032 Total Reward: 196.3669064748249 Avg_Loss: 4.247139162877026
Episode: 56: Time: 8.458593845367432 Total Reward: 277.4696356275306 Avg_Loss: 3.9663013284947692
Episode: 57: Time: 8.60712194442749 Total Reward: 367.8975265017616 Avg_Loss: 3.902098891614866
Episode: 58: Time: 8.817859411239624 Total Reward: 251.5703971119176 Avg_Loss: 4.2382583007091235
Episode: 59: Time: 8.52751088142395 Total Reward: 416.71874999999386 Avg_Loss: 4.372445583844385
Episode: 60: Time: 8.584330081939697 Total Reward: 432.47252747251923 Avg_Loss: 4.271372173513685
Episode: 61: Time: 8.777506828308105 Total Reward: 208.22580645161702 Avg_Loss: 4.964666343035818
Episode: 62: Time: 8.706783533096313 Total Reward: 324.9999999999931 Avg_Loss: 4.649392831726234
Episode: 63: Time: 8.652416944503784 Total Reward: 272.4911660777413 Avg_Loss: 4.488731570604469
Episode: 64: Time: 8.704713106155396 Total Reward: 374.79865771811393 Avg_Loss: 4.997556201049259
Episode: 65: Time: 6.977598667144775 Total Reward: -34.294936708857406 Avg_Loss: 5.018814799686273
Episode: 66: Time: 9.058687686920166 Total Reward: 224.0883190883235 Avg_Loss: 6.565468735053759
Episode: 67: Time: 8.717163324356079 Total Reward: 372.79661016948825 Avg_Loss: 6.047346511307885
Episode: 68: Time: 8.642010688781738 Total Reward: 335.6049822064028 Avg_Loss: 5.475873236395731
Episode: 69: Time: 8.551136493682861 Total Reward: 423.3673469387678 Avg_Loss: 5.211276100463226
Episode: 70: Time: 8.690770149230957 Total Reward: 597.0152091254627 Avg_Loss: 5.985399157059293
Episode: 71: Time: 8.724921464920044 Total Reward: 423.11594202898004 Avg_Loss: 5.802746419646159
Episode: 72: Time: 8.981839895248413 Total Reward: 193.40125391849966 Avg_Loss: 5.5142255064820045
Episode: 73: Time: 9.062034368515015 Total Reward: 323.0064308681567 Avg_Loss: 5.766712033948979
Episode: 74: Time: 8.97783613204956 Total Reward: 342.4999999999961 Avg_Loss: 6.773057346584416
Episode: 75: Time: 9.093238353729248 Total Reward: 226.7993079584808 Avg_Loss: 5.9324939280998805
Episode: 76: Time: 9.220790386199951 Total Reward: 214.37500000000404 Avg_Loss: 6.001623046498339
Episode: 77: Time: 8.880097150802612 Total Reward: 160.73770491803705 Avg_Loss: 6.039022809317132
Episode: 78: Time: 8.878198146820068 Total Reward: 380.86206896551266 Avg_Loss: 6.3700453022948835
Episode: 79: Time: 8.900635480880737 Total Reward: 141.666666666671 Avg_Loss: 6.540381905411472
Episode: 80: Time: 8.880651950836182 Total Reward: 317.3076923076911 Avg_Loss: 6.328245426426415
Episode: 81: Time: 8.985002040863037 Total Reward: 363.3333333333287 Avg_Loss: 6.243428697606094
Episode: 82: Time: 8.804656028747559 Total Reward: 232.46478873239914 Avg_Loss: 6.274208794120981

Episode: 83: Time: 8.676821947097778 Total Reward: 145.00000000000392 Avg_Loss: 6.163693144541829
Episode: 84: Time: 9.121578454971313 Total Reward: 169.0949554896178 Avg_Loss: 6.124674570660631
Episode: 85: Time: 9.009651899337769 Total Reward: 386.6176470588199 Avg_Loss: 6.255308812906762
Episode: 86: Time: 9.099703311920166 Total Reward: 390.34201954396707 Avg_Loss: 6.632770225781353
Episode: 87: Time: 9.040067911148071 Total Reward: 343.5964912280672 Avg_Loss: 5.933987073537682
Episode: 88: Time: 9.076050281524658 Total Reward: 333.57142857142753 Avg_Loss: 5.837284118187528
Episode: 89: Time: 8.917668104171753 Total Reward: 311.0150375939793 Avg_Loss: 5.9893299301131435
Episode: 90: Time: 9.157706022262573 Total Reward: 390.8934169278975 Avg_Loss: 5.891021738533213
Episode: 91: Time: 8.851157665252686 Total Reward: 60.03875968992677 Avg_Loss: 6.3353149961022766
Episode: 92: Time: 8.564097166061401 Total Reward: 198.04029304029655 Avg_Loss: 5.715947996668455
Episode: 93: Time: 8.902704238891602 Total Reward: 329.6575342465732 Avg_Loss: 5.826893477880654
Episode: 94: Time: 9.16528606414795 Total Reward: 372.1052631578892 Avg_Loss: 6.1401000533785135
Episode: 95: Time: 9.060025930404663 Total Reward: 433.36879432623374 Avg_Loss: 6.04397249822857
Episode: 96: Time: 9.304153680801392 Total Reward: 155.00000000000426 Avg_Loss: 6.020894598560173
Episode: 97: Time: 8.78052830696106 Total Reward: 299.366197183094 Avg_Loss: 5.930347956529184
Episode: 98: Time: 9.158217191696167 Total Reward: 216.178247734142 Avg_Loss: 6.13512929247207
Episode: 99: Time: 8.936630725860596 Total Reward: 333.07017543859183 Avg_Loss: 6.186478532662912
Validation Mean Reward: 468.2316288391237 Validation Std Reward: 164.70728540510885
Episode: 100: Time: 8.90030312538147 Total Reward: 349.06779661016714 Avg_Loss: 6.355496637460564
Episode: 101: Time: 8.818978309631348 Total Reward: 312.66550522647236 Avg_Loss: 6.088183710555069
Episode: 102: Time: 8.863079309463501 Total Reward: 328.2081911262761 Avg_Loss: 6.368436020963332
Episode: 103: Time: 8.852832078933716 Total Reward: 286.75675675675836 Avg_Loss: 6.297761362139918
Episode: 104: Time: 8.892276048660278 Total Reward: 315.59602649006126 Avg_Loss: 6.191193007621445
Episode: 105: Time: 9.028782844543457 Total Reward: 420.9010600706673 Avg_Loss: 6.383728964489047
Episode: 106: Time: 9.372996807098389 Total Reward: 364.15492957745505 Avg_Loss: 6.148156429539208
Episode: 107: Time: 8.906398296356201 Total Reward: 235.88235294118135 Avg_Loss: 6.081189411027091
Episode: 108: Time: 9.06720232963562 Total Reward: 347.56756756756425 Avg_Loss: 6.058901989159464
Episode: 109: Time: 9.074903726577759 Total Reward: 318.19444444444184 Avg_Loss: 6.383517638975833

Episode: 110: Time: 9.299425601959229 Total Reward: 269.43148688046944 Avg_Loss: 6.2037993889896805
Episode: 111: Time: 9.096110820770264 Total Reward: 343.31168831168367 Avg_Loss: 5.7615008314116665
Episode: 112: Time: 9.016019344329834 Total Reward: 342.9310344827546 Avg_Loss: 5.987511538657822
Episode: 113: Time: 9.08167815208435 Total Reward: 486.81818181746 Avg_Loss: 5.922261649820985
Episode: 114: Time: 9.247497797012329 Total Reward: 166.53846153846425 Avg_Loss: 5.830582815058091
Episode: 115: Time: 9.126161813735962 Total Reward: 162.3289902280176 Avg_Loss: 5.872409509009674
Episode: 116: Time: 9.108531951904297 Total Reward: 308.6363636363635 Avg_Loss: 6.199760917855912
Episode: 117: Time: 9.22785234451294 Total Reward: 250.276872964173 Avg_Loss: 6.120997912743512
Episode: 118: Time: 8.984291076660156 Total Reward: 336.50684931506726 Avg_Loss: 5.924262940382757
Episode: 119: Time: 8.692239046096802 Total Reward: 308.5714285714274 Avg_Loss: 6.037426803292346
Episode: 120: Time: 8.863551616668701 Total Reward: 388.55263157894274 Avg_Loss: 5.809913339735079
Episode: 121: Time: 9.175024032592773 Total Reward: 255.00000000000313 Avg_Loss: 6.130678096739184
Episode: 122: Time: 9.261340856552124 Total Reward: 379.5222929936211 Avg_Loss: 5.8207204366932395
Episode: 123: Time: 9.247202634811401 Total Reward: 292.2053872053781 Avg_Loss: 5.673971290347957
Episode: 124: Time: 9.169139623641968 Total Reward: 433.301886792442 Avg_Loss: 5.766151232879703
Episode: 125: Time: 9.141641616821289 Total Reward: 420.15151515150785 Avg_Loss: 5.365705503636048
Episode: 126: Time: 9.2326180934906 Total Reward: 388.18042813455077 Avg_Loss: 5.816592818548699
Episode: 127: Time: 9.23341989517212 Total Reward: 357.2292993630537 Avg_Loss: 5.560486525046725
Episode: 128: Time: 9.1712167263031 Total Reward: 439.6534653465266 Avg_Loss: 5.629712339709787
Episode: 129: Time: 8.97677755355835 Total Reward: 322.5438596491206 Avg_Loss: 5.774518024019835
Episode: 130: Time: 8.933695554733276 Total Reward: 173.55123674912073 Avg_Loss: 5.325212637917335
Episode: 131: Time: 9.155580759048462 Total Reward: 334.5081967213106 Avg_Loss: 5.646345194648294
Episode: 132: Time: 9.31362795829773 Total Reward: 354.0445859872519 Avg_Loss: 5.37570827047364
Episode: 133: Time: 8.975593328475952 Total Reward: 406.7667844522911 Avg_Loss: 6.10733294086296
Episode: 134: Time: 9.141383171081543 Total Reward: 541.6782006920321 Avg_Loss: 6.047196122277684
Episode: 135: Time: 8.80656099319458 Total Reward: 403.12734082396344 Avg_Loss: 6.125536714281354
Episode: 136: Time: 8.8642737865448 Total Reward: 491.61417322833813 Avg_Loss: 6.145409863536098
Episode: 137: Time: 8.561309337615967 Total Reward: 339.2629482071595 Avg_Loss: 5.861530353041256

Episode: 138: Time: 9.221336364746094 Total Reward: 298.2926829268296 Avg_Loss: 5.824019112506835
Episode: 139: Time: 9.203903675079346 Total Reward: 273.0981595092028 Avg_Loss: 5.841262102127075
Episode: 140: Time: 9.491063833236694 Total Reward: 282.7777777777788 Avg_Loss: 6.463505304160238
Episode: 141: Time: 9.082810401916504 Total Reward: 487.7586206896485 Avg_Loss: 6.063040857054606
Episode: 142: Time: 9.238645315170288 Total Reward: 411.6666666666605 Avg_Loss: 5.589861113484166
Episode: 143: Time: 9.385120391845703 Total Reward: 215.44776119403357 Avg_Loss: 6.209480169440518
Episode: 144: Time: 9.046092748641968 Total Reward: 301.61016949152094 Avg_Loss: 5.95375902712846
Episode: 145: Time: 8.845305442810059 Total Reward: 638.6065573770344 Avg_Loss: 6.3967428227432634
Episode: 146: Time: 9.395216464996338 Total Reward: 384.41176470587726 Avg_Loss: 6.040112980273592
Episode: 147: Time: 9.097713708877563 Total Reward: 485.1282051282007 Avg_Loss: 6.06953929552511
Episode: 148: Time: 9.177312850952148 Total Reward: 428.3333333333292 Avg_Loss: 5.740354018051083
Episode: 149: Time: 9.008253574371338 Total Reward: 480.97173144875495 Avg_Loss: 5.788759528087969
Validation Mean Reward: 492.63193041175725 Validation Std Reward: 162.27548870021084
Episode: 150: Time: 9.483519792556763 Total Reward: 375.98976109214163 Avg_Loss: 5.688849335958977
Episode: 151: Time: 9.114443063735962 Total Reward: 696.6666666666538 Avg_Loss: 5.845975727594199
Episode: 152: Time: 9.465603590011597 Total Reward: 412.28862973760516 Avg_Loss: 5.962439875642793
Episode: 153: Time: 8.996148586273193 Total Reward: 498.4065934065835 Avg_Loss: 6.20971939543716
Episode: 154: Time: 9.149848937988281 Total Reward: 346.3793103448271 Avg_Loss: 6.158099134429162
Episode: 155: Time: 9.045654535293579 Total Reward: 302.2602739726029 Avg_Loss: 6.31023413193326
Episode: 156: Time: 9.057140827178955 Total Reward: 446.0447761193939 Avg_Loss: 5.78962278265913
Episode: 157: Time: 9.17162275314331 Total Reward: 351.15384615384403 Avg_Loss: 5.863099718294224
Episode: 158: Time: 8.709758758544922 Total Reward: 387.26950354609664 Avg_Loss: 6.061595663303087
Episode: 159: Time: 8.776059627532959 Total Reward: 368.2352941176398 Avg_Loss: 5.9780009524161075
Episode: 160: Time: 8.953487873077393 Total Reward: 406.73010380622156 Avg_Loss: 5.7029726825842335
Episode: 161: Time: 9.117899894714355 Total Reward: 131.3513513513555 Avg_Loss: 5.791663218446138
Episode: 162: Time: 8.887683391571045 Total Reward: 421.8918918918853 Avg_Loss: 6.05809783033964
Episode: 163: Time: 8.77006983757019 Total Reward: 271.6666666666686 Avg_Loss: 5.987795321380391
Episode: 164: Time: 8.900785684585571 Total Reward: 513.9965397923806 Avg_Loss: 6.038491510543503

Episode: 165: Time: 9.246918201446533 Total Reward: 391.1538461538396 Avg_Loss: 5.911325483762917
Episode: 166: Time: 9.123233079910278 Total Reward: 375.4049844236672 Avg_Loss: 6.092941744988706
Episode: 167: Time: 8.924655675888062 Total Reward: 519.3344709897474 Avg_Loss: 6.117993170473756
Episode: 168: Time: 8.928216934204102 Total Reward: 547.1404682274164 Avg_Loss: 5.816819082789061
Episode: 169: Time: 8.831780433654785 Total Reward: 504.31506849314377 Avg_Loss: 5.991126857885794
Episode: 170: Time: 8.656875371932983 Total Reward: 543.2113821138146 Avg_Loss: 6.41523120082727
Episode: 171: Time: 9.141057014465332 Total Reward: 367.9080118694354 Avg_Loss: 5.72494311492984
Episode: 172: Time: 8.838615894317627 Total Reward: 394.2857142857093 Avg_Loss: 5.961509667524771
Episode: 173: Time: 8.741678476333618 Total Reward: 336.73431734316955 Avg_Loss: 6.000063786987497
Episode: 174: Time: 8.608460903167725 Total Reward: 353.4126984126954 Avg_Loss: 6.0212915043870945
Episode: 175: Time: 8.75484561920166 Total Reward: 373.16479400748716 Avg_Loss: 5.9598151926232985
Episode: 176: Time: 8.829013109207153 Total Reward: 356.9572953736585 Avg_Loss: 6.222374537411858
Episode: 177: Time: 8.732270002365112 Total Reward: 484.3357933579224 Avg_Loss: 5.799065840344469
Episode: 178: Time: 8.965213537216187 Total Reward: 389.94983277591035 Avg_Loss: 6.347992421198292
Episode: 179: Time: 9.032500982284546 Total Reward: 304.33993399340113 Avg_Loss: 5.816112191737199
Episode: 180: Time: 9.248551368713379 Total Reward: 343.87147335422515 Avg_Loss: 6.18525776041656
Episode: 181: Time: 8.801539421081543 Total Reward: 637.8767123287553 Avg_Loss: 5.836492621598124
Episode: 182: Time: 8.812108278274536 Total Reward: 564.7938144329794 Avg_Loss: 6.195139836864311
Episode: 183: Time: 9.2008535861969 Total Reward: 430.8620689655119 Avg_Loss: 6.221855175595324
Episode: 184: Time: 8.66279911994934 Total Reward: 526.3991769547254 Avg_Loss: 6.192809664902567
Episode: 185: Time: 8.784388542175293 Total Reward: 613.4870848708402 Avg_Loss: 6.089628597267535
Episode: 186: Time: 8.896945714950562 Total Reward: 350.85987261145846 Avg_Loss: 6.108258275424733
Episode: 187: Time: 9.187250137329102 Total Reward: 441.36363636362375 Avg_Loss: 5.892622294546175
Episode: 188: Time: 8.87038803100586 Total Reward: 314.7222222222111 Avg_Loss: 6.472536363521544
Episode: 189: Time: 8.716774940490723 Total Reward: 274.71830985915466 Avg_Loss: 6.280903064904093
Episode: 190: Time: 8.684787511825562 Total Reward: 428.4657039711135 Avg_Loss: 6.299746735757139
Episode: 191: Time: 8.902358531951904 Total Reward: 329.6575342465748 Avg_Loss: 6.346883165736158
Episode: 192: Time: 9.003032922744751 Total Reward: 323.23899371068944 Avg_Loss: 6.25149845576086

Episode: 193: Time: 8.74174165725708 Total Reward: 431.31578947367507 Avg_Loss: 6.178008529318481
 Episode: 194: Time: 9.114672899246216 Total Reward: 384.8761609907059 Avg_Loss: 6.738975046061668
 Episode: 195: Time: 8.82404637336731 Total Reward: 463.62068965516164 Avg_Loss: 6.3528156831484885
 Episode: 196: Time: 8.81563925743103 Total Reward: 488.9416058394092 Avg_Loss: 6.172656366805069
 Episode: 197: Time: 9.235471725463867 Total Reward: 263.9743589743562 Avg_Loss: 6.08975508533606
 Episode: 198: Time: 8.825702905654907 Total Reward: 343.20224719100054 Avg_Loss: 6.025180429971519
 Episode: 199: Time: 9.022519826889038 Total Reward: 328.5668789808884 Avg_Loss: 5.779696464538574
 Validation Mean Reward: 487.9650956245007 Validation Std Reward: 134.3409266799912
 Test Mean Reward: 549.5277772028743 Test Std Reward: 156.81387356367148

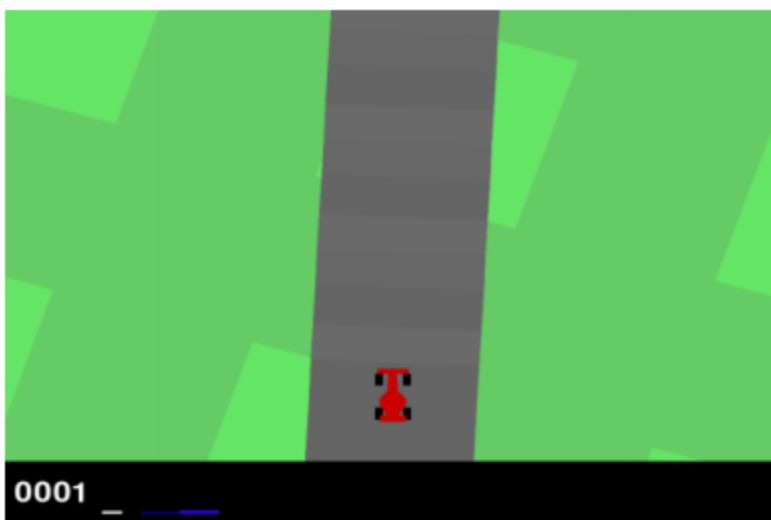
```
In [13]: img_path = os.path.join('.', 'images', 'soft_dqn_stats.png')
display(Image(filename=img_path))
```



Animation

```
In [14]: total_rewards, frames = trainSoftUpdateDQN.play_episode(0, True, 42)
anim = animate(frames)
HTML(anim.to_jshtml())
```


Out[14]:



```
In [15]: img_path = os.path.join('.', 'images', 'results_summary.png')
display(Image(filename=img_path))
```



base_DQN:

Validation Mean Reward: 623.8010792168969 **Validation Std Reward:** 201.97060741012447 **Test Mean Reward:** 605.0771519145571 **Test Std Reward:** 181.9873134653002

hard_update_DQN:

Validation Mean Reward: 513.33562338453 **Validation Std Reward:** 170.58048857484997 **Test Mean Reward:** 446.53865528587494 **Test Std Reward:** 175.3391396087704

soft_update_DQN:

Validation Mean Reward: 487.9650956245007 **Validation Std Reward:** 134.3409266799912 **Test Mean Reward:** 549.5277772028743 **Test Std Reward:** 156.81387356367148

Questions (10 points)

- Which method performed better? (5 points)

Your answer:

- Best overall was the base_DQN which had the highest test mean reward (605.08) albeit with somewhat higher variance.
- Among the variants, softupdateDQN performed better than hard updates on both test mean and variance which demonstrates more stable performance.

- If we modify the τ for soft updates **or** the C for the hard updates, how does this affect the performance of the model, come up with a intuition for this, then experimentally verify this. (5 points)

Your answer:

Intuition:

Update Type	Param	Small Value Effect	Large Value Effect	Paper Implication
Hard	C	Fast, unstable	Stable, slow	Used large C for stability
Soft	τ	Stable, slow	Fast, unstable	Not used, but similar idea

From experiments (C sweep summary - hard update DQN):

Param	mean_test_reward	std_test_reward	Notes
$C = 20$	Highest	Highest	Learns fast, but volatile
$C = 100$	Moderate	Moderate	Balanced, stable learning
$C = 500$	Slightly worse	Lower std	More stable but slower

From experiments (Tau sweep summary - hard update DQN):

Param	mean_test_reward	std_test_reward	Notes
$\tau = 0.05$	Highest	Moderate	Fast learning, decent stability
$\tau = 0.01$	Lower	Lowest	Conservative blending
$\tau = 0.001$	Slightly better than $\tau=0.01$	Highest std	Too slow, underreactive

Based on these results, our intuition held up:

- Hard update DQN performs best at moderate update frequency ($C=100$), balancing stability and learning speed; too frequent ($C=20$) causes instability, while too infrequent ($C=500$) slows convergence.
- Soft update DQN benefits from higher blending rate ($\tau=0.05$), which accelerates learning without excessive variance; very small τ (e.g., 0.001) leads to sluggish, underresponsive updates.

Tau Sweep Summary

```
In [18]: img_path = os.path.join('.', 'images', 'tau_results_summary.png')
display(Image(filename=img_path))
```



```
seed = 42,  
epsilon_scheduler = utils.exponential_decay(1, 1000, 0.1),  
save_path = utils.get_save_path("DoubleDQN_HardUpdates_test/  
trainerHardUpdateDQN.train(200, 50, 30, 50, 50)
```

saving to ./runs/DoubleDQN_HardUpdates_test/run0
Finishing previous runs because reinit is set to 'default'.

wandb: ERROR The nbformat package was not found. It is required to save notebook history.

View run **golden-sun-7** at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/4dulynil>

View project at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

Find logs at: ./wandb/run-20250608_115956-4dulynil/logs

Tracking run with wandb version 0.20.0

Run data is saved locally in

/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250608_120603-wf6uzpa1

Syncing run **lemon-gorge-8** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/wf6uzpa1>

Episode: 0: Time: 6.5495007038116455 Total Reward: -62.625899280576135 Avg_Loss: 0.0
Episode: 1: Time: 8.952810287475586 Total Reward: -43.71794871794898 Avg_Loss: 0.655454914732154
Episode: 2: Time: 9.572886228561401 Total Reward: -32.95620437956232 Avg_Loss: 0.6020928432992777
Episode: 3: Time: 9.546243667602539 Total Reward: -41.428571428571615 Avg_Loss: 0.6830792658406646
Episode: 4: Time: 9.807853698730469 Total Reward: -27.65993265993282 Avg_Loss: 0.8070681379685382
Episode: 5: Time: 9.249284267425537 Total Reward: -49.22535211267672 Avg_Loss: 0.7828165134336768
Episode: 6: Time: 8.746994256973267 Total Reward: -45.38167938931362 Avg_Loss: 0.8369419920657363
Episode: 7: Time: 9.633396863937378 Total Reward: -59.843750000000696 Avg_Loss: 0.6991477351666999
Episode: 8: Time: 9.314960956573486 Total Reward: -47.89855072463837 Avg_Loss: 0.7516652555244059
Episode: 9: Time: 8.835227012634277 Total Reward: -23.030303030303724 Avg_Loss: 0.9756142547776719
Episode: 10: Time: 9.29691767692566 Total Reward: -29.640522875817716 Avg_Loss: 0.9804354608685029
Episode: 11: Time: 9.26706600189209 Total Reward: -51.67870036101155 Avg_Loss: 1.0688985536704545
Episode: 12: Time: 9.32113242149353 Total Reward: -40.619335347432724 Avg_Loss: 1.394059543943956
Episode: 13: Time: 9.310064554214478 Total Reward: -2.849829351536486 Avg_Loss: 1.4449641145452732
Episode: 14: Time: 9.329903602600098 Total Reward: -23.813559322034653 Avg_Loss: 1.267923413016716
Episode: 15: Time: 9.483307600021362 Total Reward: -21.045016077171116 Avg_Loss: 1.1287395877121877
Episode: 16: Time: 9.304885149002075 Total Reward: 45.350877192985386 Avg_Loss: 1.4229860897136837
Episode: 17: Time: 9.840331554412842 Total Reward: 132.11864406780097 Avg_Loss: 1.5506247407874139
Episode: 18: Time: 9.935387372970581 Total Reward: 4.68847352024922 Avg_Loss: 1.9770977775714977
Episode: 19: Time: 9.667186737060547 Total Reward: -62.63754045307509 Avg_Loss: 1.6694817968151148
Episode: 20: Time: 9.90787410736084 Total Reward: 52.88732394366508 Avg_Loss: 1.5985426185261302
Episode: 21: Time: 9.96624493598938 Total Reward: 354.3243243243212 Avg_Loss: 1.7849166542291641
Episode: 22: Time: 9.457099914550781 Total Reward: 167.41134751773538 Avg_Loss: 2.6081341922032735
Episode: 23: Time: 9.894036531448364 Total Reward: 249.0860215053798 Avg_Loss: 2.8840751991051587
Episode: 24: Time: 9.857326030731201 Total Reward: -66.1538461538468 Avg_Loss: 2.7721320475850786
Episode: 25: Time: 9.692387342453003 Total Reward: 322.5824175824173 Avg_Loss: 2.283662050461569
Episode: 26: Time: 9.788154363632202 Total Reward: 111.29370629370575 Avg_Loss: 2.26493535760571
Episode: 27: Time: 9.844635248184204 Total Reward: 191.71328671328777 Avg_Loss: 2.4261854736744857

Episode: 28: Time: 9.960933446884155 Total Reward: 158.91849529780933 Avg_Loss: 2.884026718740704
Episode: 29: Time: 9.880461692810059 Total Reward: 84.5774647887329 Avg_Loss: 3.0460202603780924
Episode: 30: Time: 9.687227725982666 Total Reward: 59.12186379928757 Avg_Loss: 3.0537342213156844
Episode: 31: Time: 9.898269653320312 Total Reward: 42.50000000000008 Avg_Loss: 3.0280169098317122
Episode: 32: Time: 9.808274269104004 Total Reward: 239.6007604562787 Avg_Loss: 3.2752720342964685
Episode: 33: Time: 9.932083368301392 Total Reward: 329.8366013071907 Avg_Loss: 3.4929613744010446
Episode: 34: Time: 10.163429021835327 Total Reward: 202.46835443038498 Avg_Loss: 3.275465806241797
Episode: 35: Time: 9.943961143493652 Total Reward: 125.12578616352641 Avg_Loss: 3.405127404116783
Episode: 36: Time: 10.004844665527344 Total Reward: 300.3488372093022 Avg_Loss: 3.4263984077617904
Episode: 37: Time: 9.9406259059906 Total Reward: 219.6417445482907 Avg_Loss: 3.612440524231486
Episode: 38: Time: 9.560728549957275 Total Reward: 285.5970149253707 Avg_Loss: 4.507235213988969
Episode: 39: Time: 9.980370998382568 Total Reward: 218.29113924051083 Avg_Loss: 4.423728868740947
Episode: 40: Time: 9.88417387008667 Total Reward: 355.90909090908417 Avg_Loss: 4.487317558096237
Episode: 41: Time: 9.654994010925293 Total Reward: 385.916030534347 Avg_Loss: 4.508713839935655
Episode: 42: Time: 9.911081552505493 Total Reward: 309.6052631578914 Avg_Loss: 4.339450985443692
Episode: 43: Time: 9.783241987228394 Total Reward: 355.9090909090895 Avg_Loss: 5.586673086430846
Episode: 44: Time: 9.830735445022583 Total Reward: 335.50847457626924 Avg_Loss: 4.281008003138695
Episode: 45: Time: 10.089322805404663 Total Reward: 315.958904109591 Avg_Loss: 4.300970805793249
Episode: 46: Time: 9.723674774169922 Total Reward: 351.4285714285685 Avg_Loss: 4.577112351645942
Episode: 47: Time: 10.035784721374512 Total Reward: 400.2076677316242 Avg_Loss: 4.965706198155379
Episode: 48: Time: 9.785735845565796 Total Reward: 177.10884353741656 Avg_Loss: 4.817072315877225
Episode: 49: Time: 9.753089904785156 Total Reward: 139.69387755102017 Avg_Loss: 5.15777806574557
Validation Mean Reward: 376.79806168108155 Validation Std Reward: 162.15430051107649
Episode: 50: Time: 9.312292575836182 Total Reward: 246.2162162162207 Avg_Loss: 4.726297724647682
Episode: 51: Time: 9.159210443496704 Total Reward: 333.13455657492085 Avg_Loss: 5.195189803087411
Episode: 52: Time: 8.899564266204834 Total Reward: 313.16326530612366 Avg_Loss: 5.7420914764163875
Episode: 53: Time: 8.632073402404785 Total Reward: 180.9856630824416 Avg_Loss: 5.164476951130298
Episode: 54: Time: 8.833472490310669 Total Reward: 397.6470588235246 Avg_Loss: 6.286344790659031

Episode: 55: Time: 8.770053386688232 Total Reward: 333.0575539568341 Avg_Loss: 6.734343190153106
Episode: 56: Time: 8.744435548782349 Total Reward: 168.15789473684652 Avg_Loss: 5.154827128939268
Episode: 57: Time: 8.724367618560791 Total Reward: 396.1660777385148 Avg_Loss: 4.738316903595163
Episode: 58: Time: 8.313927412033081 Total Reward: 85.50541516245846 Avg_Loss: 4.620431852941754
Episode: 59: Time: 8.664697170257568 Total Reward: 346.40625000000006 Avg_Loss: 4.092609975017419
Episode: 60: Time: 8.624260425567627 Total Reward: 458.11355311354225 Avg_Loss: 4.444923300703032
Episode: 61: Time: 8.959390878677368 Total Reward: 208.22580645161688 Avg_Loss: 4.204730290825627
Episode: 62: Time: 8.859929084777832 Total Reward: 448.33333333332234 Avg_Loss: 3.869640484076588
Episode: 63: Time: 8.629826307296753 Total Reward: 438.5689045936297 Avg_Loss: 4.663029445319617
Episode: 64: Time: 8.674881219863892 Total Reward: 381.5100671140887 Avg_Loss: 4.830588714415286
Episode: 65: Time: 8.940448522567749 Total Reward: 363.8607594936642 Avg_Loss: 4.994564618383135
Episode: 66: Time: 9.107543706893921 Total Reward: 443.4615384615305 Avg_Loss: 5.006062305775009
Episode: 67: Time: 8.77268123626709 Total Reward: 477.88135593219675 Avg_Loss: 5.012529585040918
Episode: 68: Time: 8.731301307678223 Total Reward: 442.366548042701 Avg_Loss: 4.357497500772236
Episode: 69: Time: 8.57067608833313 Total Reward: 586.632653061213 Avg_Loss: 5.241404844933197
Episode: 70: Time: 8.815603494644165 Total Reward: 547.5855513307906 Avg_Loss: 5.064570444471696
Episode: 71: Time: 8.840604305267334 Total Reward: 412.2463768115881 Avg_Loss: 5.264630345236354
Episode: 72: Time: 9.049682140350342 Total Reward: 365.81504702193956 Avg_Loss: 5.340474135735455
Episode: 73: Time: 9.046900510787964 Total Reward: 245.83601286173968 Avg_Loss: 5.982070214107257
Episode: 74: Time: 8.800192832946777 Total Reward: 370.2777777777753 Avg_Loss: 5.518221905251511
Episode: 75: Time: 9.028656005859375 Total Reward: 368.66782006919516 Avg_Loss: 5.167770833528342
Episode: 76: Time: 9.112618446350098 Total Reward: 379.99999999999636 Avg_Loss: 5.435021523667984
Episode: 77: Time: 9.064903497695923 Total Reward: 364.01639344261827 Avg_Loss: 4.675868606367031
Episode: 78: Time: 8.93052339553833 Total Reward: 353.2758620689642 Avg_Loss: 5.005740009936966
Episode: 79: Time: 9.037742376327515 Total Reward: 291.6666666666636 Avg_Loss: 5.010442207340433
Episode: 80: Time: 9.073769330978394 Total Reward: 120.38461538461897 Avg_Loss: 5.3990987559326555
Episode: 81: Time: 8.979736804962158 Total Reward: 369.7435897435853 Avg_Loss: 4.778115665712276
Episode: 82: Time: 8.757277250289917 Total Reward: 422.6056338028118 Avg_Loss: 5.361994026087913

Episode: 83: Time: 8.740690231323242 Total Reward: 241.6666666666695 Avg_Loss: 4.844394313688038
Episode: 84: Time: 9.028905868530273 Total Reward: 507.37388724034963 Avg_Loss: 5.485330219028377
Episode: 85: Time: 8.7088942527771 Total Reward: 618.2352941176345 Avg_Loss: 4.439107630934034
Episode: 86: Time: 8.790470361709595 Total Reward: 334.96742671009406 Avg_Loss: 5.22287832338269
Episode: 87: Time: 8.84617280960083 Total Reward: 371.6666666666593 Avg_Loss: 4.951188658465858
Episode: 88: Time: 9.0144944190979 Total Reward: 353.0519480519474 Avg_Loss: 4.965974067439552
Episode: 89: Time: 8.862312078475952 Total Reward: 435.07518796991116 Avg_Loss: 5.376831813519742
Episode: 90: Time: 9.088538646697998 Total Reward: 303.1191222570494 Avg_Loss: 5.878577851447739
Episode: 91: Time: 8.677466869354248 Total Reward: 373.9922480620108 Avg_Loss: 5.384941730679584
Episode: 92: Time: 8.746922492980957 Total Reward: 549.6886446886349 Avg_Loss: 5.1039762171376655
Episode: 93: Time: 8.7941312789917 Total Reward: 524.8630136986222 Avg_Loss: 5.056091501933186
Episode: 94: Time: 8.964143991470337 Total Reward: 497.10526315788644 Avg_Loss: 5.78759402787986
Episode: 95: Time: 9.0098717212677 Total Reward: 461.737588652475 Avg_Loss: 4.84402595347717
Episode: 96: Time: 9.249739646911621 Total Reward: 355.6172839506156 Avg_Loss: 5.271173203692717
Episode: 97: Time: 8.923067808151245 Total Reward: 436.6901408450597 Avg_Loss: 5.4071093226681235
Episode: 98: Time: 9.256075859069824 Total Reward: 270.5589123867046 Avg_Loss: 5.226254833846533
Episode: 99: Time: 9.088696241378784 Total Reward: 452.3684210526226 Avg_Loss: 5.455139267845314
Validation Mean Reward: 563.1704624908373 Validation Std Reward: 112.9437492163582
Episode: 100: Time: 9.175991296768188 Total Reward: 189.74576271186888 Avg_Loss: 5.625657477298705
Episode: 101: Time: 9.045287132263184 Total Reward: 364.93031358884355 Avg_Loss: 5.7727654441064145
Episode: 102: Time: 9.133816242218018 Total Reward: 464.7269624573265 Avg_Loss: 5.691650607505767
Episode: 103: Time: 9.02588176727295 Total Reward: 303.6486486486431 Avg_Loss: 5.931471967396616
Episode: 104: Time: 9.087037324905396 Total Reward: 358.64238410595647 Avg_Loss: 5.10028039657769
Episode: 105: Time: 9.090594291687012 Total Reward: 378.49823321554436 Avg_Loss: 5.1675359020714
Episode: 106: Time: 9.312061786651611 Total Reward: 431.7605633802747 Avg_Loss: 5.218024882949701
Episode: 107: Time: 8.822723388671875 Total Reward: 452.7941176470533 Avg_Loss: 6.10216859609139
Episode: 108: Time: 8.979605436325073 Total Reward: 587.432432432424 Avg_Loss: 6.109134253333597
Episode: 109: Time: 9.101206064224243 Total Reward: 248.75000000000145 Avg_Loss: 5.827900310524371

Episode: 110: Time: 9.350907802581787 Total Reward: 161.5597667638528 Avg_Loss: 5.778121238997002
Episode: 111: Time: 9.11279010772705 Total Reward: 340.0649350649344 Avg_Loss: 5.398849215327191
Episode: 112: Time: 8.999662160873413 Total Reward: 380.8620689655113 Avg_Loss: 5.3346634872821195
Episode: 113: Time: 8.86205768585205 Total Reward: 206.81818181818596 Avg_Loss: 5.4340534385512855
Episode: 114: Time: 9.120033740997314 Total Reward: 206.53846153846553 Avg_Loss: 5.9793353311154025
Episode: 115: Time: 9.068922281265259 Total Reward: 501.0912052117196 Avg_Loss: 5.137741150976229
Episode: 116: Time: 8.813103914260864 Total Reward: 337.72727272727195 Avg_Loss: 5.255498219938839
Episode: 117: Time: 8.978428363800049 Total Reward: 214.4462540716658 Avg_Loss: 5.806656791382477
Episode: 118: Time: 8.81026029586792 Total Reward: 387.87671232875994 Avg_Loss: 5.547824719372918
Episode: 119: Time: 8.731915950775146 Total Reward: 508.5714285714171 Avg_Loss: 6.229006130655272
Episode: 120: Time: 8.770632028579712 Total Reward: 414.86842105262787 Avg_Loss: 5.742734715718181
Episode: 121: Time: 9.07791781425476 Total Reward: 419.7058823529335 Avg_Loss: 5.670873563329713
Episode: 122: Time: 9.034256935119629 Total Reward: 465.5095541401133 Avg_Loss: 5.577330562246948
Episode: 123: Time: 9.021939277648926 Total Reward: 295.5723905723843 Avg_Loss: 5.57174618805156
Episode: 124: Time: 8.923874855041504 Total Reward: 530.7861635220012 Avg_Loss: 5.487766927029906
Episode: 125: Time: 8.801369667053223 Total Reward: 319.1414141414166 Avg_Loss: 5.74230088606602
Episode: 126: Time: 8.905743598937988 Total Reward: 333.13455657491903 Avg_Loss: 5.3411065400147635
Episode: 127: Time: 8.98410677909851 Total Reward: 274.42675159234847 Avg_Loss: 6.29606388897455
Episode: 128: Time: 8.90984058380127 Total Reward: 350.54455445544113 Avg_Loss: 5.732430378929908
Episode: 129: Time: 8.69679880142212 Total Reward: 522.5438596491148 Avg_Loss: 5.53923446490985
Episode: 130: Time: 8.749695539474487 Total Reward: 583.445229681969 Avg_Loss: 5.584022810479172
Episode: 131: Time: 8.929392576217651 Total Reward: 459.09836065573 Avg_Loss: 5.822902506139098
Episode: 132: Time: 9.021062850952148 Total Reward: 264.87261146496985 Avg_Loss: 5.616636204118488
Episode: 133: Time: 8.618757486343384 Total Reward: 424.4346289752594 Avg_Loss: 5.854095436444803
Episode: 134: Time: 8.837188959121704 Total Reward: 406.7301038062186 Avg_Loss: 5.4884401369495555
Episode: 135: Time: 8.621813535690308 Total Reward: 358.1835205992434 Avg_Loss: 5.204802545178838
Episode: 136: Time: 8.673527956008911 Total Reward: 853.8188976377841 Avg_Loss: 6.0595745900098015
Episode: 137: Time: 8.592684745788574 Total Reward: 486.673306772902 Avg_Loss: 6.09539388458268

Episode: 138: Time: 9.170080423355103 Total Reward: 484.2682926829189 Avg_Loss: 5.750886100680888
Episode: 139: Time: 8.99093508720398 Total Reward: 365.12269938650195 Avg_Loss: 6.117053145120124
Episode: 140: Time: 9.292836666107178 Total Reward: 346.6666666666607 Avg_Loss: 5.611735011850085
Episode: 141: Time: 8.828991651535034 Total Reward: 484.31034482757775 Avg_Loss: 5.85176264939188
Episode: 142: Time: 9.030216932296753 Total Reward: 501.6666666666574 Avg_Loss: 5.839071276808987
Episode: 143: Time: 9.201725721359253 Total Reward: 319.92537313432695 Avg_Loss: 6.134836316108704
Episode: 144: Time: 8.839284658432007 Total Reward: 338.89830508473875 Avg_Loss: 6.0243308123420265
Episode: 145: Time: 8.683513402938843 Total Reward: 564.8360655737574 Avg_Loss: 5.785106777143078
Episode: 146: Time: 9.341193199157715 Total Reward: 499.1176470588143 Avg_Loss: 5.490948895446393
Episode: 147: Time: 9.074527740478516 Total Reward: 417.82051282050793 Avg_Loss: 6.078458813058228
Episode: 148: Time: 9.171911478042603 Total Reward: 528.3333333333211 Avg_Loss: 5.8246965423351575
Episode: 149: Time: 9.049257516860962 Total Reward: 378.4982332155387 Avg_Loss: 5.284171712498705
Validation Mean Reward: 475.74485229204504 Validation Std Reward: 211.1280351266616
Episode: 150: Time: 9.455951690673828 Total Reward: 416.94539249146254 Avg_Loss: 5.6597360432648856
Episode: 151: Time: 9.274663925170898 Total Reward: 476.9696969696909 Avg_Loss: 6.213148237276478
Episode: 152: Time: 9.412887573242188 Total Reward: 208.20699708455106 Avg_Loss: 6.432444860454367
Episode: 153: Time: 8.942827463150024 Total Reward: 571.6666666666586 Avg_Loss: 6.348932544223401
Episode: 154: Time: 9.127218008041382 Total Reward: 460.1724137930977 Avg_Loss: 5.56984365487299
Episode: 155: Time: 9.083600759506226 Total Reward: 288.56164383561315 Avg_Loss: 6.004022382888474
Episode: 156: Time: 9.098911762237549 Total Reward: 502.01492537312697 Avg_Loss: 5.746527579652161
Episode: 157: Time: 9.077198028564453 Total Reward: 169.6153846153885 Avg_Loss: 6.018673110408943
Episode: 158: Time: 8.935819387435913 Total Reward: 518.4751773049586 Avg_Loss: 5.434718360420034
Episode: 159: Time: 8.880932092666626 Total Reward: 397.6470588235231 Avg_Loss: 6.029916905555405
Episode: 160: Time: 9.13607144355774 Total Reward: 486.3148788927235 Avg_Loss: 6.062339624436963
Episode: 161: Time: 9.218574523925781 Total Reward: 418.51351351350985 Avg_Loss: 5.245007499927232
Episode: 162: Time: 9.075580596923828 Total Reward: 269.8648648648677 Avg_Loss: 6.514530604126072
Episode: 163: Time: 9.109979391098022 Total Reward: 488.33333333332223 Avg_Loss: 5.979726872524293
Episode: 164: Time: 9.067340850830078 Total Reward: 486.31487889272614 Avg_Loss: 5.827657920973642

Episode: 165: Time: 9.432322978973389 Total Reward: 360.38461538460604 Avg_Loss: 6.214004869721517
Episode: 166: Time: 9.178732872009277 Total Reward: 534.2834890965626 Avg_Loss: 7.3079368757600545
Episode: 167: Time: 9.19777226448059 Total Reward: 423.7713310580111 Avg_Loss: 6.774131379708522
Episode: 168: Time: 9.09201455116272 Total Reward: 500.3177257525025 Avg_Loss: 6.630632963501105
Episode: 169: Time: 9.066387414932251 Total Reward: 343.3561643835604 Avg_Loss: 6.031928031384444
Episode: 170: Time: 8.991393327713013 Total Reward: 531.0162601625879 Avg_Loss: 5.949311156232818
Episode: 171: Time: 9.33280324935913 Total Reward: -3.0118694362024083 Avg_Loss: 5.9842137258593775
Episode: 172: Time: 9.114694118499756 Total Reward: 590.7142857142743 Avg_Loss: 5.711179994735398
Episode: 173: Time: 8.994828224182129 Total Reward: 499.09594095940224 Avg_Loss: 5.87244689564745
Episode: 174: Time: 8.893805742263794 Total Reward: 615.3174603174494 Avg_Loss: 5.599986789106321
Episode: 175: Time: 9.00933313369751 Total Reward: 826.3483146067322 Avg_Loss: 5.94033504834696
Episode: 176: Time: 9.15973448753357 Total Reward: 595.391459074724 Avg_Loss: 5.89575094735923
Episode: 177: Time: 9.070490598678589 Total Reward: 628.2472324723176 Avg_Loss: 6.7693416641539885
Episode: 178: Time: 9.305131912231445 Total Reward: 359.84949832775476 Avg_Loss: 6.0214857454059505
Episode: 179: Time: 9.260103464126587 Total Reward: 479.25742574256924 Avg_Loss: 6.315702683785382
Episode: 180: Time: 9.540346145629883 Total Reward: 312.5235109717791 Avg_Loss: 6.721808438541508
Episode: 181: Time: 8.951114654541016 Total Reward: 651.575342465743 Avg_Loss: 6.179653552399964
Episode: 182: Time: 9.023867130279541 Total Reward: 547.6116838487935 Avg_Loss: 6.049509453172443
Episode: 183: Time: 9.413198471069336 Total Reward: 272.81609195401387 Avg_Loss: 6.1105186748905345
Episode: 184: Time: 8.820467948913574 Total Reward: 509.93827160493083 Avg_Loss: 6.494148974659062
Episode: 185: Time: 9.031449794769287 Total Reward: 650.387453874529 Avg_Loss: 6.452501959159594
Episode: 186: Time: 9.496684789657593 Total Reward: 427.2929936305645 Avg_Loss: 6.522313643904293
Episode: 187: Time: 9.415345191955566 Total Reward: 404.9999999999937 Avg_Loss: 7.05122645061557
Episode: 188: Time: 9.147208213806152 Total Reward: 502.222222222121 Avg_Loss: 6.277145289573349
Episode: 189: Time: 9.179821014404297 Total Reward: 545.8450704225239 Avg_Loss: 6.659776904001958
Episode: 190: Time: 8.973827600479126 Total Reward: 266.01083032491255 Avg_Loss: 6.169942862847272
Episode: 191: Time: 9.324206352233887 Total Reward: 305.6849315068482 Avg_Loss: 6.523924737417397
Episode: 192: Time: 9.245073318481445 Total Reward: 269.7798742138383 Avg_Loss: 6.584982914083144

Episode: 193: Time: 9.061000108718872 Total Reward: 442.5939849624005 Avg_Loss: 7.013026714324951
 Episode: 194: Time: 9.383429050445557 Total Reward: 428.2198142414835 Avg_Loss: 7.568753500946429
 Episode: 195: Time: 9.190256595611572 Total Reward: 470.51724137929637 Avg_Loss: 6.755424574643624
 Episode: 196: Time: 9.157865762710571 Total Reward: 419.5985401459824 Avg_Loss: 7.006165170869908
 Episode: 197: Time: 9.459447860717773 Total Reward: 204.145299145304 Avg_Loss: 6.930825941702899
 Episode: 198: Time: 9.065754175186157 Total Reward: 635.3370786516743 Avg_Loss: 7.33348697373847
 Episode: 199: Time: 9.326997756958008 Total Reward: 280.7961783439492 Avg_Loss: 7.329927698904727
 Validation Mean Reward: 667.1884506460161 Validation Std Reward: 163.57638182437205
 Test Mean Reward: 615.0866983203043 Test Std Reward: 181.67272041497947

```

In [13]: trainerHardUpdateDQN = DQN.HardUpdateDQN(EnvWrapper(env),
          model.Nature_Paper_Conv,
          update_freq = 500,                                # Change this
          lr = 0.00025,
          gamma = 0.95,
          buffer_size = 100000,
          batch_size = 32,
          loss_fn = "mse_loss",
          use_wandb = True,
          device = 'cuda',
          seed = 42,
          epsilon_scheduler = utils.exponential_decay(1, 1000, 0.1),
          save_path = utils.get_save_path("DoubleDQN_HardUpdates_test/

trainerHardUpdateDQN.train(200, 50, 30, 50, 50)
  
```

saving to ./runs/DoubleDQN_HardUpdates_test/run1
 Finishing previous runs because reinit is set to 'default'.

Run history:



Run summary:

loss	7.32993
mean_reward	667.18845
mean_test_reward	615.0867
std_reward	163.57638
std_test_reward	181.67272
total_reward	280.79618

View run **lemon-gorge-8** at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/wf6uzpa1>

View project at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

Find logs at: ./wandb/run-20250608_120603-wf6uzpa1/logs

Tracking run with wandb version 0.20.0

Run data is saved locally in

/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250608_130613-j8bv0sy5

Syncing run **polar-snowball-9** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/j8bv0sy5>

Episode: 0: Time: 6.4389026165008545 Total Reward: -52.031250000000355 Avg_Loss: 0.0
Episode: 1: Time: 8.255250930786133 Total Reward: -31.39575971731456 Avg_Loss: 0.7503820823588568
Episode: 2: Time: 8.446129560470581 Total Reward: -30.018050541516196 Avg_Loss: 0.9030566392701214
Episode: 3: Time: 8.575628519058228 Total Reward: -61.777408637874416 Avg_Loss: 0.7929545947734047
Episode: 4: Time: 8.874367952346802 Total Reward: -62.21311475409895 Avg_Loss: 0.6885187760339815
Episode: 5: Time: 8.308270692825317 Total Reward: -32.69470404984493 Avg_Loss: 0.7193250393322786
Episode: 6: Time: 8.785995721817017 Total Reward: -40.255474452555205 Avg_Loss: 0.6230055453380993
Episode: 7: Time: 8.85183048248291 Total Reward: -76.07255520504734 Avg_Loss: 0.5502683583240048
Episode: 8: Time: 8.453720569610596 Total Reward: -28.333333333333783 Avg_Loss: 0.609753114069585
Episode: 9: Time: 8.93119192123413 Total Reward: 10.431309904152254 Avg_Loss: 0.5872186506088791
Episode: 10: Time: 8.756636619567871 Total Reward: -57.06896551724209 Avg_Loss: 0.6360740925365386
Episode: 11: Time: 9.088459968566895 Total Reward: 94.60244648318341 Avg_Loss: 0.6494725785773843
Episode: 12: Time: 8.619980573654175 Total Reward: 85.64516129032583 Avg_Loss: 0.7254321983694529
Episode: 13: Time: 8.178656339645386 Total Reward: -52.894736842105914 Avg_Loss: 0.7065953154022954
Episode: 14: Time: 8.790500402450562 Total Reward: 16.782477341389598 Avg_Loss: 0.7206518024581821
Episode: 15: Time: 8.151129722595215 Total Reward: -23.31541218638067 Avg_Loss: 0.6803926380366838
Episode: 16: Time: 8.215713024139404 Total Reward: -17.178988326848778 Avg_Loss: 0.7411864735968724
Episode: 17: Time: 9.105412006378174 Total Reward: 1.3855421686740428 Avg_Loss: 0.658576260375626
Episode: 18: Time: 8.943377494812012 Total Reward: 52.39884393063487 Avg_Loss: 0.7845273032596632
Episode: 19: Time: 8.635081768035889 Total Reward: 267.59541984731777 Avg_Loss: 0.8970033222762477
Episode: 20: Time: 8.745373725891113 Total Reward: 37.075471698114995 Avg_Loss: 0.9536759690577242
Episode: 21: Time: 8.909645318984985 Total Reward: 113.07453416149343 Avg_Loss: 0.9308270411644163
Episode: 22: Time: 8.613789796829224 Total Reward: 432.55905511810096 Avg_Loss: 1.1908370278963523
Episode: 23: Time: 8.76596736907959 Total Reward: 125.1492537313477 Avg_Loss: 1.3503276608696504
Episode: 24: Time: 8.57244324684143 Total Reward: 226.67832167832591 Avg_Loss: 1.4928266106783843
Episode: 25: Time: 8.492033243179321 Total Reward: 136.9391634981026 Avg_Loss: 1.5238783766492074
Episode: 26: Time: 8.745815515518188 Total Reward: 41.18677042801822 Avg_Loss: 1.6012702897316267
Episode: 27: Time: 8.710015535354614 Total Reward: 459.34782608695195 Avg_Loss: 1.6722903855207587

Episode: 28: Time: 8.557602882385254 Total Reward: 8.806228373701645 Avg_Loss: 1.6337596179307008
Episode: 29: Time: 8.523100852966309 Total Reward: 15.687022900763038 Avg_Loss: 1.9579324306560164
Episode: 30: Time: 8.52834677696228 Total Reward: 234.41176470588397 Avg_Loss: 1.718363927066827
Episode: 31: Time: 8.631817102432251 Total Reward: -36.605839416058856 Avg_Loss: 2.0488568672112057
Episode: 32: Time: 8.816261053085327 Total Reward: 97.72727272727688 Avg_Loss: 2.1250939324122515
Episode: 33: Time: 8.746047019958496 Total Reward: 595.0369003689955 Avg_Loss: 2.3316052912914453
Episode: 34: Time: 8.917491912841797 Total Reward: 33.37837837837956 Avg_Loss: 2.3205370159209275
Episode: 35: Time: 8.718326568603516 Total Reward: 168.88888888889323 Avg_Loss: 2.50243496419001
Episode: 36: Time: 8.639442920684814 Total Reward: 343.4057971014438 Avg_Loss: 2.378616181491804
Episode: 37: Time: 8.743472337722778 Total Reward: 220.78947368421342 Avg_Loss: 2.737858040242636
Episode: 38: Time: 8.643094301223755 Total Reward: 147.10526315789866 Avg_Loss: 2.640550151592543
Episode: 39: Time: 8.513446569442749 Total Reward: 154.1582491582534 Avg_Loss: 2.931183828776624
Episode: 40: Time: 8.700028896331787 Total Reward: 192.71929824561482 Avg_Loss: 2.716281865324293
Episode: 41: Time: 8.761762857437134 Total Reward: 278.98373983739873 Avg_Loss: 2.848226918893702
Episode: 42: Time: 8.951642751693726 Total Reward: 212.16723549488492 Avg_Loss: 3.0264076938649187
Episode: 43: Time: 8.73553204536438 Total Reward: 143.59649122807326 Avg_Loss: 3.3743660364832198
Episode: 44: Time: 8.783403873443604 Total Reward: 116.72638436482461 Avg_Loss: 3.2453789039820182
Episode: 45: Time: 8.832040071487427 Total Reward: 196.9463087248351 Avg_Loss: 3.293572170894687
Episode: 46: Time: 8.768513917922974 Total Reward: 3.3050847457626666 Avg_Loss: 3.2602851881700405
Episode: 47: Time: 8.928743124008179 Total Reward: 408.24675324674485 Avg_Loss: 3.308495701611543
Episode: 48: Time: 8.572720766067505 Total Reward: 543.4615384615325 Avg_Loss: 3.5660885782802807
Episode: 49: Time: 8.73019552230835 Total Reward: 199.7761194029885 Avg_Loss: 3.6786377925832734
Validation Mean Reward: 308.80541492418496 Validation Std Reward: 192.87331113915556
Episode: 50: Time: 8.96383023262024 Total Reward: 306.8987341772157 Avg_Loss: 3.9369510199342455
Episode: 51: Time: 9.097306728363037 Total Reward: 276.42857142857224 Avg_Loss: 3.969771259973029
Episode: 52: Time: 8.670980215072632 Total Reward: 439.79853479852886 Avg_Loss: 4.198375666341862
Episode: 53: Time: 8.824498414993286 Total Reward: 236.16883116883463 Avg_Loss: 3.8009322445933558
Episode: 54: Time: 8.888934135437012 Total Reward: 350.57823129251426 Avg_Loss: 4.201924719229466

Episode: 55: Time: 8.649335384368896 Total Reward: 217.50000000000443 Avg_Loss: 4.043002549339743
Episode: 56: Time: 8.901369094848633 Total Reward: 331.2820512820494 Avg_Loss: 4.155640004061851
Episode: 57: Time: 8.756320714950562 Total Reward: 282.2241992882573 Avg_Loss: 4.172755420207977
Episode: 58: Time: 9.091580867767334 Total Reward: 237.29813664596705 Avg_Loss: 4.17998033862154
Episode: 59: Time: 9.157258987426758 Total Reward: 261.90235690235147 Avg_Loss: 3.899202047275896
Episode: 60: Time: 8.972252368927002 Total Reward: 260.7046979865729 Avg_Loss: 4.366603489182577
Episode: 61: Time: 8.809324502944946 Total Reward: 318.3858267716523 Avg_Loss: 4.317523904207374
Episode: 62: Time: 8.959850311279297 Total Reward: 493.0398671096244 Avg_Loss: 4.65768139903285
Episode: 63: Time: 8.919708967208862 Total Reward: 302.5155279503108 Avg_Loss: 4.74199727102488
Episode: 64: Time: 9.087061882019043 Total Reward: 245.62500000000438 Avg_Loss: 4.743659382106877
Episode: 65: Time: 8.787028789520264 Total Reward: 522.8571428571348 Avg_Loss: 4.8720313110271425
Episode: 66: Time: 8.899693250656128 Total Reward: 306.9933554817272 Avg_Loss: 5.648883237558253
Episode: 67: Time: 8.737599611282349 Total Reward: 275.74829931972124 Avg_Loss: 4.876691952973855
Episode: 68: Time: 8.78323769569397 Total Reward: 424.57295373664664 Avg_Loss: 4.882070067049074
Episode: 69: Time: 8.673351764678955 Total Reward: 285.13698630136304 Avg_Loss: 5.128196862565369
Episode: 70: Time: 8.953622579574585 Total Reward: 506.91082802546754 Avg_Loss: 5.4938379315768975
Episode: 71: Time: 8.79429292678833 Total Reward: 465.7843137254881 Avg_Loss: 5.2126343230239485
Episode: 72: Time: 9.069419860839844 Total Reward: 214.85915492958168 Avg_Loss: 5.348371759182265
Episode: 73: Time: 8.82794451713562 Total Reward: 97.72727272727664 Avg_Loss: 5.278079893909583
Episode: 74: Time: 8.928473472595215 Total Reward: 175.35830618892896 Avg_Loss: 4.95597223524286
Episode: 75: Time: 8.698085308074951 Total Reward: 428.80952380951703 Avg_Loss: 5.297351434451191
Episode: 76: Time: 8.75436520576477 Total Reward: 24.13357400722159 Avg_Loss: 5.157387792062359
Episode: 77: Time: 8.867250680923462 Total Reward: 156.61290322580695 Avg_Loss: 5.382217375671162
Episode: 78: Time: 9.028779745101929 Total Reward: 101.8253968253982 Avg_Loss: 5.600969541974428
Episode: 79: Time: 8.837549686431885 Total Reward: 278.7373737373746 Avg_Loss: 5.558953583741388
Episode: 80: Time: 8.902163982391357 Total Reward: 413.53242320818373 Avg_Loss: 5.318848365495185
Episode: 81: Time: 8.761207818984985 Total Reward: 105.68027210884371 Avg_Loss: 5.703598594966055
Episode: 82: Time: 8.873214960098267 Total Reward: 251.5346534653489 Avg_Loss: 5.269073863991168

Episode: 83: Time: 8.787366151809692 Total Reward: 413.5324232081833 Avg_Loss: 5.759369785044374
Episode: 84: Time: 8.777789115905762 Total Reward: 505.7326007325937 Avg_Loss: 5.735346744541361
Episode: 85: Time: 8.821728229522705 Total Reward: 268.32179930795223 Avg_Loss: 5.1897350369381305
Episode: 86: Time: 8.768955945968628 Total Reward: 363.9041095890372 Avg_Loss: 5.332773652397284
Episode: 87: Time: 2.9660143852233887 Total Reward: -68.67414965986427 Avg_Loss: 5.299276261102586
Episode: 88: Time: 9.26429533958435 Total Reward: 110.60747663551435 Avg_Loss: 5.436146127576587
Episode: 89: Time: 8.91404128074646 Total Reward: 454.3421052631532 Avg_Loss: 5.455363784016681
Episode: 90: Time: 8.862672567367554 Total Reward: 367.40601503759166 Avg_Loss: 5.523359656333923
Episode: 91: Time: 8.848169326782227 Total Reward: 336.5789473684166 Avg_Loss: 7.499766533114329
Episode: 92: Time: 9.18509817123413 Total Reward: 207.4691358024739 Avg_Loss: 5.680491362299238
Episode: 93: Time: 8.867718696594238 Total Reward: 208.03030303030764 Avg_Loss: 7.239896753755938
Episode: 94: Time: 8.95173978805542 Total Reward: 341.4820846905474 Avg_Loss: 5.952149255936887
Episode: 95: Time: 4.971399784088135 Total Reward: -15.486138613859538 Avg_Loss: 6.0562012107933265
Episode: 96: Time: 9.133865594863892 Total Reward: 171.4576802507869 Avg_Loss: 5.701533408725963
Episode: 97: Time: 8.792855739593506 Total Reward: 228.741007194249 Avg_Loss: 5.860407322394748
Episode: 98: Time: 9.228764772415161 Total Reward: 322.9104477611941 Avg_Loss: 5.703316621419762
Episode: 99: Time: 8.741441249847412 Total Reward: 325.195439739412 Avg_Loss: 5.6725949038978385
Validation Mean Reward: 399.953492478727 Validation Std Reward: 103.92373208145169
Episode: 100: Time: 9.559770107269287 Total Reward: 401.64429530200823 Avg_Loss: 9.542456163578674
Episode: 101: Time: 9.135803461074829 Total Reward: 369.5390070921955 Avg_Loss: 10.14269595586953
Episode: 102: Time: 8.748307704925537 Total Reward: 148.24324324324758 Avg_Loss: 5.902889458071284
Episode: 103: Time: 9.367921352386475 Total Reward: 139.0425531914924 Avg_Loss: 6.969742566597562
Episode: 104: Time: 9.106154203414917 Total Reward: 421.5562913907214 Avg_Loss: 8.414267972236923
Episode: 105: Time: 9.006991147994995 Total Reward: 442.87878787878174 Avg_Loss: 5.8561245032719205
Episode: 106: Time: 8.838223457336426 Total Reward: 689.5528455284452 Avg_Loss: 6.419700958147771
Episode: 107: Time: 9.122624158859253 Total Reward: 340.81081081080754 Avg_Loss: 6.365661749819748
Episode: 108: Time: 9.0559401512146 Total Reward: 433.0528052805246 Avg_Loss: 5.858990708319079
Episode: 109: Time: 9.366560459136963 Total Reward: 290.5799373040736 Avg_Loss: 5.95559626266736

Episode: 110: Time: 9.023941040039062 Total Reward: 284.79094076655304 Avg_Loss: 6.491036235785284
Episode: 111: Time: 9.108924150466919 Total Reward: 316.5755627009614 Avg_Loss: 6.073557118407819
Episode: 112: Time: 9.001177787780762 Total Reward: 425.1342281879173 Avg_Loss: 5.878028917713325
Episode: 113: Time: 4.197131872177124 Total Reward: -80.31036789297718 Avg_Loss: 5.892341864209215
Episode: 114: Time: 9.027549743652344 Total Reward: 380.2475247524716 Avg_Loss: 6.117070763051009
Episode: 115: Time: 8.951233863830566 Total Reward: 336.5068493150623 Avg_Loss: 5.954145828215014
Episode: 116: Time: 8.763563871383667 Total Reward: 325.0743494423716 Avg_Loss: 6.583058362247563
Episode: 117: Time: 8.902001857757568 Total Reward: 325.8754208754161 Avg_Loss: 6.251108603818076
Episode: 118: Time: 8.879004955291748 Total Reward: 435.68592057761094 Avg_Loss: 6.310373771090467
Episode: 119: Time: 8.568905115127563 Total Reward: 105.000000000000317 Avg_Loss: 6.346694855129018
Episode: 120: Time: 8.933384656906128 Total Reward: 483.7671232876668 Avg_Loss: 5.795152756346374
Episode: 121: Time: 9.003633499145508 Total Reward: 360.12820512820093 Avg_Loss: 5.796678053230798
Episode: 122: Time: 9.093908309936523 Total Reward: 257.9411764705838 Avg_Loss: 6.019389757589132
Episode: 123: Time: 9.092081785202026 Total Reward: 388.76623376623087 Avg_Loss: 6.167869306412063
Episode: 124: Time: 8.994673013687134 Total Reward: 515.5263157894676 Avg_Loss: 6.119218816276358
Episode: 125: Time: 8.962742328643799 Total Reward: 598.3797909407556 Avg_Loss: 5.968296114136191
Episode: 126: Time: 9.110531091690063 Total Reward: 369.6153846153795 Avg_Loss: 5.9150729479909945
Episode: 127: Time: 9.132225513458252 Total Reward: 332.1186440677867 Avg_Loss: 6.159886123252516
Episode: 128: Time: 8.839430570602417 Total Reward: 395.56603773584726 Avg_Loss: 5.84495517285932
Episode: 129: Time: 8.951565504074097 Total Reward: 503.6622073578539 Avg_Loss: 6.183743254477236
Episode: 130: Time: 9.263639450073242 Total Reward: 442.95379537953227 Avg_Loss: 8.447138345542074
Episode: 131: Time: 9.205625772476196 Total Reward: 244.74358974358293 Avg_Loss: 6.9202536099097305
Episode: 132: Time: 9.005797624588013 Total Reward: 275.6293706293605 Avg_Loss: 6.309264933862606
Episode: 133: Time: 8.86644721031189 Total Reward: 560.5555555555427 Avg_Loss: 6.6000614106154245
Episode: 134: Time: 9.246344089508057 Total Reward: 404.9999999999994 Avg_Loss: 6.233511380287779
Episode: 135: Time: 8.701882123947144 Total Reward: 145.000000000000432 Avg_Loss: 7.2314020806000014
Episode: 136: Time: 8.901700735092163 Total Reward: 569.4736842105164 Avg_Loss: 6.776984288912861
Episode: 137: Time: 8.778560638427734 Total Reward: 444.5189003436351 Avg_Loss: 7.470519756068702

Episode: 138: Time: 8.978103399276733 Total Reward: 348.39622641508686 Avg_Loss: 7.101317074118542
Episode: 139: Time: 9.019516706466675 Total Reward: 370.67164179104054 Avg_Loss: 7.422649572877323
Episode: 140: Time: 8.87307095527649 Total Reward: 386.4814814814756 Avg_Loss: 7.388053959157286
Episode: 141: Time: 9.17137885093689 Total Reward: 343.87147335423145 Avg_Loss: 7.413561042617349
Episode: 142: Time: 8.836299419403076 Total Reward: 478.7051792828612 Avg_Loss: 7.547081336253831
Episode: 143: Time: 9.087306499481201 Total Reward: 406.69491525423456 Avg_Loss: 7.550261470450073
Episode: 144: Time: 8.974781036376953 Total Reward: 492.59124087590226 Avg_Loss: 7.601143587537172
Episode: 145: Time: 9.30847978591919 Total Reward: 343.0952380952359 Avg_Loss: 7.635033052508571
Episode: 146: Time: 9.134598016738892 Total Reward: 220.2542372881376 Avg_Loss: 7.478524926329861
Episode: 147: Time: 9.131514310836792 Total Reward: 595.4024767801782 Avg_Loss: 7.7643446050772145
Episode: 148: Time: 9.070313930511475 Total Reward: 623.7499999999874 Avg_Loss: 7.490222710521281
Episode: 149: Time: 8.98642349243164 Total Reward: 495.27777777777106 Avg_Loss: 7.457038367495818
Validation Mean Reward: 509.00265141542593 Validation Std Reward: 173.32193682814162
Episode: 150: Time: 9.444414138793945 Total Reward: 508.38983050846457 Avg_Loss: 7.764348589071707
Episode: 151: Time: 8.913782835006714 Total Reward: 418.51351351350604 Avg_Loss: 7.443209107182607
Episode: 152: Time: 8.96460747718811 Total Reward: 255.16835016834006 Avg_Loss: 7.017907185714786
Episode: 153: Time: 8.903270483016968 Total Reward: 508.6363636363543 Avg_Loss: 6.912559228784898
Episode: 154: Time: 8.77627944946289 Total Reward: 358.90070921985324 Avg_Loss: 7.115573615587058
Episode: 155: Time: 8.837299823760986 Total Reward: 365.78431372547897 Avg_Loss: 6.983125225836489
Episode: 156: Time: 8.786910772323608 Total Reward: 401.42857142856235 Avg_Loss: 7.007669453861332
Episode: 157: Time: 8.698958396911621 Total Reward: 596.7562724014226 Avg_Loss: 7.372478809677252
Episode: 158: Time: 8.959820985794067 Total Reward: 677.0588235294005 Avg_Loss: 7.2097254680986165
Episode: 159: Time: 8.954644441604614 Total Reward: 511.0606060605973 Avg_Loss: 7.322837492998908
Episode: 160: Time: 8.913387537002563 Total Reward: 366.81818181816834 Avg_Loss: 7.137689558397822
Episode: 161: Time: 8.946526050567627 Total Reward: 444.7350993377444 Avg_Loss: 7.458194183702228
Episode: 162: Time: 9.125279188156128 Total Reward: 227.2591362126267 Avg_Loss: 7.787750853209936
Episode: 163: Time: 9.280099868774414 Total Reward: 369.05228758169375 Avg_Loss: 8.116740091508175
Episode: 164: Time: 9.155835628509521 Total Reward: 432.77777777777305 Avg_Loss: 8.060613038159218

Episode: 165: Time: 9.168130159378052 Total Reward: 403.3606557377032 Avg_Loss: 7.6325031949692415
Episode: 166: Time: 9.248130559921265 Total Reward: 483.61635220124896 Avg_Loss: 8.00687194972479
Episode: 167: Time: 9.076822996139526 Total Reward: 564.6491228070048 Avg_Loss: 7.86160772387721
Episode: 168: Time: 9.073077917098999 Total Reward: 318.37386018236015 Avg_Loss: 7.346193640171981
Episode: 169: Time: 9.1148362159729 Total Reward: 629.9070631970125 Avg_Loss: 7.333574999280336
Episode: 170: Time: 9.073678493499756 Total Reward: 415.3448275862029 Avg_Loss: 7.215462563418541
Episode: 171: Time: 9.155794858932495 Total Reward: 370.5172413793081 Avg_Loss: 6.846945997045822
Episode: 172: Time: 8.945671319961548 Total Reward: 331.5232974910399 Avg_Loss: 7.162883289721834
Episode: 173: Time: 9.187593460083008 Total Reward: 587.4324324324211 Avg_Loss: 6.784620489392962
Episode: 174: Time: 9.163450717926025 Total Reward: 536.067961165037 Avg_Loss: 6.914489320346287
Episode: 175: Time: 9.022217512130737 Total Reward: 243.70967741935652 Avg_Loss: 7.836787996171903
Episode: 176: Time: 8.867122650146484 Total Reward: 123.30985915493375 Avg_Loss: 7.433313488960266
Episode: 177: Time: 9.216876745223999 Total Reward: 440.16819571864914 Avg_Loss: 7.811799956970856
Episode: 178: Time: 8.929874181747437 Total Reward: 619.8014440433145 Avg_Loss: 7.281484885376041
Episode: 179: Time: 9.04206657409668 Total Reward: 553.8294314381201 Avg_Loss: 7.10034277960032
Episode: 180: Time: 8.899816989898682 Total Reward: 740.2059925093529 Avg_Loss: 7.4507116289699775
Episode: 181: Time: 9.142321586608887 Total Reward: 555.6849315068409 Avg_Loss: 7.868258554394505
Episode: 182: Time: 9.208608388900757 Total Reward: 234.07348242811457 Avg_Loss: 7.330141170686033
Episode: 183: Time: 8.878854513168335 Total Reward: 330.08710801392766 Avg_Loss: 7.551546584658262
Episode: 184: Time: 9.078736305236816 Total Reward: 554.9999999999897 Avg_Loss: 7.854627152450946
Episode: 185: Time: 9.15530514717102 Total Reward: 504.99999999999034 Avg_Loss: 7.513438275882176
Episode: 186: Time: 9.100300312042236 Total Reward: 589.2105263157791 Avg_Loss: 7.7216315690208885
Episode: 187: Time: 8.997204780578613 Total Reward: 490.18518518517726 Avg_Loss: 7.6643460999016
Episode: 188: Time: 9.108789443969727 Total Reward: 691.6666666666657 Avg_Loss: 7.477719801814616
Episode: 189: Time: 8.923763036727905 Total Reward: 532.4509803921495 Avg_Loss: 7.521722342787671
Episode: 190: Time: 8.96717619895935 Total Reward: 423.27242524916335 Avg_Loss: 7.1623254323206025
Episode: 191: Time: 8.888558387756348 Total Reward: 207.08333333333826 Avg_Loss: 7.012135899367452
Episode: 192: Time: 8.914247751235962 Total Reward: 292.5432525951522 Avg_Loss: 6.838759916169303

Episode: 193: Time: 9.114394187927246 Total Reward: 304.3902439024409 Avg_Loss: 6.558910922843869
Episode: 194: Time: 9.194780588150024 Total Reward: 483.61635220124845 Avg_Loss: 7.0837289736050515
Episode: 195: Time: 9.031086444854736 Total Reward: 251.15384615384247 Avg_Loss: 6.922670248175869
Episode: 196: Time: 9.175132513046265 Total Reward: 406.5974440894551 Avg_Loss: 6.82564138464567
Episode: 197: Time: 8.831825971603394 Total Reward: 641.059479553895 Avg_Loss: 6.861193116973428
Episode: 198: Time: 9.296360731124878 Total Reward: 528.417721518976 Avg_Loss: 6.389542445415208
Episode: 199: Time: 9.182455062866211 Total Reward: 544.1752577319517 Avg_Loss: 6.282278945466049
Validation Mean Reward: 503.15601352258574 Validation Std Reward: 227.35055488897302
Test Mean Reward: 506.88676710695654 Test Std Reward: 150.87997229532925

```
In [14]: traineSoftUpdateDQN = DQN.SoftUpdateDQN(EnvWrapper(env),
        model.Nature_Paper_Conv,
        tau = 0.001,
        update_freq = 1,
        lr = 0.00025,
        gamma = 0.95,
        buffer_size = 100000,
        batch_size = 32,
        loss_fn = "mse_loss",
        use_wandb = True,
        device = 'cuda',
        seed = 42,
        epsilon_scheduler = utils.exponential_decay(1, 1000, 0.1),
        save_path = utils.get_save_path("DoubleDQN_SoftUpdates_test")

traineSoftUpdateDQN.train(200, 50, 30, 50, 50)
```

saving to ./runs/DoubleDQN_SoftUpdates_test/run0

Finishing previous runs because reinit is set to 'default'.

wandb: ERROR The nbformat package was not found. It is required to save notebook history.

Run history:



Run summary:

loss	6.28228
mean_reward	503.15601
mean_test_reward	506.88677
std_reward	227.35055
std_test_reward	150.87997
total_reward	544.17526

View run **polar-snowball-9** at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/j8bv0sy5>

View project at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

Find logs at: ./wandb/run-20250608_130613-j8bv0sy5/logs

Tracking run with wandb version 0.20.0

Run data is saved locally in

/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250608_143427-op4iy3re

Syncing run **rare-music-10** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/op4iy3re>

Episode: 0: Time: 6.517987012863159 Total Reward: -33.110749185668126 Avg_Loss: 0.0
Episode: 1: Time: 7.89872670173645 Total Reward: -36.37931034482813 Avg_Loss: 0.7851506260455035
Episode: 2: Time: 8.406813859939575 Total Reward: -52.894736842105694 Avg_Loss: 0.7723749261192915
Episode: 3: Time: 7.627444505691528 Total Reward: -150.57936507936546 Avg_Loss: 0.7820082961572944
Episode: 4: Time: 8.5784752368927 Total Reward: -59.70588235294164 Avg_Loss: 5.518818340882533
Episode: 5: Time: 8.780124187469482 Total Reward: -35.74074074074096 Avg_Loss: 1.8013605856015777
Episode: 6: Time: 8.761979818344116 Total Reward: -40.76271186440751 Avg_Loss: 1.4816426527099449
Episode: 7: Time: 8.791294574737549 Total Reward: -74.31034482758632 Avg_Loss: 0.6449806437585033
Episode: 8: Time: 9.162508726119995 Total Reward: -41.02272727272788 Avg_Loss: 0.6105091005657651
Episode: 9: Time: 8.279945373535156 Total Reward: -48.73665480427111 Avg_Loss: 0.5733335452122974
Episode: 10: Time: 4.0604541301727295 Total Reward: -87.2475728155345 Avg_Loss: 0.5731748041549796
Episode: 11: Time: 8.413454532623291 Total Reward: -58.36996336996401 Avg_Loss: 6.155274676863386
Episode: 12: Time: 8.865975856781006 Total Reward: -35.298507462686985 Avg_Loss: 1.2802112975603892
Episode: 13: Time: 8.626904487609863 Total Reward: -31.93693693693762 Avg_Loss: 3.1157351944032086
Episode: 14: Time: 8.994204998016357 Total Reward: -37.95302013422889 Avg_Loss: 1.2262572649115024
Episode: 15: Time: 8.59404730796814 Total Reward: -73.33935018050563 Avg_Loss: 1.7025969403506327
Episode: 16: Time: 8.83869481086731 Total Reward: -66.73144876325152 Avg_Loss: 0.9387197666840643
Episode: 17: Time: 2.5893099308013916 Total Reward: -55.181250000000304 Avg_Loss: 0.8022984611625606
Episode: 18: Time: 8.60946774482727 Total Reward: -32.67705382436333 Avg_Loss: 3.073452651735489
Episode: 19: Time: 8.456228971481323 Total Reward: -19.39862542955393 Avg_Loss: 2.0617988645529546
Episode: 20: Time: 9.285467386245728 Total Reward: -38.85964912280774 Avg_Loss: 1.3439402021780735
Episode: 21: Time: 8.557771682739258 Total Reward: -45.16611295681136 Avg_Loss: 1.2497807737408566
Episode: 22: Time: 8.205527305603027 Total Reward: 47.34875444840226 Avg_Loss: 1.155052278347376
Episode: 23: Time: 2.9363057613372803 Total Reward: -65.76219081272126 Avg_Loss: 0.8559398705431496
Episode: 24: Time: 9.178642988204956 Total Reward: 14.540636042403662 Avg_Loss: 4.961100143763949
Episode: 25: Time: 8.336953401565552 Total Reward: -27.8621908127216 Avg_Loss: 2.710484579956832
Episode: 26: Time: 8.804186344146729 Total Reward: -14.708029197080894 Avg_Loss: 1.834161223233247
Episode: 27: Time: 8.608407020568848 Total Reward: 14.848484848483857 Avg_Loss: 2.061193415002913

Episode: 28: Time: 6.344376802444458 Total Reward: -81.18066914498118 Avg_Loss: 3.122769036872239
Episode: 29: Time: 8.502626895904541 Total Reward: -28.12101910828097 Avg_Loss: 2.6663702936889995
Episode: 30: Time: 3.3635952472686768 Total Reward: -85.03086419753129 Avg_Loss: 2.4562880610848996
Episode: 31: Time: 9.579005718231201 Total Reward: 73.11594202898989 Avg_Loss: 4.355222101659835
Episode: 32: Time: 8.967965126037598 Total Reward: -35.397350993378225 Avg_Loss: 3.08605646246559
Episode: 33: Time: 4.661568880081177 Total Reward: -91.67341772151953 Avg_Loss: 2.8642523208489785
Episode: 34: Time: 8.824270248413086 Total Reward: -34.297124600639705 Avg_Loss: 1.2039457387408288
Episode: 35: Time: 7.468776702880859 Total Reward: -114.85873015873071 Avg_Loss: 1.297338062932469
Episode: 36: Time: 6.392802953720093 Total Reward: -110.60417827298095 Avg_Loss: 3.7508960936455367
Episode: 37: Time: 8.817289590835571 Total Reward: 25.437956204378708 Avg_Loss: 2.4056795520567094
Episode: 38: Time: 8.988691091537476 Total Reward: -9.179104477612084 Avg_Loss: 2.9149884203088883
Episode: 39: Time: 8.462157011032104 Total Reward: 50.66929133858654 Avg_Loss: 2.209058712072232
Episode: 40: Time: 8.555100202560425 Total Reward: -32.304075235110446 Avg_Loss: 1.6133782704274695
Episode: 41: Time: 8.434928894042969 Total Reward: -34.71631205673833 Avg_Loss: 1.5523683160793882
Episode: 42: Time: 8.55157470703125 Total Reward: 8.58565737051862 Avg_Loss: 1.0518532904381512
Episode: 43: Time: 8.546396017074585 Total Reward: -36.780821917808936 Avg_Loss: 1.3177424684761703
Episode: 44: Time: 8.936720371246338 Total Reward: -33.08049535603783 Avg_Loss: 1.1752360825810362
Episode: 45: Time: 8.750353336334229 Total Reward: -3.08823529411856 Avg_Loss: 1.1914309178236402
Episode: 46: Time: 8.392256736755371 Total Reward: -33.11074918566848 Avg_Loss: 0.7892094627335793
Episode: 47: Time: 9.207422971725464 Total Reward: -29.027777777778486 Avg_Loss: 0.8087933770060038
Episode: 48: Time: 8.354970455169678 Total Reward: 41.51877133106104 Avg_Loss: 1.0428775866758548
Episode: 49: Time: 9.20979928970337 Total Reward: -38.03797468354503 Avg_Loss: 0.8431559107227236
Validation Mean Reward: -30.395542588485878 Validation Std Reward: 35.249983303610755
Episode: 50: Time: 2.5512702465057373 Total Reward: -69.65280235988229 Avg_Loss: 0.8448163924308923
Episode: 51: Time: 8.596171617507935 Total Reward: 54.60629921260235 Avg_Loss: 1.5621256378455823
Episode: 52: Time: 8.477054834365845 Total Reward: 46.22137404580468 Avg_Loss: 1.0773265699879462
Episode: 53: Time: 8.840584516525269 Total Reward: 65.27874564460113 Avg_Loss: 1.0643694870674811
Episode: 54: Time: 8.746686697006226 Total Reward: 285.2816901408432 Avg_Loss: 1.1013203216638385

Episode: 55: Time: 8.965197801589966 Total Reward: -31.305732484077105 Avg_Loss: 1.1149191288512295
Episode: 56: Time: 8.752743244171143 Total Reward: 71.07773851590551 Avg_Loss: 0.9149605137836031
Episode: 57: Time: 8.560729026794434 Total Reward: -21.17449664429604 Avg_Loss: 1.1706459996830516
Episode: 58: Time: 8.736500978469849 Total Reward: 29.08759124087819 Avg_Loss: 1.1797618339730411
Episode: 59: Time: 8.63001823425293 Total Reward: 16.91335740072188 Avg_Loss: 1.1156484010402394
Episode: 60: Time: 8.656843423843384 Total Reward: -39.44444444444514 Avg_Loss: 1.366432961045193
Episode: 61: Time: 8.57627820968628 Total Reward: 142.5886524822744 Avg_Loss: 1.3999129339050846
Episode: 62: Time: 9.07403016090393 Total Reward: 167.34567901235008 Avg_Loss: 1.3096264031629603
Episode: 63: Time: 9.175196647644043 Total Reward: 173.29268292683298 Avg_Loss: 1.3561335961733545
Episode: 64: Time: 9.181767702102661 Total Reward: 318.33333333332473 Avg_Loss: 1.359421288704171
Episode: 65: Time: 8.766611337661743 Total Reward: 174.78417266187543 Avg_Loss: 1.3818133976153966
Episode: 66: Time: 8.889705419540405 Total Reward: 194.0365448505026 Avg_Loss: 1.5251703956986176
Episode: 67: Time: 8.853414297103882 Total Reward: 28.52941176470729 Avg_Loss: 1.3811524383911566
Episode: 68: Time: 8.980798721313477 Total Reward: 364.9999999999324 Avg_Loss: 1.3680101309535133
Episode: 69: Time: 9.423115730285645 Total Reward: 109.67836257310285 Avg_Loss: 1.3324687889143199
Episode: 70: Time: 9.288426637649536 Total Reward: 393.2154882154789 Avg_Loss: 1.523715219953481
Episode: 71: Time: 9.020573854446411 Total Reward: 316.96013289035614 Avg_Loss: 1.6482161964688982
Episode: 72: Time: 8.670556545257568 Total Reward: 834.9610894941513 Avg_Loss: 1.6566407242242027
Episode: 73: Time: 8.661933898925781 Total Reward: 689.38661710036 Avg_Loss: 1.7840057524437665
Episode: 74: Time: 9.187336921691895 Total Reward: 261.0371517027793 Avg_Loss: 2.1845736336432586
Episode: 75: Time: 8.773205757141113 Total Reward: 411.9930069929986 Avg_Loss: 1.9527600480227911
Episode: 76: Time: 8.941222667694092 Total Reward: 99.52887537994351 Avg_Loss: 2.2367359675279186
Episode: 77: Time: 9.017112493515015 Total Reward: 378.684210526312 Avg_Loss: 2.3142987254287015
Episode: 78: Time: 9.139372825622559 Total Reward: 459.38596491227014 Avg_Loss: 2.24416000602626
Episode: 79: Time: 9.178406000137329 Total Reward: 65.40955631399751 Avg_Loss: 2.100681341996714
Episode: 80: Time: 9.107641696929932 Total Reward: 421.5562913907244 Avg_Loss: 2.4464964468439088
Episode: 81: Time: 8.741514921188354 Total Reward: 248.6426116838406 Avg_Loss: 2.411141144628284
Episode: 82: Time: 8.930920839309692 Total Reward: 232.64505119453395 Avg_Loss: 2.4928486920705364

Episode: 83: Time: 9.185783624649048 Total Reward: 301.7213114753995 Avg_Loss: 2.3302837630279924
Episode: 84: Time: 8.978396654129028 Total Reward: 48.812709030103214 Avg_Loss: 2.386736288917165
Episode: 85: Time: 8.96529245376587 Total Reward: 663.2781456953567 Avg_Loss: 2.6775375900148344
Episode: 86: Time: 8.805817604064941 Total Reward: 722.8438661709902 Avg_Loss: 2.7402261699698554
Episode: 87: Time: 8.94369387626648 Total Reward: 487.1917808219083 Avg_Loss: 3.1315693682482264
Episode: 88: Time: 8.66255784034729 Total Reward: 258.8461538461441 Avg_Loss: 2.6272487880802955
Episode: 89: Time: 8.487017631530762 Total Reward: 256.25448028672554 Avg_Loss: 2.6942917904933963
Episode: 90: Time: 8.747597455978394 Total Reward: 142.45819397993782 Avg_Loss: 2.8162439687161887
Episode: 91: Time: 9.322379350662231 Total Reward: 288.3333333333335 Avg_Loss: 2.9239785344410345
Episode: 92: Time: 9.063215255737305 Total Reward: 398.24324324323874 Avg_Loss: 2.9655535797111128
Episode: 93: Time: 9.1397864818573 Total Reward: 299.23076923076013 Avg_Loss: 3.1471211320462347
Episode: 94: Time: 8.954813718795776 Total Reward: 575.0680272108776 Avg_Loss: 3.243391263885658
Episode: 95: Time: 9.008143424987793 Total Reward: 584.7153024910975 Avg_Loss: 3.3149573084186104
Episode: 96: Time: 9.473471403121948 Total Reward: 165.58631921824562 Avg_Loss: 3.6435212933716654
Episode: 97: Time: 9.491592645645142 Total Reward: 568.2302405498194 Avg_Loss: 3.276320558865683
Episode: 98: Time: 9.113019704818726 Total Reward: 573.9895470383158 Avg_Loss: 3.2391942849680153
Episode: 99: Time: 8.636539459228516 Total Reward: 156.7730496453949 Avg_Loss: 3.102412067791995
Validation Mean Reward: 454.26080010167226 Validation Std Reward: 223.5620900201326
Episode: 100: Time: 9.28870153427124 Total Reward: 379.04844290656627 Avg_Loss: 3.3026807051496347
Episode: 101: Time: 9.251375675201416 Total Reward: 454.01960784313235 Avg_Loss: 3.4858227822960925
Episode: 102: Time: 9.069865226745605 Total Reward: 513.843537414956 Avg_Loss: 3.5525216382591664
Episode: 103: Time: 9.063061952590942 Total Reward: 635.3370786516746 Avg_Loss: 3.4396573565587274
Episode: 104: Time: 9.019737720489502 Total Reward: 515.1694915254114 Avg_Loss: 3.402727768200786
Episode: 105: Time: 9.454664945602417 Total Reward: 474.07894736841155 Avg_Loss: 3.572247917411708
Episode: 106: Time: 9.2510404586792 Total Reward: 508.77358490564905 Avg_Loss: 3.61997254280483
Episode: 107: Time: 9.045000076293945 Total Reward: 406.7421602787431 Avg_Loss: 3.9296244898764026
Episode: 108: Time: 9.255696535110474 Total Reward: 422.96407185628186 Avg_Loss: 3.6150570058522105
Episode: 109: Time: 8.964823484420776 Total Reward: 420.35836177473743 Avg_Loss: 3.7626925055720224

Episode: 110: Time: 9.548035621643066 Total Reward: 358.41614906831694 Avg_Loss: 4.054413089481723
Episode: 111: Time: 9.129107475280762 Total Reward: 141.84210526316173 Avg_Loss: 3.909037183062369
Episode: 112: Time: 9.20482325553894 Total Reward: 92.50000000000392 Avg_Loss: 3.89728384583938
Episode: 113: Time: 8.996813297271729 Total Reward: 473.70229007632855 Avg_Loss: 4.073739424973977
Episode: 114: Time: 8.90030813217163 Total Reward: 678.9463601532473 Avg_Loss: 3.8067799420917736
Episode: 115: Time: 9.136903047561646 Total Reward: 460.92105263157225 Avg_Loss: 3.7114007250100625
Episode: 116: Time: 9.119772672653198 Total Reward: 450.1505016722304 Avg_Loss: 4.6282954278613335
Episode: 117: Time: 9.182605028152466 Total Reward: 470.37102473497913 Avg_Loss: 4.088028818619351
Episode: 118: Time: 9.479071378707886 Total Reward: 234.03225806452014 Avg_Loss: 4.100485504174433
Episode: 119: Time: 9.25711989402771 Total Reward: 410.6179775280872 Avg_Loss: 4.128257314197156
Episode: 120: Time: 9.209830284118652 Total Reward: 529.1379310344713 Avg_Loss: 4.394224421567276
Episode: 121: Time: 9.110737085342407 Total Reward: 522.3285198555856 Avg_Loss: 4.633405098394186
Episode: 122: Time: 9.304264783859253 Total Reward: 408.0674846625718 Avg_Loss: 4.405876796786525
Episode: 123: Time: 9.52597951889038 Total Reward: 347.1364985163172 Avg_Loss: 4.721594299588885
Episode: 124: Time: 9.194671392440796 Total Reward: 426.4285714285676 Avg_Loss: 5.026639160989713
Episode: 125: Time: 8.944711208343506 Total Reward: 430.4237288135582 Avg_Loss: 4.30609350294626
Episode: 126: Time: 8.883428812026978 Total Reward: 654.9999999999883 Avg_Loss: 4.777582610354704
Episode: 127: Time: 8.949157476425171 Total Reward: 208.3333333333373 Avg_Loss: 4.859015448253696
Episode: 128: Time: 9.001689672470093 Total Reward: 197.06349206349552 Avg_Loss: 5.050806759285326
Episode: 129: Time: 9.017389059066772 Total Reward: 252.2222222222632 Avg_Loss: 5.319938200862468
Episode: 130: Time: 9.048270225524902 Total Reward: 452.4452554744458 Avg_Loss: 5.236895769083199
Episode: 131: Time: 9.259659051895142 Total Reward: 291.9863013698598 Avg_Loss: 4.748771575569105
Episode: 132: Time: 9.427180051803589 Total Reward: 321.40378548895893 Avg_Loss: 4.716913819313049
Episode: 133: Time: 8.80291485786438 Total Reward: 520.999999999955 Avg_Loss: 5.16191780717433
Episode: 134: Time: 9.02002215385437 Total Reward: 116.22112211221537 Avg_Loss: 5.409420298177655
Episode: 135: Time: 8.903197288513184 Total Reward: 431.6903914590704 Avg_Loss: 5.074378925211289
Episode: 136: Time: 9.180867910385132 Total Reward: 384.5918367346901 Avg_Loss: 5.139073545692348
Episode: 137: Time: 9.282783269882202 Total Reward: 311.8965517241377 Avg_Loss: 5.61262704194093

Episode: 138: Time: 8.976841449737549 Total Reward: 368.70967741935334 Avg_Loss: 5.146897232332149
Episode: 139: Time: 9.194902181625366 Total Reward: 446.40127388534137 Avg_Loss: 5.221278950446794
Episode: 140: Time: 9.025276184082031 Total Reward: 389.94983277591587 Avg_Loss: 5.5969773751347
Episode: 141: Time: 8.869401693344116 Total Reward: 453.38709677419087 Avg_Loss: 6.0143230427213075
Episode: 142: Time: 8.974151372909546 Total Reward: 432.02702702702015 Avg_Loss: 5.96090510061809
Episode: 143: Time: 9.187320947647095 Total Reward: 395.13157894735855 Avg_Loss: 5.28618519496517
Episode: 144: Time: 9.192233562469482 Total Reward: 416.78451178450484 Avg_Loss: 5.496456217865984
Episode: 145: Time: 9.319454193115234 Total Reward: 429.4755244755207 Avg_Loss: 5.602028929385819
Episode: 146: Time: 9.17629861831665 Total Reward: 155.85910652921297 Avg_Loss: 5.29569058658696
Episode: 147: Time: 9.01144814491272 Total Reward: 465.43956043955 Avg_Loss: 5.597525611645033
Episode: 148: Time: 8.744663715362549 Total Reward: 421.5289256198314 Avg_Loss: 5.93818889746145
Episode: 149: Time: 8.963855266571045 Total Reward: 286.4814814814747 Avg_Loss: 6.24158089601693
Validation Mean Reward: 391.2284518133914 Validation Std Reward: 142.9146211749332
Episode: 150: Time: 9.814082145690918 Total Reward: 243.70967741935812 Avg_Loss: 5.7365383891498345
Episode: 151: Time: 9.350341558456421 Total Reward: 496.3621262458365 Avg_Loss: 5.595381116165834
Episode: 152: Time: 9.142569303512573 Total Reward: 337.5259515570873 Avg_Loss: 5.5067505340616245
Episode: 153: Time: 9.199182987213135 Total Reward: 398.1034482758548 Avg_Loss: 6.179093718528748
Episode: 154: Time: 9.215584516525269 Total Reward: 457.8169014084449 Avg_Loss: 5.800200292543203
Episode: 155: Time: 8.805477619171143 Total Reward: 479.0740740740648 Avg_Loss: 5.706302278682966
Episode: 156: Time: 9.116844415664673 Total Reward: 263.49056603773727 Avg_Loss: 5.750262961167247
Episode: 157: Time: 9.178764820098877 Total Reward: 362.2368421052579 Avg_Loss: 6.2516690672946575
Episode: 158: Time: 9.194191932678223 Total Reward: 370.51724137930125 Avg_Loss: 5.955229728161788
Episode: 159: Time: 9.15082836151123 Total Reward: 90.05338078292243 Avg_Loss: 5.73282740246348
Episode: 160: Time: 9.131704568862915 Total Reward: 377.72727272726695 Avg_Loss: 5.715432045840416
Episode: 161: Time: 9.167474746704102 Total Reward: 475.44673539518135 Avg_Loss: 6.15544828697413
Episode: 162: Time: 9.068080425262451 Total Reward: 167.13592233010144 Avg_Loss: 5.459136962890625
Episode: 163: Time: 9.010035514831543 Total Reward: 420.2542372881285 Avg_Loss: 6.718961082586722
Episode: 164: Time: 8.829501867294312 Total Reward: 130.45454545454967 Avg_Loss: 5.950427332846057

Episode: 165: Time: 9.189848899841309 Total Reward: 452.29729729729115 Avg_Loss: 5.995124981182964
Episode: 166: Time: 9.37351393699646 Total Reward: 119.28571428571718 Avg_Loss: 6.048812803601017
Episode: 167: Time: 9.249087572097778 Total Reward: 430.17985611510204 Avg_Loss: 5.8580608207638525
Episode: 168: Time: 9.12891435623169 Total Reward: 376.6312056737532 Avg_Loss: 6.36778466641402
Episode: 169: Time: 9.02166199684143 Total Reward: 423.7969924811945 Avg_Loss: 6.318931608640847
Episode: 170: Time: 8.912148237228394 Total Reward: 144.726027397262 Avg_Loss: 6.208930236451766
Episode: 171: Time: 9.1906578540802 Total Reward: 218.38028169014274 Avg_Loss: 6.2227847616211704
Episode: 172: Time: 9.24285626411438 Total Reward: 311.0150375939852 Avg_Loss: 6.180407596235516
Episode: 173: Time: 9.561874866485596 Total Reward: 344.25233644859713 Avg_Loss: 6.180716964877954
Episode: 174: Time: 9.134507179260254 Total Reward: 382.1929824561354 Avg_Loss: 6.754353652481272
Episode: 175: Time: 9.329571723937988 Total Reward: 275.491803278685 Avg_Loss: 6.554105173639891
Episode: 176: Time: 8.8794846534729 Total Reward: 327.53521126760796 Avg_Loss: 6.0867763647512225
Episode: 177: Time: 9.047541379928589 Total Reward: 483.39721254354697 Avg_Loss: 6.517616270970898
Episode: 178: Time: 9.112586736679077 Total Reward: 263.1560283687867 Avg_Loss: 6.816468802820735
Episode: 179: Time: 9.236074924468994 Total Reward: 5.671140939596884 Avg_Loss: 6.644019722938538
Episode: 180: Time: 9.227322101593018 Total Reward: 410.4945054945032 Avg_Loss: 6.681969551479115
Episode: 181: Time: 9.382993221282959 Total Reward: 119.52145214521877 Avg_Loss: 6.37745388616033
Episode: 182: Time: 9.139262914657593 Total Reward: 184.4612794612838 Avg_Loss: 6.22804204155417
Episode: 183: Time: 9.086684942245483 Total Reward: 385.96885813148685 Avg_Loss: 6.279455515516906
Episode: 184: Time: 9.327415227890015 Total Reward: 327.2222222222194 Avg_Loss: 6.790889425938871
Episode: 185: Time: 9.418649196624756 Total Reward: 355.16077170417753 Avg_Loss: 7.0678726334531765
Episode: 186: Time: 9.743006944656372 Total Reward: 122.51412429378951 Avg_Loss: 6.3968217593281205
Episode: 187: Time: 8.970724821090698 Total Reward: 438.86454183266375 Avg_Loss: 6.428489553828199
Episode: 188: Time: 9.458844423294067 Total Reward: 327.15568862275495 Avg_Loss: 6.630683366991892
Episode: 189: Time: 9.166199922561646 Total Reward: 419.56310679610885 Avg_Loss: 6.660175645050883
Episode: 190: Time: 9.04369044303894 Total Reward: 82.35849056604162 Avg_Loss: 6.475533809481549
Episode: 191: Time: 9.155906438827515 Total Reward: 216.59420289855439 Avg_Loss: 7.096500553002878
Episode: 192: Time: 9.153875827789307 Total Reward: 489.5070422535139 Avg_Loss: 6.447372699985985

Episode: 193: Time: 9.19782042503357 Total Reward: 190.71428571429067 Avg_Loss: 6.795293039634448
Episode: 194: Time: 9.489245653152466 Total Reward: 290.3820598006638 Avg_Loss: 7.24889570224185
Episode: 195: Time: 9.215915203094482 Total Reward: 399.7735191637586 Avg_Loss: 6.9011933047230505
Episode: 196: Time: 9.207491636276245 Total Reward: 438.3333333333235 Avg_Loss: 6.567090383096903
Episode: 197: Time: 9.46213412284851 Total Reward: 520.6156156156085 Avg_Loss: 6.876639698733802
Episode: 198: Time: 9.258939027786255 Total Reward: 399.6996466431057 Avg_Loss: 7.130074843639085
Episode: 199: Time: 9.312360763549805 Total Reward: 507.1897810218919 Avg_Loss: 6.531013310456476
Validation Mean Reward: 577.6487311267821 Validation Std Reward: 140.61054013639088
Test Mean Reward: 574.9394309586244 Test Std Reward: 161.74126340008897

```
In [15]: traineSoftUpdateDQN = DQN.SoftUpdateDQN(EnvWrapper(env),
        model.Nature_Paper_Conv,
        tau = 0.05,
        update_freq = 1,
        lr = 0.00025,
        gamma = 0.95,
        buffer_size = 100000,
        batch_size = 32,
        loss_fn = "mse_loss",
        use_wandb = True,
        device = 'cuda',
        seed = 42,
        epsilon_scheduler = utils.exponential_decay(1, 1000, 0.1),
        save_path = utils.get_save_path("DoubleDQN_SoftUpdates_test")

    traineSoftUpdateDQN.train(200, 50, 30, 50, 50)
```

saving to ./runs/DoubleDQN_SoftUpdates_test/run1
Finishing previous runs because reinit is set to 'default'.

Run history:



Run summary:

loss	6.53101
mean_reward	577.64873
mean_test_reward	574.93943
std_reward	140.61054
std_test_reward	161.74126
total_reward	507.18978

View run **rare-music-10** at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/op4iy3re>

View project at: <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)

Find logs at: `./wandb/run-20250608_143427-op4iy3re/logs`

Tracking run with wandb version 0.20.0

Run data is saved locally in

`/home/tilboon/Documents/239AS_2/Project_4/wandb/run-20250608_152510-c8s95aqq`

Syncing run **fine-gorge-11** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/kbaseba-ucla/racing-car-dqn>

View run at <https://wandb.ai/kbaseba-ucla/racing-car-dqn/runs/c8s95aqq>

Episode: 0: Time: 6.55631422996521 Total Reward: -54.86622073578655 Avg_Loss: 0.0
Episode: 1: Time: 7.581755876541138 Total Reward: -68.28244274809225 Avg_Loss: 0.4865546637914695
Episode: 2: Time: 8.656254053115845 Total Reward: -55.65573770491892 Avg_Loss: 0.5086549636379605
Episode: 3: Time: 8.724108457565308 Total Reward: -71.27118644067843 Avg_Loss: 0.4809999328044032
Episode: 4: Time: 8.742109060287476 Total Reward: -72.1241830065363 Avg_Loss: 0.43843170103984846
Episode: 5: Time: 8.42009162902832 Total Reward: -22.083333333334075 Avg_Loss: 0.5805974589316409
Episode: 6: Time: 8.606217861175537 Total Reward: -13.644067796610983 Avg_Loss: 0.5989418487620204
Episode: 7: Time: 8.704542636871338 Total Reward: -71.94524495677253 Avg_Loss: 0.5599531452758473
Episode: 8: Time: 9.113034725189209 Total Reward: -66.8750000000006 Avg_Loss: 0.493291992287175
Episode: 9: Time: 8.81141471862793 Total Reward: -8.621262458472435 Avg_Loss: 0.6571037256855423
Episode: 10: Time: 8.22707724571228 Total Reward: -25.65693430657005 Avg_Loss: 0.6525096923989409
Episode: 11: Time: 8.38829517364502 Total Reward: -25.693069306931417 Avg_Loss: 0.6255317283402971
Episode: 12: Time: 8.871688842773438 Total Reward: -16.824104234528445 Avg_Loss: 0.772825041551049
Episode: 13: Time: 8.55030083656311 Total Reward: -36.97952218430106 Avg_Loss: 0.8928053650159796
Episode: 14: Time: 8.642497062683105 Total Reward: 103.52941176471026 Avg_Loss: 0.9824710191548371
Episode: 15: Time: 8.895750999450684 Total Reward: 6.010101010101762 Avg_Loss: 1.0852697646292317
Episode: 16: Time: 8.733076572418213 Total Reward: -20.80645161290385 Avg_Loss: 1.0438092677157467
Episode: 17: Time: 8.475907564163208 Total Reward: 79.90494296578422 Avg_Loss: 1.0586559718020825
Episode: 18: Time: 8.296856880187988 Total Reward: -29.210526315790187 Avg_Loss: 1.2190330320296168
Episode: 19: Time: 8.771589040756226 Total Reward: 13.433734939759745 Avg_Loss: 1.3291235137261261
Episode: 20: Time: 8.712218761444092 Total Reward: 137.08191126280315 Avg_Loss: 1.4172903011326028
Episode: 21: Time: 8.55316948890686 Total Reward: 24.718309859156484 Avg_Loss: 1.461368911531793
Episode: 22: Time: 8.505423784255981 Total Reward: 41.0544217687104 Avg_Loss: 1.695549615604036
Episode: 23: Time: 8.857015371322632 Total Reward: 576.1409395973062 Avg_Loss: 1.7074172242599375
Episode: 24: Time: 8.569616794586182 Total Reward: 59.882154882158886 Avg_Loss: 2.302375654963886
Episode: 25: Time: 8.973933458328247 Total Reward: 14.929078014184945 Avg_Loss: 2.1217133114818765
Episode: 26: Time: 8.579491138458252 Total Reward: 302.8873239436549 Avg_Loss: 2.426853459046668
Episode: 27: Time: 8.743886232376099 Total Reward: -17.297297297297842 Avg_Loss: 2.6968440588782814

Episode: 28: Time: 9.076586484909058 Total Reward: 410.26315789472335 Avg_Loss: 3.1910811737805855
Episode: 29: Time: 9.137084245681763 Total Reward: 540.514018691581 Avg_Loss: 3.1396390307350317
Episode: 30: Time: 8.925475120544434 Total Reward: 550.0381679389219 Avg_Loss: 3.927446303748283
Episode: 31: Time: 8.71179461479187 Total Reward: 379.8201438848848 Avg_Loss: 4.670656949031253
Episode: 32: Time: 8.862555980682373 Total Reward: 325.3389830508378 Avg_Loss: 4.554288895691142
Episode: 33: Time: 8.81320858001709 Total Reward: 601.5517241379213 Avg_Loss: 4.833331258607512
Episode: 34: Time: 9.507601261138916 Total Reward: 380.72815533979724 Avg_Loss: 4.29290444209796
Episode: 35: Time: 9.097410678863525 Total Reward: 300.77039274923635 Avg_Loss: 4.6109382094455365
Episode: 36: Time: 8.74570631980896 Total Reward: 41.94267515923875 Avg_Loss: 4.961462226485004
Episode: 37: Time: 8.309059619903564 Total Reward: 20.25423728813709 Avg_Loss: 4.859638847723729
Episode: 38: Time: 8.812207698822021 Total Reward: 527.1374045801467 Avg_Loss: 4.757881412486069
Episode: 39: Time: 9.319167137145996 Total Reward: 289.6153846153864 Avg_Loss: 5.481468550297392
Episode: 40: Time: 9.206023931503296 Total Reward: 175.1149425287366 Avg_Loss: 4.8679322825760405
Episode: 41: Time: 8.916181087493896 Total Reward: 392.8048780487742 Avg_Loss: 4.853203788024037
Episode: 42: Time: 9.14850926399231 Total Reward: 266.37071651090446 Avg_Loss: 4.893166213977237
Episode: 43: Time: 8.99984073638916 Total Reward: 636.6176470588127 Avg_Loss: 5.495840977720854
Episode: 44: Time: 8.740896701812744 Total Reward: 395.63670411985055 Avg_Loss: 5.9712739116003535
Episode: 45: Time: 8.9917631149292 Total Reward: 442.76435045316174 Avg_Loss: 6.005145690521272
Episode: 46: Time: 8.592645406723022 Total Reward: 298.7282229965132 Avg_Loss: 5.790172118599675
Episode: 47: Time: 8.44641661643982 Total Reward: 76.32867132867563 Avg_Loss: 6.541955533648739
Episode: 48: Time: 9.14127516746521 Total Reward: 304.2805755395703 Avg_Loss: 5.46043711950799
Episode: 49: Time: 8.963106155395508 Total Reward: 499.3060498220573 Avg_Loss: 5.89729894810364
Validation Mean Reward: 58.31206389510172 Validation Std Reward: 286.7354936594495
Episode: 50: Time: 9.658407211303711 Total Reward: 433.8461538461494 Avg_Loss: 6.053715921750589
Episode: 51: Time: 8.60307765007019 Total Reward: 314.9616858237438 Avg_Loss: 6.300067284527947
Episode: 52: Time: 9.194609880447388 Total Reward: 451.0317460317408 Avg_Loss: 5.95908453684895
Episode: 53: Time: 9.252808809280396 Total Reward: 467.9139072847596 Avg_Loss: 5.79165279815177
Episode: 54: Time: 9.05179214477539 Total Reward: 202.29729729730002 Avg_Loss: 6.425098375612948

Episode: 55: Time: 8.673473119735718 Total Reward: 105.00000000000452 Avg_Loss: 6.589431604918311
Episode: 56: Time: 8.764201879501343 Total Reward: 316.55234657039824 Avg_Loss: 6.908313648039553
Episode: 57: Time: 8.801118850708008 Total Reward: 341.17021276595756 Avg_Loss: 6.4400429520286435
Episode: 58: Time: 9.116507530212402 Total Reward: 313.4084084084079 Avg_Loss: 5.991703991629496
Episode: 59: Time: 8.80799126625061 Total Reward: 63.73015873016092 Avg_Loss: 6.325787195137569
Episode: 60: Time: 9.07243299484253 Total Reward: 316.3475177304961 Avg_Loss: 5.935520343920764
Episode: 61: Time: 8.851253986358643 Total Reward: 269.62093862815897 Avg_Loss: 6.414194993612145
Episode: 62: Time: 8.957344055175781 Total Reward: 344.9999999999977 Avg_Loss: 6.391266602428019
Episode: 63: Time: 8.931047439575195 Total Reward: 239.43708609271107 Avg_Loss: 6.1813709891142965
Episode: 64: Time: 8.74915099143982 Total Reward: 273.6274509803913 Avg_Loss: 5.9635495102706075
Episode: 65: Time: 8.793562650680542 Total Reward: 249.8275862068998 Avg_Loss: 6.390672708759789
Episode: 66: Time: 8.815327644348145 Total Reward: 304.99999999999693 Avg_Loss: 6.091458806470663
Episode: 67: Time: 8.58182430267334 Total Reward: 70.56291390728924 Avg_Loss: 6.00324061588079
Episode: 68: Time: 9.027546882629395 Total Reward: 426.8978102189732 Avg_Loss: 6.32195696259747
Episode: 69: Time: 9.03364896774292 Total Reward: 342.2623574144475 Avg_Loss: 5.976115789734015
Episode: 70: Time: 9.46207046508789 Total Reward: 155.00000000000387 Avg_Loss: 6.172614700653973
Episode: 71: Time: 9.06264328956604 Total Reward: 443.9610389610336 Avg_Loss: 5.4264433839741875
Episode: 72: Time: 8.89584231376648 Total Reward: 245.35087719298434 Avg_Loss: 5.861206529521141
Episode: 73: Time: 8.893415927886963 Total Reward: 255.99337748344706 Avg_Loss: 5.880333795266993
Episode: 74: Time: 9.067109107971191 Total Reward: 383.4172661870447 Avg_Loss: 5.518294702057077
Episode: 75: Time: 9.276307106018066 Total Reward: 346.1764705882313 Avg_Loss: 6.075960108212063
Episode: 76: Time: 9.055240869522095 Total Reward: 387.31511254019176 Avg_Loss: 5.383485968874282
Episode: 77: Time: 8.838793992996216 Total Reward: 252.51773049645885 Avg_Loss: 5.840982933505242
Episode: 78: Time: 8.845920085906982 Total Reward: 288.606557377053 Avg_Loss: 5.550864078918425
Episode: 79: Time: 8.684959650039673 Total Reward: 469.20233463034197 Avg_Loss: 5.976971710429472
Episode: 80: Time: 8.906686305999756 Total Reward: 482.1812080536823 Avg_Loss: 6.225578040635886
Episode: 81: Time: 8.982645988464355 Total Reward: 455.335570469791 Avg_Loss: 5.969168678051283
Episode: 82: Time: 9.05826449394226 Total Reward: 301.2264150943397 Avg_Loss: 6.014311863594696

Episode: 83: Time: 9.054595470428467 Total Reward: 476.92982456139373 Avg_Loss: 5.649929322114511
Episode: 84: Time: 8.913748979568481 Total Reward: 560.4054054053931 Avg_Loss: 6.14131827965504
Episode: 85: Time: 8.974913358688354 Total Reward: 468.1768953068487 Avg_Loss: 5.958060878665507
Episode: 86: Time: 8.837276935577393 Total Reward: 194.38906752411998 Avg_Loss: 5.73548565291557
Episode: 87: Time: 8.935795783996582 Total Reward: 277.34042553191773 Avg_Loss: 5.6888635288767455
Episode: 88: Time: 9.145404815673828 Total Reward: 298.65079365078464 Avg_Loss: 5.915154963982205
Episode: 89: Time: 3.037879705429077 Total Reward: 11.194533762058825 Avg_Loss: 5.631167059852963
Episode: 90: Time: 8.942649841308594 Total Reward: 418.88888888888516 Avg_Loss: 5.824968677108028
Episode: 91: Time: 8.945473670959473 Total Reward: 439.96503496503146 Avg_Loss: 5.774096612168961
Episode: 92: Time: 8.977162837982178 Total Reward: 387.9931972789103 Avg_Loss: 5.530826104789221
Episode: 93: Time: 8.797797918319702 Total Reward: 496.6030534351044 Avg_Loss: 6.007252356084455
Episode: 94: Time: 8.790629863739014 Total Reward: 439.0909090909032 Avg_Loss: 6.177214349017424
Episode: 95: Time: 9.273578882217407 Total Reward: 359.8286604361312 Avg_Loss: 5.9552068855582165
Episode: 96: Time: 9.403635501861572 Total Reward: 398.54838709676767 Avg_Loss: 5.848199717637871
Episode: 97: Time: 9.27015209197998 Total Reward: 358.674121405748 Avg_Loss: 7.359165480156906
Episode: 98: Time: 8.919915676116943 Total Reward: 357.2968197879827 Avg_Loss: 6.189526237359567
Episode: 99: Time: 9.031726121902466 Total Reward: 518.63636363636 Avg_Loss: 6.111856684965246
Validation Mean Reward: 359.17731124489706 Validation Std Reward: 80.71327194746027
Episode: 100: Time: 9.359397649765015 Total Reward: 168.77952755905966 Avg_Loss: 7.008666909041525
Episode: 101: Time: 9.502763748168945 Total Reward: 437.30769230768806 Avg_Loss: 6.392469161197919
Episode: 102: Time: 9.276285409927368 Total Reward: 303.75389408098636 Avg_Loss: 5.725183223976808
Episode: 103: Time: 9.051869869232178 Total Reward: 525.0716845878071 Avg_Loss: 5.876977665083749
Episode: 104: Time: 8.97771406173706 Total Reward: 525.9150326797306 Avg_Loss: 5.852052958071733
Episode: 105: Time: 8.826242685317993 Total Reward: 538.9622641509325 Avg_Loss: 5.8595993143169816
Episode: 106: Time: 8.954313278198242 Total Reward: 525.209059233439 Avg_Loss: 5.860063932523006
Episode: 107: Time: 9.129685163497925 Total Reward: 233.94736842105627 Avg_Loss: 6.080731180535645
Episode: 108: Time: 9.092817306518555 Total Reward: 794.3129770992251 Avg_Loss: 5.4701331192705815
Episode: 109: Time: 9.06210970878601 Total Reward: 311.59340659339284 Avg_Loss: 5.812028173138113

Episode: 110: Time: 8.97760534286499 Total Reward: 502.3597359735924 Avg_Loss: 5.393580956619327
Episode: 111: Time: 8.957183837890625 Total Reward: 479.25742574256685 Avg_Loss: 5.957861848238136
Episode: 112: Time: 9.271071910858154 Total Reward: 295.5325443786918 Avg_Loss: 6.192001383845546
Episode: 113: Time: 8.876284122467041 Total Reward: 566.9718309859048 Avg_Loss: 6.3718596466449124
Episode: 114: Time: 9.145821332931519 Total Reward: 624.1780821917682 Avg_Loss: 5.804842846734183
Episode: 115: Time: 9.122658967971802 Total Reward: 438.80782918149373 Avg_Loss: 6.335016512069382
Episode: 116: Time: 9.548054456710815 Total Reward: 424.2307692307583 Avg_Loss: 5.997030410446039
Episode: 117: Time: 9.375480890274048 Total Reward: 473.2539682539638 Avg_Loss: 5.956745285947783
Episode: 118: Time: 8.997241735458374 Total Reward: 95.00000000000395 Avg_Loss: 6.0423396675526595
Episode: 119: Time: 8.928192377090454 Total Reward: 368.576158940395 Avg_Loss: 6.381139435687987
Episode: 120: Time: 9.169574499130249 Total Reward: 562.7181208053605 Avg_Loss: 6.2024202266661055
Episode: 121: Time: 9.23650860786438 Total Reward: 569.2599277978225 Avg_Loss: 5.798326379110833
Episode: 122: Time: 9.257285833358765 Total Reward: 654.0909090908967 Avg_Loss: 6.12752515027503
Episode: 123: Time: 9.254431009292603 Total Reward: 334.60288808663387 Avg_Loss: 6.29556497505733
Episode: 124: Time: 9.289996862411499 Total Reward: 431.8138801261716 Avg_Loss: 6.252889081209648
Episode: 125: Time: 9.2678382396698 Total Reward: 467.4999999999545 Avg_Loss: 6.792087465775113
Episode: 126: Time: 8.944157600402832 Total Reward: 647.2680412371033 Avg_Loss: 6.228049789657112
Episode: 127: Time: 8.924472093582153 Total Reward: 470.8362989323758 Avg_Loss: 6.642333588680299
Episode: 128: Time: 8.995054960250854 Total Reward: 742.6383763837558 Avg_Loss: 6.108020710344074
Episode: 129: Time: 9.782647609710693 Total Reward: 297.67015706806217 Avg_Loss: 6.197192994486384
Episode: 130: Time: 9.555768728256226 Total Reward: 439.9544072948234 Avg_Loss: 6.340594404885749
Episode: 131: Time: 9.08438515663147 Total Reward: 397.90780141843845 Avg_Loss: 5.982351170868433
Episode: 132: Time: 9.027397394180298 Total Reward: 466.64383561643047 Avg_Loss: 5.744589814618856
Episode: 133: Time: 8.702022552490234 Total Reward: 492.54863813229167 Avg_Loss: 6.2613062007086615
Episode: 134: Time: 9.332598209381104 Total Reward: 342.888198757763 Avg_Loss: 6.794724473432333
Episode: 135: Time: 9.123998403549194 Total Reward: 380.52447552447114 Avg_Loss: 5.528870404267511
Episode: 136: Time: 9.27253532409668 Total Reward: 575.2508960573357 Avg_Loss: 5.924862504005432
Episode: 137: Time: 9.009866714477539 Total Reward: 666.3636363636247 Avg_Loss: 6.5243209580413435

Episode: 138: Time: 9.150161743164062 Total Reward: 418.15789473683725 Avg_Loss: 6.575525171616498
Episode: 139: Time: 8.928948402404785 Total Reward: 596.5254237288032 Avg_Loss: 6.027219917093005
Episode: 140: Time: 8.987720012664795 Total Reward: 642.762237762226 Avg_Loss: 6.3318118688439124
Episode: 141: Time: 9.00813913345337 Total Reward: 616.8055555555452 Avg_Loss: 6.603807534490313
Episode: 142: Time: 9.115390539169312 Total Reward: 579.9116607773775 Avg_Loss: 6.52598726098277
Episode: 143: Time: 9.207422494888306 Total Reward: 485.6451612903141 Avg_Loss: 6.654843094969998
Episode: 144: Time: 9.366041660308838 Total Reward: 163.06451612903646 Avg_Loss: 6.61655907470639
Episode: 145: Time: 3.6324942111968994 Total Reward: -8.788475836429754 Avg_Loss: 6.5973257312969285
Episode: 146: Time: 9.06598448753357 Total Reward: 691.2318840579638 Avg_Loss: 6.710681978394003
Episode: 147: Time: 9.002128601074219 Total Reward: 471.5529010238816 Avg_Loss: 6.325907439744773
Episode: 148: Time: 9.054104328155518 Total Reward: 663.0645161290242 Avg_Loss: 6.059993513992855
Episode: 149: Time: 9.109380006790161 Total Reward: 580.9581881532996 Avg_Loss: 6.3760623090407424
Validation Mean Reward: 711.2303030879721 Validation Std Reward: 80.55127160966934
Episode: 150: Time: 9.619236469268799 Total Reward: 547.8571428571347 Avg_Loss: 7.041560387411037
Episode: 151: Time: 9.219735383987427 Total Reward: 477.9537366547968 Avg_Loss: 6.785124844863635
Episode: 152: Time: 9.139271020889282 Total Reward: 529.5353159851177 Avg_Loss: 6.361135151205945
Episode: 153: Time: 9.236505508422852 Total Reward: 502.9381443298912 Avg_Loss: 6.251157969987693
Episode: 154: Time: 9.163758277893066 Total Reward: 604.2481203007388 Avg_Loss: 6.389032204611962
Episode: 155: Time: 9.723037958145142 Total Reward: 308.50877192981704 Avg_Loss: 8.795837664804539
Episode: 156: Time: 9.465022563934326 Total Reward: 570.5405405405338 Avg_Loss: 7.40439294766979
Episode: 157: Time: 9.205631017684937 Total Reward: 751.9387755101959 Avg_Loss: 7.215280253346227
Episode: 158: Time: 9.296576261520386 Total Reward: 529.6056782334269 Avg_Loss: 6.811880622591291
Episode: 159: Time: 9.044585704803467 Total Reward: 605.7299270072876 Avg_Loss: 7.655168439660754
Episode: 160: Time: 9.250665187835693 Total Reward: 598.8775510203991 Avg_Loss: 8.002801392258716
Episode: 161: Time: 9.125159978866577 Total Reward: 469.5756457564505 Avg_Loss: 7.047252157155206
Episode: 162: Time: 9.319249629974365 Total Reward: 438.80782918148964 Avg_Loss: 6.59282609745234
Episode: 163: Time: 9.383734703063965 Total Reward: 332.0833333333305 Avg_Loss: 6.674469880697106
Episode: 164: Time: 9.15494680404663 Total Reward: 440.5932203389743 Avg_Loss: 6.264924726566346

Episode: 165: Time: 9.238390445709229 Total Reward: 399.8096885813127 Avg_Loss: 6.876724563225978
Episode: 166: Time: 9.162024021148682 Total Reward: 682.4193548387013 Avg_Loss: 6.842188826128214
Episode: 167: Time: 9.239818096160889 Total Reward: 605.3367003366866 Avg_Loss: 6.901169469376572
Episode: 168: Time: 9.151572227478027 Total Reward: 606.438848920852 Avg_Loss: 6.5898542854966236
Episode: 169: Time: 9.183410882949829 Total Reward: 675.3180212014009 Avg_Loss: 6.455259120764852
Episode: 170: Time: 9.35300874710083 Total Reward: 431.49006622515856 Avg_Loss: 7.010238892891827
Episode: 171: Time: 9.117608547210693 Total Reward: 350.9930313588752 Avg_Loss: 7.185561111995152
Episode: 172: Time: 9.059525728225708 Total Reward: 542.3239436619638 Avg_Loss: 6.87105826870734
Episode: 173: Time: 9.28116250038147 Total Reward: 342.299035369772 Avg_Loss: 6.718605231838066
Episode: 174: Time: 9.198558807373047 Total Reward: 680.3623188405702 Avg_Loss: 7.989525622680407
Episode: 175: Time: 9.22857403755188 Total Reward: 551.3878326996114 Avg_Loss: 7.088277621429508
Episode: 176: Time: 9.231263875961304 Total Reward: 677.2419928825508 Avg_Loss: 7.23193033703235
Episode: 177: Time: 9.193994522094727 Total Reward: 642.5886524822591 Avg_Loss: 7.762847203166545
Episode: 178: Time: 9.241966724395752 Total Reward: 559.6052631578873 Avg_Loss: 7.360956490540705
Episode: 179: Time: 9.092979431152344 Total Reward: 613.9041095890284 Avg_Loss: 7.084643033372254
Episode: 180: Time: 9.30953860282898 Total Reward: 627.22222222213 Avg_Loss: 7.261575407340747
Episode: 181: Time: 9.40207815170288 Total Reward: 426.12676056337136 Avg_Loss: 7.1180779443067665
Episode: 182: Time: 9.3632173538208 Total Reward: 677.1088435374066 Avg_Loss: 7.171522392946131
Episode: 183: Time: 9.220801830291748 Total Reward: 634.4520547945087 Avg_Loss: 8.891976547842265
Episode: 184: Time: 9.10243535041809 Total Reward: 571.6666666666579 Avg_Loss: 7.501421173079675
Episode: 185: Time: 9.325447797775269 Total Reward: 228.03370786517002 Avg_Loss: 7.462831434081583
Episode: 186: Time: 9.430377006530762 Total Reward: 298.5860058309054 Avg_Loss: 7.543626017931129
Episode: 187: Time: 9.33642029762268 Total Reward: 249.72049689440084 Avg_Loss: 7.096131737492666
Episode: 188: Time: 9.254774808883667 Total Reward: 535.508474576262 Avg_Loss: 7.2828914508098315
Episode: 189: Time: 9.368557453155518 Total Reward: 575.1030927834935 Avg_Loss: 8.295567695834055
Episode: 190: Time: 9.326198101043701 Total Reward: 557.777777777678 Avg_Loss: 6.974707424139776
Episode: 191: Time: 9.064075946807861 Total Reward: 602.0802919707938 Avg_Loss: 7.795918186171716
Episode: 192: Time: 8.915533781051636 Total Reward: 663.4905660377241 Avg_Loss: 7.021428292538939

Episode: 193: Time: 9.186847686767578 Total Reward: 507.1897810218916 Avg_Loss: 7.711164915261149
Episode: 194: Time: 9.376569509506226 Total Reward: 434.78056426331705 Avg_Loss: 7.624573596385347
Episode: 195: Time: 9.396681547164917 Total Reward: 553.2758620689541 Avg_Loss: 8.298684581988999
Episode: 196: Time: 9.380217552185059 Total Reward: 357.3076923076892 Avg_Loss: 7.318462701404796
Episode: 197: Time: 9.642503261566162 Total Reward: 345.9221902017235 Avg_Loss: 6.913225777008954
Episode: 198: Time: 8.861536264419556 Total Reward: 646.3127413127265 Avg_Loss: 7.168872042864311
Episode: 199: Time: 8.84920334815979 Total Reward: 674.2307692307588 Avg_Loss: 7.7386015302994675
Validation Mean Reward: 580.7310101725287 Validation Std Reward: 158.73226598640662
Test Mean Reward: 704.801135613288 Test Std Reward: 107.43199837224009

This notebook was converted with convert.ploomber.io


```

import cv2
import numpy as np
import gymnasium as gym
import matplotlib.pyplot as plt
from utils import preprocess #this is a helper function that may be useful to grayscale and crop the image

class EnvWrapper(gym.Wrapper):
    def __init__(
        self,
        env:gym.Env,
        skip_frames:int=4,
        stack_frames:int=4,
        initial_no_op:int=50,
        do_nothing_action:int=0,
        **kwargs
    ):
        """
        Gym Wrapper for CarRacing-v3 that adds frame-skipping, frame-stacking,
        and an initial no-op period to facilitate training.

        Args:
            env (gym.Env): the original environment
            skip_frames (int, optional): the number of frames to skip, in other words we will
            repeat the same action for `skip_frames` steps.
            stack_frames (int, optional): the number of frames to stack as a state, we stack
            `stack_frames` frames to form the state and allow agent understand the motion of
            the car. Defaults to 4.
            initial_no_op (int, optional): the initial number of no-op steps to do nothing at
            the beginning of the episode. Defaults to 50.
            do_nothing_action (int, optional): the action index for doing nothing. Defaults to
            0, which should be correct unless you have modified the
            discretization of the action space.
        """
        super().__init__(env, **kwargs)
        self.initial_no_op = initial_no_op
        self.skip_frames = skip_frames
        self.stack_frames = stack_frames
        self.observation_space = gym.spaces.Box(
            low=0,
            high=1,
            shape=(stack_frames, 84, 84),
            dtype=np.float32
        )
        self.do_nothing_action = do_nothing_action

    def reset(self, **kwargs):
        """
        Reset the environment and perform a sequence of no-op actions before
        returning the initial stacked state.

        Returns:
            stacked_state (np.ndarray): Array shape (stack_frames,84,84).
            info (dict): Env-provided info dict.
        """
        # ===== YOUR CODE HERE =====
        # TODO:
        # 1. call the enviroment reset
        # 2. do nothing for the next self.initial_no_op` steps
        # 3. crop and resize the final frame
        # 4. stack the frames to form the initial state
        # =====

        # step 1

```

```

obs, info = self.env.reset(**kwargs)

# step 2
for _ in range(self.initial_no_op):
    obs, _, terminated, truncated, _ = self.env.step(self.do_nothing_action)
    if terminated or truncated:
        obs, info = self.env.reset(**kwargs)

# step 3
processed_frame = preprocess(obs)  # Assume this returns (84, 84) grayscale float32
np.array

# step 4
frames = []
for _ in range(self.stack_frames):
    frames.append(processed_frame)

self.stacked_state = np.stack(frames, axis=0)

# ===== YOUR CODE ENDS =====

return self.stacked_state, info

def step(self, action):
    """
    Apply the given action with frame-skipping, accumulate rewards,
    and update the frame stack.

    Args:
        action (int): Discrete action index to execute.

    Returns:
        stacked_state (np.ndarray): Updated state with recent `stack_frames` frames.
        reward (float): Sum of rewards over skipped frames.
        terminated (bool): True if env episode ended.
        truncated (bool): True if env truncated.
        info (dict): Env-provided info dict.
    """
    # ===== YOUR CODE HERE =====
    # TODO:
    # 1. take step(action) on underlying env for `self.skip_frames` steps.
    # 2. sum the immediate rewards
    # 3. preprocess the final observed frame.
    # 4. append new frame to `self.stacked_state` and remove oldest.
    # =====

    total_reward = 0.0
    terminated = False
    truncated = False
    info = {}

    # step 1 and step 2
    for _ in range(self.skip_frames):
        obs, reward, terminated, truncated, info = self.env.step(action)
        total_reward += reward
        if terminated or truncated:
            break

    # step 3
    processed_frame = preprocess(obs)

    # step 4
    self.stacked_state = np.concatenate(
        [self.stacked_state[1:], np.expand_dims(processed_frame, axis=0)],
        axis=0
    )

```

```
# ===== YOUR CODE ENDS =====
```

```
return self.stacked_state, total_reward, terminated, truncated, info
```

```
import torch as torch
import torch.nn as nn
```

```
import torch
import torch.nn as nn
import numpy as np
```

```
class MLP(nn.Module):
    def __init__(self, input_size:int, action_size:int,
hidden_size:int=256,non_linear:nn.Module=nn.ReLU):
    """
    input: tuple[int]
        The input size of the image, of shape (channels, height, width)
    action_size: int
        The number of possible actions
    hidden_size: int
        The number of neurons in the hidden layer

    This is a separate class because it may be useful for the bonus questions
    """
    super(MLP, self).__init__()
    # ===== YOUR CODE HERE =====
    # TODO:
    # self.linear1 =
    # self.output =
    # self.non_linear =
    # =====

    self.linear1 = nn.Linear(input_size, hidden_size)
    self.output = nn.Linear(hidden_size, action_size)
    self.non_linear = non_linear()

    # ===== YOUR CODE ENDS =====

    def forward(self, x:torch.Tensor)->torch.Tensor:
        # ===== YOUR CODE HERE =====

        x = self.non_linear(self.linear1(x))
        x = self.output(x)

        # ===== YOUR CODE ENDS =====
        return x
```

```
class Nature_Paper_Conv(nn.Module):
    """
    A class that defines a neural network with the following architecture:
    - 1 convolutional layer with 32 8x8 kernels with a stride of 4x4 w/ ReLU activation
    - 1 convolutional layer with 64 4x4 kernels with a stride of 2x2 w/ ReLU activation
    - 1 convolutional layer with 64 3x3 kernels with a stride of 1x1 w/ ReLU activation
    - 1 fully connected layer with 512 neurons and ReLU activation.
    Based on 2015 paper 'Human-level control through deep reinforcement learning' by Mnih et al
    """
    def __init__(self, input_size:tuple[int], action_size:int,**kwargs):
        """
        input: tuple[int]
            The input size of the image, of shape (channels, height, width)
        action_size: int
            The number of possible actions
        **kwargs: dict
            additional kwargs to pass for stuff like dropout, etc if you would want to
            implement it
        """
        super(Nature_Paper_Conv, self).__init__()
        # ===== YOUR CODE HERE =====
```

```

c, h, w = input_size

self.CNN = nn.Sequential(
    nn.Conv2d(c, 32, kernel_size=8, stride=4),
    nn.ReLU(),
    nn.Conv2d(32, 64, kernel_size=4, stride=2),
    nn.ReLU(),
    nn.Conv2d(64, 64, kernel_size=3, stride=1),
    nn.ReLU()
)

# Compute output size after conv layers
def conv2d_out(size, kernel, stride):
    return (size - kernel) // stride + 1

convw = conv2d_out(conv2d_out(conv2d_out(w, 8, 4), 4, 2), 3, 1)
convh = conv2d_out(conv2d_out(conv2d_out(h, 8, 4), 4, 2), 3, 1)
linear_input_size = convw * convh * 64

# Must match weight file: 512 hidden units
self.MLP = MLP(input_size=linear_input_size, action_size=action_size, hidden_size=512)

# ===== YOUR CODE ENDS =====

def forward(self, x:torch.Tensor)->torch.Tensor:
    # ===== YOUR CODE HERE =====

    x = self.CNN(x)
    x = x.view(x.size(0), -1)
    x = self.MLP(x)

    # ===== YOUR CODE ENDS =====
    return x

```

```
import random
import torch
import numpy as np
```

```
class ReplayBufferDQN:
```

```
    def __init__(self, buffer_size:int, seed:int=42):
        self.buffer_size = buffer_size
        self.seed = seed
        self.buffer = []
        random.seed(self.seed)
```

```
    def add(self, state:np.ndarray, action:int, reward:float, next_state:np.ndarray
        , done:bool):
```

```
        """
```

```
        Add a new experience to the buffer
```

```
        Args:
```

```
            state (np.ndarray): the current state of shape [n_c,h,w]
```

```
            action (int): the action taken
```

```
            reward (float): the reward received
```

```
            next_state (np.ndarray): the next state of shape [n_c,h,w]
```

```
            done (bool): whether the episode is done
```

```
        """
```

```
        self.buffer.append((state, action, reward, next_state, done))
```

```
        if len(self.buffer) > self.buffer_size:
```

```
            self.buffer.pop(0)
```

```
    def sample(self, batch_size:int, device='cpu'):
```

```
        """
```

```
        Randomly sample a batch of experiences from the replay buffer.
```

```
        Args:
```

```
            batch_size (int): the number of samples to take
```

```
        Returns:
```

```
            states (torch.Tensor): Tensor of shape (batch_size, n_channels, height, width),
```

```
            dtype torch.float32.
```

```
            actions (torch.Tensor): Tensor of shape (batch_size,), dtype torch.int64
```

```
(converted via `.long()`).
```

```
            rewards (torch.Tensor): Tensor of shape (batch_size,), dtype torch.float32.
```

```
            next_states (torch.Tensor): Tensor of shape (batch_size, n_channels, height,
```

```
            width), dtype torch.float32.
```

```
            dones (torch.Tensor): Tensor of shape (batch_size,), dtype torch.bool.
```

```
        Notes:
```

```
            1. Use `random.sample` for uniform sampling without replacement.
```

```
            2. Convert NumPy arrays to torch tensors with the correct dtype before moving to
```

```
`device`.
```

```
            3. Use `torch.stack` to combine individual tensors into a batch dimension.
```

```
            4. Keep the output shapes and dtypes consistent.
```

```
        """
```

```
        # ===== YOUR CODE HERE =====
```

```
        # TODO:
```

```
        # 1. sample random indices
```

```
        # 2. collect experiences using the sampled indices
```

```
        # 3. stack and move batches to the specified device, making sure to convert to the
        correct dtype
```

```
        # =====
```

```
        # step 1
```

```
        batch = random.sample(self.buffer, batch_size)
```

```
        # step 2
```

```
        states, actions, rewards, next_states, dones = zip(*batch)
```

```
# step 3
states = torch.stack([torch.tensor(s, dtype=torch.float32) for s in
states]).to(device)
actions = torch.tensor(actions, dtype=torch.int64).to(device)
rewards = torch.tensor(rewards, dtype=torch.float32).to(device)
next_states = torch.stack([torch.tensor(ns, dtype=torch.float32) for ns in
next_states]).to(device)
dones = torch.tensor(dones, dtype=torch.bool).to(device)

# ===== YOUR CODE ENDS =====

return states, actions, rewards, next_states, dones

def __len__(self):
    return len(self.buffer)
```

```

import torch
import torch.optim as optim
import torch.nn.functional as F
import torch.nn
import gymnasium as gym
from replay_buffer import ReplayBufferDQN
import wandb
import random
import numpy as np
import os
import time
from utils import exponential_decay
import typing

```

TODO: change the logging here if you don't like wandb

```
class DQN:
```

```

    def __init__(self, env:typing.Union[gym.Env,gym.Wrapper],
        #model params
        model:torch.nn.Module,
        model_kwargs:dict = {},
        #overall hyperparams
        lr:float = 0.001, gamma:float = 0.99,
        buffer_size:int = 10000, batch_size:int = 32,
        loss_fn:str = 'mse_loss',
        use_wandb:bool = False,
        device:str = 'cpu',
        seed:int = 42,
        epsilon_scheduler = exponential_decay(1,700,0.1),
        save_path:str = None):

```

"""Initializes the DQN algorithm

Args:

env (gym.Env/gym.Wrapper): the environment to train on
model (torch.nn.Module): the model to train
model_kwargs (dict, optional): the keyword arguments to pass to the model. Defaults

to {}.

lr (float, optional): the learning rate to use in the optimizer. Defaults to 0.001.
gamma (float, optional): discount factor. Defaults to 0.99.
buffer_size (int, optional): the size of the replay buffer. Defaults to 10000.
batch_size (int, optional): the batch size. Defaults to 32.
loss_fn (str, optional): the name of the loss function to use. Defaults to

'mse_loss'.

use_wandb (bool, optional): _description_. Defaults to False.
device (str, optional): _description_. Defaults to 'cpu'.
seed (int, optional): the seed to use for reproducibility. Defaults to 42.
epsilon_scheduler ([type], optional): the epsilon scheduler to use, must have a

__call__ method that returns a float between 0 and 1

save_path (str, optional): _description_. Defaults to None.

Raises:

ValueError: _description_

"""

```
self.env = env
```

```
self._set_seed(seed)
```

```
self.observation_space = self.env.observation_space.shape
```

```

self.model = model(
    self.observation_space,
    self.env.action_space.n, **model_kwargs
).to(device)
self.model.train()

```

```
self.optimizer = optim.Adam(self.model.parameters(), lr = lr)
```

```
self.gamma = gamma
```

```
self.replay_buffer = ReplayBufferDQN(buffer_size)
```



```

self.batch_size = batch_size
self.i_update = 0
self.device = device
self.epsilon_decay = epsilon_scheduler
self.save_path = save_path if save_path is not None else "./"

#set the loss function
if loss_fn == 'smooth_l1_loss':
    self.loss_fn = F.smooth_l1_loss
elif loss_fn == 'mse_loss':
    self.loss_fn = F.mse_loss
else:
    raise ValueError('loss_fn must be either smooth_l1_loss or mse_loss')

self.wandb = use_wandb
if self.wandb:
    wandb.init(project = 'racing-car-dqn')
    #log the hyperparameters
    wandb.config.update({
        'lr': lr,
        'gamma': gamma,
        'buffer_size': buffer_size,
        'batch_size': batch_size,
        'loss_fn': loss_fn,
        'device': device,
        'seed': seed,
        'save_path': save_path
    })

def train(self, n_episodes:int = 1000,validate_every:int = 100, n_validation_episodes:int
= 10, n_test_episodes:int = 10,save_every:int = 100):
    os.makedirs(self.save_path, exist_ok = True)
    best_val_reward = -np.inf

    for episode in range(n_episodes):
        state,_ = self.env.reset()
        done = False
        truncated = False
        total_reward = 0
        i = 0
        loss = 0
        start_time = time.time()
        epsilon = self.epsilon_decay()
        while (not done) and (not truncated):
            action = self._sample_action(state, epsilon)
            next_state, reward, done, truncated, _ = self.env.step(action)
            self.replay_buffer.add(state, action, reward, next_state, done)
            total_reward += reward
            state = next_state

            not_warm_starting,l = self._optimize_model()
            if not_warm_starting:
                loss += l
                epsilon = self.epsilon_decay()
                i += 1

        if i == 0:
            avg_loss = loss / (i+1)
        else:
            avg_loss = loss/i
        if self.wandb:
            wandb.log({'total_reward': total_reward, 'loss': avg_loss})
        print(f"Episode: {episode}: Time: {time.time() - start_time} Total Reward:
{total_reward} Avg_Loss: {avg_loss}")
        if episode % validate_every == validate_every - 1:
            mean_reward, std_reward = self.validate(n_validation_episodes)
            if self.wandb:
                wandb.log({'mean_reward': mean_reward, 'std_reward': std_reward})

```

```

        print("Validation Mean Reward: {}".format(mean_reward))
        print("Validation Std Reward: {}".format(std_reward))
        if mean_reward > best_val_reward:
            best_val_reward = mean_reward
            self._save('best')

    if episode % save_every == save_every - 1:
        self._save(str(episode))

    self._save('final')
    self.load_model('best')
    mean_reward, std_reward = self.validate(n_test_episodes)
    if self.wandb:
        wandb.log({'mean_test_reward': mean_reward, 'std_test_reward': std_reward})
    print("Test Mean Reward: {} Test Std Reward: {}".format(mean_reward, std_reward))

def _optimize_model(self):
    """
    Performs one optimization step on the DQN.

    Returns:
        bool: whether we have enough samples to optimize the model, which we define as
        having at least 10*batch_size samples
        float: the loss, if we do not have enough samples, we return 0
    """
    # ===== YOUR CODE HERE =====
    # TODO: some hints to get you started:
    # 1. check if the replay buffer has enough samples
    # 2. sample a minibatch
    # 3. No-op
    # 4. compute current Q: q_values = ...
    # 5. compute target Q
    # 6. compute loss between current Q and target Q
    # 7. backprop
    # =====

    # step 1
    if len(self.replay_buffer) < (10 * self.batch_size):
        return False, 0.0

    # step 2
    states, actions, rewards, next_states, dones =
self.replay_buffer.sample(self.batch_size, device=self.device)

    # step 3
    # No-op

    # step 4
    # Select Q-values for the taken actions
    q_values = self.model(states) # shape: (batch_size, num_actions)

    # step 5: target_q
    with torch.no_grad():
        next_q_values = self.model(next_states)
        max_next_q_values = next_q_values.max(dim=1)[0]
        if dones.all():
            targets = rewards
        else:
            targets = rewards + self.gamma * max_next_q_values

    # step 6
    q_values_given_action = q_values.gather(1, actions.unsqueeze(1)).squeeze(1) # shape:
(batch_size,)
    loss = self.loss_fn(q_values_given_action, targets)

    # step 7
    self.optimizer.zero_grad()

```

```

loss.backward()
self.optimizer.step()

return True, loss.item()

# ===== YOUR CODE ENDS =====

def _sample_action(self, state:np.ndarray
                  , epsilon:float = 0.1)->int:
    """
    Samples an action from the model

    Args:
        state (np.ndarray): the state, of shape [n_c,h,w]
        epsilon (float, optional): the epsilon for epsilon greedy. Defaults to 0.1.

    Returns:
        int: the index of the action to take
    """
    # ===== YOUR CODE HERE =====
    # TODO:
    # Epsilon-greedy action selection:
    # - if probability epsilon: random action
    # - else: greedy action
    # =====

    if random.random() < epsilon:
        index = self.env.action_space.sample()
    else:
        state_tensor = torch.tensor(state, dtype=torch.float32,
device=self.device).unsqueeze(0)
        with torch.no_grad():
            q_values = self.model(state_tensor)
            index = q_values.argmax(dim=1).item()

    # ===== YOUR CODE ENDS =====
    return index

def _set_seed(self, seed:int):
    random.seed(seed)
    np.random.seed(seed)
    self.seed = seed
    torch.manual_seed(seed)
    torch.cuda.manual_seed(seed)
    torch.backends.cudnn.deterministic = True
    gym.utils.seeding.np_random(seed)

def _validate_once(self):
    state,_ = self.env.reset()
    done = False
    truncated = False
    total_reward = 0
    i = 0
    # epsilon = self.epsilon_decay()
    while (not done) and (not truncated):
        action = self._sample_action(state, 0)
        # out = self.env.step(action)
        next_state, reward, done, truncated, _ = self.env.step(action)
        # next_state = np.array(state_buffer[-self.n_frames:])
        total_reward += reward
        state = next_state
    return total_reward

def validate(self, n_episodes:int = 10):
    # self.model.eval()
    rewards_per_episode = []

```

```

for _ in range(n_episodes):
    rewards_per_episode.append(self._validate_once())
    # self.model.train()
    return np.mean(rewards_per_episode), np.std(rewards_per_episode)

def load_model(self, suffix:str = ''):
    self.model.load_state_dict(torch.load(os.path.join(self.save_path,
f'model_{suffix}.pt'))))

def _save(self, suffix:str = ''):
    torch.save(self.model.state_dict(), os.path.join(self.save_path,
f'model_{suffix}.pt'))

def play_episode(self, epsilon:float = 0, return_frames:bool = True, seed:int = None):
    """Plays an episode of the environment

    Args:
        epsilon (float, optional): the epsilon for epsilon greedy. Defaults to 0.
        return_frames (bool, optional): whether we should return frames. Defaults to True.
        seed (int, optional): the seed for the enviroment. Defaults to None.

    Returns:
        if return_frames is True, returns the total reward and the frames
        if return_frames is False, returns the total reward
    """
    if seed is not None:
        state, _ = self.env.reset(seed = seed)
    else:
        state, _ = self.env.reset()

    done = False
    total_reward = 0
    if return_frames:
        frames = []

    with torch.no_grad():
        while not done:
            action = self._sample_action(state, epsilon)
            next_state, reward, terminated, truncated, _ = self.env.step(action)
            total_reward += reward
            done = terminated or truncated
            if return_frames:
                frames.append(self.env.render())
            state = next_state

    if return_frames:
        return total_reward, frames

    return total_reward

```

```

class HardUpdateDQN(DQN):

```

```

    def __init__(self, env, model, model_kwargs:dict = {},
        update_freq:int = 5, *args, **kwargs):
        super().__init__(env, model, model_kwargs, *args, **kwargs)
        # ===== YOUR CODE HERE =====
        # TODO:
        # fill in the initialization and synchronization of the target model weights
        # =====

        # Initialize target network with same architecture
        self.target_model = model(
            self.observation_space,
            self.env.action_space.n,
            **model_kwargs

```

```

).to(self.device)

# Copy initial weights from model to target_model
self.target_model.load_state_dict(self.model.state_dict())

# Set target network to eval mode (not training)
self.target_model.eval()

# Store update frequency for hard target updates
self.update_freq = update_freq

# ===== YOUR CODE ENDS =====

def _optimize_model(self):
    """Optimizes the model

    Returns:
        bool: whether we have enough samples to optimize the model, which we define as
        having at least 10*batch_size samples
        float: the loss, if we do not have enough samples, we return 0
    """
    # ===== YOUR CODE HERE =====
    # TODO:
    # hint: you can copy over most of the code from the parent class
    # and only change one line
    # =====

    if len(self.replay_buffer) < (10 * self.batch_size):
        return False, 0.0

    # Sample batch
    states, actions, rewards, next_states, dones =
self.replay_buffer.sample(self.batch_size, device=self.device)

    # Compute Q(s, a) from current model
    q_values = self.model(states)
    q_values_given_action = q_values.gather(1, actions.unsqueeze(1)).squeeze(1)

    # Compute target Q using target network
    with torch.no_grad():
        next_q_values = self.target_model(next_states) # changed from self.model
        max_next_q_values = next_q_values.max(dim=1)[0]
        targets = rewards + self.gamma * max_next_q_values * (~dones)

    # Compute loss and update
    loss = self.loss_fn(q_values_given_action, targets)
    self.optimizer.zero_grad()
    loss.backward()
    self.optimizer.step()

    # Update target model if needed
    self._update_model()

    return True, loss.item()

# ===== YOUR CODE ENDS =====

def _update_model(self):
    self.i_update += 1
    if self.i_update % self.update_freq == 0:
        self.target_model.load_state_dict(self.model.state_dict())

def _save(self, suffix:str = ''):
    torch.save(self.model.state_dict(), os.path.join(self.save_path,
f'model_{suffix}.pt'))
    torch.save(self.target_model.state_dict(), os.path.join(self.save_path,
f'target_model_{suffix}.pt'))

```

```

def load_model(self, suffix:str = ''):
    self.model.load_state_dict(torch.load(os.path.join(self.save_path,
f'model_{suffix}.pt'))))
    self.target_model.load_state_dict(torch.load(os.path.join(self.save_path,
f'target_model_{suffix}.pt'))))

class SoftUpdateDQN(HardUpdateDQN):
    def __init__(self,env,model,model_kwargs:dict = {},
        tau:float = 0.01,*args,**kwargs):
        super().__init__(env,model,model_kwargs,*args,**kwargs)
        self.tau = tau

    def _update_model(self):
        """
        Soft updates the target model
        """
        # ===== YOUR CODE HERE =====
        # TODO
        # =====

        for target_param, model_param in zip(self.target_model.parameters(),
self.model.parameters()):
            target_param.data.copy_(self.tau * model_param.data + (1.0 - self.tau) *
target_param.data)

        # ===== YOUR CODE ENDS =====

```