EXPERIMENTING WITH INPUT TO THE DQN

Ayan Abhiranya Singh

CS 298

Gym's Atari environment and the DeepMind wrapper

- The env.reset() function resets the environment to the initial state (start of the game). The DeepMind wrapper creates a stack of frames by duplicating the first frame 4 times.

- The env.step(action) function plays a move within the environment and returns a single observation, i.e., 1 frame per step.

- The DeepMind wrapper is stacking the last k number of frames, in this case, 4. The stack uses the deque data structure.

- So, FrameStack's step function takes one action and then appends the latest frame to the current queue of frames.
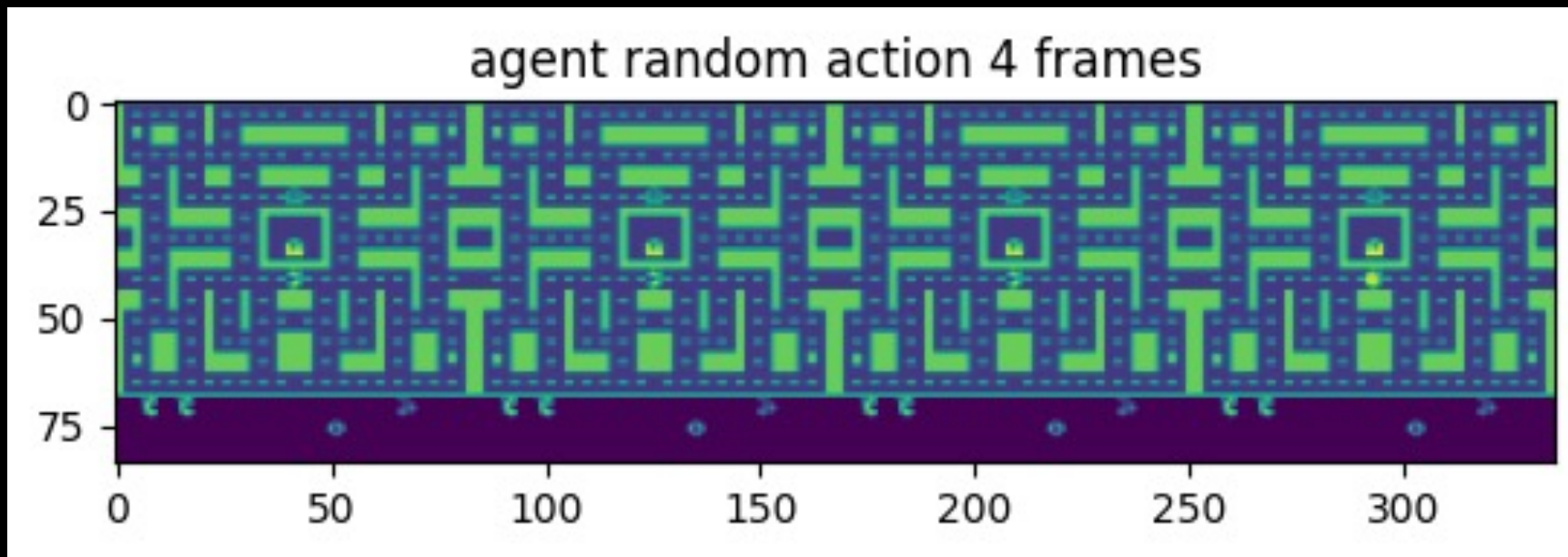
Visualizing the frame stack

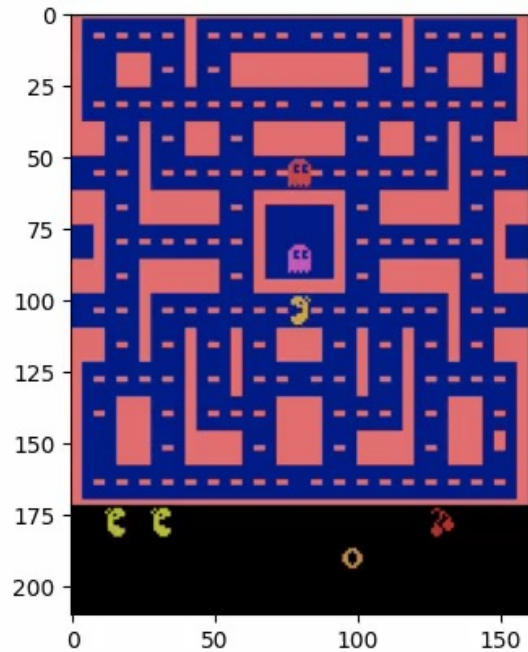- Game loaded, initial observation (1st frame duplicated 4 times)
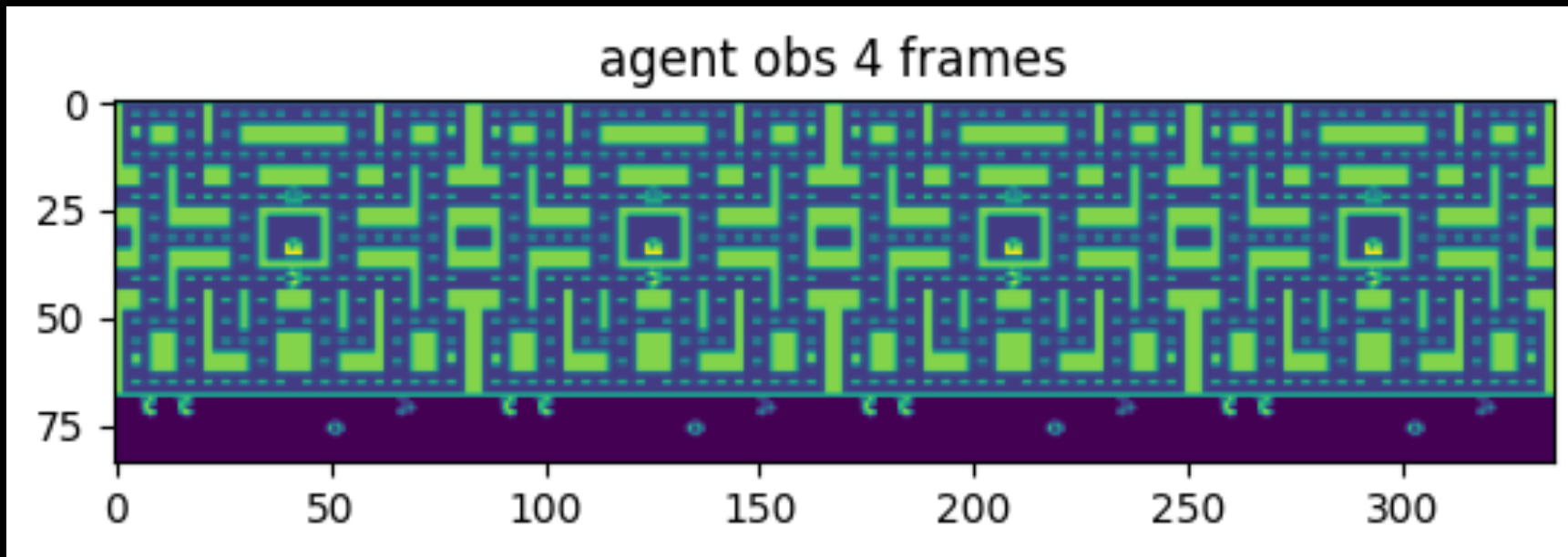
Visualizing the frame stack

- One action executed, single new frame appended to end of the stack

Measuring time elapsed per frame

- Approach: Time gameplay in the Stella emulator versus how many frames the agent in the Gym environment takes to move the same distance.

- It takes 3.70s to move from bottom left to bottom right (45 frames, around 1 frame per 0.0822 of a second).

- It takes 5.70s from game launch to reach the first left wall (68 frames, around 1 frame per 0.0853 of a second).

- One frame represents approximately 0.085 seconds.

Agent plays Ms. PacMan, trained on 5000 episodes
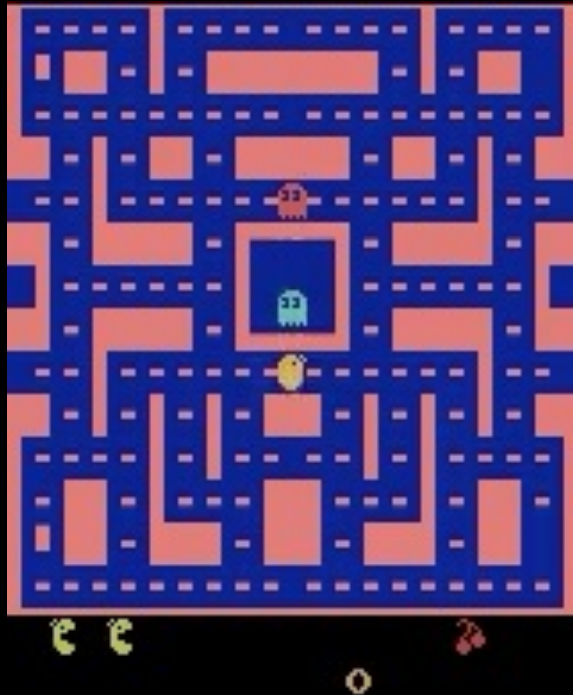
- In this experiment, input is provided to the net in the same manner as training. If frame 1 is the observation for the first time step, frame 2 is the observation for the second time step, then the progression of input to the net looks like [1111] ➔ [1112] ➔ [1123] ➔ [1234], and so on.

- The agent is trained over 5000 episodes, working on the reduced move set (NO-OP, UP, DOWN, LEFT, RIGHT)

- The agent scores 850 points, a decent return after training the net for 21 minutes and 35 seconds.

Experiment #1: Duplicating frames in input
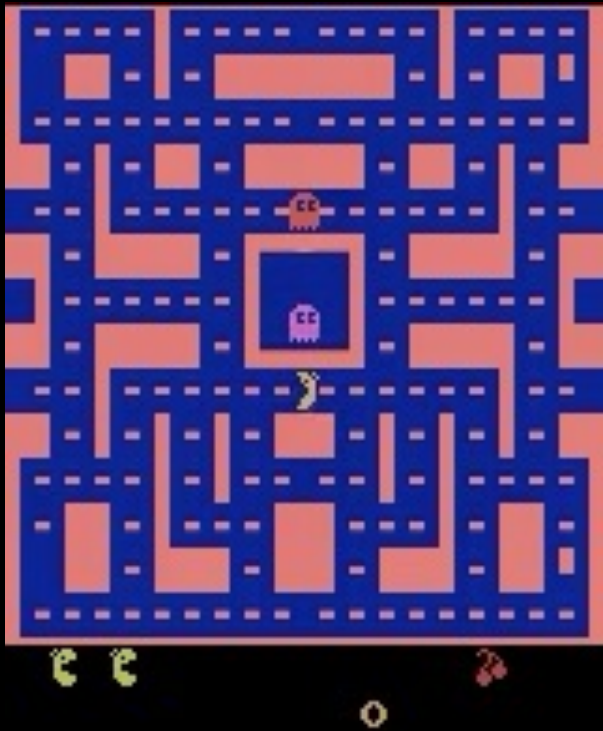
- In this experiment, we duplicate frames for every pair of 2 moves. If frame 1 is the observation for the first time step, frame 2 is the observation for the second time step, then the progression of input to the net looks like [1111] → [1122] → [2233] → [3344], and so on.

- The agent is trained over 5000 episodes, working on the reduced move set (NO-OP, UP, DOWN, LEFT, RIGHT)

- The agent scores 530 points, lower than the previous score of 850.

Notes

- The behavior of the yellow ghost appears to be deterministic.

- Possible bug as the agent uses the "FIRE" command to reset (FIRE may have mapped to the UP command in the net's move choice list).