SOC_25-Final Report
RL: Race And Learning

Train a deep reinforcement learning (RL) agent using Proximal Policy Optimization (PPO) to solve the CarRacing-v3 environment using raw visual input only, with the goal of achieving stable driving behavior and a high cumulative reward.

Preprocessing Pipeline

- **Frame Format:** RGB frames converted to grayscale
- **Resized To:** 96 x 96 pixels
- **Frame Stack:** Last 4 frames stacked channel-wise → input shape: (4, 96, 96)
- **Purpose:** Captures temporal dependencies like speed and turns using recent visual history
- **Wrapper Used:** Framestackwrapper — handles grayscale conversion, resizing, and stacking

PPO Policy Architecture

- **Base Network:** Custom CNN with 3 convolutional layers

Heads:

- **Actor Head:** Linear → Relu → Linear → Softmax over 5 discrete actions
- **Critic Head:** Linear → Relu → Linear → Scalar Value

Action Space: Discrete (5 actions mapped to combinations of steer, gas, brake)

Normalization: Inputs scaled to [0, 1] by dividing by 255

Training Process

- **Training Status:** Current code performs inference only, not training
- **Missing Components:** No reward tracking, no episode logging, no PPO loss/backprop
- **To be added:**
  - Buffer for trajectories (states, actions, rewards)
  - Policy & value loss functions with PPO clipping
  - Reward curves (episodic returns) for evaluation

Reward Shaping

- No custom reward shaping is applied in the current inference loop.
- The environment’s native reward is used directly.
- In future, shaping ideas:
  - Penalize going off-track
  - Bonus for sustained acceleration or staying within boundaries

Observations & Insights

- **Challenges:**
  - Sparse rewards for off-road or early crashes
  - Visual noise from background and track lighting
  - Requires long-term temporal memory, difficult with pure CNN
- **Interesting Behaviors:**
  - Learned policies often spin or drift when overfitting
  - Steering too aggressively can break the lap flow
- **Overfitting:**
  - Without proper regularization or diverse seeds, the policy may overfit to early tracks or corner cases
- **Generalization:**
  - Frame stacking and grayscale preprocessing help generalize across different turns and lighting
  - Further improvement requires randomizing track seeds during training

Output

- A video of agent inference is saved as **car_racing_run.mp4**