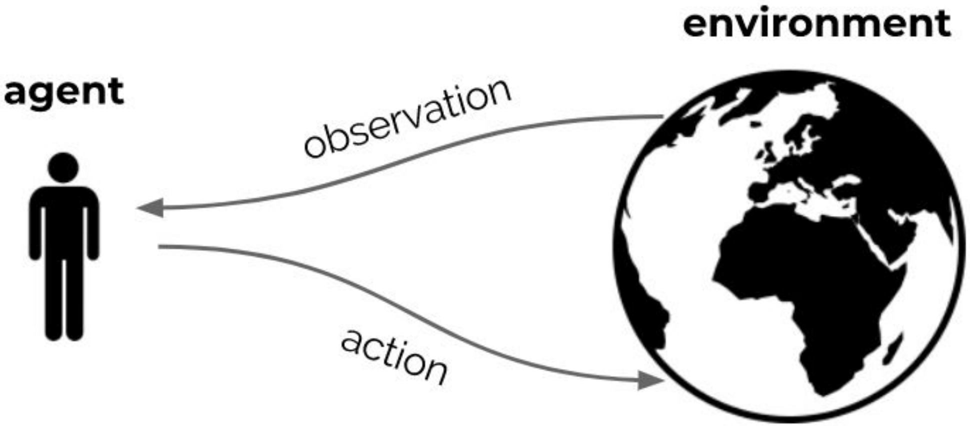
Game Playing Agent using Deep Reinforcement Learning

Supervisor - Dr.Marikkannan
Team members - Anbuselvam, Gobinath, Balaji
Government College of Engineering, Erode.

Reinforcement Learning Framework



Goal : Optimize the sum of rewards by taking good actions

Reward is enough

Intelligence and its associated abilities can be understood as subserving the maximisation of reward. So powerful reinforcement learning agents could constitute a solution to AGI.

Paper - https://www.sciencedirect.com/science/article/pii/S0004370221000862

RL as a fine-tuning paradigm

Fine-tuning large language models using RL

- ➔ https://openai.com/blog/instruction-following/
- ➔ https://openai.com/blog/summarizing-books/

Blog - https://ankeshanand.com/blog/2022/01/08/rl-fine-tuning.html

RL in the Real world

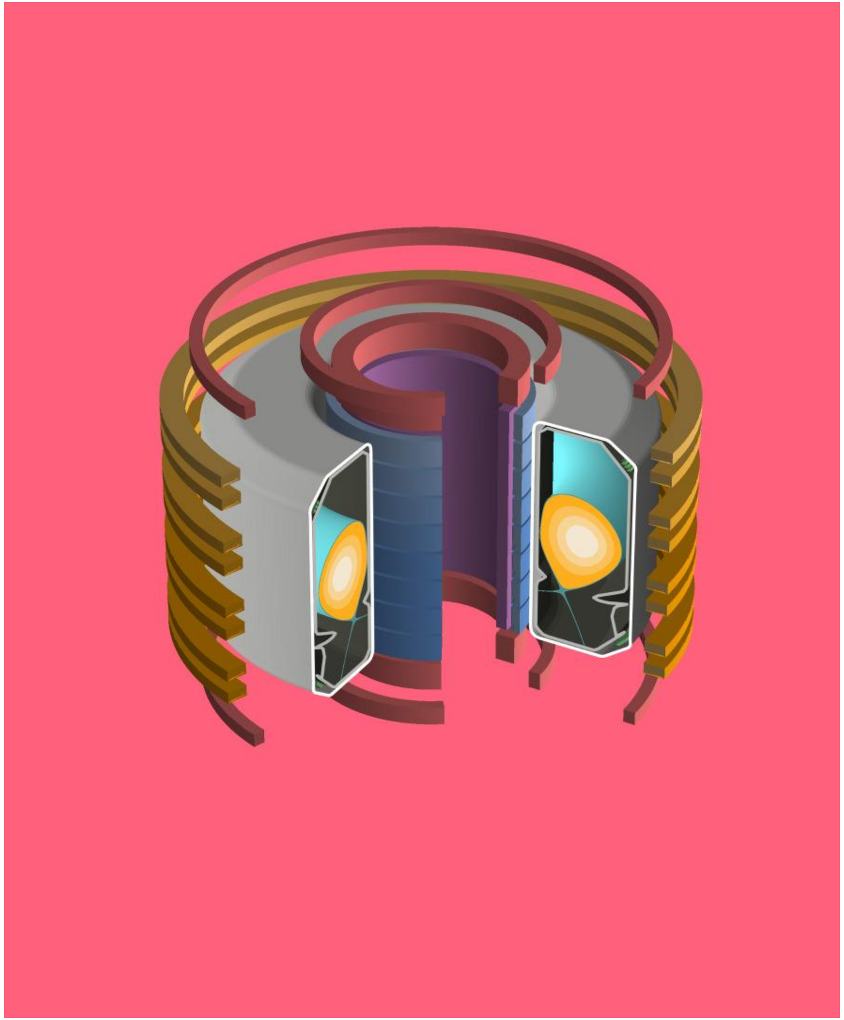
- Robotics
- Control problems
- Self driving vehicles
- Chip Design - https://ai.googleblog.com/2020/04/chip-design-with-deep-reinforcement.html
- Drug discovery - mila.quebec/en/ai-society/exascale-search-of-molecules/
- Video compression - https://www.deepmind.com/blog/muzeros-first-step-from-research-into-the-real-world

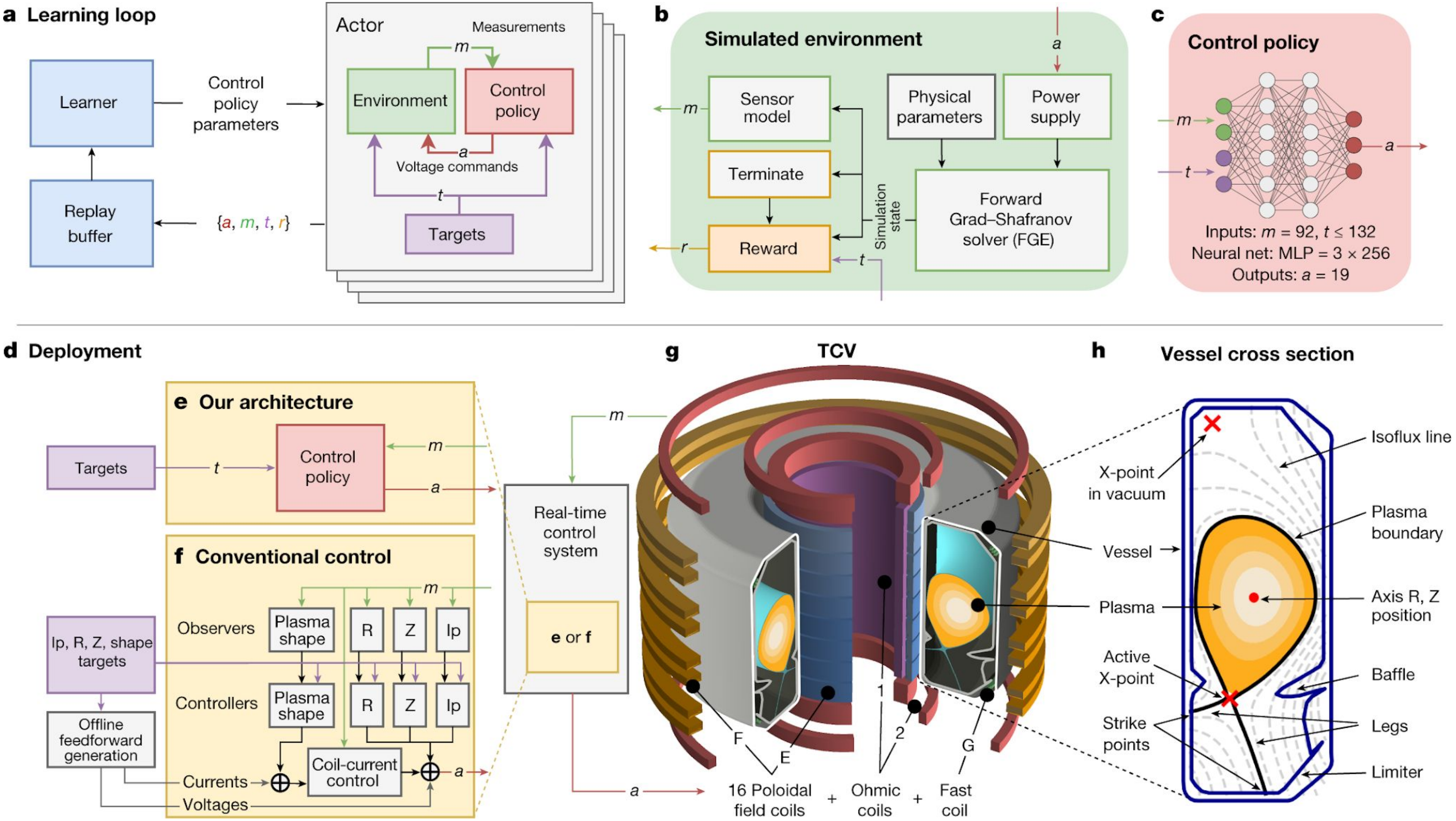
and many more…

Case Study - Controlling Fusion Reactor using DeepRL

References :

Blog - deepmind.com/blog/accelerating-fusion-science-through-learned-plasma-control
Paper - https://www.nature.com/articles/s41586-021-04301-9



a Learning loop
Learner
Control policy parameters
Replay buffer
{a, m, t, r}
Actor
Measurements
m
Environment
Control policy
a
Voltage commands
t
Targets
b
Simulated environment
a
m
Sensor model
Physical parameters
Power supply
Terminate
Simulation state
Forward Grad–Shafranov solver (FGE)
r
Reward
t
c
Control policy
m
t
a
Inputs: m = 92, t ≤ 132
Neural net: MLP = 3 × 256
Outputs: a = 19
d Deployment
e Our architecture
Targets
t
Control policy
m
a
f Conventional control
m
Observers
Plasma shape
R
Z
Ip
Ip, R, Z, shape targets
Controllers
Plasma shape
R
Z
Ip
Offline feedforward generation
Currents
Coil-current control
a
Voltages
m
Real-time control system
e or f
a
g
TCV
F
E
1
2
G
16 Poloidal field coils
+
Ohmic coils
+
Fast coil
h
Vessel cross section
X-point in vacuum
Isoflux line
Plasma boundary
Vessel
Plasma
Axis R, Z position
Active X-point
Baffle
Strike points
Legs
Limiter

Why Games?

Games are like microcosmos of the outside world.
If we look at the history of AI, Games have been the testing ground for AI algorithms.

And there are success stories of AI beating Humans in games.

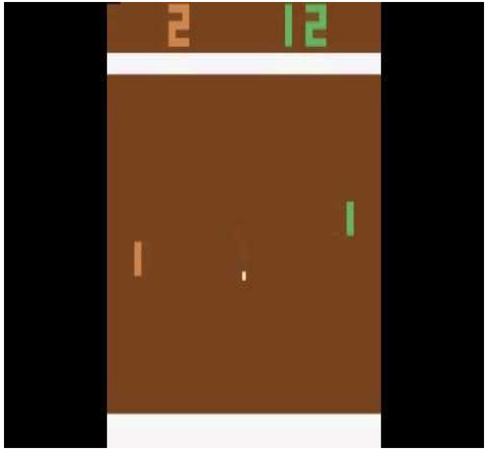
Game playing has no practical applications, but we can use the algorithm behind game playing, for practical applications like driving a car, walking a robot or controlling fusion reactors.

Project Description

To implement a Deep Reinforcement learning algorithm to achieve superhuman performance at Atari Breakout, Space Invaders and Pong.

Our goal

❖ To learn about different RL algorithms & validate it on an environment using OpenAI Gym.
❖ To develop a DeepRL algorithm to play Breakout, Space Invaders, Pong.



What we’ll be using

- Python
- PyTorch
- OpenAI Gym

Our Innovation

1. We'll be using Vision Transformers (ViT) to play the games.

Learning Resources

1. Deepmind RL lectures 2021
2. Coursera RL Specialization
3. RL book - Richard Sutton and Andrew Barto
4. Foundations of DeepRL - Laura Graesser and Wah Loon Keng
5. Spinning up in DeepRL - OpenAI
6. RAIL course on DeepRL - UC Berkeley

What we did

We implemented REINFORCE (vanilla policy gradient algorithm) and DQN algorithm.

First we tested both on CartPole environment, then we used DQN algorithm to train on Breakout, Space Invaders, Pong.

DQN algorithm

Paper - https://www.nature.com/articles/nature14236

Algorithm 1: deep Q-learning with experience replay.

Initialize replay memory D to capacity N

Initialize action-value function Q with random weights θ

Initialize target action-value function $\hat{Q}$ with weights $\theta^{-} = \theta$

For episode = 1, M **do**

  Initialize sequence $s_1 = \{x_1\}$ and preprocessed sequence $\phi_1 = \phi(s_1)$

  For $t = 1$,T **do**

    With probability ε select a random action a_t

    otherwise select $a_t = \text{argmax}_a Q(\phi(s_t), a; \theta)$

    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 D

    Sample random minibatch of transitions $\left(\phi_j, a_j, r_j, \phi_{j+1}\right)$ from D

    Set $y_j = \begin{cases} r_j & \text{if episode terminates at step j+1} \\ r_j + \gamma \max_{a'} \hat{Q}\left(\phi_{j+1}, a'; \theta^{-}\right) & \text{otherwise} \end{cases}$

    Perform a gradient descent step on $\left(y_j - Q\left(\phi_j, a_j; \theta\right)\right)^2$ with respect to the network parameters θ

    Every C steps reset $\hat{Q} = Q$

  End For

End For

Convolutional neural network

```
self.features = nn.Sequential(
                    nn.Conv2d(input_shape[0], 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() ) )

self.fc = nn.Sequential(
                    nn.Linear(self.feature_size(), 512), nn.ReLU(),
                    nn.Linear(512, self.num_actions) )
```

Vision Transformer

Paper - https://openreview.net/pdf?id=YicbFdNTTy

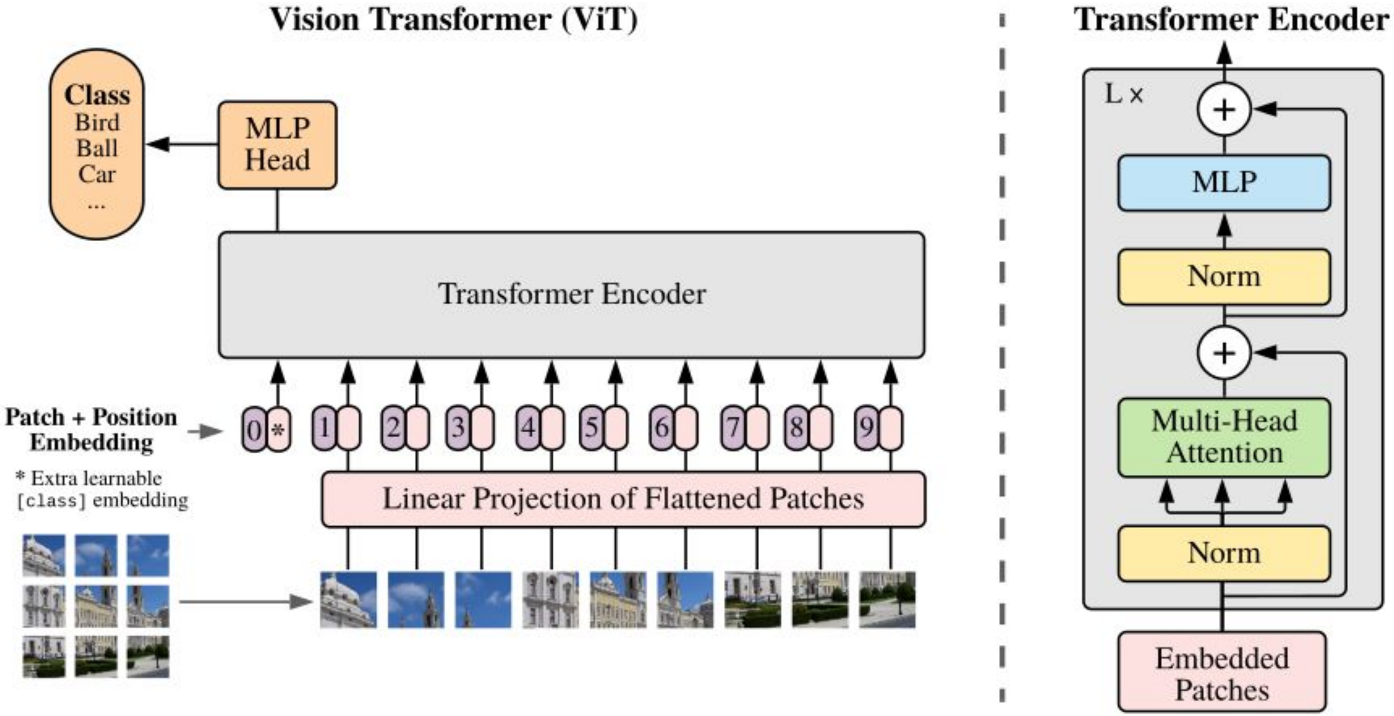
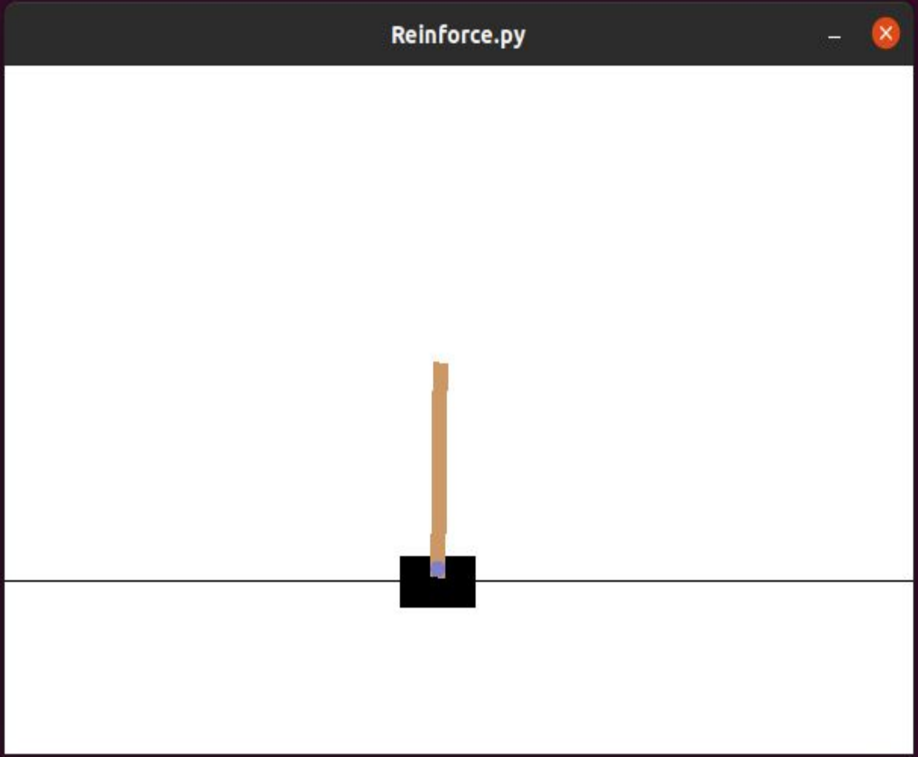


Figure 1: Model overview. We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, we use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

Results on CartPole

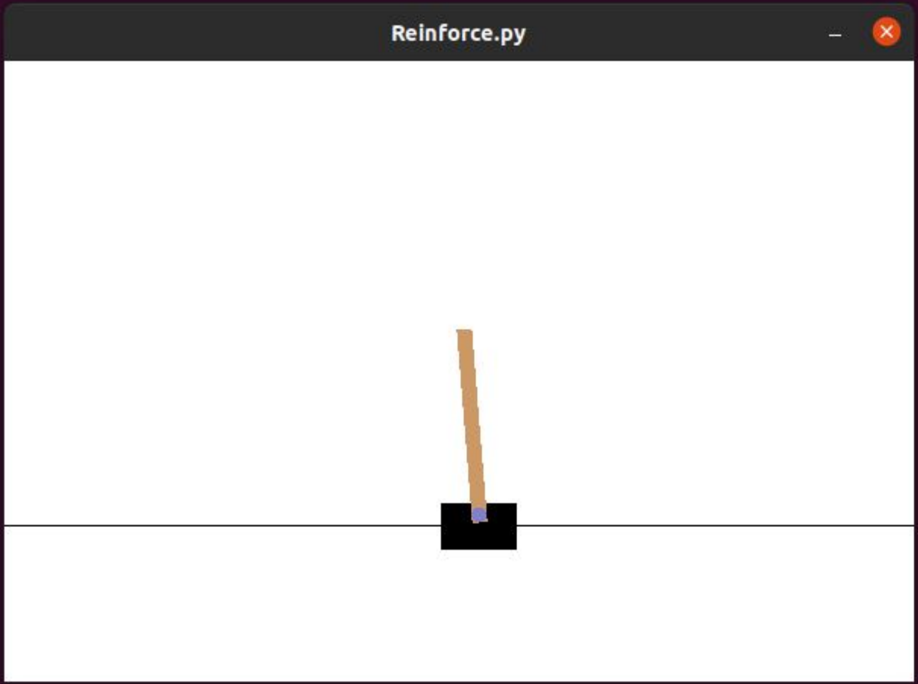
balaji@balaji-pc: ~/Downloads

```
Episode 47, Loss: 1507.42310, Total_reward: 81.0, Solved: False
Episode 48, Loss: 694.03735, Total_reward: 49.0, Solved: False
Episode 49, Loss: 191.11530, Total_reward: 25.0, Solved: False
Episode 50, Loss: 127.59194, Total_reward: 20.0, Solved: False
Episode 51, Loss: 1318.04370, Total_reward: 76.0, Solved: False
Episode 52, Loss: 1121.55579, Total_reward: 67.0, Solved: False
Episode 53, Loss: 996.42065, Total_reward: 65.0, Solved: False
Episode 54, Loss: 150.20213, Total_reward: 21.0, Solved: False
Episode 55, Loss: 855.16724, Total_reward: 60.0, Solved: False
Episode 56, Loss: 479.91821, Total_reward: 44.0, Solved: False
Episode 57, Loss: 3659.35938, Total_reward: 142.0, Solved: False
Episode 58, Loss: 102.46311, Total_reward: 16.0, Solved: False
Episode 59, Loss: 2931.17944, Total_reward: 127.0, Solved: False
Episode 60, Loss: 926.02765, Total_reward: 65.0, Solved: False
Episode 61, Loss: 966.83362, Total_reward: 69.0, Solved: False
Episode 62, Loss: 3552.70557, Total_reward: 154.0, Solved: False
Episode 63, Loss: 1105.90613, Total_reward: 80.0, Solved: False
Episode 64, Loss: 470.19693, Total_reward: 47.0, Solved: False
Episode 65, Loss: 1000.59070, Total_reward: 82.0, Solved: False
Episode 66, Loss: 783.85223, Total_reward: 73.0, Solved: False
Episode 67, Loss: 1587.05054, Total_reward: 98.0, Solved: False
Episode 68, Loss: 1468.55261, Total_reward: 98.0, Solved: False
Episode 69, Loss: 479.69556, Total_reward: 57.0, Solved: False
Episode 70, Loss: 834.55756, Total_reward: 80.0, Solved: False
Episode 71, Loss: 1172.16736, Total_reward: 87.0, Solved: False
Episode 72, Loss: 1609.24512, Total_reward: 119.0, Solved: False
Episode 73, Loss: 740.49292, Total_reward: 80.0, Solved: False
Episode 74, Loss: 2634.03662, Total_reward: 153.0, Solved: False
Episode 75, Loss: 429.32166, Total_reward: 58.0, Solved: False
Episode 76, Loss: 870.25714, Total_reward: 90.0, Solved: False
Episode 77, Loss: 1132.46838, Total_reward: 89.0, Solved: False
Episode 78, Loss: 761.31311, Total_reward: 80.0, Solved: False
```

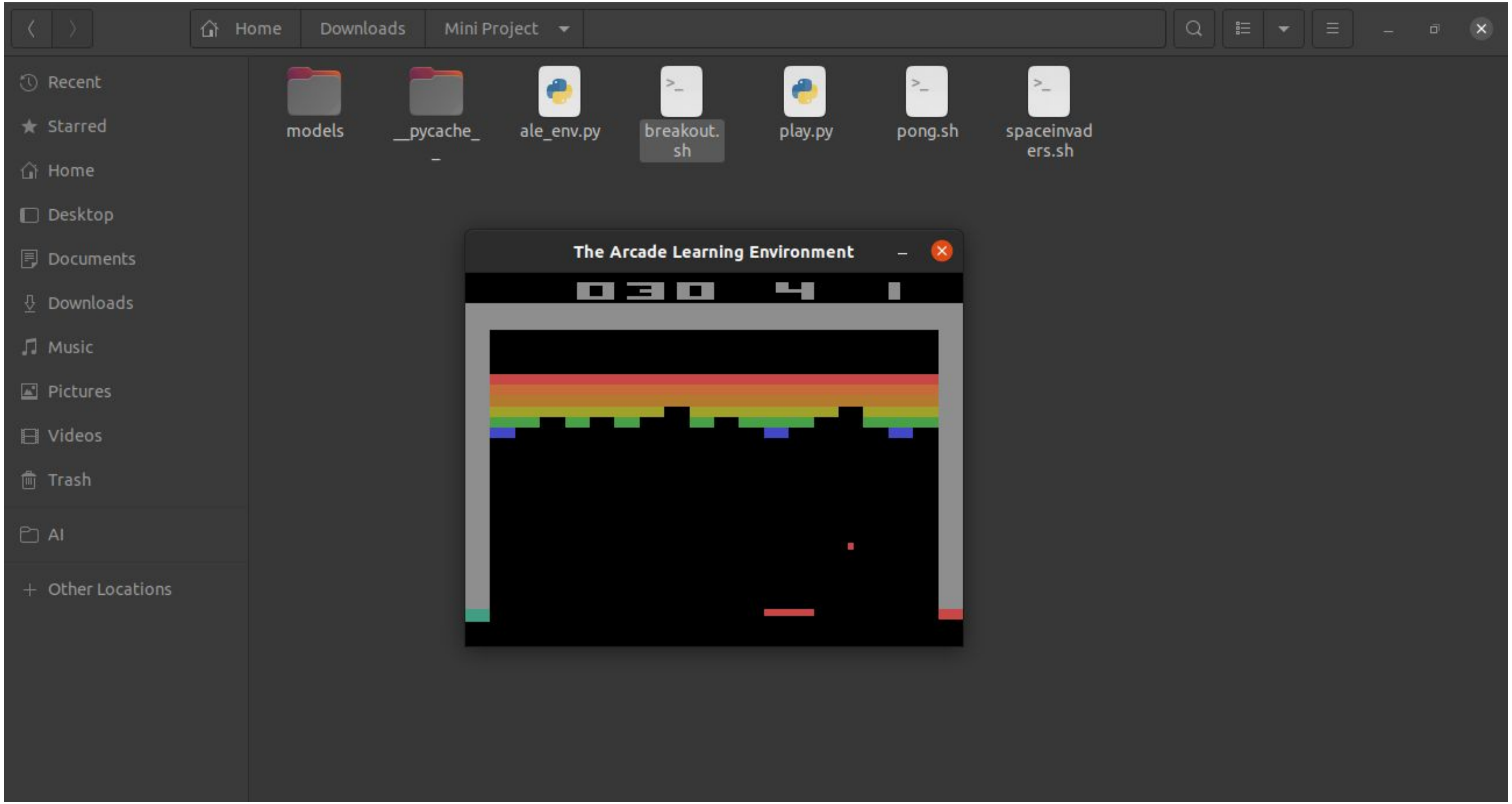


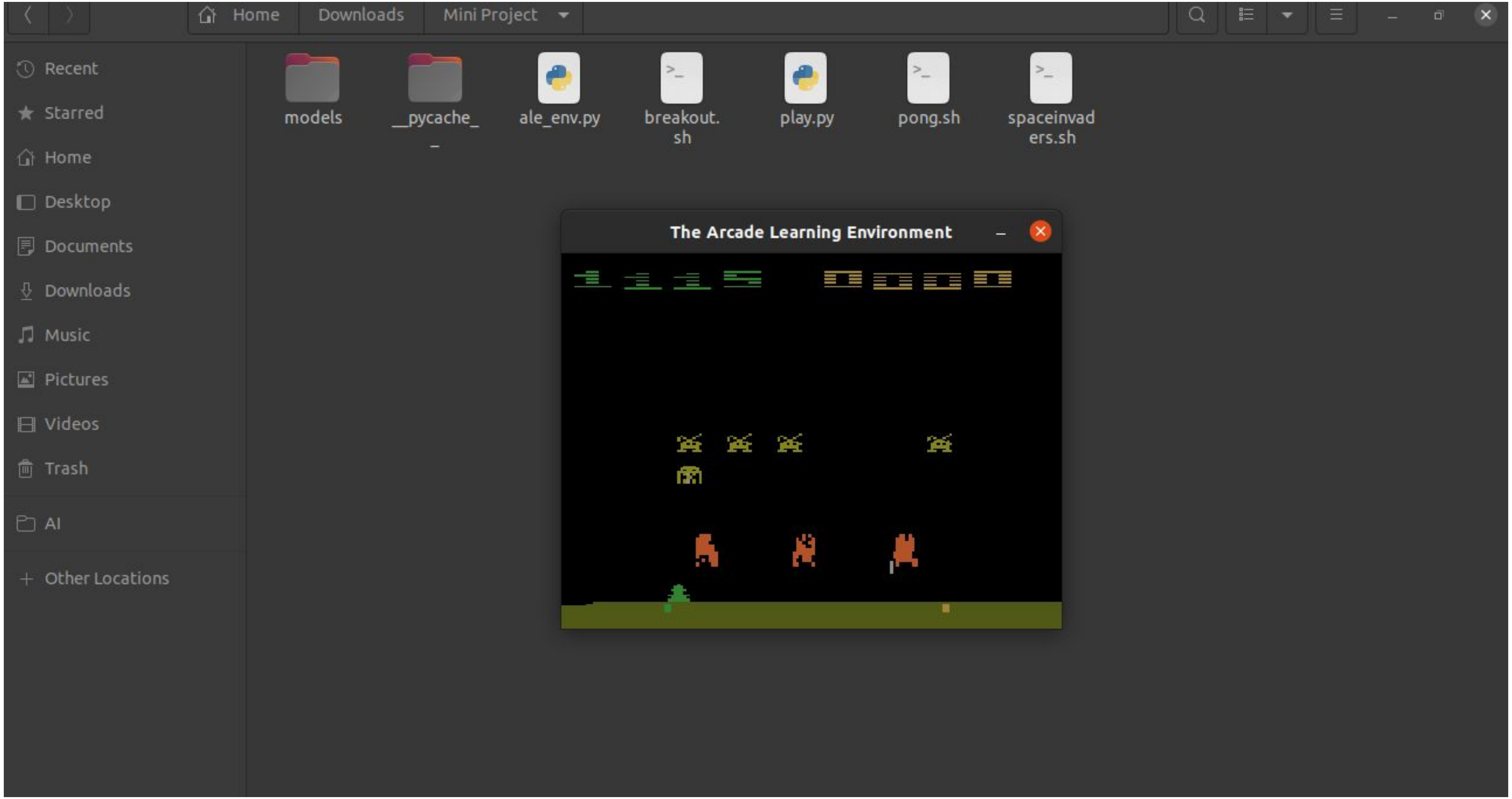
balaji@balaji-pc: ~/Downloads

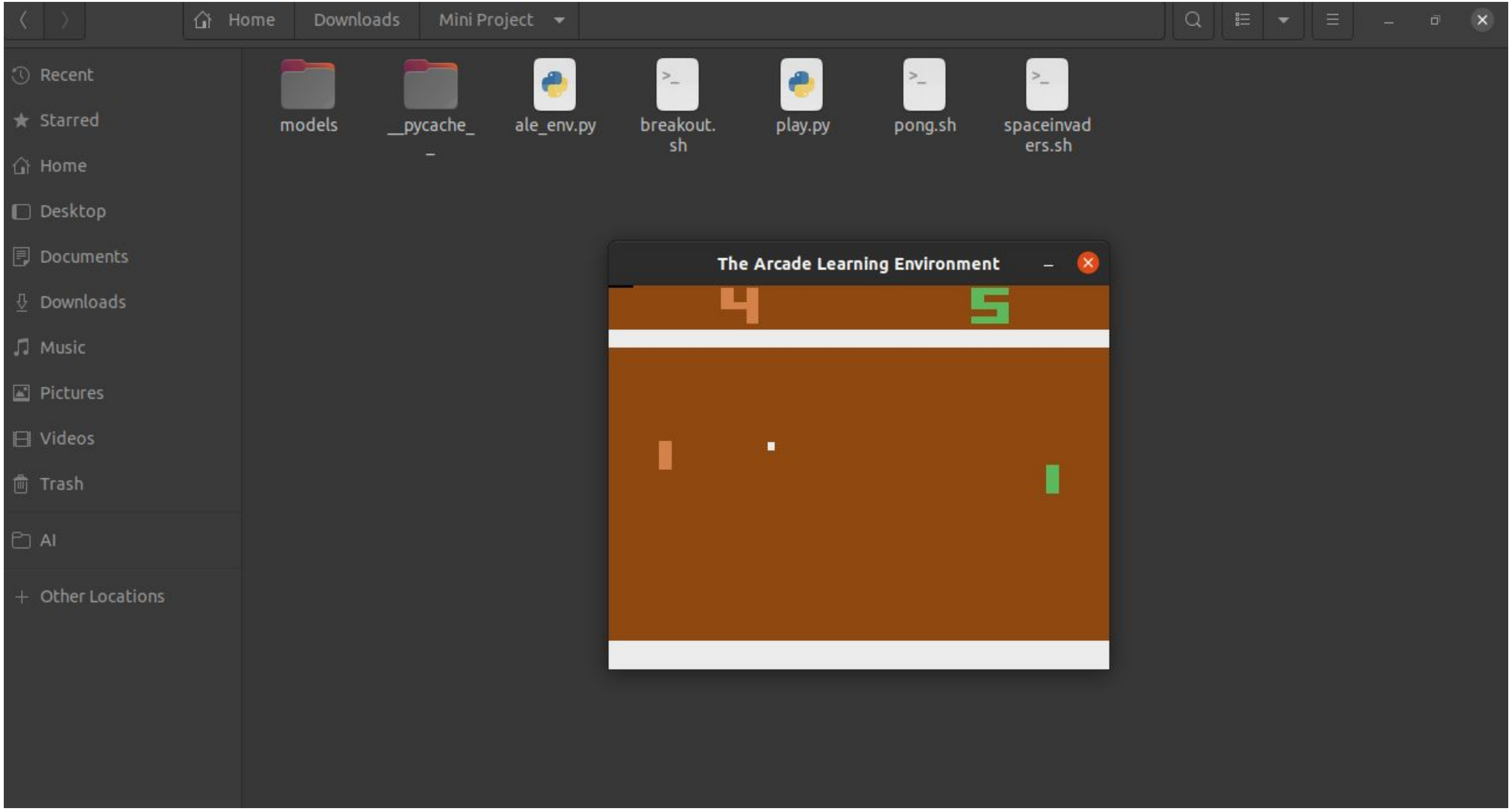
```
Episode 224, Loss: 475.19894, Total_reward: 58.0, Solved: False
Episode 225, Loss: 433.42072, Total_reward: 57.0, Solved: False
Episode 226, Loss: 808.86121, Total_reward: 61.0, Solved: False
Episode 227, Loss: 393.08423, Total_reward: 53.0, Solved: False
Episode 228, Loss: 902.64355, Total_reward: 69.0, Solved: False
Episode 229, Loss: 538.74207, Total_reward: 57.0, Solved: False
Episode 230, Loss: 546.34021, Total_reward: 62.0, Solved: False
Episode 231, Loss: 520.94275, Total_reward: 53.0, Solved: False
Episode 232, Loss: 1208.93298, Total_reward: 88.0, Solved: False
Episode 233, Loss: 413.22754, Total_reward: 50.0, Solved: False
Episode 234, Loss: 100.12862, Total_reward: 24.0, Solved: False
Episode 235, Loss: 403.75287, Total_reward: 45.0, Solved: False
Episode 236, Loss: 1092.58765, Total_reward: 80.0, Solved: False
Episode 237, Loss: 497.25482, Total_reward: 49.0, Solved: False
Episode 238, Loss: 492.30002, Total_reward: 49.0, Solved: False
Episode 239, Loss: 57.37786, Total_reward: 15.0, Solved: False
Episode 240, Loss: 503.06451, Total_reward: 51.0, Solved: False
Episode 241, Loss: 42.22906, Total_reward: 12.0, Solved: False
Episode 242, Loss: 1035.26660, Total_reward: 77.0, Solved: False
Episode 243, Loss: 351.20273, Total_reward: 45.0, Solved: False
Episode 244, Loss: 439.07465, Total_reward: 45.0, Solved: False
Episode 245, Loss: 295.87741, Total_reward: 40.0, Solved: False
Episode 246, Loss: 687.17200, Total_reward: 56.0, Solved: False
Episode 247, Loss: 1136.34241, Total_reward: 73.0, Solved: False
Episode 248, Loss: 429.46368, Total_reward: 45.0, Solved: False
Episode 249, Loss: 417.31152, Total_reward: 44.0, Solved: False
Episode 250, Loss: 555.87665, Total_reward: 55.0, Solved: False
Episode 251, Loss: 2257.04272, Total_reward: 114.0, Solved: False
Episode 252, Loss: 1303.30481, Total_reward: 81.0, Solved: False
Episode 253, Loss: 1697.67712, Total_reward: 97.0, Solved: False
Episode 254, Loss: 2372.98291, Total_reward: 120.0, Solved: False
Episode 255, Loss: 3087.16821, Total_reward: 137.0, Solved: False
Episode 256, Loss: 1349.73828, Total_reward: 89.0, Solved: False
Episode 257, Loss: 2768.69434, Total_reward: 141.0, Solved: False
Episode 258, Loss: 4067.86987, Total_reward: 170.0, Solved: False
Episode 259, Loss: 5216.43945, Total_reward: 196.0, Solved: True
```



Final Results


Home
Downloads
Mini Project
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
AI
Other Locations
models
__pycache__
ale_env.py
breakout.sh
play.py
pong.sh
spaceinvaders.sh
The Arcade Learning Environment


Home
Downloads
Mini Project
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
AI
Other Locations
models
__pycache__
ale_env.py
breakout.sh
play.py
pong.sh
spaceinvaders.sh
The Arcade Learning Environment



Home
Downloads
Mini Project
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
AI
Other Locations
models
__pycache__
ale_env.py
breakout.sh
play.py
pong.sh
spaceinvaders.sh
The Arcade Learning Environment
4
5

Future work

- ➢ How a single agent can learn to play all these games - https://www.deepmind.com/publications/a-generalist-agent
- ➢ How to make these algorithms sample-efficient ?

Any Questions?

Thank you !