16_Reinforcement_Learning

May 10, 2018

1 TensorFlow Tutorial #16

2 Reinforcement Learning (Q-Learning)

by Magnus Erik Hvass Pedersen / GitHub / Videos on YouTube

2.1 Introduction

This tutorial is about so-called Reinforcement Learning in which an agent is learning how to navigate some environment, in this case Atari games from the 1970-80's. The agent does not know anything about the game and must learn how to play it from trial and error. The only information that is available to the agent is the screen output of the game, and whether the previous action resulted in a reward or penalty.

This is a very difficult problem in Machine Learning / Artificial Intelligence, because the agent must both learn to distinguish features in the game-images, and then connect the occurence of certain features in the game-images with its own actions and a reward or penalty that may be deferred many steps into the future.

This problem was first solved by the researchers from Google DeepMind. This tutorial is based on the main ideas from their early research papers (especially this and this), although we make several changes because the original DeepMind algorithm was awkward and over-complicated in some ways. But it turns out that you still need several tricks in order to stabilize the training of the agent, so the implementation in this tutorial is unfortunately also somewhat complicated.

The basic idea is to have the agent estimate so-called Q-values whenever it sees an image from the game-environment. The Q-values tell the agent which action is most likely to lead to the highest cumulative reward in the future. The problem is then reduced to finding these Q-values and storing them for later retrieval using a function approximator.

This builds on some of the previous tutorials. You should be familiar with TensorFlow and Convolutional Neural Networks from Tutorial #01 and #02. It will also be helpful if you are familiar with one of the builder APIs in Tutorials #03 or #03-B.

2.2 The Problem

This tutorial uses the Atari game Breakout, where the player or agent is supposed to hit a ball with a paddle, thus avoiding death while scoring points when the ball smashes pieces of a wall.

When a human learns to play a game like this, the first thing to figure out is what part of the game environment you are controlling - in this case the paddle at the bottom. If you move right on the joystick then the paddle moves right and vice versa. The next thing is to figure out what the

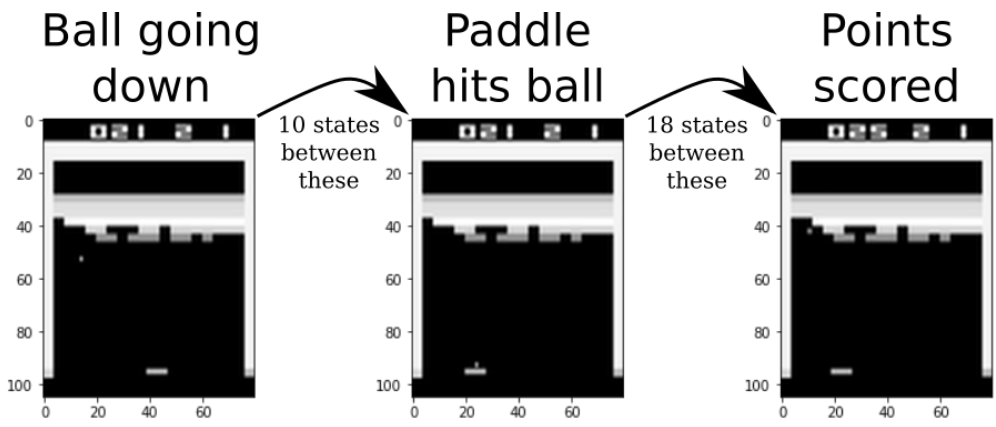


Illustration of the problem

goal of the game is - in this case to smash as many bricks in the wall as possible so as to maximize the score. Finally you need to learn what to avoid - in this case you must avoid dying by letting the ball pass beside the paddle.

Below are shown 3 images from the game that demonstrate what we need our agent to learn. In the image to the left, the ball is going downwards and the agent must learn to move the paddle so as to hit the ball and avoid death. The image in the middle shows the paddle hitting the ball, which eventually leads to the image on the right where the ball smashes some bricks and scores points. The ball then continues downwards and the process repeats.

The problem is that there are 10 states between the ball going downwards and the paddle hitting the ball, and there are an additional 18 states before the reward is obtained when the ball hits the wall and smashes some bricks. How can we teach an agent to connect these three situations and generalize to similar situations? The answer is to use so-called Reinforcement Learning with a Neural Network, as shown in this tutorial.

2.3 Q-Learning

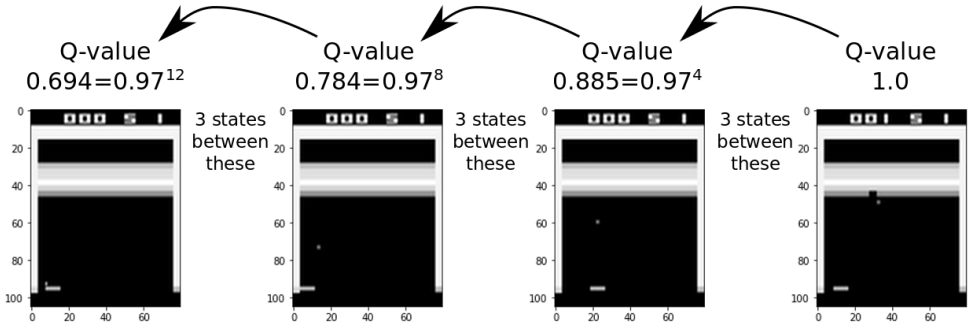
One of the simplest ways of doing Reinforcement Learning is called Q-learning. Here we want to estimate so-called Q-values which are also called action-values, because they map a state of the game-environment to a numerical value for each possible action that the agent may take. The Q-values indicate which action is expected to result in the highest future reward, thus telling the agent which action to take.

Unfortunately we do not know what the Q-values are supposed to be, so we have to estimate them somehow. The Q-values are all initialized to zero and then updated repeatedly as new information is collected from the agent playing the game. When the agent scores a point then the Q-value must be updated with the new information.

There are different formulas for updating Q-values, but the simplest is to set the new Q-value to the reward that was observed, plus the maximum Q-value for the following state of the game. This gives the total reward that the agent can expect from the current game-state and onwards. Typically we also multiply the max Q-value for the following state by a so-called discount-factor slightly below 1. This causes more distant rewards to contribute less to the Q-value, thus making the agent favour rewards that are closer in time.

The formula for updating the Q-value is:

```
Q-value for state and action = reward + discount * max Q-value for next state
```



Q-values Simple Example

In academic papers, this is typically written with mathematical symbols like this:

$$Q(s_t, a_t) \leftarrow \underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of future rewards}}$$

Furthermore, when the agent loses a life, then we know that the future reward is zero because the agent is dead, so we set the Q-value for that state to zero.

2.3.1 Simple Example

The images below demonstrate how Q-values are updated in a backwards sweep through the game-states that have previously been visited. In this simple example we assume all Q-values have been initialized to zero. The agent gets a reward of 1 point in the right-most image. This reward is then propagated backwards to the previous game-states, so when we see similar game-states in the future, we know that the given actions resulted in that reward.

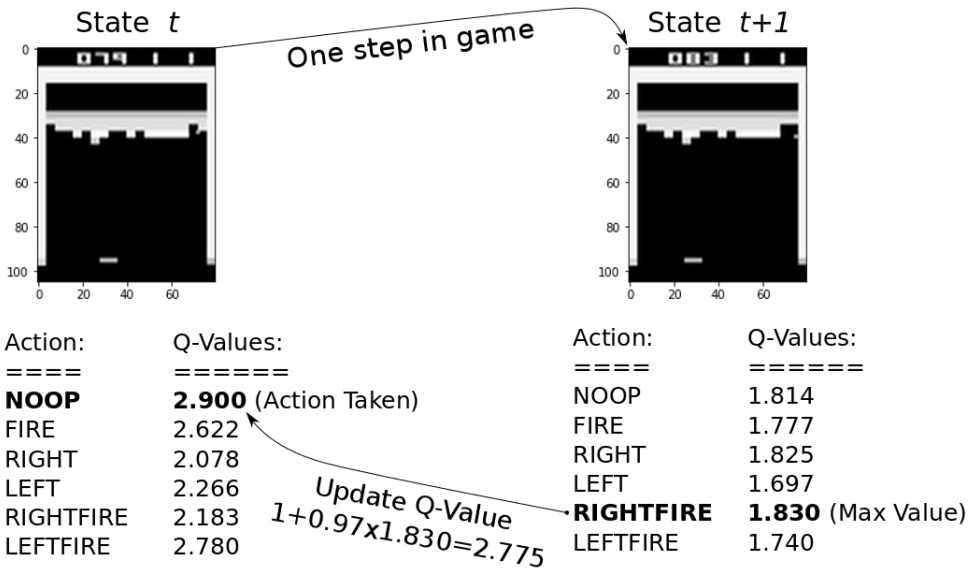
The discounting is an exponentially decreasing function. This example uses a discount-factor of 0.97 so the Q-value for the 3rd image is about $0.885 \simeq 0.97^4$ because it is 4 states prior to the state that actually received the reward. Similarly for the other states. This example only shows one Q-value per state, but in reality there is one Q-value for each possible action in the state, and the Q-values are updated in a backwards-sweep using the formula above. This is shown in the next section.

2.3.2 Detailed Example

This is a more detailed example showing the Q-values for two successive states of the game-environment and how to update them.

The Q-values for the possible actions have been estimated by a Neural Network. For the action NOOP in state *t* the Q-value is estimated to be 2.900, which is the highest Q-value for that state so the agent takes that action, i.e. the agent does not do anything between state *t* and *t+1* because NOOP means "No Operation".

In state *t+1* the agent scores 4 points, but this is limited to 1 point in this implementation so as to stabilize the training. The maximum Q-value for state *t+1* is 1.830 for the action RIGHTFIRE. So if we select that action and continue to select the actions proposed by the Q-values estimated by the Neural Network, then the discounted sum of all the future rewards is expected to be 1.830.



Q-values Detailed Example

Now that we know the reward of taking the NOOP action from state *t* to *t+1*, we can update the Q-value to incorporate this new information. This uses the formula above:

$$Q(state_t, NOOP) \leftarrow \underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount}} \cdot \underbrace{\max_a Q(state_{t+1}, a)}_{\text{estimate of future rewards}} = 1.0 + 0.97 \cdot 1.830 \simeq 2.775$$

The new Q-value is 2.775 which is slightly lower than the previous estimate of 2.900. This Neural Network has already been trained for 150 hours so it is quite good at estimating Q-values, but earlier during the training, the estimated Q-values would be more different.

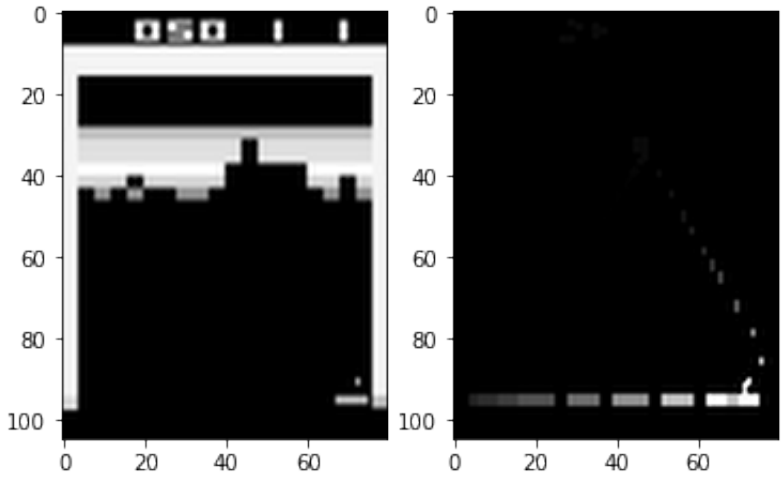
The idea is to have the agent play many, many games and repeatedly update the estimates of the Q-values as more information about rewards and penalties becomes available. This will eventually lead to good estimates of the Q-values, provided the training is numerically stable, as discussed further below. By doing this, we create a connection between rewards and prior actions.

2.4 Motion Trace

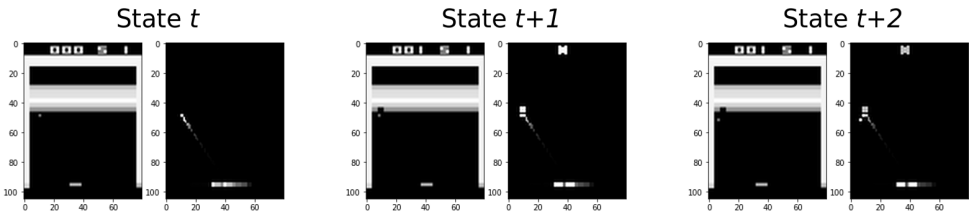
If we only use a single image from the game-environment then we cannot tell which direction the ball is moving. The typical solution is to use multiple consecutive images to represent the state of the game-environment.

This implementation uses another approach by processing the images from the game-environment in a motion-tracer that outputs two images as shown below. The left image is from the game-environment and the right image is the processed image, which shows traces of recent movements in the game-environment. In this case we can see that the ball is going downwards and has bounced off the right wall, and that the paddle has moved from the left to the right side of the screen.

Note that the motion-tracer has only been tested for Breakout and partially tested for Space Invaders, so it may not work for games with more complicated graphics such as Doom.



Motion Trace



Training Stability

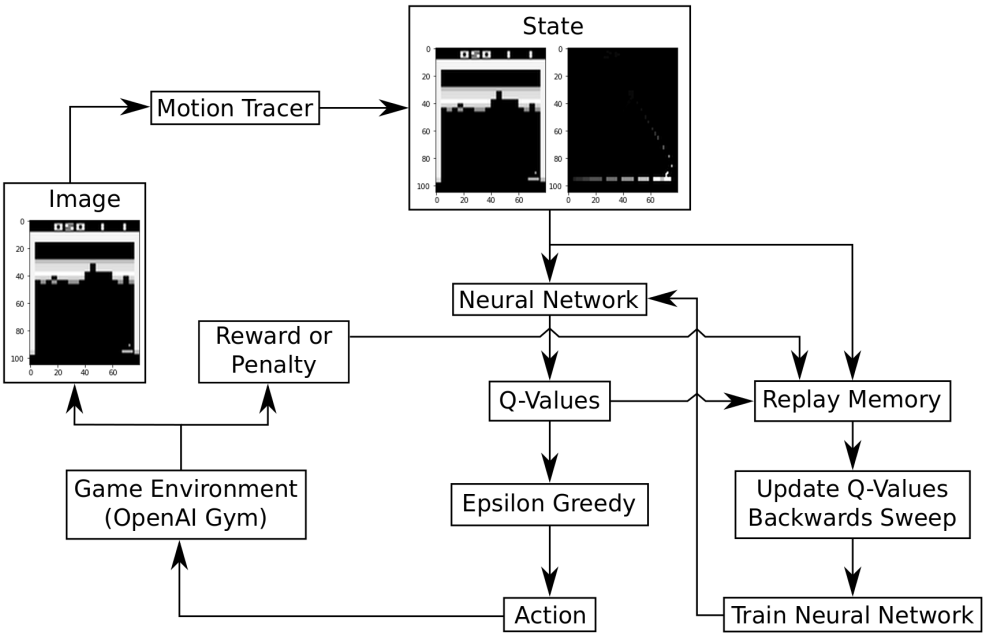
2.5 Training Stability

We need a function approximator that can take a state of the game-environment as input and produce as output an estimate of the Q-values for that state. We will use a Convolutional Neural Network for this. Although they have achieved great fame in recent years, they are actually a quite old technologies with many problems - one of which is training stability. A significant part of the research for this tutorial was spent on tuning and stabilizing the training of the Neural Network.

To understand why training stability is a problem, consider the 3 images below which show the game-environment in 3 consecutive states. At state t the agent is about to score a point, which happens in the following state $t + 1$. Assuming all Q-values were zero prior to this, we should now set the Q-value for state $t + 1$ to be 1.0 and it should be 0.97 for state t if the discount-value is 0.97, according to the formula above for updating Q-values.

If we were to train a Neural Network to estimate the Q-values for the two states t and $t + 1$ with Q-values 0.97 and 1.0, respectively, then the Neural Network will most likely be unable to distinguish properly between the images of these two states. As a result the Neural Network will also estimate a Q-value near 1.0 for state $t + 2$ because the images are so similar. But this is clearly wrong because the Q-values for state $t + 2$ should be zero as we do not know anything about future rewards at this point, and that is what the Q-values are supposed to estimate.

If this is continued and the Neural Network is trained after every new game-state is observed, then it will quickly cause the estimated Q-values to explode. This is an artifact of training Neural Networks which must have sufficiently large and diverse training-sets. For this reason we will use a so-called Replay Memory so we can gather a large number of game-states and shuffle them



Flowchart

during training of the Neural Network.

2.6 Flowchart

This flowchart shows roughly how Reinforcement Learning is implemented in this tutorial. There are two main loops which are run sequentially until the Neural Network is sufficiently accurate at estimating Q-values.

The first loop is for playing the game and recording data. This uses the Neural Network to estimate Q-values from a game-state. It then stores the game-state along with the corresponding Q-values and reward/penalty in the Replay Memory for later use.

The other loop is activated when the Replay Memory is sufficiently full. First it makes a full backwards sweep through the Replay Memory to update the Q-values with the new rewards and penalties that have been observed. Then it performs an optimization run so as to train the Neural Network to better estimate these updated Q-values.

There are many more details in the implementation, such as decreasing the learning-rate and increasing the fraction of the Replay Memory being used during training, but this flowchart shows the main ideas.

2.7 Neural Network Architecture

The Neural Network used in this implementation has 3 convolutional layers, all of which have filter-size 3x3. The layers have 16, 32, and 64 output channels, respectively. The stride is 2 in the first two convolutional layers and 1 in the last layer.

Following the 3 convolutional layers there are 4 fully-connected layers each with 1024 units and ReLU-activation. Then there is a single fully-connected layer with linear activation used as the output of the Neural Network.

This architecture is different from those typically used in research papers from DeepMind and others. They often have large convolutional filter-sizes of 8x8 and 4x4 with high stride-values. This causes more aggressive down-sampling of the game-state images. They also typically have only a single fully-connected layer with 256 or 512 ReLU units.

During the research for this tutorial, it was found that smaller filter-sizes and strides in the convolutional layers, combined with several fully-connected layers having more units, were necessary in order to have sufficiently accurate Q-values. The Neural Network architectures originally used by DeepMind appear to distort the Q-values quite significantly. A reason that their approach still worked, is possibly due to their use of a very large Replay Memory with 1 million states, and that the Neural Network did one mini-batch of training for each step of the game-environment, and some other tricks.

The architecture used here is probably excessive but it takes several days of training to test each architecture, so it is left as an exercise for the reader to try and find a smaller Neural Network architecture that still performs well.

2.8 Installation

The documentation for OpenAI Gym currently suggests that you need to build it in order to install it. But if you just want to install the Atari games, then you only need to install a single pip-package by typing the following commands in a terminal.

- conda create --name tf-gym --clone tf
- source activate tf-gym
- pip install gym[atari]

This assumes you already have an Anaconda environment named `tf` which has TensorFlow installed, it will then be cloned to another environment named `tf-gym` where OpenAI Gym is also installed. This allows you to easily switch between your normal TensorFlow environment and another one which also contains OpenAI Gym.

You can also have two environments named `tf-gpu` and `tf-gpu-gym` for the GPU versions of TensorFlow.

2.9 Imports

```
In [1]: %matplotlib inline
        import matplotlib.pyplot as plt
        import tensorflow as tf
        import gym
        import numpy as np
        import math
```

The main source-code for Reinforcement Learning is located in the following module:

```
In [2]: import reinforcement_learning as rl
```

This was developed using Python 3.6.0 (Anaconda) with package versions:

```
In [3]: # TensorFlow
        tf.__version__
```

```
Out[3]: '1.1.0'
```

```
In [4]: # OpenAI Gym
        gym.__version__
```

```
Out[4]: '0.8.1'
```

2.10 Game Environment

This is the name of the game-environment that we want to use in OpenAI Gym.

```
In [5]: env_name = 'Breakout-v0'
        # env_name = 'SpaceInvaders-v0'
```

This is the base-directory for the TensorFlow checkpoints as well as various log-files.

```
In [6]: rl.checkpoint_base_dir = 'checkpoints_tutorial16/'
```

Once the base-dir has been set, you need to call this function to set all the paths that will be used. This will also create the checkpoint-dir if it does not already exist.

```
In [7]: rl.update_paths(env_name=env_name)
```

2.11 Download Pre-Trained Model

You can download a TensorFlow checkpoint which holds all the pre-trained variables for the Neural Network. Two checkpoints are provided, one for Breakout and one for Space Invaders. They were both trained for about 150 hours on a laptop with 2.6 GHz CPU and a GTX 1070 GPU.

COMPATIBILITY ISSUES These TensorFlow checkpoints were developed with OpenAI gym v. 0.8.1 and atari-py v. 0.0.19 which had unused / redundant actions as noted above. There appears to have been a change in the gym API since then, as the unused actions are no longer present. This means the vectors with actions and Q-values now only contain 4 elements instead of the 6 shown here. This also means that the TensorFlow checkpoints cannot be used with newer versions of gym and atari-py, so in order to use these pre-trained checkpoints you need to install the older versions of gym and atari-py - or you can just train a new model yourself so you get a new TensorFlow checkpoint.

WARNING! These checkpoints are 280-360 MB each. They are currently hosted on the webserver I use for www.hvass-labs.org because it is awkward to automatically download large files on Google Drive. To lower the traffic on my webserver, this line has been commented out, so you have to activate it manually. You are welcome to download it, I just don't want it to download automatically for everyone who only wants to run this Notebook briefly.

```
In [8]: # rl.maybe_download_checkpoint(env_name=env_name)
```

I believe the webserver is located in Denmark. If you are having problems downloading the files using the automatic function above, then you can try and download the files manually in a webbrowser or using `wget` or `curl`. Or you can download from Google Drive, where you will get an anti-virus warning that is awkward to bypass automatically:

- Download Breakout Checkpoint from Google Drive
- Download Space Invaders Checkpoint from Google Drive

You can use the checksum to ensure the downloaded files are complete:

- SHA256 Checksum

2.12 Create Agent

The Agent-class implements the main loop for playing the game, recording data and optimizing the Neural Network. We create an object-instance and need to set `training=True` because we want to use the replay-memory to record states and Q-values for plotting further below. We disable logging so this does not corrupt the logs from the actual training that was done previously. We can also set `render=True` but it will have no effect as long as `training==True`.

```
In [9]: agent = rl.Agent(env_name=env_name,
                         training=True,
                         render=True,
                         use_logging=False)
```

```
[2017-05-15 15:48:47,348] Making new env: Breakout-v0
```

```
Trying to restore last checkpoint ...
INFO:tensorflow:Restoring parameters from checkpoints_tutorial16/Breakout-v0/checkpoint-127639
```

```
[2017-05-15 15:48:47,868] Restoring parameters from checkpoints_tutorial16/Breakout-v0/checkpo
```

```
Restored checkpoint from: checkpoints_tutorial16/Breakout-v0/checkpoint-127639066
```

The Neural Network is automatically instantiated by the Agent-class. We will create a direct reference for convenience.

```
In [10]: model = agent.model
```

Similarly, the Agent-class also allocates the replay-memory when `training==True`. The replay-memory will require more than 3 GB of RAM, so it should only be allocated when needed. We will need the replay-memory in this Notebook to record the states and Q-values we observe, so they can be plotted further below.

```
In [11]: replay_memory = agent.replay_memory
```

2.13 Training

The agent's `run()` function is used to play the game. This uses the Neural Network to estimate Q-values and hence determine the agent's actions. If `training==True` then it will also gather states and Q-values in the replay-memory and train the Neural Network when the replay-memory is sufficiently full. You can set `num_episodes=None` if you want an infinite loop that you would stop manually with `ctrl-c`. In this case we just set `num_episodes=1` because we are not actually interested in training the Neural Network any further, we merely want to collect some states and Q-values in the replay-memory so we can plot them below.

```
In [12]: agent.run(num_episodes=1)
```

```
87584:127639721         Epsilon: 0.10         Reward: 12.0         Episode Mean: 12.0
```

In training-mode, this function will output a line for each episode. The first counter is for the number of episodes that have been processed. The second counter is for the number of states that have been processed. These two counters are stored in the TensorFlow checkpoint along with the weights of the Neural Network, so you can restart the training e.g. if you only have one computer and need to train during the night.

Note that the number of episodes is almost 90k. It is impractical to print that many lines in this Notebook, so the training is better done in a terminal window by running the following commands:

```
source activate tf-gpu-gym  # Activate your Python environment with TF and Gym.
python reinforcement-learning.py --env Breakout-v0 --training
```

2.14 Training Progress

Data is being logged during training so we can plot the progress afterwards. The reward for each episode and a running mean of the last 30 episodes are logged to file. Basic statistics for the Q-values in the replay-memory are also logged to file before each optimization run.

This could be logged using TensorFlow and TensorBoard, but they were designed for logging variables of the TensorFlow graph and data that flows through the graph. In this case the data we want logged does not reside in the graph, so it becomes a bit awkward to use TensorFlow to log this data.

We have therefore implemented a few small classes that can write and read these logs.

```
In [13]: log_q_values = rl.LogQValues()
         log_reward = rl.LogReward()
```

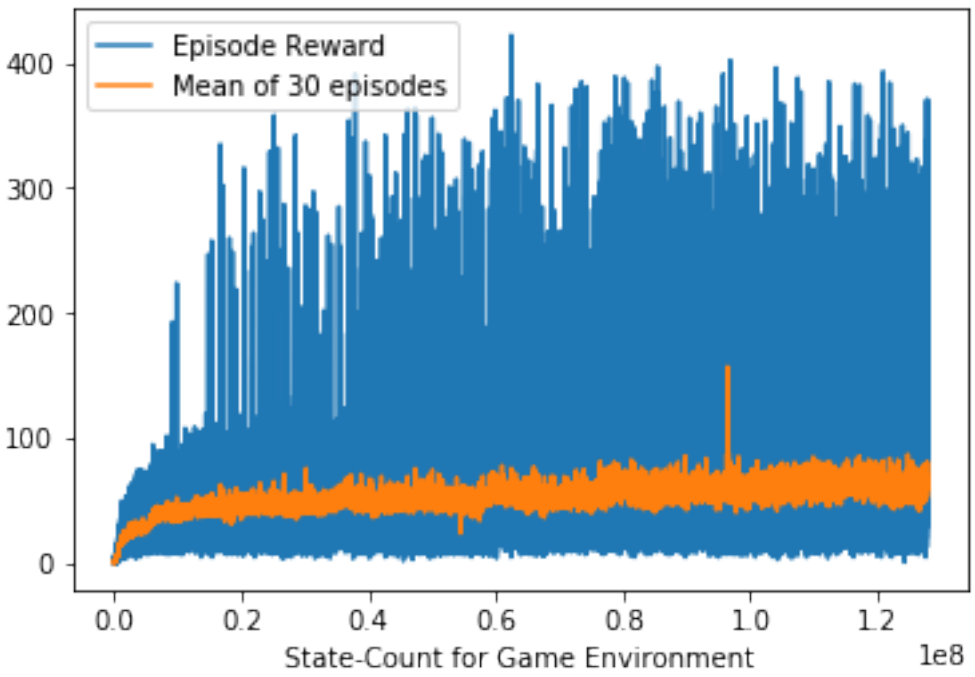
We can now read the logs from file:

```
In [14]: log_q_values.read()
         log_reward.read()
```

2.14.1 Training Progress: Reward

This plot shows the reward for each episode during training, as well as the running mean of the last 30 episodes. Note how the reward varies greatly from one episode to the next, so it is difficult to say from this plot alone whether the agent is really improving during the training, although the running mean does appear to trend upwards slightly.

```
In [15]: plt.plot(log_reward.count_states, log_reward.episode, label='Episode Reward')
         plt.plot(log_reward.count_states, log_reward.mean, label='Mean of 30 episodes')
         plt.xlabel('State-Count for Game Environment')
         plt.legend()
         plt.show()
```



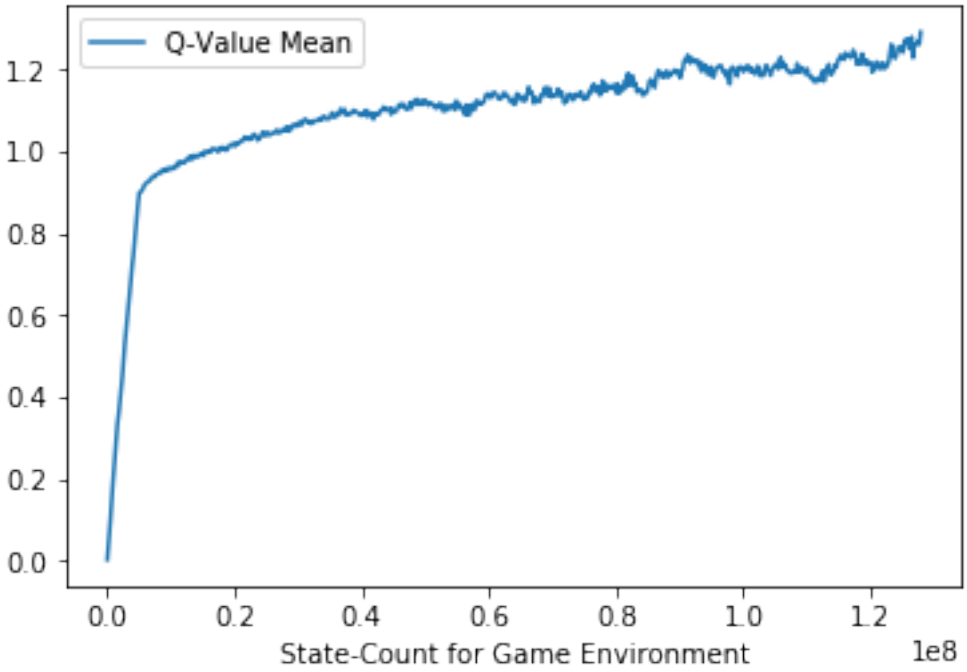
2.14.2 Training Progress: Q-Values

The following plot shows the mean Q-values from the replay-memory prior to each run of the optimizer for the Neural Network. Note how the mean Q-values increase rapidly in the beginning and then they increase fairly steadily for 40 million states, after which they still trend upwards but somewhat more irregularly.

The fast improvement in the beginning is probably due to (1) the use of a smaller replay-memory early in training so the Neural Network is optimized more often and the new information is used faster, (2) the backwards-sweeping of the replay-memory so the rewards are used to update the Q-values for many of the states, instead of just updating the Q-values for a single state, and (3) the replay-memory is balanced so at least half of each mini-batch contains states whose Q-values have high estimation-errors for the Neural Network.

The original paper from DeepMind showed much slower progress in the first phase of training, see Figure 2 in that paper but note that the Q-values are not directly comparable, possibly because they used a higher discount factor of 0.99 while we only used 0.97 here.

```
In [16]: plt.plot(log_q_values.count_states, log_q_values.mean, label='Q-Value Mean')
         plt.xlabel('State-Count for Game Environment')
         plt.legend()
         plt.show()
```



2.15 Testing

When the agent and Neural Network is being trained, the so-called epsilon-probability is typically decreased from 1.0 to 0.1 over a large number of steps, after which the probability is held fixed at 0.1. This means the probability is 0.1 or 10% that the agent will select a random action in each step, otherwise it will select the action that has the highest Q-value. This is known as the epsilon-greedy policy. The choice of 0.1 for the epsilon-probability is a compromise between taking the actions that are already known to be good, versus exploring new actions that might lead to even higher rewards or might lead to death of the agent.

During testing it is common to lower the epsilon-probability even further. We have set it to 0.01 as shown here:

```
In [17]: agent.epsilon_greedy.epsilon_testing
```

```
Out[17]: 0.01
```

We will now instruct the agent that it should no longer perform training by setting this boolean:

```
In [18]: agent.training = False
```

We also reset the previous episode rewards.

```
In [19]: agent.reset_episode_rewards()
```

We can render the game-environment to screen so we can see the agent playing the game, by setting this boolean:

```
In [20]: agent.render = True
```

We can now run a single episode by calling the `run()` function again. This should open a new window that shows the game being played by the agent. At the time of this writing, it was not possible to resize this tiny window, and the developers at OpenAI did not seem to care about this feature which should obviously be there.

```
In [21]: agent.run(num_episodes=1)
```

```
87586:127639767        Q-min: 1.765        Q-max: 1.783        Lives: 5        Reward: 1.0
87586:127639820        Q-min: 1.608        Q-max: 1.619        Lives: 5        Reward: 2.0
87586:127639882        Q-min: 1.712        Q-max: 1.734        Lives: 5        Reward: 3.0
87586:127639931        Q-min: 1.968        Q-max: 1.998        Lives: 5        Reward: 4.0
87586:127639963        Q-min: 1.953        Q-max: 1.988        Lives: 5        Reward: 5.0
87586:127639985        Q-min: 0.013        Q-max: 0.184        Lives: 4        Reward: 5.0
87586:127640039        Q-min: 1.651        Q-max: 1.664        Lives: 4        Reward: 6.0
87586:127640090        Q-min: 1.902        Q-max: 1.919        Lives: 4        Reward: 7.0
87586:127640130        Q-min: 1.960        Q-max: 1.968        Lives: 4        Reward: 8.0
87586:127640166        Q-min: 1.915        Q-max: 1.929        Lives: 4        Reward: 9.0
87586:127640197        Q-min: 2.002        Q-max: 2.022        Lives: 4        Reward: 10.0
87586:127640228        Q-min: 1.952        Q-max: 1.982        Lives: 4        Reward: 11.0
87586:127640260        Q-min: 2.031        Q-max: 2.050        Lives: 4        Reward: 12.0
87586:127640306        Q-min: 1.682        Q-max: 1.737        Lives: 4        Reward: 13.0
87586:127640371        Q-min: 1.700        Q-max: 1.726        Lives: 4        Reward: 14.0
87586:127640439        Q-min: 1.555        Q-max: 1.665        Lives: 4        Reward: 15.0
87586:127640510        Q-min: 1.619        Q-max: 1.699        Lives: 4        Reward: 16.0
87586:127640552        Q-min: -0.068        Q-max: 0.219        Lives: 3        Reward: 16.0
87586:127640595        Q-min: 1.868        Q-max: 1.893        Lives: 3        Reward: 17.0
87586:127640639        Q-min: 1.975        Q-max: 1.996        Lives: 3        Reward: 18.0
87586:127640681        Q-min: 1.918        Q-max: 1.947        Lives: 3        Reward: 19.0
87586:127640718        Q-min: 2.025        Q-max: 2.090        Lives: 3        Reward: 20.0
87586:127640751        Q-min: 1.981        Q-max: 2.006        Lives: 3        Reward: 21.0
87586:127640785        Q-min: 2.041        Q-max: 2.072        Lives: 3        Reward: 25.0
87586:127640818        Q-min: 2.052        Q-max: 2.329        Lives: 3        Reward: 29.0
87586:127640840        Q-min: 2.298        Q-max: 2.444        Lives: 3        Reward: 30.0
87586:127640860        Q-min: 2.400        Q-max: 2.477        Lives: 3        Reward: 34.0
87586:127640882        Q-min: 2.344        Q-max: 2.398        Lives: 3        Reward: 35.0
87586:127640906        Q-min: 2.314        Q-max: 2.418        Lives: 3        Reward: 39.0
87586:127640927        Q-min: 2.211        Q-max: 2.266        Lives: 3        Reward: 40.0
87586:127640947        Q-min: 2.433        Q-max: 2.514        Lives: 3        Reward: 41.0
```

```
87586:127640968       Q-min: 2.259       Q-max: 2.518       Lives: 3       Reward: 45.0
87586:127640990       Q-min: 2.381       Q-max: 2.445       Lives: 3       Reward: 49.0
87586:127641011       Q-min: 2.299       Q-max: 2.477       Lives: 3       Reward: 53.0
87586:127641032       Q-min: 2.431       Q-max: 2.521       Lives: 3       Reward: 54.0
87586:127641053       Q-min: 2.292       Q-max: 2.394       Lives: 3       Reward: 55.0
87586:127641074       Q-min: 2.312       Q-max: 2.515       Lives: 3       Reward: 56.0
87586:127641094       Q-min: 2.310       Q-max: 2.421       Lives: 3       Reward: 60.0
87586:127641117       Q-min: 2.284       Q-max: 2.431       Lives: 3       Reward: 64.0
87586:127641137       Q-min: 2.328       Q-max: 2.442       Lives: 3       Reward: 65.0
87586:127641156       Q-min: 2.411       Q-max: 2.459       Lives: 3       Reward: 66.0
87586:127641178       Q-min: 1.457       Q-max: 2.612       Lives: 3       Reward: 73.0
87586:127641192       Q-min: -0.155       Q-max: 0.483       Lives: 2       Reward: 73.0
87586:127641236       Q-min: 2.176       Q-max: 2.289       Lives: 2       Reward: 74.0
87586:127641282       Q-min: 2.060       Q-max: 2.132       Lives: 2       Reward: 78.0
87586:127641340       Q-min: 1.806       Q-max: 1.967       Lives: 2       Reward: 79.0
87586:127641389       Q-min: 2.202       Q-max: 2.385       Lives: 2       Reward: 80.0
87586:127641418       Q-min: 2.359       Q-max: 2.446       Lives: 2       Reward: 81.0
87586:127641454       Q-min: 2.278       Q-max: 2.435       Lives: 2       Reward: 85.0
87586:127641487       Q-min: 2.157       Q-max: 2.391       Lives: 2       Reward: 86.0
87586:127641546       Q-min: 1.722       Q-max: 2.306       Lives: 2       Reward: 90.0
87586:127641570       Q-min: 2.165       Q-max: 2.662       Lives: 2       Reward: 94.0
87586:127641591       Q-min: 2.422       Q-max: 2.789       Lives: 2       Reward: 98.0
87586:127641605       Q-min: 0.044       Q-max: 0.432       Lives: 1       Reward: 98.0
87586:127641664       Q-min: 1.532       Q-max: 2.163       Lives: 1       Reward: 102.0
87586:127641723       Q-min: 2.338       Q-max: 2.518       Lives: 1       Reward: 106.0
87586:127641783       Q-min: 1.870       Q-max: 2.321       Lives: 1       Reward: 110.0
87586:127641830       Q-min: 2.606       Q-max: 2.781       Lives: 1       Reward: 114.0
87586:127641852       Q-min: -0.278       Q-max: 0.069       Lives: 0       Reward: 114.0
```

2.15.1 Mean Reward

The game-play is slightly random, both with regard to selecting actions using the epsilon-greedy policy, but also because the OpenAI Gym environment will repeat any action between 2-4 times, with the number chosen at random. So the reward of one episode is not an accurate estimate of the reward that can be expected in general from this agent.

We need to run 30 or even 50 episodes to get a more accurate estimate of the reward that can be expected.

We will first reset the previous episode rewards.

```
In [22]: agent.reset_episode_rewards()
```

We disable the screen-rendering so the game-environment runs much faster.

```
In [23]: agent.render = False
```

We can now run 30 episodes. This records the rewards for each episode. It might have been a good idea to disable the output so it does not print all these lines - you can do this as an exercise.

```
In [24]: agent.run(num_episodes=30)
```

```
87588:127641897       Q-min: 1.755       Q-max: 1.774       Lives: 5       Reward: 1.0
87588:127641950       Q-min: 1.634       Q-max: 1.650       Lives: 5       Reward: 2.0
87588:127642002       Q-min: 1.849       Q-max: 1.872       Lives: 5       Reward: 3.0
87588:127642037       Q-min: 1.930       Q-max: 1.966       Lives: 5       Reward: 4.0
87588:127642067       Q-min: 1.936       Q-max: 1.970       Lives: 5       Reward: 5.0
87588:127642101       Q-min: 1.950       Q-max: 1.963       Lives: 5       Reward: 9.0
87588:127642136       Q-min: 2.189       Q-max: 2.341       Lives: 5       Reward: 13.0
87588:127642159       Q-min: 1.926       Q-max: 2.292       Lives: 5       Reward: 14.0
87588:127642178       Q-min: 1.976       Q-max: 2.286       Lives: 5       Reward: 15.0
87588:127642199       Q-min: 2.169       Q-max: 2.290       Lives: 5       Reward: 16.0
87588:127642218       Q-min: 2.243       Q-max: 2.338       Lives: 5       Reward: 17.0
87588:127642240       Q-min: 2.127       Q-max: 2.307       Lives: 5       Reward: 24.0
87588:127642261       Q-min: 2.328       Q-max: 2.408       Lives: 5       Reward: 25.0
87588:127642280       Q-min: 2.272       Q-max: 2.454       Lives: 5       Reward: 26.0
87588:127642302       Q-min: 2.251       Q-max: 2.401       Lives: 5       Reward: 27.0
87588:127642323       Q-min: 2.339       Q-max: 2.423       Lives: 5       Reward: 31.0
87588:127642343       Q-min: 2.365       Q-max: 2.458       Lives: 5       Reward: 32.0
87588:127642364       Q-min: 2.278       Q-max: 2.398       Lives: 5       Reward: 33.0
87588:127642382       Q-min: 2.226       Q-max: 2.399       Lives: 5       Reward: 34.0
87588:127642396       Q-min: -0.085       Q-max: 0.443       Lives: 4       Reward: 34.0
87588:127642437       Q-min: 1.988       Q-max: 2.028       Lives: 4       Reward: 35.0
87588:127642478       Q-min: 1.929       Q-max: 2.025       Lives: 4       Reward: 36.0
87588:127642522       Q-min: 2.039       Q-max: 2.062       Lives: 4       Reward: 37.0
87588:127642559       Q-min: 2.125       Q-max: 2.207       Lives: 4       Reward: 38.0
87588:127642595       Q-min: 2.249       Q-max: 2.385       Lives: 4       Reward: 42.0
87588:127642632       Q-min: 2.044       Q-max: 2.165       Lives: 4       Reward: 43.0
87588:127642666       Q-min: 2.204       Q-max: 2.507       Lives: 4       Reward: 47.0
87588:127642685       Q-min: 2.409       Q-max: 2.493       Lives: 4       Reward: 48.0
87588:127642703       Q-min: 2.254       Q-max: 2.433       Lives: 4       Reward: 49.0
87588:127642716       Q-min: -0.187       Q-max: 0.041       Lives: 3       Reward: 49.0
87588:127642759       Q-min: 2.005       Q-max: 2.038       Lives: 3       Reward: 50.0
87588:127642805       Q-min: 2.048       Q-max: 2.137       Lives: 3       Reward: 54.0
87588:127642854       Q-min: 2.376       Q-max: 2.597       Lives: 3       Reward: 58.0
87588:127642875       Q-min: 2.335       Q-max: 2.511       Lives: 3       Reward: 59.0
87588:127642897       Q-min: 2.480       Q-max: 2.536       Lives: 3       Reward: 63.0
87588:127642910       Q-min: 0.048       Q-max: 0.153       Lives: 2       Reward: 63.0
87588:127642960       Q-min: 2.096       Q-max: 2.239       Lives: 2       Reward: 64.0
87588:127643019       Q-min: 1.741       Q-max: 1.920       Lives: 2       Reward: 68.0
87588:127643087       Q-min: 1.986       Q-max: 2.044       Lives: 2       Reward: 69.0
87588:127643135       Q-min: 2.272       Q-max: 2.352       Lives: 2       Reward: 70.0
87588:127643168       Q-min: 2.441       Q-max: 2.554       Lives: 2       Reward: 74.0
87588:127643202       Q-min: 2.076       Q-max: 2.292       Lives: 2       Reward: 78.0
87588:127643225       Q-min: -0.176       Q-max: 0.268       Lives: 1       Reward: 78.0
87588:127643281       Q-min: 1.966       Q-max: 2.221       Lives: 1       Reward: 82.0
87588:127643349       Q-min: 1.627       Q-max: 2.724       Lives: 1       Reward: 86.0
87588:127643370       Q-min: 2.374       Q-max: 2.479       Lives: 1       Reward: 93.0
```

```
87588:127643390       Q-min: 2.446       Q-max: 2.602       Lives: 1       Reward: 94.0
87588:127643412       Q-min: 1.203       Q-max: 1.788       Lives: 1       Reward: 98.0
87588:127643435       Q-min: 2.395       Q-max: 2.539       Lives: 1       Reward: 102.0
87588:127643448       Q-min: -0.182       Q-max: 0.072       Lives: 0       Reward: 102.0
87589:127643490       Q-min: 1.753       Q-max: 1.763       Lives: 5       Reward: 1.0
87589:127643540       Q-min: 1.656       Q-max: 1.658       Lives: 5       Reward: 2.0
87589:127643604       Q-min: 1.705       Q-max: 1.724       Lives: 5       Reward: 3.0
87589:127643649       Q-min: 1.970       Q-max: 1.978       Lives: 5       Reward: 4.0
87589:127643683       Q-min: 1.972       Q-max: 2.004       Lives: 5       Reward: 5.0
87589:127643716       Q-min: 1.970       Q-max: 1.998       Lives: 5       Reward: 6.0
87589:127643751       Q-min: 1.816       Q-max: 1.837       Lives: 5       Reward: 7.0
87589:127643802       Q-min: 1.631       Q-max: 1.676       Lives: 5       Reward: 8.0
87589:127643865       Q-min: 1.726       Q-max: 1.738       Lives: 5       Reward: 9.0
87589:127643930       Q-min: 1.688       Q-max: 1.705       Lives: 5       Reward: 10.0
87589:127643992       Q-min: 1.655       Q-max: 1.675       Lives: 5       Reward: 11.0
87589:127644039       Q-min: 1.976       Q-max: 1.991       Lives: 5       Reward: 12.0
87589:127644071       Q-min: 1.870       Q-max: 1.899       Lives: 5       Reward: 13.0
87589:127644104       Q-min: 1.978       Q-max: 2.014       Lives: 5       Reward: 14.0
87589:127644140       Q-min: 2.065       Q-max: 2.085       Lives: 5       Reward: 18.0
87589:127644176       Q-min: 2.025       Q-max: 2.114       Lives: 5       Reward: 19.0
87589:127644209       Q-min: 2.068       Q-max: 2.149       Lives: 5       Reward: 23.0
87589:127644246       Q-min: 2.119       Q-max: 2.161       Lives: 5       Reward: 24.0
87589:127644276       Q-min: 2.175       Q-max: 2.211       Lives: 5       Reward: 25.0
87589:127644310       Q-min: 2.073       Q-max: 2.097       Lives: 5       Reward: 26.0
87589:127644341       Q-min: 2.106       Q-max: 2.172       Lives: 5       Reward: 30.0
87589:127644377       Q-min: 2.312       Q-max: 2.571       Lives: 5       Reward: 34.0
87589:127644392       Q-min: 0.043       Q-max: 0.267       Lives: 4       Reward: 34.0
87589:127644446       Q-min: 1.782       Q-max: 1.803       Lives: 4       Reward: 35.0
87589:127644499       Q-min: 1.961       Q-max: 2.130       Lives: 4       Reward: 36.0
87589:127644540       Q-min: 2.092       Q-max: 2.265       Lives: 4       Reward: 37.0
87589:127644581       Q-min: 2.203       Q-max: 2.238       Lives: 4       Reward: 38.0
87589:127644611       Q-min: 2.169       Q-max: 2.270       Lives: 4       Reward: 39.0
87589:127644644       Q-min: 2.176       Q-max: 2.307       Lives: 4       Reward: 43.0
87589:127644677       Q-min: 2.138       Q-max: 2.240       Lives: 4       Reward: 44.0
87589:127644725       Q-min: 1.803       Q-max: 1.827       Lives: 4       Reward: 45.0
87589:127644792       Q-min: 1.820       Q-max: 1.887       Lives: 4       Reward: 46.0
87589:127644855       Q-min: 1.821       Q-max: 1.853       Lives: 4       Reward: 47.0
87589:127644924       Q-min: 1.803       Q-max: 1.938       Lives: 4       Reward: 48.0
87589:127644973       Q-min: 2.182       Q-max: 2.255       Lives: 4       Reward: 52.0
87589:127645008       Q-min: 2.057       Q-max: 2.107       Lives: 4       Reward: 53.0
87589:127645029       Q-min: -0.272       Q-max: 0.309       Lives: 3       Reward: 53.0
87589:127645074       Q-min: 1.963       Q-max: 2.158       Lives: 3       Reward: 57.0
87589:127645121       Q-min: 2.300       Q-max: 2.361       Lives: 3       Reward: 58.0
87589:127645164       Q-min: 2.119       Q-max: 2.211       Lives: 3       Reward: 59.0
87589:127645204       Q-min: 2.328       Q-max: 2.377       Lives: 3       Reward: 63.0
87589:127645242       Q-min: 1.591       Q-max: 2.503       Lives: 3       Reward: 70.0
87589:127645265       Q-min: 1.942       Q-max: 2.711       Lives: 3       Reward: 74.0
87589:127645289       Q-min: 1.739       Q-max: 3.524       Lives: 3       Reward: 81.0
```

```
87589:127645319       Q-min: 1.548       Q-max: 5.599       Lives: 3       Reward: 88.0
87589:127645326       Q-min: 3.214       Q-max: 6.187       Lives: 3       Reward: 95.0
87589:127645332       Q-min: 4.149       Q-max: 7.073       Lives: 3       Reward: 102.0
87589:127645338       Q-min: 2.279       Q-max: 6.700       Lives: 3       Reward: 109.0
87589:127645344       Q-min: 3.218       Q-max: 6.832       Lives: 3       Reward: 116.0
87589:127645348       Q-min: 3.802       Q-max: 5.502       Lives: 3       Reward: 123.0
87589:127645354       Q-min: 1.270       Q-max: 6.387       Lives: 3       Reward: 130.0
87589:127645360       Q-min: 2.805       Q-max: 6.095       Lives: 3       Reward: 137.0
87589:127645397       Q-min: 2.879       Q-max: 6.591       Lives: 3       Reward: 144.0
87589:127645404       Q-min: 3.505       Q-max: 6.818       Lives: 3       Reward: 151.0
87589:127645410       Q-min: 3.764       Q-max: 6.270       Lives: 3       Reward: 158.0
87589:127645415       Q-min: 3.677       Q-max: 5.796       Lives: 3       Reward: 165.0
87589:127645421       Q-min: 2.668       Q-max: 5.257       Lives: 3       Reward: 172.0
87589:127645427       Q-min: 3.923       Q-max: 5.098       Lives: 3       Reward: 179.0
87589:127645432       Q-min: 1.970       Q-max: 5.844       Lives: 3       Reward: 186.0
87589:127645437       Q-min: 2.983       Q-max: 5.170       Lives: 3       Reward: 193.0
87589:127645442       Q-min: 0.927       Q-max: 5.070       Lives: 3       Reward: 200.0
87589:127645449       Q-min: 2.823       Q-max: 4.550       Lives: 3       Reward: 207.0
87589:127645457       Q-min: 2.697       Q-max: 4.629       Lives: 3       Reward: 214.0
87589:127645463       Q-min: 2.279       Q-max: 3.921       Lives: 3       Reward: 221.0
87589:127645469       Q-min: 1.649       Q-max: 4.535       Lives: 3       Reward: 228.0
87589:127645477       Q-min: 1.761       Q-max: 4.724       Lives: 3       Reward: 235.0
87589:127645485       Q-min: 2.276       Q-max: 4.794       Lives: 3       Reward: 242.0
87589:127645493       Q-min: 1.740       Q-max: 4.088       Lives: 3       Reward: 246.0
87589:127645500       Q-min: 2.749       Q-max: 4.242       Lives: 3       Reward: 253.0
87589:127645507       Q-min: 1.777       Q-max: 4.064       Lives: 3       Reward: 260.0
87589:127645516       Q-min: 1.485       Q-max: 3.484       Lives: 3       Reward: 267.0
87589:127645523       Q-min: 2.242       Q-max: 4.177       Lives: 3       Reward: 274.0
87589:127645531       Q-min: 2.252       Q-max: 3.996       Lives: 3       Reward: 281.0
87589:127645566       Q-min: 1.331       Q-max: 4.973       Lives: 3       Reward: 285.0
87589:127645575       Q-min: 1.970       Q-max: 3.440       Lives: 3       Reward: 289.0
87589:127645584       Q-min: 1.505       Q-max: 3.210       Lives: 3       Reward: 293.0
87589:127645592       Q-min: 1.477       Q-max: 3.720       Lives: 3       Reward: 300.0
87589:127645600       Q-min: 2.563       Q-max: 3.410       Lives: 3       Reward: 304.0
87589:127645608       Q-min: 1.711       Q-max: 3.448       Lives: 3       Reward: 311.0
87589:127645615       Q-min: 2.012       Q-max: 3.991       Lives: 3       Reward: 318.0
87589:127645624       Q-min: 1.686       Q-max: 3.728       Lives: 3       Reward: 325.0
87589:127645632       Q-min: 1.994       Q-max: 3.683       Lives: 3       Reward: 329.0
87589:127645638       Q-min: 2.120       Q-max: 4.264       Lives: 3       Reward: 336.0
87589:127645646       Q-min: 2.023       Q-max: 4.184       Lives: 3       Reward: 340.0
87589:127645655       Q-min: 2.003       Q-max: 2.833       Lives: 3       Reward: 344.0
87589:127645665       Q-min: -0.107       Q-max: 1.473       Lives: 3       Reward: 345.0
87589:127645674       Q-min: 2.355       Q-max: 4.078       Lives: 3       Reward: 349.0
87589:127645713       Q-min: 1.669       Q-max: 2.720       Lives: 3       Reward: 353.0
87589:127645721       Q-min: 2.809       Q-max: 4.243       Lives: 3       Reward: 357.0
87589:127645753       Q-min: 1.922       Q-max: 2.913       Lives: 3       Reward: 361.0
87589:127645772       Q-min: 2.140       Q-max: 2.795       Lives: 3       Reward: 362.0
87589:127645783       Q-min: 0.064       Q-max: 0.209       Lives: 2       Reward: 362.0
```

```
87589:127645841       Q-min: 1.845       Q-max: 2.442       Lives: 2       Reward: 366.0
87589:127645919       Q-min: 0.936       Q-max: 2.705       Lives: 2       Reward: 370.0
87589:127645949       Q-min: 1.205       Q-max: 3.975       Lives: 2       Reward: 374.0
87589:127645956       Q-min: 3.382       Q-max: 4.514       Lives: 2       Reward: 381.0
87589:127645979       Q-min: 0.130       Q-max: 0.357       Lives: 1       Reward: 381.0
87589:127646023       Q-min: 2.012       Q-max: 2.867       Lives: 1       Reward: 385.0
87589:127646070       Q-min: 2.544       Q-max: 2.705       Lives: 1       Reward: 386.0
87589:127646095       Q-min: 0.001       Q-max: 0.201       Lives: 0       Reward: 386.0
87590:127646139       Q-min: 1.787       Q-max: 1.799       Lives: 5       Reward: 1.0
87590:127646182       Q-min: 1.817       Q-max: 1.831       Lives: 5       Reward: 2.0
87590:127646227       Q-min: 1.924       Q-max: 1.953       Lives: 5       Reward: 3.0
87590:127646262       Q-min: 2.057       Q-max: 2.093       Lives: 5       Reward: 4.0
87590:127646293       Q-min: 1.956       Q-max: 1.980       Lives: 5       Reward: 5.0
87590:127646324       Q-min: 1.892       Q-max: 1.911       Lives: 5       Reward: 6.0
87590:127646358       Q-min: 1.726       Q-max: 1.840       Lives: 5       Reward: 7.0
87590:127646406       Q-min: 1.681       Q-max: 1.705       Lives: 5       Reward: 8.0
87590:127646469       Q-min: 1.489       Q-max: 1.679       Lives: 5       Reward: 9.0
87590:127646534       Q-min: 1.689       Q-max: 1.710       Lives: 5       Reward: 10.0
87590:127646601       Q-min: 1.625       Q-max: 1.658       Lives: 5       Reward: 11.0
87590:127646650       Q-min: 1.949       Q-max: 1.968       Lives: 5       Reward: 12.0
87590:127646683       Q-min: 1.941       Q-max: 1.961       Lives: 5       Reward: 13.0
87590:127646715       Q-min: 1.990       Q-max: 2.067       Lives: 5       Reward: 14.0
87590:127646738       Q-min: -0.265       Q-max: 0.133       Lives: 4       Reward: 14.0
87590:127646782       Q-min: 1.799       Q-max: 1.840       Lives: 4       Reward: 15.0
87590:127646825       Q-min: 1.948       Q-max: 1.967       Lives: 4       Reward: 16.0
87590:127646871       Q-min: 1.927       Q-max: 1.959       Lives: 4       Reward: 17.0
87590:127646904       Q-min: 2.043       Q-max: 2.058       Lives: 4       Reward: 18.0
87590:127646935       Q-min: 1.972       Q-max: 2.004       Lives: 4       Reward: 19.0
87590:127646966       Q-min: 2.014       Q-max: 2.062       Lives: 4       Reward: 20.0
87590:127646985       Q-min: -0.062       Q-max: 0.107       Lives: 3       Reward: 20.0
87590:127647040       Q-min: 1.724       Q-max: 1.740       Lives: 3       Reward: 21.0
87590:127647095       Q-min: 1.943       Q-max: 1.955       Lives: 3       Reward: 22.0
87590:127647142       Q-min: 2.024       Q-max: 2.103       Lives: 3       Reward: 26.0
87590:127647182       Q-min: 1.959       Q-max: 2.022       Lives: 3       Reward: 27.0
87590:127647218       Q-min: 2.041       Q-max: 2.136       Lives: 3       Reward: 31.0
87590:127647255       Q-min: 2.036       Q-max: 2.070       Lives: 3       Reward: 32.0
87590:127647290       Q-min: 1.980       Q-max: 2.191       Lives: 3       Reward: 36.0
87590:127647342       Q-min: 1.706       Q-max: 1.736       Lives: 3       Reward: 37.0
87590:127647409       Q-min: 1.795       Q-max: 1.842       Lives: 3       Reward: 38.0
87590:127647473       Q-min: 1.780       Q-max: 1.805       Lives: 3       Reward: 39.0
87590:127647542       Q-min: 1.660       Q-max: 1.909       Lives: 3       Reward: 40.0
87590:127647597       Q-min: 2.079       Q-max: 2.107       Lives: 3       Reward: 41.0
87590:127647629       Q-min: 1.994       Q-max: 2.072       Lives: 3       Reward: 42.0
87590:127647665       Q-min: 2.155       Q-max: 2.173       Lives: 3       Reward: 43.0
87590:127647699       Q-min: 2.047       Q-max: 2.189       Lives: 3       Reward: 47.0
87590:127647736       Q-min: 2.219       Q-max: 2.495       Lives: 3       Reward: 51.0
87590:127647756       Q-min: 2.399       Q-max: 2.565       Lives: 3       Reward: 55.0
87590:127647778       Q-min: 2.284       Q-max: 2.583       Lives: 3       Reward: 59.0
```

```
87590:127647800       Q-min: 2.406       Q-max: 2.620       Lives: 3       Reward: 63.0
87590:127647822       Q-min: 2.359       Q-max: 2.564       Lives: 3       Reward: 67.0
87590:127647833       Q-min: -0.036       Q-max: 0.223       Lives: 2       Reward: 67.0
87590:127647884       Q-min: 1.975       Q-max: 2.198       Lives: 2       Reward: 71.0
87590:127647929       Q-min: 2.128       Q-max: 2.484       Lives: 2       Reward: 75.0
87590:127647972       Q-min: 2.320       Q-max: 2.380       Lives: 2       Reward: 76.0
87590:127648014       Q-min: 2.009       Q-max: 2.463       Lives: 2       Reward: 80.0
87590:127648049       Q-min: 2.390       Q-max: 2.437       Lives: 2       Reward: 81.0
87590:127648081       Q-min: 2.299       Q-max: 2.460       Lives: 2       Reward: 82.0
87590:127648112       Q-min: 2.319       Q-max: 2.551       Lives: 2       Reward: 86.0
87590:127648165       Q-min: 1.847       Q-max: 2.040       Lives: 2       Reward: 87.0
87590:127648241       Q-min: 1.213       Q-max: 2.371       Lives: 2       Reward: 91.0
87590:127648256       Q-min: 0.018       Q-max: 0.151       Lives: 1       Reward: 91.0
87590:127648314       Q-min: 2.263       Q-max: 2.612       Lives: 1       Reward: 95.0
87590:127648335       Q-min: 2.463       Q-max: 2.630       Lives: 1       Reward: 96.0
87590:127648357       Q-min: 2.585       Q-max: 2.698       Lives: 1       Reward: 100.0
87590:127648378       Q-min: 2.491       Q-max: 2.898       Lives: 1       Reward: 104.0
87590:127648400       Q-min: 2.554       Q-max: 2.852       Lives: 1       Reward: 108.0
87590:127648421       Q-min: 2.414       Q-max: 2.958       Lives: 1       Reward: 115.0
87590:127648445       Q-min: 2.390       Q-max: 3.135       Lives: 1       Reward: 119.0
87590:127648468       Q-min: 2.370       Q-max: 3.299       Lives: 1       Reward: 126.0
87590:127648494       Q-min: 2.755       Q-max: 3.314       Lives: 1       Reward: 133.0
87590:127648509       Q-min: -0.140       Q-max: 0.334       Lives: 0       Reward: 133.0
87591:127648551       Q-min: 1.768       Q-max: 1.783       Lives: 5       Reward: 1.0
87591:127648601       Q-min: 1.657       Q-max: 1.664       Lives: 5       Reward: 2.0
87591:127648652       Q-min: 1.851       Q-max: 1.862       Lives: 5       Reward: 3.0
87591:127648690       Q-min: 2.005       Q-max: 2.023       Lives: 5       Reward: 4.0
87591:127648723       Q-min: 1.894       Q-max: 1.913       Lives: 5       Reward: 5.0
87591:127648758       Q-min: 1.980       Q-max: 2.014       Lives: 5       Reward: 6.0
87591:127648791       Q-min: 1.818       Q-max: 1.831       Lives: 5       Reward: 7.0
87591:127648813       Q-min: -0.174       Q-max: 0.101       Lives: 4       Reward: 7.0
87591:127648866       Q-min: 1.657       Q-max: 1.681       Lives: 4       Reward: 8.0
87591:127648929       Q-min: 1.683       Q-max: 1.697       Lives: 4       Reward: 9.0
87591:127648981       Q-min: 1.868       Q-max: 1.886       Lives: 4       Reward: 10.0
87591:127649018       Q-min: 1.920       Q-max: 2.034       Lives: 4       Reward: 11.0
87591:127649052       Q-min: 1.979       Q-max: 2.015       Lives: 4       Reward: 15.0
87591:127649089       Q-min: 1.940       Q-max: 2.027       Lives: 4       Reward: 19.0
87591:127649121       Q-min: 1.964       Q-max: 2.039       Lives: 4       Reward: 20.0
87591:127649168       Q-min: 1.749       Q-max: 1.771       Lives: 4       Reward: 21.0
87591:127649232       Q-min: 1.642       Q-max: 1.716       Lives: 4       Reward: 22.0
87591:127649294       Q-min: 1.752       Q-max: 1.774       Lives: 4       Reward: 23.0
87591:127649359       Q-min: 1.725       Q-max: 1.774       Lives: 4       Reward: 24.0
87591:127649412       Q-min: 1.957       Q-max: 1.990       Lives: 4       Reward: 25.0
87591:127649447       Q-min: 2.097       Q-max: 2.117       Lives: 4       Reward: 26.0
87591:127649478       Q-min: 2.056       Q-max: 2.085       Lives: 4       Reward: 27.0
87591:127649509       Q-min: 2.032       Q-max: 2.131       Lives: 4       Reward: 31.0
87591:127649544       Q-min: 2.193       Q-max: 2.302       Lives: 4       Reward: 35.0
87591:127649565       Q-min: 2.264       Q-max: 2.424       Lives: 4       Reward: 36.0
```

```
87591:127649584       Q-min: 2.031       Q-max: 2.287       Lives: 4       Reward: 37.0
87591:127649596       Q-min: -0.144       Q-max: 0.076       Lives: 3       Reward: 37.0
87591:127649649       Q-min: 1.760       Q-max: 1.798       Lives: 3       Reward: 38.0
87591:127649719       Q-min: 1.564       Q-max: 1.715       Lives: 3       Reward: 42.0
87591:127649777       Q-min: 2.073       Q-max: 2.168       Lives: 3       Reward: 46.0
87591:127649821       Q-min: 2.199       Q-max: 2.354       Lives: 3       Reward: 50.0
87591:127649856       Q-min: 2.322       Q-max: 2.456       Lives: 3       Reward: 54.0
87591:127649869       Q-min: 0.025       Q-max: 0.418       Lives: 2       Reward: 54.0
87591:127649930       Q-min: 1.381       Q-max: 1.778       Lives: 2       Reward: 58.0
87591:127650007       Q-min: 1.768       Q-max: 2.462       Lives: 2       Reward: 62.0
87591:127650027       Q-min: 2.428       Q-max: 2.487       Lives: 2       Reward: 63.0
87591:127650047       Q-min: 2.191       Q-max: 2.541       Lives: 2       Reward: 64.0
87591:127650066       Q-min: 2.439       Q-max: 2.499       Lives: 2       Reward: 65.0
87591:127650086       Q-min: 2.355       Q-max: 2.534       Lives: 2       Reward: 66.0
87591:127650107       Q-min: 2.535       Q-max: 2.644       Lives: 2       Reward: 67.0
87591:127650128       Q-min: 2.424       Q-max: 2.611       Lives: 2       Reward: 71.0
87591:127650149       Q-min: 2.291       Q-max: 2.567       Lives: 2       Reward: 72.0
87591:127650173       Q-min: 2.357       Q-max: 2.671       Lives: 2       Reward: 76.0
87591:127650198       Q-min: 1.918       Q-max: 3.150       Lives: 2       Reward: 83.0
87591:127650223       Q-min: 1.593       Q-max: 3.739       Lives: 2       Reward: 90.0
87591:127650253       Q-min: 2.184       Q-max: 6.902       Lives: 2       Reward: 97.0
87591:127650258       Q-min: 2.303       Q-max: 6.611       Lives: 2       Reward: 104.0
87591:127650263       Q-min: 3.708       Q-max: 6.076       Lives: 2       Reward: 111.0
87591:127650267       Q-min: 2.672       Q-max: 6.862       Lives: 2       Reward: 118.0
87591:127650272       Q-min: 1.783       Q-max: 6.216       Lives: 2       Reward: 125.0
87591:127650277       Q-min: 2.125       Q-max: 5.576       Lives: 2       Reward: 132.0
87591:127650283       Q-min: 3.175       Q-max: 5.794       Lives: 2       Reward: 139.0
87591:127650288       Q-min: 4.208       Q-max: 6.206       Lives: 2       Reward: 146.0
87591:127650293       Q-min: 3.793       Q-max: 5.610       Lives: 2       Reward: 153.0
87591:127650298       Q-min: 3.057       Q-max: 5.366       Lives: 2       Reward: 160.0
87591:127650304       Q-min: 3.419       Q-max: 5.766       Lives: 2       Reward: 167.0
87591:127650308       Q-min: 2.737       Q-max: 6.274       Lives: 2       Reward: 174.0
87591:127650313       Q-min: 4.419       Q-max: 6.346       Lives: 2       Reward: 181.0
87591:127650319       Q-min: 2.684       Q-max: 5.516       Lives: 2       Reward: 188.0
87591:127650322       Q-min: 3.003       Q-max: 4.892       Lives: 2       Reward: 195.0
87591:127650328       Q-min: 3.717       Q-max: 5.497       Lives: 2       Reward: 202.0
87591:127650334       Q-min: 3.530       Q-max: 5.096       Lives: 2       Reward: 209.0
87591:127650341       Q-min: 3.669       Q-max: 5.190       Lives: 2       Reward: 216.0
87591:127650346       Q-min: 2.570       Q-max: 3.876       Lives: 2       Reward: 223.0
87591:127650353       Q-min: 3.737       Q-max: 5.336       Lives: 2       Reward: 230.0
87591:127650360       Q-min: 2.802       Q-max: 5.215       Lives: 2       Reward: 237.0
87591:127650366       Q-min: 1.970       Q-max: 5.470       Lives: 2       Reward: 244.0
87591:127650373       Q-min: 2.147       Q-max: 5.573       Lives: 2       Reward: 251.0
87591:127650379       Q-min: 2.106       Q-max: 4.824       Lives: 2       Reward: 258.0
87591:127650386       Q-min: 2.746       Q-max: 4.343       Lives: 2       Reward: 265.0
87591:127650425       Q-min: 1.870       Q-max: 5.265       Lives: 2       Reward: 272.0
87591:127650432       Q-min: 3.059       Q-max: 4.681       Lives: 2       Reward: 279.0
87591:127650438       Q-min: 2.537       Q-max: 4.428       Lives: 2       Reward: 286.0
```

```
87591:127650445       Q-min: 1.591       Q-max: 4.291       Lives: 2       Reward: 290.0
87591:127650454       Q-min: 2.578       Q-max: 3.873       Lives: 2       Reward: 294.0
87591:127650462       Q-min: 3.137       Q-max: 4.334       Lives: 2       Reward: 301.0
87591:127650469       Q-min: 2.537       Q-max: 3.330       Lives: 2       Reward: 305.0
87591:127650491       Q-min: -0.535       Q-max: 0.537       Lives: 1       Reward: 305.0
87591:127650552       Q-min: 1.948       Q-max: 2.805       Lives: 1       Reward: 309.0
87591:127650573       Q-min: 2.012       Q-max: 3.112       Lives: 1       Reward: 313.0
87591:127650586       Q-min: -0.178       Q-max: 0.125       Lives: 0       Reward: 313.0
87592:127650631       Q-min: 1.738       Q-max: 1.756       Lives: 5       Reward: 1.0
87592:127650681       Q-min: 1.608       Q-max: 1.681       Lives: 5       Reward: 2.0
87592:127650745       Q-min: 1.693       Q-max: 1.710       Lives: 5       Reward: 3.0
87592:127650791       Q-min: 2.007       Q-max: 2.043       Lives: 5       Reward: 4.0
87592:127650823       Q-min: 1.982       Q-max: 1.999       Lives: 5       Reward: 5.0
87592:127650855       Q-min: 1.947       Q-max: 1.978       Lives: 5       Reward: 6.0
87592:127650891       Q-min: 1.860       Q-max: 1.888       Lives: 5       Reward: 7.0
87592:127650936       Q-min: 1.647       Q-max: 1.686       Lives: 5       Reward: 8.0
87592:127650996       Q-min: 1.627       Q-max: 1.706       Lives: 5       Reward: 9.0
87592:127651058       Q-min: 1.615       Q-max: 1.658       Lives: 5       Reward: 10.0
87592:127651120       Q-min: 1.606       Q-max: 1.666       Lives: 5       Reward: 11.0
87592:127651171       Q-min: 2.078       Q-max: 2.122       Lives: 5       Reward: 15.0
87592:127651194       Q-min: -0.342       Q-max: 0.180       Lives: 4       Reward: 15.0
87592:127651249       Q-min: 1.698       Q-max: 1.723       Lives: 4       Reward: 16.0
87592:127651314       Q-min: 1.764       Q-max: 1.785       Lives: 4       Reward: 17.0
87592:127651374       Q-min: 1.671       Q-max: 1.684       Lives: 4       Reward: 18.0
87592:127651421       Q-min: 1.929       Q-max: 1.945       Lives: 4       Reward: 19.0
87592:127651453       Q-min: 1.996       Q-max: 2.010       Lives: 4       Reward: 20.0
87592:127651485       Q-min: 1.951       Q-max: 2.018       Lives: 4       Reward: 21.0
87592:127651521       Q-min: 2.054       Q-max: 2.101       Lives: 4       Reward: 25.0
87592:127651571       Q-min: 1.697       Q-max: 1.773       Lives: 4       Reward: 26.0
87592:127651639       Q-min: 1.759       Q-max: 1.799       Lives: 4       Reward: 30.0
87592:127651707       Q-min: 1.742       Q-max: 1.765       Lives: 4       Reward: 31.0
87592:127651752       Q-min: 0.088       Q-max: 0.286       Lives: 3       Reward: 31.0
87592:127651796       Q-min: 1.841       Q-max: 2.045       Lives: 3       Reward: 35.0
87592:127651844       Q-min: 2.039       Q-max: 2.095       Lives: 3       Reward: 39.0
87592:127651899       Q-min: 1.763       Q-max: 1.800       Lives: 3       Reward: 40.0
87592:127651950       Q-min: 2.205       Q-max: 2.345       Lives: 3       Reward: 44.0
87592:127651973       Q-min: 2.190       Q-max: 2.513       Lives: 3       Reward: 48.0
87592:127651997       Q-min: 2.491       Q-max: 2.628       Lives: 3       Reward: 52.0
87592:127652020       Q-min: 2.352       Q-max: 2.592       Lives: 3       Reward: 53.0
87592:127652042       Q-min: 2.402       Q-max: 2.633       Lives: 3       Reward: 57.0
87592:127652066       Q-min: 2.368       Q-max: 2.562       Lives: 3       Reward: 61.0
87592:127652088       Q-min: 2.472       Q-max: 2.575       Lives: 3       Reward: 62.0
87592:127652110       Q-min: 2.395       Q-max: 2.561       Lives: 3       Reward: 63.0
87592:127652132       Q-min: 2.375       Q-max: 2.585       Lives: 3       Reward: 67.0
87592:127652157       Q-min: 2.473       Q-max: 2.893       Lives: 3       Reward: 71.0
87592:127652176       Q-min: 2.306       Q-max: 2.499       Lives: 3       Reward: 72.0
87592:127652197       Q-min: 2.349       Q-max: 2.556       Lives: 3       Reward: 76.0
87592:127652210       Q-min: -0.203       Q-max: 0.078       Lives: 2       Reward: 76.0
```

```
87592:127652270       Q-min: 2.092       Q-max: 2.818       Lives: 2       Reward: 80.0
87592:127652291       Q-min: 2.315       Q-max: 2.478       Lives: 2       Reward: 81.0
87592:127652312       Q-min: 2.388       Q-max: 2.683       Lives: 2       Reward: 85.0
87592:127652333       Q-min: 2.325       Q-max: 2.544       Lives: 2       Reward: 86.0
87592:127652356       Q-min: 2.417       Q-max: 2.561       Lives: 2       Reward: 93.0
87592:127652380       Q-min: 2.176       Q-max: 2.591       Lives: 2       Reward: 97.0
87592:127652400       Q-min: 2.240       Q-max: 2.582       Lives: 2       Reward: 104.0
87592:127652423       Q-min: 2.176       Q-max: 2.865       Lives: 2       Reward: 111.0
87592:127652447       Q-min: 2.034       Q-max: 3.135       Lives: 2       Reward: 118.0
87592:127652472       Q-min: 2.173       Q-max: 3.220       Lives: 2       Reward: 125.0
87592:127652494       Q-min: 2.264       Q-max: 3.115       Lives: 2       Reward: 126.0
87592:127652514       Q-min: 2.601       Q-max: 3.100       Lives: 2       Reward: 127.0
87592:127652534       Q-min: 2.156       Q-max: 2.808       Lives: 2       Reward: 131.0
87592:127652555       Q-min: 2.541       Q-max: 2.808       Lives: 2       Reward: 132.0
87592:127652577       Q-min: 2.678       Q-max: 3.090       Lives: 2       Reward: 136.0
87592:127652590       Q-min: 0.011       Q-max: 0.967       Lives: 1       Reward: 136.0
87592:127652640       Q-min: 2.543       Q-max: 3.653       Lives: 1       Reward: 143.0
87592:127652665       Q-min: 2.447       Q-max: 4.720       Lives: 1       Reward: 150.0
87592:127652689       Q-min: 2.704       Q-max: 4.175       Lives: 1       Reward: 157.0
87592:127652711       Q-min: 3.109       Q-max: 4.645       Lives: 1       Reward: 161.0
87592:127652725       Q-min: -0.014       Q-max: 0.422       Lives: 0       Reward: 161.0
87593:127652768       Q-min: 1.803       Q-max: 1.816       Lives: 5       Reward: 1.0
87593:127652826       Q-min: 1.633       Q-max: 1.654       Lives: 5       Reward: 2.0
87593:127652889       Q-min: 1.706       Q-max: 1.731       Lives: 5       Reward: 3.0
87593:127652934       Q-min: 1.998       Q-max: 2.011       Lives: 5       Reward: 4.0
87593:127652964       Q-min: 2.007       Q-max: 2.038       Lives: 5       Reward: 5.0
87593:127652996       Q-min: 1.879       Q-max: 1.935       Lives: 5       Reward: 6.0
87593:127653026       Q-min: 1.763       Q-max: 1.822       Lives: 5       Reward: 7.0
87593:127653049       Q-min: -0.249       Q-max: 0.235       Lives: 4       Reward: 7.0
87593:127653091       Q-min: 1.865       Q-max: 1.886       Lives: 4       Reward: 8.0
87593:127653140       Q-min: 1.693       Q-max: 1.709       Lives: 4       Reward: 9.0
87593:127653201       Q-min: 1.706       Q-max: 1.724       Lives: 4       Reward: 10.0
87593:127653248       Q-min: 1.998       Q-max: 2.038       Lives: 4       Reward: 11.0
87593:127653281       Q-min: 1.933       Q-max: 2.007       Lives: 4       Reward: 12.0
87593:127653315       Q-min: 1.936       Q-max: 1.983       Lives: 4       Reward: 16.0
87593:127653350       Q-min: 2.174       Q-max: 2.322       Lives: 4       Reward: 20.0
87593:127653365       Q-min: 0.036       Q-max: 0.426       Lives: 3       Reward: 20.0
87593:127653408       Q-min: 1.847       Q-max: 1.862       Lives: 3       Reward: 21.0
87593:127653462       Q-min: 1.701       Q-max: 1.795       Lives: 3       Reward: 22.0
87593:127653529       Q-min: 1.778       Q-max: 1.801       Lives: 3       Reward: 23.0
87593:127653578       Q-min: 2.010       Q-max: 2.020       Lives: 3       Reward: 24.0
87593:127653611       Q-min: 2.053       Q-max: 2.078       Lives: 3       Reward: 25.0
87593:127653647       Q-min: 1.991       Q-max: 2.196       Lives: 3       Reward: 29.0
87593:127653682       Q-min: 2.111       Q-max: 2.268       Lives: 3       Reward: 30.0
87593:127653729       Q-min: 1.767       Q-max: 1.797       Lives: 3       Reward: 31.0
87593:127653799       Q-min: 1.709       Q-max: 1.831       Lives: 3       Reward: 32.0
87593:127653866       Q-min: 1.755       Q-max: 1.803       Lives: 3       Reward: 33.0
87593:127653933       Q-min: 1.789       Q-max: 1.811       Lives: 3       Reward: 34.0
```

```
87593:127653987   Q-min: 2.132   Q-max: 2.257   Lives: 3   Reward: 38.0
87593:127654008   Q-min: 2.290   Q-max: 2.414   Lives: 3   Reward: 39.0
87593:127654029   Q-min: 2.187   Q-max: 2.466   Lives: 3   Reward: 43.0
87593:127654049   Q-min: 2.475   Q-max: 2.549   Lives: 3   Reward: 47.0
87593:127654072   Q-min: 2.380   Q-max: 2.476   Lives: 3   Reward: 48.0
87593:127654092   Q-min: 2.384   Q-max: 2.493   Lives: 3   Reward: 52.0
87593:127654105   Q-min: 0.072   Q-max: 0.224   Lives: 2   Reward: 52.0
87593:127654162   Q-min: 1.802   Q-max: 1.834   Lives: 2   Reward: 53.0
87593:127654229   Q-min: 1.875   Q-max: 1.913   Lives: 2   Reward: 54.0
87593:127654301   Q-min: 1.697   Q-max: 1.889   Lives: 2   Reward: 58.0
87593:127654351   Q-min: 2.277   Q-max: 2.401   Lives: 2   Reward: 59.0
87593:127654375   Q-min: 0.016   Q-max: 0.213   Lives: 1   Reward: 59.0
87593:127654435   Q-min: 1.763   Q-max: 1.906   Lives: 1   Reward: 60.0
87593:127654499   Q-min: 1.852   Q-max: 1.911   Lives: 1   Reward: 61.0
87593:127654566   Q-min: 1.833   Q-max: 1.903   Lives: 1   Reward: 62.0
87593:127654618   Q-min: 1.870   Q-max: 2.419   Lives: 1   Reward: 66.0
87593:127654631   Q-min: -0.017   Q-max: 0.137   Lives: 0   Reward: 66.0
87594:127654685   Q-min: 1.651   Q-max: 1.682   Lives: 5   Reward: 1.0
87594:127654736   Q-min: 1.850   Q-max: 1.864   Lives: 5   Reward: 2.0
87594:127654780   Q-min: 1.891   Q-max: 1.934   Lives: 5   Reward: 3.0
87594:127654817   Q-min: 1.999   Q-max: 2.044   Lives: 5   Reward: 4.0
87594:127654850   Q-min: 1.971   Q-max: 1.982   Lives: 5   Reward: 5.0
87594:127654882   Q-min: 1.880   Q-max: 1.913   Lives: 5   Reward: 6.0
87594:127654916   Q-min: 1.829   Q-max: 1.878   Lives: 5   Reward: 10.0
87594:127654937   Q-min: -0.237   Q-max: 0.162   Lives: 4   Reward: 10.0
87594:127654990   Q-min: 1.683   Q-max: 1.710   Lives: 4   Reward: 11.0
87594:127655040   Q-min: 1.909   Q-max: 2.004   Lives: 4   Reward: 12.0
87594:127655083   Q-min: 2.002   Q-max: 2.042   Lives: 4   Reward: 16.0
87594:127655127   Q-min: 1.953   Q-max: 1.996   Lives: 4   Reward: 20.0
87594:127655164   Q-min: 2.028   Q-max: 2.074   Lives: 4   Reward: 21.0
87594:127655196   Q-min: 2.035   Q-max: 2.054   Lives: 4   Reward: 22.0
87594:127655229   Q-min: 2.120   Q-max: 2.405   Lives: 4   Reward: 26.0
87594:127655251   Q-min: 2.246   Q-max: 2.396   Lives: 4   Reward: 27.0
87594:127655268   Q-min: 2.243   Q-max: 2.392   Lives: 4   Reward: 28.0
87594:127655286   Q-min: 2.245   Q-max: 2.390   Lives: 4   Reward: 29.0
87594:127655308   Q-min: 2.304   Q-max: 2.416   Lives: 4   Reward: 30.0
87594:127655330   Q-min: 2.293   Q-max: 2.345   Lives: 4   Reward: 34.0
87594:127655343   Q-min: -0.039   Q-max: 0.290   Lives: 3   Reward: 34.0
87594:127655383   Q-min: 2.074   Q-max: 2.111   Lives: 3   Reward: 35.0
87594:127655434   Q-min: 1.858   Q-max: 1.963   Lives: 3   Reward: 36.0
87594:127655487   Q-min: 2.120   Q-max: 2.189   Lives: 3   Reward: 37.0
87594:127655527   Q-min: 2.145   Q-max: 2.264   Lives: 3   Reward: 41.0
87594:127655559   Q-min: 2.238   Q-max: 2.304   Lives: 3   Reward: 42.0
87594:127655580   Q-min: -0.035   Q-max: 0.260   Lives: 2   Reward: 42.0
87594:127655636   Q-min: 1.821   Q-max: 2.036   Lives: 2   Reward: 46.0
87594:127655705   Q-min: 1.870   Q-max: 1.949   Lives: 2   Reward: 47.0
87594:127655772   Q-min: 1.753   Q-max: 1.917   Lives: 2   Reward: 48.0
87594:127655821   Q-min: 2.071   Q-max: 2.226   Lives: 2   Reward: 49.0
```

```
87594:127655852       Q-min: 2.208       Q-max: 2.255       Lives: 2       Reward: 50.0
87594:127655885       Q-min: 2.251       Q-max: 2.297       Lives: 2       Reward: 51.0
87594:127655906       Q-min: -0.061       Q-max: 0.177       Lives: 1       Reward: 51.0
87594:127655952       Q-min: 2.156       Q-max: 2.214       Lives: 1       Reward: 52.0
87594:127655991       Q-min: 2.186       Q-max: 2.215       Lives: 1       Reward: 53.0
87594:127656034       Q-min: 2.292       Q-max: 2.545       Lives: 1       Reward: 57.0
87594:127656055       Q-min: 2.133       Q-max: 2.445       Lives: 1       Reward: 64.0
87594:127656076       Q-min: 2.376       Q-max: 2.508       Lives: 1       Reward: 65.0
87594:127656099       Q-min: 2.311       Q-max: 2.453       Lives: 1       Reward: 69.0
87594:127656122       Q-min: 2.194       Q-max: 2.632       Lives: 1       Reward: 73.0
87594:127656135       Q-min: 0.041       Q-max: 0.202       Lives: 0       Reward: 73.0
87595:127656179       Q-min: 1.787       Q-max: 1.804       Lives: 5       Reward: 1.0
87595:127656232       Q-min: 1.614       Q-max: 1.645       Lives: 5       Reward: 2.0
87595:127656293       Q-min: 1.658       Q-max: 1.683       Lives: 5       Reward: 3.0
87595:127656342       Q-min: 2.007       Q-max: 2.039       Lives: 5       Reward: 4.0
87595:127656371       Q-min: 1.982       Q-max: 2.006       Lives: 5       Reward: 5.0
87595:127656402       Q-min: 1.909       Q-max: 1.933       Lives: 5       Reward: 6.0
87595:127656436       Q-min: 1.743       Q-max: 1.803       Lives: 5       Reward: 7.0
87595:127656456       Q-min: -0.313       Q-max: 0.038       Lives: 4       Reward: 7.0
87595:127656499       Q-min: 1.838       Q-max: 1.848       Lives: 4       Reward: 8.0
87595:127656537       Q-min: 1.963       Q-max: 1.978       Lives: 4       Reward: 9.0
87595:127656590       Q-min: 1.693       Q-max: 1.720       Lives: 4       Reward: 10.0
87595:127656638       Q-min: 1.960       Q-max: 1.993       Lives: 4       Reward: 11.0
87595:127656657       Q-min: -0.459       Q-max: 0.010       Lives: 3       Reward: 11.0
87595:127656706       Q-min: 1.820       Q-max: 1.841       Lives: 3       Reward: 12.0
87595:127656748       Q-min: 1.909       Q-max: 1.929       Lives: 3       Reward: 13.0
87595:127656793       Q-min: 1.927       Q-max: 1.964       Lives: 3       Reward: 14.0
87595:127656833       Q-min: 2.082       Q-max: 2.122       Lives: 3       Reward: 18.0
87595:127656868       Q-min: 1.989       Q-max: 2.011       Lives: 3       Reward: 19.0
87595:127656900       Q-min: 2.109       Q-max: 2.127       Lives: 3       Reward: 23.0
87595:127656934       Q-min: 2.041       Q-max: 2.196       Lives: 3       Reward: 24.0
87595:127656983       Q-min: 1.759       Q-max: 1.890       Lives: 3       Reward: 25.0
87595:127657045       Q-min: 1.616       Q-max: 1.682       Lives: 3       Reward: 26.0
87595:127657107       Q-min: 1.712       Q-max: 1.735       Lives: 3       Reward: 27.0
87595:127657169       Q-min: 1.645       Q-max: 1.729       Lives: 3       Reward: 28.0
87595:127657218       Q-min: 2.122       Q-max: 2.151       Lives: 3       Reward: 29.0
87595:127657251       Q-min: 2.031       Q-max: 2.106       Lives: 3       Reward: 30.0
87595:127657285       Q-min: 2.027       Q-max: 2.106       Lives: 3       Reward: 34.0
87595:127657318       Q-min: 2.200       Q-max: 2.330       Lives: 3       Reward: 38.0
87595:127657356       Q-min: 2.014       Q-max: 2.550       Lives: 3       Reward: 42.0
87595:127657379       Q-min: 2.378       Q-max: 2.420       Lives: 3       Reward: 46.0
87595:127657402       Q-min: 2.039       Q-max: 2.519       Lives: 3       Reward: 50.0
87595:127657423       Q-min: 2.143       Q-max: 2.505       Lives: 3       Reward: 54.0
87595:127657436       Q-min: 0.032       Q-max: 0.222       Lives: 2       Reward: 54.0
87595:127657488       Q-min: 1.849       Q-max: 1.884       Lives: 2       Reward: 55.0
87595:127657542       Q-min: 2.099       Q-max: 2.417       Lives: 2       Reward: 62.0
87595:127657558       Q-min: 0.040       Q-max: 0.255       Lives: 1       Reward: 62.0
87595:127657615       Q-min: 1.877       Q-max: 1.897       Lives: 1       Reward: 63.0
```

```
87595:127657683       Q-min: 1.949       Q-max: 2.054       Lives: 1       Reward: 64.0
87595:127657735       Q-min: 2.190       Q-max: 2.218       Lives: 1       Reward: 65.0
87595:127657773       Q-min: 2.213       Q-max: 2.361       Lives: 1       Reward: 69.0
87595:127657812       Q-min: 2.257       Q-max: 2.287       Lives: 1       Reward: 70.0
87595:127657846       Q-min: 2.189       Q-max: 2.253       Lives: 1       Reward: 71.0
87595:127657881       Q-min: 2.073       Q-max: 2.765       Lives: 1       Reward: 78.0
87595:127657904       Q-min: 2.204       Q-max: 2.673       Lives: 1       Reward: 85.0
87595:127657925       Q-min: 2.340       Q-max: 2.609       Lives: 1       Reward: 89.0
87595:127657946       Q-min: 2.491       Q-max: 2.714       Lives: 1       Reward: 96.0
87595:127657963       Q-min: -0.053       Q-max: 0.182       Lives: 0       Reward: 96.0
87596:127657995       Q-min: 0.108       Q-max: 0.173       Lives: 4       Reward: 0.0
87596:127658040       Q-min: 1.857       Q-max: 1.868       Lives: 4       Reward: 1.0
87596:127658091       Q-min: 1.656       Q-max: 1.679       Lives: 4       Reward: 2.0
87596:127658146       Q-min: 1.914       Q-max: 1.934       Lives: 4       Reward: 3.0
87596:127658180       Q-min: 1.994       Q-max: 2.027       Lives: 4       Reward: 4.0
87596:127658215       Q-min: 1.923       Q-max: 1.957       Lives: 4       Reward: 5.0
87596:127658246       Q-min: 1.904       Q-max: 1.943       Lives: 4       Reward: 6.0
87596:127658277       Q-min: 1.882       Q-max: 1.923       Lives: 4       Reward: 7.0
87596:127658297       Q-min: -0.045       Q-max: 0.174       Lives: 3       Reward: 7.0
87596:127658341       Q-min: 1.873       Q-max: 1.892       Lives: 3       Reward: 8.0
87596:127658396       Q-min: 1.697       Q-max: 1.740       Lives: 3       Reward: 9.0
87596:127658450       Q-min: 1.920       Q-max: 1.935       Lives: 3       Reward: 10.0
87596:127658492       Q-min: 1.899       Q-max: 1.933       Lives: 3       Reward: 11.0
87596:127658523       Q-min: 1.970       Q-max: 2.013       Lives: 3       Reward: 12.0
87596:127658557       Q-min: 2.045       Q-max: 2.082       Lives: 3       Reward: 16.0
87596:127658589       Q-min: 2.011       Q-max: 2.075       Lives: 3       Reward: 17.0
87596:127658635       Q-min: 1.676       Q-max: 1.695       Lives: 3       Reward: 18.0
87596:127658696       Q-min: 1.684       Q-max: 1.778       Lives: 3       Reward: 19.0
87596:127658764       Q-min: 1.651       Q-max: 1.702       Lives: 3       Reward: 20.0
87596:127658827       Q-min: 1.536       Q-max: 1.713       Lives: 3       Reward: 21.0
87596:127658880       Q-min: 1.987       Q-max: 2.016       Lives: 3       Reward: 22.0
87596:127658911       Q-min: 2.022       Q-max: 2.084       Lives: 3       Reward: 23.0
87596:127658941       Q-min: 2.085       Q-max: 2.137       Lives: 3       Reward: 24.0
87596:127658975       Q-min: 2.006       Q-max: 2.028       Lives: 3       Reward: 25.0
87596:127659009       Q-min: 1.986       Q-max: 2.021       Lives: 3       Reward: 26.0
87596:127659040       Q-min: 1.968       Q-max: 2.005       Lives: 3       Reward: 27.0
87596:127659073       Q-min: 1.963       Q-max: 2.024       Lives: 3       Reward: 31.0
87596:127659106       Q-min: 2.057       Q-max: 2.103       Lives: 3       Reward: 32.0
87596:127659126       Q-min: -0.127       Q-max: 0.197       Lives: 2       Reward: 32.0
87596:127659171       Q-min: 1.811       Q-max: 1.927       Lives: 2       Reward: 36.0
87596:127659220       Q-min: 2.315       Q-max: 2.631       Lives: 2       Reward: 40.0
87596:127659241       Q-min: 2.398       Q-max: 2.459       Lives: 2       Reward: 44.0
87596:127659263       Q-min: 2.251       Q-max: 2.506       Lives: 2       Reward: 48.0
87596:127659284       Q-min: 2.396       Q-max: 2.576       Lives: 2       Reward: 55.0
87596:127659310       Q-min: 2.400       Q-max: 2.546       Lives: 2       Reward: 59.0
87596:127659333       Q-min: 2.413       Q-max: 2.656       Lives: 2       Reward: 63.0
87596:127659353       Q-min: 2.393       Q-max: 2.510       Lives: 2       Reward: 64.0
87596:127659367       Q-min: 0.114       Q-max: 0.326       Lives: 1       Reward: 64.0
```

```
87596:127659412       Q-min: 2.210       Q-max: 2.251       Lives: 1       Reward: 65.0
87596:127659466       Q-min: 1.759       Q-max: 1.894       Lives: 1       Reward: 66.0
87596:127659525       Q-min: 2.140       Q-max: 2.177       Lives: 1       Reward: 67.0
87596:127659555       Q-min: -0.112       Q-max: 0.265       Lives: 0       Reward: 67.0
87597:127659610       Q-min: 1.648       Q-max: 1.659       Lives: 5       Reward: 1.0
87597:127659675       Q-min: 1.680       Q-max: 1.706       Lives: 5       Reward: 2.0
87597:127659734       Q-min: 1.666       Q-max: 1.694       Lives: 5       Reward: 3.0
87597:127659781       Q-min: 1.980       Q-max: 2.012       Lives: 5       Reward: 4.0
87597:127659813       Q-min: 1.925       Q-max: 1.954       Lives: 5       Reward: 5.0
87597:127659847       Q-min: 1.937       Q-max: 1.965       Lives: 5       Reward: 6.0
87597:127659880       Q-min: 1.783       Q-max: 1.804       Lives: 5       Reward: 7.0
87597:127659930       Q-min: 1.663       Q-max: 1.684       Lives: 5       Reward: 8.0
87597:127659994       Q-min: 1.693       Q-max: 1.731       Lives: 5       Reward: 9.0
87597:127660061       Q-min: 1.694       Q-max: 1.711       Lives: 5       Reward: 10.0
87597:127660122       Q-min: 1.646       Q-max: 1.711       Lives: 5       Reward: 11.0
87597:127660171       Q-min: 1.918       Q-max: 1.954       Lives: 5       Reward: 12.0
87597:127660203       Q-min: 1.916       Q-max: 1.961       Lives: 5       Reward: 13.0
87597:127660239       Q-min: 1.961       Q-max: 2.006       Lives: 5       Reward: 17.0
87597:127660273       Q-min: 1.975       Q-max: 2.003       Lives: 5       Reward: 18.0
87597:127660309       Q-min: 1.822       Q-max: 1.941       Lives: 5       Reward: 19.0
87597:127660343       Q-min: 2.056       Q-max: 2.079       Lives: 5       Reward: 20.0
87597:127660382       Q-min: 1.792       Q-max: 2.034       Lives: 5       Reward: 24.0
87597:127660416       Q-min: 2.032       Q-max: 2.174       Lives: 5       Reward: 25.0
87597:127660446       Q-min: 2.066       Q-max: 2.134       Lives: 5       Reward: 26.0
87597:127660470       Q-min: -0.343       Q-max: -0.004       Lives: 4       Reward: 26.0
87597:127660516       Q-min: 1.972       Q-max: 2.023       Lives: 4       Reward: 27.0
87597:127660567       Q-min: 1.783       Q-max: 1.794       Lives: 4       Reward: 28.0
87597:127660620       Q-min: 2.041       Q-max: 2.067       Lives: 4       Reward: 29.0
87597:127660663       Q-min: 1.952       Q-max: 2.087       Lives: 4       Reward: 30.0
87597:127660699       Q-min: 2.036       Q-max: 2.089       Lives: 4       Reward: 31.0
87597:127660735       Q-min: 2.154       Q-max: 2.557       Lives: 4       Reward: 35.0
87597:127660758       Q-min: 2.147       Q-max: 2.391       Lives: 4       Reward: 36.0
87597:127660778       Q-min: 2.132       Q-max: 2.248       Lives: 4       Reward: 37.0
87597:127660790       Q-min: -0.046       Q-max: 0.138       Lives: 3       Reward: 37.0
87597:127660845       Q-min: 1.827       Q-max: 1.883       Lives: 3       Reward: 38.0
87597:127660908       Q-min: 1.814       Q-max: 1.831       Lives: 3       Reward: 39.0
87597:127660964       Q-min: 2.022       Q-max: 2.082       Lives: 3       Reward: 40.0
87597:127661005       Q-min: 2.147       Q-max: 2.237       Lives: 3       Reward: 44.0
87597:127661038       Q-min: 2.118       Q-max: 2.151       Lives: 3       Reward: 45.0
87597:127661061       Q-min: -1.010       Q-max: 0.276       Lives: 2       Reward: 45.0
87597:127661107       Q-min: 2.039       Q-max: 2.158       Lives: 2       Reward: 49.0
87597:127661158       Q-min: 1.998       Q-max: 2.660       Lives: 2       Reward: 53.0
87597:127661181       Q-min: 2.162       Q-max: 2.353       Lives: 2       Reward: 57.0
87597:127661194       Q-min: 0.028       Q-max: 0.147       Lives: 1       Reward: 57.0
87597:127661243       Q-min: 1.971       Q-max: 2.025       Lives: 1       Reward: 61.0
87597:127661293       Q-min: 2.387       Q-max: 2.612       Lives: 1       Reward: 65.0
87597:127661307       Q-min: -0.086       Q-max: 0.145       Lives: 0       Reward: 65.0
87598:127661349       Q-min: 1.720       Q-max: 1.728       Lives: 5       Reward: 1.0
```

```
87598:127661392       Q-min: 1.806       Q-max: 1.829       Lives: 5       Reward: 2.0
87598:127661434       Q-min: 1.861       Q-max: 1.879       Lives: 5       Reward: 3.0
87598:127661471       Q-min: 1.963       Q-max: 1.981       Lives: 5       Reward: 4.0
87598:127661506       Q-min: 1.988       Q-max: 2.017       Lives: 5       Reward: 5.0
87598:127661527       Q-min: -0.344       Q-max: 0.013       Lives: 4       Reward: 5.0
87598:127661571       Q-min: 1.917       Q-max: 1.947       Lives: 4       Reward: 6.0
87598:127661611       Q-min: 1.861       Q-max: 1.881       Lives: 4       Reward: 7.0
87598:127661665       Q-min: 1.710       Q-max: 1.722       Lives: 4       Reward: 8.0
87598:127661713       Q-min: 1.970       Q-max: 2.007       Lives: 4       Reward: 9.0
87598:127661746       Q-min: 1.936       Q-max: 1.950       Lives: 4       Reward: 10.0
87598:127661767       Q-min: 0.092       Q-max: 0.242       Lives: 3       Reward: 10.0
87598:127661820       Q-min: 1.649       Q-max: 1.680       Lives: 3       Reward: 11.0
87598:127661876       Q-min: 1.952       Q-max: 2.013       Lives: 3       Reward: 12.0
87598:127661899       Q-min: -0.122       Q-max: 0.316       Lives: 2       Reward: 12.0
87598:127661943       Q-min: 1.914       Q-max: 1.941       Lives: 2       Reward: 13.0
87598:127661993       Q-min: 1.659       Q-max: 1.671       Lives: 2       Reward: 14.0
87598:127662061       Q-min: 1.709       Q-max: 1.782       Lives: 2       Reward: 15.0
87598:127662109       Q-min: 1.955       Q-max: 1.971       Lives: 2       Reward: 16.0
87598:127662142       Q-min: 1.949       Q-max: 1.971       Lives: 2       Reward: 17.0
87598:127662173       Q-min: 1.925       Q-max: 1.955       Lives: 2       Reward: 18.0
87598:127662208       Q-min: 2.023       Q-max: 2.068       Lives: 2       Reward: 22.0
87598:127662256       Q-min: 1.701       Q-max: 1.749       Lives: 2       Reward: 23.0
87598:127662319       Q-min: 1.701       Q-max: 1.724       Lives: 2       Reward: 24.0
87598:127662392       Q-min: 1.767       Q-max: 1.814       Lives: 2       Reward: 28.0
87598:127662461       Q-min: 1.711       Q-max: 1.729       Lives: 2       Reward: 29.0
87598:127662513       Q-min: 2.016       Q-max: 2.035       Lives: 2       Reward: 30.0
87598:127662549       Q-min: 2.117       Q-max: 2.186       Lives: 2       Reward: 31.0
87598:127662581       Q-min: 2.076       Q-max: 2.144       Lives: 2       Reward: 35.0
87598:127662616       Q-min: 2.132       Q-max: 2.203       Lives: 2       Reward: 36.0
87598:127662651       Q-min: 1.996       Q-max: 2.062       Lives: 2       Reward: 37.0
87598:127662685       Q-min: 2.059       Q-max: 2.100       Lives: 2       Reward: 38.0
87598:127662718       Q-min: 2.218       Q-max: 2.223       Lives: 2       Reward: 42.0
87598:127662753       Q-min: 2.221       Q-max: 2.278       Lives: 2       Reward: 43.0
87598:127662787       Q-min: 2.014       Q-max: 2.063       Lives: 2       Reward: 47.0
87598:127662822       Q-min: 2.088       Q-max: 2.438       Lives: 2       Reward: 51.0
87598:127662844       Q-min: 2.259       Q-max: 2.415       Lives: 2       Reward: 52.0
87598:127662864       Q-min: 2.365       Q-max: 2.445       Lives: 2       Reward: 56.0
87598:127662884       Q-min: 2.356       Q-max: 2.439       Lives: 2       Reward: 60.0
87598:127662905       Q-min: 2.422       Q-max: 2.500       Lives: 2       Reward: 64.0
87598:127662919       Q-min: 0.224       Q-max: 0.494       Lives: 1       Reward: 64.0
87598:127662966       Q-min: 2.476       Q-max: 2.743       Lives: 1       Reward: 68.0
87598:127662989       Q-min: 2.496       Q-max: 2.813       Lives: 1       Reward: 69.0
87598:127663010       Q-min: 2.405       Q-max: 2.518       Lives: 1       Reward: 73.0
87598:127663024       Q-min: -0.284       Q-max: 0.232       Lives: 0       Reward: 73.0
87599:127663067       Q-min: 1.753       Q-max: 1.762       Lives: 5       Reward: 1.0
87599:127663109       Q-min: 1.800       Q-max: 1.813       Lives: 5       Reward: 2.0
87599:127663160       Q-min: 1.662       Q-max: 1.687       Lives: 5       Reward: 3.0
87599:127663207       Q-min: 1.974       Q-max: 1.987       Lives: 5       Reward: 4.0
```

```
87599:127663242       Q-min: 1.966       Q-max: 1.994       Lives: 5       Reward: 5.0
87599:127663274       Q-min: 1.901       Q-max: 1.921       Lives: 5       Reward: 6.0
87599:127663308       Q-min: 1.814       Q-max: 1.833       Lives: 5       Reward: 7.0
87599:127663357       Q-min: 1.644       Q-max: 1.661       Lives: 5       Reward: 8.0
87599:127663424       Q-min: 1.685       Q-max: 1.703       Lives: 5       Reward: 9.0
87599:127663486       Q-min: 1.663       Q-max: 1.695       Lives: 5       Reward: 10.0
87599:127663556       Q-min: 1.684       Q-max: 1.699       Lives: 5       Reward: 11.0
87599:127663606       Q-min: 2.030       Q-max: 2.084       Lives: 5       Reward: 15.0
87599:127663628       Q-min: -0.448       Q-max: 0.082       Lives: 4       Reward: 15.0
87599:127663684       Q-min: 1.722       Q-max: 1.736       Lives: 4       Reward: 16.0
87599:127663748       Q-min: 1.720       Q-max: 1.772       Lives: 4       Reward: 17.0
87599:127663813       Q-min: 1.686       Q-max: 1.717       Lives: 4       Reward: 18.0
87599:127663861       Q-min: 2.008       Q-max: 2.113       Lives: 4       Reward: 22.0
87599:127663896       Q-min: 2.009       Q-max: 2.052       Lives: 4       Reward: 23.0
87599:127663929       Q-min: 2.039       Q-max: 2.057       Lives: 4       Reward: 24.0
87599:127663961       Q-min: 2.076       Q-max: 2.184       Lives: 4       Reward: 25.0
87599:127664008       Q-min: 1.692       Q-max: 1.707       Lives: 4       Reward: 26.0
87599:127664077       Q-min: 1.739       Q-max: 1.818       Lives: 4       Reward: 27.0
87599:127664144       Q-min: 1.713       Q-max: 1.725       Lives: 4       Reward: 28.0
87599:127664208       Q-min: 1.743       Q-max: 1.789       Lives: 4       Reward: 29.0
87599:127664255       Q-min: 2.060       Q-max: 2.139       Lives: 4       Reward: 30.0
87599:127664291       Q-min: 2.060       Q-max: 2.154       Lives: 4       Reward: 34.0
87599:127664325       Q-min: 2.065       Q-max: 2.095       Lives: 4       Reward: 35.0
87599:127664358       Q-min: 2.066       Q-max: 2.166       Lives: 4       Reward: 36.0
87599:127664391       Q-min: 1.992       Q-max: 2.050       Lives: 4       Reward: 37.0
87599:127664420       Q-min: 2.124       Q-max: 2.158       Lives: 4       Reward: 38.0
87599:127664455       Q-min: 2.036       Q-max: 2.213       Lives: 4       Reward: 42.0
87599:127664490       Q-min: 2.089       Q-max: 2.205       Lives: 4       Reward: 43.0
87599:127664513       Q-min: -0.521       Q-max: -0.111       Lives: 3       Reward: 43.0
87599:127664559       Q-min: 1.911       Q-max: 1.972       Lives: 3       Reward: 44.0
87599:127664607       Q-min: 2.068       Q-max: 2.141       Lives: 3       Reward: 45.0
87599:127664656       Q-min: 2.022       Q-max: 2.095       Lives: 3       Reward: 49.0
87599:127664699       Q-min: 2.059       Q-max: 2.756       Lives: 3       Reward: 53.0
87599:127664719       Q-min: 2.275       Q-max: 2.449       Lives: 3       Reward: 54.0
87599:127664732       Q-min: 0.060       Q-max: 0.301       Lives: 2       Reward: 54.0
87599:127664778       Q-min: 2.109       Q-max: 2.157       Lives: 2       Reward: 58.0
87599:127664820       Q-min: 2.384       Q-max: 2.539       Lives: 2       Reward: 62.0
87599:127664842       Q-min: 2.198       Q-max: 2.455       Lives: 2       Reward: 66.0
87599:127664864       Q-min: 2.476       Q-max: 2.600       Lives: 2       Reward: 70.0
87599:127664886       Q-min: 2.233       Q-max: 2.521       Lives: 2       Reward: 74.0
87599:127664907       Q-min: 2.210       Q-max: 2.531       Lives: 2       Reward: 78.0
87599:127664929       Q-min: 2.236       Q-max: 2.470       Lives: 2       Reward: 82.0
87599:127664951       Q-min: 2.313       Q-max: 2.487       Lives: 2       Reward: 83.0
87599:127664973       Q-min: 2.177       Q-max: 2.553       Lives: 2       Reward: 87.0
87599:127664987       Q-min: 0.015       Q-max: 0.203       Lives: 1       Reward: 87.0
87599:127665035       Q-min: 2.091       Q-max: 2.143       Lives: 1       Reward: 91.0
87599:127665089       Q-min: 1.995       Q-max: 2.141       Lives: 1       Reward: 92.0
87599:127665161       Q-min: 1.824       Q-max: 2.351       Lives: 1       Reward: 96.0
```

```
87599:127665203        Q-min: -0.230        Q-max: -0.095        Lives: 0        Reward: 96.0
87600:127665258        Q-min: 1.623        Q-max: 1.681        Lives: 5        Reward: 1.0
87600:127665323        Q-min: 1.673        Q-max: 1.694        Lives: 5        Reward: 2.0
87600:127665382        Q-min: 1.696        Q-max: 1.725        Lives: 5        Reward: 3.0
87600:127665429        Q-min: 1.982        Q-max: 2.019        Lives: 5        Reward: 4.0
87600:127665458        Q-min: 1.942        Q-max: 1.968        Lives: 5        Reward: 5.0
87600:127665490        Q-min: 2.015        Q-max: 2.054        Lives: 5        Reward: 6.0
87600:127665526        Q-min: 1.801        Q-max: 1.817        Lives: 5        Reward: 7.0
87600:127665574        Q-min: 1.669        Q-max: 1.709        Lives: 5        Reward: 8.0
87600:127665615        Q-min: -0.153        Q-max: 0.143        Lives: 4        Reward: 8.0
87600:127665660        Q-min: 1.875        Q-max: 1.899        Lives: 4        Reward: 9.0
87600:127665705        Q-min: 1.866        Q-max: 1.891        Lives: 4        Reward: 10.0
87600:127665747        Q-min: 1.927        Q-max: 1.947        Lives: 4        Reward: 11.0
87600:127665784        Q-min: 1.979        Q-max: 2.002        Lives: 4        Reward: 12.0
87600:127665815        Q-min: 1.992        Q-max: 2.034        Lives: 4        Reward: 13.0
87600:127665849        Q-min: 1.990        Q-max: 2.037        Lives: 4        Reward: 17.0
87600:127665882        Q-min: 1.974        Q-max: 2.028        Lives: 4        Reward: 18.0
87600:127665933        Q-min: 1.660        Q-max: 1.804        Lives: 4        Reward: 19.0
87600:127665995        Q-min: 1.712        Q-max: 1.726        Lives: 4        Reward: 20.0
87600:127666054        Q-min: 1.695        Q-max: 1.713        Lives: 4        Reward: 21.0
87600:127666119        Q-min: 1.669        Q-max: 1.699        Lives: 4        Reward: 22.0
87600:127666168        Q-min: 1.967        Q-max: 2.047        Lives: 4        Reward: 23.0
87600:127666189        Q-min: 0.049        Q-max: 0.201        Lives: 3        Reward: 23.0
87600:127666233        Q-min: 1.899        Q-max: 1.930        Lives: 3        Reward: 24.0
87600:127666287        Q-min: 1.692        Q-max: 1.740        Lives: 3        Reward: 25.0
87600:127666345        Q-min: 1.973        Q-max: 2.027        Lives: 3        Reward: 29.0
87600:127666390        Q-min: 2.341        Q-max: 2.606        Lives: 3        Reward: 33.0
87600:127666409        Q-min: 2.472        Q-max: 2.503        Lives: 3        Reward: 34.0
87600:127666429        Q-min: 2.264        Q-max: 2.515        Lives: 3        Reward: 38.0
87600:127666447        Q-min: 2.428        Q-max: 2.572        Lives: 3        Reward: 39.0
87600:127666466        Q-min: 2.377        Q-max: 2.562        Lives: 3        Reward: 43.0
87600:127666488        Q-min: 2.328        Q-max: 2.461        Lives: 3        Reward: 47.0
87600:127666510        Q-min: 2.380        Q-max: 2.607        Lives: 3        Reward: 51.0
87600:127666533        Q-min: 2.363        Q-max: 2.568        Lives: 3        Reward: 55.0
87600:127666553        Q-min: 2.401        Q-max: 2.500        Lives: 3        Reward: 56.0
87600:127666575        Q-min: 2.114        Q-max: 2.592        Lives: 3        Reward: 60.0
87600:127666598        Q-min: 2.280        Q-max: 2.538        Lives: 3        Reward: 64.0
87600:127666620        Q-min: 2.412        Q-max: 2.506        Lives: 3        Reward: 65.0
87600:127666632        Q-min: 0.158        Q-max: 0.395        Lives: 2        Reward: 65.0
87600:127666681        Q-min: 2.151        Q-max: 2.262        Lives: 2        Reward: 69.0
87600:127666728        Q-min: 2.302        Q-max: 2.573        Lives: 2        Reward: 73.0
87600:127666748        Q-min: 2.387        Q-max: 2.597        Lives: 2        Reward: 77.0
87600:127666769        Q-min: 2.162        Q-max: 2.485        Lives: 2        Reward: 81.0
87600:127666791        Q-min: 2.353        Q-max: 2.633        Lives: 2        Reward: 85.0
87600:127666811        Q-min: 2.475        Q-max: 2.603        Lives: 2        Reward: 86.0
87600:127666832        Q-min: 1.714        Q-max: 2.607        Lives: 2        Reward: 93.0
87600:127666847        Q-min: 0.025        Q-max: 0.334        Lives: 1        Reward: 93.0
87600:127666884        Q-min: -0.352        Q-max: 0.459        Lives: 0        Reward: 93.0
```

```
87601:127666925       Q-min: 1.749       Q-max: 1.771       Lives: 5       Reward: 1.0
87601:127666967       Q-min: 1.770       Q-max: 1.802       Lives: 5       Reward: 2.0
87601:127667011       Q-min: 1.867       Q-max: 1.932       Lives: 5       Reward: 3.0
87601:127667047       Q-min: 1.946       Q-max: 1.969       Lives: 5       Reward: 4.0
87601:127667077       Q-min: 1.974       Q-max: 2.042       Lives: 5       Reward: 5.0
87601:127667111       Q-min: 1.965       Q-max: 1.987       Lives: 5       Reward: 6.0
87601:127667142       Q-min: 1.820       Q-max: 1.833       Lives: 5       Reward: 7.0
87601:127667162       Q-min: -0.263       Q-max: 0.170       Lives: 4       Reward: 7.0
87601:127667206       Q-min: 1.866       Q-max: 1.882       Lives: 4       Reward: 8.0
87601:127667250       Q-min: 1.988       Q-max: 2.027       Lives: 4       Reward: 9.0
87601:127667294       Q-min: 1.814       Q-max: 1.851       Lives: 4       Reward: 10.0
87601:127667331       Q-min: 1.933       Q-max: 1.976       Lives: 4       Reward: 11.0
87601:127667363       Q-min: 1.970       Q-max: 2.004       Lives: 4       Reward: 12.0
87601:127667397       Q-min: 1.976       Q-max: 2.032       Lives: 4       Reward: 16.0
87601:127667432       Q-min: 1.974       Q-max: 2.020       Lives: 4       Reward: 17.0
87601:127667478       Q-min: 1.677       Q-max: 1.694       Lives: 4       Reward: 18.0
87601:127667544       Q-min: 1.690       Q-max: 1.733       Lives: 4       Reward: 19.0
87601:127667606       Q-min: 1.794       Q-max: 1.832       Lives: 4       Reward: 20.0
87601:127667671       Q-min: 1.756       Q-max: 1.784       Lives: 4       Reward: 21.0
87601:127667718       Q-min: 2.080       Q-max: 2.130       Lives: 4       Reward: 22.0
87601:127667751       Q-min: 2.029       Q-max: 2.068       Lives: 4       Reward: 23.0
87601:127667781       Q-min: 2.010       Q-max: 2.046       Lives: 4       Reward: 24.0
87601:127667812       Q-min: 2.073       Q-max: 2.228       Lives: 4       Reward: 28.0
87601:127667835       Q-min: -0.028       Q-max: 0.177       Lives: 3       Reward: 28.0
87601:127667891       Q-min: 1.767       Q-max: 1.842       Lives: 3       Reward: 29.0
87601:127667946       Q-min: 1.970       Q-max: 2.021       Lives: 3       Reward: 30.0
87601:127667991       Q-min: 2.062       Q-max: 2.180       Lives: 3       Reward: 31.0
87601:127668030       Q-min: 2.031       Q-max: 2.068       Lives: 3       Reward: 35.0
87601:127668063       Q-min: 2.128       Q-max: 2.416       Lives: 3       Reward: 39.0
87601:127668085       Q-min: 2.294       Q-max: 2.450       Lives: 3       Reward: 43.0
87601:127668108       Q-min: 2.309       Q-max: 2.417       Lives: 3       Reward: 47.0
87601:127668130       Q-min: 2.342       Q-max: 2.450       Lives: 3       Reward: 54.0
87601:127668154       Q-min: 2.212       Q-max: 2.443       Lives: 3       Reward: 58.0
87601:127668176       Q-min: 2.301       Q-max: 2.701       Lives: 3       Reward: 65.0
87601:127668199       Q-min: 2.417       Q-max: 2.585       Lives: 3       Reward: 69.0
87601:127668226       Q-min: 2.306       Q-max: 2.719       Lives: 3       Reward: 76.0
87601:127668247       Q-min: 2.650       Q-max: 3.426       Lives: 3       Reward: 80.0
87601:127668270       Q-min: 1.950       Q-max: 3.292       Lives: 3       Reward: 84.0
87601:127668284       Q-min: 0.406       Q-max: 0.723       Lives: 2       Reward: 84.0
87601:127668333       Q-min: 2.601       Q-max: 3.435       Lives: 2       Reward: 91.0
87601:127668364       Q-min: 2.556       Q-max: 5.411       Lives: 2       Reward: 98.0
87601:127668369       Q-min: 2.444       Q-max: 5.953       Lives: 2       Reward: 105.0
87601:127668374       Q-min: 3.765       Q-max: 5.552       Lives: 2       Reward: 112.0
87601:127668378       Q-min: 3.498       Q-max: 6.066       Lives: 2       Reward: 119.0
87601:127668384       Q-min: 3.315       Q-max: 6.838       Lives: 2       Reward: 126.0
87601:127668388       Q-min: 3.629       Q-max: 6.227       Lives: 2       Reward: 133.0
87601:127668394       Q-min: 3.125       Q-max: 5.766       Lives: 2       Reward: 140.0
87601:127668399       Q-min: 1.872       Q-max: 5.354       Lives: 2       Reward: 147.0
```

```
87601:127668405       Q-min: 1.891       Q-max: 5.260       Lives: 2       Reward: 154.0
87601:127668444       Q-min: 3.421       Q-max: 5.474       Lives: 2       Reward: 161.0
87601:127668450       Q-min: 3.133       Q-max: 5.921       Lives: 2       Reward: 168.0
87601:127668454       Q-min: 1.863       Q-max: 5.537       Lives: 2       Reward: 175.0
87601:127668459       Q-min: 1.319       Q-max: 5.649       Lives: 2       Reward: 182.0
87601:127668463       Q-min: 3.745       Q-max: 5.352       Lives: 2       Reward: 189.0
87601:127668468       Q-min: 3.931       Q-max: 5.100       Lives: 2       Reward: 196.0
87601:127668473       Q-min: 3.457       Q-max: 5.721       Lives: 2       Reward: 203.0
87601:127668479       Q-min: 2.423       Q-max: 3.470       Lives: 2       Reward: 210.0
87601:127668488       Q-min: 1.242       Q-max: 2.573       Lives: 2       Reward: 211.0
87601:127668496       Q-min: 1.938       Q-max: 3.770       Lives: 2       Reward: 218.0
87601:127668503       Q-min: 2.297       Q-max: 3.498       Lives: 2       Reward: 222.0
87601:127668511       Q-min: 1.545       Q-max: 3.395       Lives: 2       Reward: 226.0
87601:127668520       Q-min: 1.693       Q-max: 3.364       Lives: 2       Reward: 233.0
87601:127668527       Q-min: 1.683       Q-max: 3.130       Lives: 2       Reward: 240.0
87601:127668565       Q-min: 1.745       Q-max: 5.647       Lives: 2       Reward: 247.0
87601:127668571       Q-min: 1.912       Q-max: 3.448       Lives: 2       Reward: 251.0
87601:127668611       Q-min: 1.719       Q-max: 3.410       Lives: 2       Reward: 255.0
87601:127668619       Q-min: 2.959       Q-max: 4.002       Lives: 2       Reward: 259.0
87601:127668655       Q-min: 0.753       Q-max: 4.262       Lives: 2       Reward: 263.0
87601:127668663       Q-min: 2.337       Q-max: 4.141       Lives: 2       Reward: 267.0
87601:127668715       Q-min: 0.070       Q-max: 0.779       Lives: 1       Reward: 267.0
87601:127668780       Q-min: 0.650       Q-max: 2.531       Lives: 1       Reward: 274.0
87601:127668794       Q-min: 0.176       Q-max: 0.366       Lives: 0       Reward: 274.0
87602:127668849       Q-min: 1.660       Q-max: 1.689       Lives: 5       Reward: 1.0
87602:127668901       Q-min: 1.762       Q-max: 1.772       Lives: 5       Reward: 2.0
87602:127668955       Q-min: 1.668       Q-max: 1.699       Lives: 5       Reward: 3.0
87602:127669000       Q-min: 1.975       Q-max: 2.019       Lives: 5       Reward: 4.0
87602:127669020       Q-min: -0.215       Q-max: 0.113       Lives: 4       Reward: 4.0
87602:127669074       Q-min: 1.651       Q-max: 1.672       Lives: 4       Reward: 5.0
87602:127669129       Q-min: 1.880       Q-max: 1.922       Lives: 4       Reward: 6.0
87602:127669171       Q-min: 1.946       Q-max: 1.986       Lives: 4       Reward: 10.0
87602:127669212       Q-min: 1.944       Q-max: 2.002       Lives: 4       Reward: 11.0
87602:127669242       Q-min: 1.973       Q-max: 2.019       Lives: 4       Reward: 12.0
87602:127669274       Q-min: 1.926       Q-max: 1.975       Lives: 4       Reward: 13.0
87602:127669307       Q-min: 2.105       Q-max: 2.278       Lives: 4       Reward: 17.0
87602:127669357       Q-min: 1.747       Q-max: 1.796       Lives: 4       Reward: 18.0
87602:127669417       Q-min: 1.715       Q-max: 1.727       Lives: 4       Reward: 19.0
87602:127669482       Q-min: 1.681       Q-max: 1.718       Lives: 4       Reward: 20.0
87602:127669553       Q-min: 1.743       Q-max: 1.765       Lives: 4       Reward: 21.0
87602:127669603       Q-min: 2.072       Q-max: 2.117       Lives: 4       Reward: 22.0
87602:127669636       Q-min: 2.084       Q-max: 2.119       Lives: 4       Reward: 23.0
87602:127669670       Q-min: 2.010       Q-max: 2.060       Lives: 4       Reward: 24.0
87602:127669704       Q-min: 1.898       Q-max: 1.979       Lives: 4       Reward: 28.0
87602:127669738       Q-min: 2.096       Q-max: 2.180       Lives: 4       Reward: 29.0
87602:127669772       Q-min: 2.142       Q-max: 2.170       Lives: 4       Reward: 33.0
87602:127669807       Q-min: 2.072       Q-max: 2.183       Lives: 4       Reward: 34.0
87602:127669841       Q-min: 2.174       Q-max: 2.224       Lives: 4       Reward: 38.0
```

```
87602:127669873       Q-min: 2.093       Q-max: 2.191       Lives: 4       Reward: 39.0
87602:127669910       Q-min: 2.197       Q-max: 2.437       Lives: 4       Reward: 43.0
87602:127669931       Q-min: 2.435       Q-max: 2.479       Lives: 4       Reward: 44.0
87602:127669952       Q-min: 2.251       Q-max: 2.526       Lives: 4       Reward: 45.0
87602:127669974       Q-min: 2.238       Q-max: 2.452       Lives: 4       Reward: 46.0
87602:127669993       Q-min: 2.353       Q-max: 2.468       Lives: 4       Reward: 47.0
87602:127670013       Q-min: 2.226       Q-max: 2.469       Lives: 4       Reward: 51.0
87602:127670035       Q-min: 2.220       Q-max: 2.600       Lives: 4       Reward: 55.0
87602:127670058       Q-min: 2.358       Q-max: 2.584       Lives: 4       Reward: 59.0
87602:127670082       Q-min: 2.341       Q-max: 2.521       Lives: 4       Reward: 63.0
87602:127670103       Q-min: 2.369       Q-max: 2.522       Lives: 4       Reward: 67.0
87602:127670126       Q-min: 1.917       Q-max: 2.413       Lives: 4       Reward: 68.0
87602:127670145       Q-min: 2.342       Q-max: 2.609       Lives: 4       Reward: 69.0
87602:127670158       Q-min: -0.303       Q-max: 0.424       Lives: 3       Reward: 69.0
87602:127670202       Q-min: 2.325       Q-max: 2.465       Lives: 3       Reward: 73.0
87602:127670224       Q-min: 2.308       Q-max: 2.635       Lives: 3       Reward: 74.0
87602:127670245       Q-min: 2.460       Q-max: 2.646       Lives: 3       Reward: 78.0
87602:127670269       Q-min: 2.455       Q-max: 2.712       Lives: 3       Reward: 85.0
87602:127670292       Q-min: 2.240       Q-max: 2.845       Lives: 3       Reward: 92.0
87602:127670315       Q-min: 2.779       Q-max: 3.246       Lives: 3       Reward: 96.0
87602:127670328       Q-min: 0.010       Q-max: 0.268       Lives: 2       Reward: 96.0
87602:127670382       Q-min: 0.438       Q-max: 2.764       Lives: 2       Reward: 103.0
87602:127670406       Q-min: 2.238       Q-max: 4.566       Lives: 2       Reward: 110.0
87602:127670429       Q-min: 2.586       Q-max: 3.575       Lives: 2       Reward: 114.0
87602:127670457       Q-min: 2.830       Q-max: 6.067       Lives: 2       Reward: 121.0
87602:127670462       Q-min: 3.539       Q-max: 6.227       Lives: 2       Reward: 128.0
87602:127670467       Q-min: 2.655       Q-max: 5.876       Lives: 2       Reward: 135.0
87602:127670472       Q-min: 2.309       Q-max: 6.446       Lives: 2       Reward: 142.0
87602:127670478       Q-min: 2.977       Q-max: 6.508       Lives: 2       Reward: 149.0
87602:127670483       Q-min: 2.145       Q-max: 5.609       Lives: 2       Reward: 156.0
87602:127670487       Q-min: 1.552       Q-max: 5.483       Lives: 2       Reward: 163.0
87602:127670493       Q-min: 0.675       Q-max: 6.091       Lives: 2       Reward: 170.0
87602:127670499       Q-min: 3.525       Q-max: 5.526       Lives: 2       Reward: 177.0
87602:127670506       Q-min: 3.911       Q-max: 5.610       Lives: 2       Reward: 184.0
87602:127670513       Q-min: 3.107       Q-max: 4.286       Lives: 2       Reward: 191.0
87602:127670519       Q-min: 3.875       Q-max: 5.261       Lives: 2       Reward: 198.0
87602:127670524       Q-min: 3.681       Q-max: 4.988       Lives: 2       Reward: 205.0
87602:127670530       Q-min: 3.829       Q-max: 5.252       Lives: 2       Reward: 212.0
87602:127670537       Q-min: 3.661       Q-max: 5.947       Lives: 2       Reward: 219.0
87602:127670543       Q-min: 0.818       Q-max: 6.100       Lives: 2       Reward: 226.0
87602:127670550       Q-min: 1.349       Q-max: 4.426       Lives: 2       Reward: 233.0
87602:127670556       Q-min: 3.577       Q-max: 4.942       Lives: 2       Reward: 240.0
87602:127670563       Q-min: 1.653       Q-max: 3.518       Lives: 2       Reward: 244.0
87602:127670569       Q-min: 1.497       Q-max: 4.850       Lives: 2       Reward: 251.0
87602:127670606       Q-min: 3.148       Q-max: 6.722       Lives: 2       Reward: 258.0
87602:127670611       Q-min: 3.101       Q-max: 4.130       Lives: 2       Reward: 265.0
87602:127670616       Q-min: 4.503       Q-max: 5.602       Lives: 2       Reward: 272.0
87602:127670623       Q-min: 2.842       Q-max: 5.075       Lives: 2       Reward: 279.0
```

```
87602:127670629       Q-min: 2.633       Q-max: 5.970       Lives: 2       Reward: 286.0
87602:127670635       Q-min: 2.501       Q-max: 5.630       Lives: 2       Reward: 293.0
87602:127670641       Q-min: 2.565       Q-max: 5.003       Lives: 2       Reward: 300.0
87602:127670648       Q-min: 1.812       Q-max: 2.868       Lives: 2       Reward: 304.0
87602:127670655       Q-min: 2.751       Q-max: 3.801       Lives: 2       Reward: 311.0
87602:127670661       Q-min: 2.372       Q-max: 4.040       Lives: 2       Reward: 318.0
87602:127670682       Q-min: -0.290       Q-max: 0.641       Lives: 1       Reward: 318.0
87602:127670749       Q-min: 1.195       Q-max: 2.524       Lives: 1       Reward: 322.0
87602:127670757       Q-min: 1.663       Q-max: 3.343       Lives: 1       Reward: 326.0
87602:127670781       Q-min: -0.047       Q-max: 0.144       Lives: 0       Reward: 326.0
87603:127670833       Q-min: 1.652       Q-max: 1.663       Lives: 5       Reward: 1.0
87603:127670884       Q-min: 1.845       Q-max: 1.871       Lives: 5       Reward: 2.0
87603:127670925       Q-min: 1.900       Q-max: 1.927       Lives: 5       Reward: 3.0
87603:127670963       Q-min: 2.001       Q-max: 2.066       Lives: 5       Reward: 4.0
87603:127670996       Q-min: 1.983       Q-max: 2.002       Lives: 5       Reward: 5.0
87603:127671027       Q-min: 1.927       Q-max: 1.968       Lives: 5       Reward: 6.0
87603:127671060       Q-min: 1.748       Q-max: 1.767       Lives: 5       Reward: 7.0
87603:127671081       Q-min: -0.046       Q-max: 0.168       Lives: 4       Reward: 7.0
87603:127671130       Q-min: 1.708       Q-max: 1.726       Lives: 4       Reward: 8.0
87603:127671184       Q-min: 1.876       Q-max: 1.902       Lives: 4       Reward: 9.0
87603:127671225       Q-min: 1.927       Q-max: 1.958       Lives: 4       Reward: 10.0
87603:127671266       Q-min: 2.034       Q-max: 2.070       Lives: 4       Reward: 11.0
87603:127671303       Q-min: 1.926       Q-max: 1.949       Lives: 4       Reward: 12.0
87603:127671336       Q-min: 2.010       Q-max: 2.036       Lives: 4       Reward: 16.0
87603:127671366       Q-min: 1.950       Q-max: 1.971       Lives: 4       Reward: 17.0
87603:127671409       Q-min: 1.646       Q-max: 1.664       Lives: 4       Reward: 18.0
87603:127671468       Q-min: 1.675       Q-max: 1.699       Lives: 4       Reward: 19.0
87603:127671535       Q-min: 1.714       Q-max: 1.759       Lives: 4       Reward: 20.0
87603:127671597       Q-min: 1.722       Q-max: 1.766       Lives: 4       Reward: 21.0
87603:127671642       Q-min: 1.999       Q-max: 2.018       Lives: 4       Reward: 22.0
87603:127671664       Q-min: -0.050       Q-max: 0.210       Lives: 3       Reward: 22.0
87603:127671719       Q-min: 1.703       Q-max: 1.715       Lives: 3       Reward: 23.0
87603:127671777       Q-min: 1.927       Q-max: 2.036       Lives: 3       Reward: 27.0
87603:127671830       Q-min: 1.748       Q-max: 1.798       Lives: 3       Reward: 28.0
87603:127671884       Q-min: 1.995       Q-max: 2.128       Lives: 3       Reward: 32.0
87603:127671907       Q-min: -0.121       Q-max: 0.204       Lives: 2       Reward: 32.0
87603:127671947       Q-min: 2.019       Q-max: 2.045       Lives: 2       Reward: 33.0
87603:127672004       Q-min: 1.658       Q-max: 1.772       Lives: 2       Reward: 34.0
87603:127672069       Q-min: 1.687       Q-max: 1.790       Lives: 2       Reward: 35.0
87603:127672121       Q-min: 2.131       Q-max: 2.251       Lives: 2       Reward: 36.0
87603:127672157       Q-min: 2.166       Q-max: 2.329       Lives: 2       Reward: 37.0
87603:127672190       Q-min: 2.115       Q-max: 2.157       Lives: 2       Reward: 41.0
87603:127672224       Q-min: 2.188       Q-max: 2.245       Lives: 2       Reward: 42.0
87603:127672270       Q-min: 1.796       Q-max: 1.822       Lives: 2       Reward: 43.0
87603:127672330       Q-min: 1.802       Q-max: 1.857       Lives: 2       Reward: 44.0
87603:127672396       Q-min: 1.711       Q-max: 1.784       Lives: 2       Reward: 45.0
87603:127672464       Q-min: 1.600       Q-max: 1.845       Lives: 2       Reward: 49.0
87603:127672520       Q-min: 2.268       Q-max: 2.461       Lives: 2       Reward: 53.0
```

```
87603:127672540       Q-min: 2.428       Q-max: 2.548       Lives: 2       Reward: 54.0
87603:127672554       Q-min: 0.115       Q-max: 0.197       Lives: 1       Reward: 54.0
87603:127672609       Q-min: 1.839       Q-max: 1.941       Lives: 1       Reward: 58.0
87603:127672663       Q-min: 2.221       Q-max: 2.305       Lives: 1       Reward: 59.0
87603:127672717       Q-min: 2.086       Q-max: 2.380       Lives: 1       Reward: 63.0
87603:127672771       Q-min: 2.328       Q-max: 2.475       Lives: 1       Reward: 67.0
87603:127672793       Q-min: 2.290       Q-max: 2.543       Lives: 1       Reward: 71.0
87603:127672813       Q-min: 2.222       Q-max: 2.463       Lives: 1       Reward: 75.0
87603:127672827       Q-min: 0.138       Q-max: 0.258       Lives: 0       Reward: 75.0
87604:127672885       Q-min: 1.693       Q-max: 1.714       Lives: 5       Reward: 1.0
87604:127672948       Q-min: 1.663       Q-max: 1.679       Lives: 5       Reward: 2.0
87604:127673000       Q-min: 1.873       Q-max: 1.875       Lives: 5       Reward: 3.0
87604:127673035       Q-min: 2.045       Q-max: 2.064       Lives: 5       Reward: 4.0
87604:127673066       Q-min: 1.967       Q-max: 1.997       Lives: 5       Reward: 5.0
87604:127673101       Q-min: 1.931       Q-max: 1.941       Lives: 5       Reward: 6.0
87604:127673134       Q-min: 1.836       Q-max: 1.872       Lives: 5       Reward: 7.0
87604:127673182       Q-min: 1.651       Q-max: 1.662       Lives: 5       Reward: 8.0
87604:127673246       Q-min: 1.671       Q-max: 1.690       Lives: 5       Reward: 9.0
87604:127673310       Q-min: 1.638       Q-max: 1.679       Lives: 5       Reward: 10.0
87604:127673378       Q-min: 1.585       Q-max: 1.661       Lives: 5       Reward: 11.0
87604:127673428       Q-min: 1.931       Q-max: 1.982       Lives: 5       Reward: 12.0
87604:127673461       Q-min: 1.894       Q-max: 1.993       Lives: 5       Reward: 16.0
87604:127673496       Q-min: 1.928       Q-max: 1.944       Lives: 5       Reward: 17.0
87604:127673529       Q-min: 2.050       Q-max: 2.108       Lives: 5       Reward: 18.0
87604:127673550       Q-min: -0.158       Q-max: 0.070       Lives: 4       Reward: 18.0
87604:127673605       Q-min: 1.720       Q-max: 1.787       Lives: 4       Reward: 19.0
87604:127673657       Q-min: 1.970       Q-max: 2.001       Lives: 4       Reward: 20.0
87604:127673701       Q-min: 1.980       Q-max: 2.030       Lives: 4       Reward: 21.0
87604:127673738       Q-min: 2.030       Q-max: 2.057       Lives: 4       Reward: 22.0
87604:127673774       Q-min: 2.097       Q-max: 2.124       Lives: 4       Reward: 23.0
87604:127673807       Q-min: 2.042       Q-max: 2.060       Lives: 4       Reward: 27.0
87604:127673841       Q-min: 2.120       Q-max: 2.154       Lives: 4       Reward: 31.0
87604:127673889       Q-min: 1.773       Q-max: 1.815       Lives: 4       Reward: 32.0
87604:127673952       Q-min: 1.786       Q-max: 1.805       Lives: 4       Reward: 33.0
87604:127674015       Q-min: 1.828       Q-max: 1.850       Lives: 4       Reward: 34.0
87604:127674080       Q-min: 1.758       Q-max: 1.826       Lives: 4       Reward: 35.0
87604:127674132       Q-min: 2.034       Q-max: 2.092       Lives: 4       Reward: 36.0
87604:127674166       Q-min: 2.050       Q-max: 2.077       Lives: 4       Reward: 37.0
87604:127674199       Q-min: 2.194       Q-max: 2.223       Lives: 4       Reward: 38.0
87604:127674237       Q-min: 2.105       Q-max: 2.220       Lives: 4       Reward: 42.0
87604:127674268       Q-min: 2.101       Q-max: 2.207       Lives: 4       Reward: 43.0
87604:127674302       Q-min: 2.200       Q-max: 2.251       Lives: 4       Reward: 44.0
87604:127674332       Q-min: 2.115       Q-max: 2.179       Lives: 4       Reward: 45.0
87604:127674364       Q-min: 2.114       Q-max: 2.247       Lives: 4       Reward: 49.0
87604:127674400       Q-min: 2.137       Q-max: 2.320       Lives: 4       Reward: 50.0
87604:127674435       Q-min: 2.079       Q-max: 2.216       Lives: 4       Reward: 54.0
87604:127674475       Q-min: 2.200       Q-max: 2.377       Lives: 4       Reward: 58.0
87604:127674499       Q-min: 2.350       Q-max: 2.537       Lives: 4       Reward: 62.0
```

```
87604:127674520       Q-min: 2.408       Q-max: 2.597       Lives: 4       Reward: 66.0
87604:127674543       Q-min: 2.412       Q-max: 2.525       Lives: 4       Reward: 67.0
87604:127674565       Q-min: 2.482       Q-max: 2.556       Lives: 4       Reward: 68.0
87604:127674585       Q-min: 2.394       Q-max: 2.539       Lives: 4       Reward: 72.0
87604:127674600       Q-min: 0.053       Q-max: 0.197       Lives: 3       Reward: 72.0
87604:127674656       Q-min: 1.584       Q-max: 1.884       Lives: 3       Reward: 76.0
87604:127674714       Q-min: 2.218       Q-max: 2.353       Lives: 3       Reward: 80.0
87604:127674759       Q-min: 2.189       Q-max: 2.243       Lives: 3       Reward: 81.0
87604:127674802       Q-min: 2.331       Q-max: 2.365       Lives: 3       Reward: 85.0
87604:127674843       Q-min: 2.299       Q-max: 2.737       Lives: 3       Reward: 92.0
87604:127674865       Q-min: 2.367       Q-max: 2.514       Lives: 3       Reward: 96.0
87604:127674889       Q-min: 2.248       Q-max: 2.649       Lives: 3       Reward: 103.0
87604:127674912       Q-min: 2.439       Q-max: 3.002       Lives: 3       Reward: 107.0
87604:127674938       Q-min: 2.280       Q-max: 2.987       Lives: 3       Reward: 114.0
87604:127674963       Q-min: 2.794       Q-max: 3.503       Lives: 3       Reward: 121.0
87604:127674991       Q-min: 2.500       Q-max: 7.708       Lives: 3       Reward: 128.0
87604:127674997       Q-min: 2.156       Q-max: 7.000       Lives: 3       Reward: 135.0
87604:127675002       Q-min: 2.477       Q-max: 6.028       Lives: 3       Reward: 142.0
87604:127675007       Q-min: 4.239       Q-max: 6.303       Lives: 3       Reward: 149.0
87604:127675011       Q-min: 3.918       Q-max: 7.023       Lives: 3       Reward: 156.0
87604:127675017       Q-min: 3.540       Q-max: 6.474       Lives: 3       Reward: 163.0
87604:127675023       Q-min: 1.202       Q-max: 6.183       Lives: 3       Reward: 170.0
87604:127675028       Q-min: 0.965       Q-max: 5.638       Lives: 3       Reward: 177.0
87604:127675035       Q-min: 2.349       Q-max: 4.229       Lives: 3       Reward: 181.0
87604:127675042       Q-min: 2.396       Q-max: 5.087       Lives: 3       Reward: 188.0
87604:127675048       Q-min: 3.254       Q-max: 4.975       Lives: 3       Reward: 195.0
87604:127675054       Q-min: 3.008       Q-max: 5.374       Lives: 3       Reward: 202.0
87604:127675060       Q-min: 3.376       Q-max: 5.494       Lives: 3       Reward: 209.0
87604:127675065       Q-min: 3.540       Q-max: 5.774       Lives: 3       Reward: 216.0
87604:127675070       Q-min: 3.060       Q-max: 4.781       Lives: 3       Reward: 223.0
87604:127675077       Q-min: 1.581       Q-max: 5.604       Lives: 3       Reward: 230.0
87604:127675081       Q-min: 1.432       Q-max: 4.240       Lives: 3       Reward: 237.0
87604:127675086       Q-min: 3.265       Q-max: 4.325       Lives: 3       Reward: 244.0
87604:127675092       Q-min: 2.793       Q-max: 4.364       Lives: 3       Reward: 251.0
87604:127675098       Q-min: 1.659       Q-max: 4.788       Lives: 3       Reward: 258.0
87604:127675104       Q-min: 2.185       Q-max: 3.692       Lives: 3       Reward: 265.0
87604:127675110       Q-min: 3.059       Q-max: 5.168       Lives: 3       Reward: 272.0
87604:127675116       Q-min: 2.764       Q-max: 3.735       Lives: 3       Reward: 279.0
87604:127675154       Q-min: 2.274       Q-max: 4.377       Lives: 3       Reward: 283.0
87604:127675162       Q-min: 2.393       Q-max: 5.473       Lives: 3       Reward: 287.0
87604:127675199       Q-min: 2.164       Q-max: 3.879       Lives: 3       Reward: 294.0
87604:127675206       Q-min: 1.621       Q-max: 4.900       Lives: 3       Reward: 301.0
87604:127675213       Q-min: 2.385       Q-max: 4.757       Lives: 3       Reward: 308.0
87604:127675220       Q-min: 3.074       Q-max: 4.138       Lives: 3       Reward: 315.0
87604:127675226       Q-min: 1.986       Q-max: 2.857       Lives: 3       Reward: 319.0
87604:127675250       Q-min: -0.037       Q-max: 0.340       Lives: 2       Reward: 319.0
87604:127675303       Q-min: 1.394       Q-max: 2.884       Lives: 2       Reward: 326.0
87604:127675327       Q-min: 1.153       Q-max: 2.299       Lives: 2       Reward: 330.0
```

```
87604:127675348       Q-min: 1.783       Q-max: 2.660       Lives: 2       Reward: 334.0
87604:127675371       Q-min: 2.439       Q-max: 3.126       Lives: 2       Reward: 338.0
87604:127675391       Q-min: 2.384       Q-max: 2.837       Lives: 2       Reward: 339.0
87604:127675412       Q-min: 1.705       Q-max: 4.095       Lives: 2       Reward: 343.0
87604:127675431       Q-min: 1.738       Q-max: 3.458       Lives: 2       Reward: 347.0
87604:127675462       Q-min: 2.482       Q-max: 4.338       Lives: 2       Reward: 354.0
87604:127675485       Q-min: 0.097       Q-max: 0.350       Lives: 1       Reward: 354.0
87604:127675583       Q-min: -0.169       Q-max: 0.298       Lives: 0       Reward: 354.0
87605:127675625       Q-min: 1.748       Q-max: 1.756       Lives: 5       Reward: 1.0
87605:127675666       Q-min: 1.814       Q-max: 1.842       Lives: 5       Reward: 2.0
87605:127675705       Q-min: 1.902       Q-max: 1.919       Lives: 5       Reward: 3.0
87605:127675743       Q-min: 1.980       Q-max: 2.007       Lives: 5       Reward: 4.0
87605:127675778       Q-min: 1.964       Q-max: 1.994       Lives: 5       Reward: 5.0
87605:127675810       Q-min: 1.965       Q-max: 2.003       Lives: 5       Reward: 6.0
87605:127675843       Q-min: 1.766       Q-max: 1.801       Lives: 5       Reward: 7.0
87605:127675891       Q-min: 1.605       Q-max: 1.707       Lives: 5       Reward: 8.0
87605:127675954       Q-min: 1.706       Q-max: 1.718       Lives: 5       Reward: 9.0
87605:127676023       Q-min: 1.680       Q-max: 1.760       Lives: 5       Reward: 10.0
87605:127676089       Q-min: 1.653       Q-max: 1.678       Lives: 5       Reward: 11.0
87605:127676138       Q-min: 1.868       Q-max: 2.061       Lives: 5       Reward: 12.0
87605:127676172       Q-min: 1.976       Q-max: 1.995       Lives: 5       Reward: 13.0
87605:127676193       Q-min: -0.059       Q-max: 0.295       Lives: 4       Reward: 13.0
87605:127676235       Q-min: 1.906       Q-max: 1.942       Lives: 4       Reward: 14.0
87605:127676288       Q-min: 1.763       Q-max: 1.829       Lives: 4       Reward: 15.0
87605:127676346       Q-min: 2.014       Q-max: 2.055       Lives: 4       Reward: 19.0
87605:127676385       Q-min: 2.041       Q-max: 2.080       Lives: 4       Reward: 20.0
87605:127676418       Q-min: 2.044       Q-max: 2.170       Lives: 4       Reward: 21.0
87605:127676451       Q-min: 2.075       Q-max: 2.162       Lives: 4       Reward: 25.0
87605:127676475       Q-min: -0.099       Q-max: 0.215       Lives: 3       Reward: 25.0
87605:127676530       Q-min: 1.829       Q-max: 1.890       Lives: 3       Reward: 29.0
87605:127676595       Q-min: 1.862       Q-max: 1.896       Lives: 3       Reward: 30.0
87605:127676649       Q-min: 1.995       Q-max: 2.130       Lives: 3       Reward: 31.0
87605:127676685       Q-min: 2.184       Q-max: 2.242       Lives: 3       Reward: 32.0
87605:127676718       Q-min: 2.100       Q-max: 2.182       Lives: 3       Reward: 33.0
87605:127676752       Q-min: 2.150       Q-max: 2.231       Lives: 3       Reward: 34.0
87605:127676785       Q-min: 2.192       Q-max: 2.229       Lives: 3       Reward: 35.0
87605:127676836       Q-min: 1.745       Q-max: 1.770       Lives: 3       Reward: 36.0
87605:127676902       Q-min: 1.591       Q-max: 1.756       Lives: 3       Reward: 37.0
87605:127676973       Q-min: 1.822       Q-max: 1.921       Lives: 3       Reward: 41.0
87605:127677046       Q-min: 1.600       Q-max: 2.286       Lives: 3       Reward: 45.0
87605:127677099       Q-min: 2.053       Q-max: 2.102       Lives: 3       Reward: 46.0
87605:127677134       Q-min: 2.109       Q-max: 2.140       Lives: 3       Reward: 47.0
87605:127677167       Q-min: 2.327       Q-max: 2.429       Lives: 3       Reward: 51.0
87605:127677189       Q-min: -0.038       Q-max: 0.213       Lives: 2       Reward: 51.0
87605:127677235       Q-min: 2.000       Q-max: 2.037       Lives: 2       Reward: 52.0
87605:127677280       Q-min: 2.027       Q-max: 2.063       Lives: 2       Reward: 53.0
87605:127677324       Q-min: 2.117       Q-max: 2.134       Lives: 2       Reward: 54.0
87605:127677362       Q-min: 2.116       Q-max: 2.305       Lives: 2       Reward: 58.0
```

```
87605:127677399       Q-min: 2.268       Q-max: 2.319       Lives: 2       Reward: 62.0
87605:127677433       Q-min: 2.207       Q-max: 2.433       Lives: 2       Reward: 66.0
87605:127677453       Q-min: 2.363       Q-max: 2.507       Lives: 2       Reward: 67.0
87605:127677474       Q-min: 2.441       Q-max: 2.623       Lives: 2       Reward: 71.0
87605:127677487       Q-min: 0.035       Q-max: 0.164       Lives: 1       Reward: 71.0
87605:127677533       Q-min: 2.462       Q-max: 2.580       Lives: 1       Reward: 75.0
87605:127677558       Q-min: 2.145       Q-max: 2.673       Lives: 1       Reward: 82.0
87605:127677581       Q-min: 2.687       Q-max: 2.993       Lives: 1       Reward: 86.0
87605:127677602       Q-min: 2.374       Q-max: 2.936       Lives: 1       Reward: 90.0
87605:127677625       Q-min: 2.129       Q-max: 2.879       Lives: 1       Reward: 97.0
87605:127677647       Q-min: 2.395       Q-max: 2.794       Lives: 1       Reward: 101.0
87605:127677663       Q-min: 0.015       Q-max: 0.272       Lives: 0       Reward: 101.0
87606:127677716       Q-min: 1.678       Q-max: 1.684       Lives: 5       Reward: 1.0
87606:127677767       Q-min: 1.834       Q-max: 1.844       Lives: 5       Reward: 2.0
87606:127677810       Q-min: 1.907       Q-max: 1.941       Lives: 5       Reward: 3.0
87606:127677848       Q-min: 1.949       Q-max: 2.005       Lives: 5       Reward: 4.0
87606:127677881       Q-min: 1.976       Q-max: 2.021       Lives: 5       Reward: 5.0
87606:127677913       Q-min: 1.949       Q-max: 1.972       Lives: 5       Reward: 6.0
87606:127677945       Q-min: 1.789       Q-max: 1.822       Lives: 5       Reward: 7.0
87606:127677991       Q-min: 1.646       Q-max: 1.659       Lives: 5       Reward: 8.0
87606:127678056       Q-min: 1.690       Q-max: 1.730       Lives: 5       Reward: 9.0
87606:127678100       Q-min: -0.085       Q-max: 0.158       Lives: 4       Reward: 9.0
87606:127678143       Q-min: 1.937       Q-max: 1.976       Lives: 4       Reward: 10.0
87606:127678186       Q-min: 1.940       Q-max: 1.950       Lives: 4       Reward: 11.0
87606:127678243       Q-min: 1.645       Q-max: 1.722       Lives: 4       Reward: 12.0
87606:127678289       Q-min: 2.028       Q-max: 2.049       Lives: 4       Reward: 13.0
87606:127678321       Q-min: 1.960       Q-max: 1.993       Lives: 4       Reward: 14.0
87606:127678357       Q-min: 2.025       Q-max: 2.069       Lives: 4       Reward: 18.0
87606:127678393       Q-min: 1.958       Q-max: 2.482       Lives: 4       Reward: 22.0
87606:127678413       Q-min: 2.310       Q-max: 2.430       Lives: 4       Reward: 23.0
87606:127678436       Q-min: 2.148       Q-max: 2.430       Lives: 4       Reward: 27.0
87606:127678449       Q-min: -0.273       Q-max: 0.286       Lives: 3       Reward: 27.0
87606:127678495       Q-min: 1.984       Q-max: 2.000       Lives: 3       Reward: 28.0
87606:127678538       Q-min: 2.054       Q-max: 2.119       Lives: 3       Reward: 29.0
87606:127678592       Q-min: 1.758       Q-max: 1.815       Lives: 3       Reward: 30.0
87606:127678635       Q-min: -0.036       Q-max: 0.414       Lives: 2       Reward: 30.0
87606:127678679       Q-min: 1.944       Q-max: 1.981       Lives: 2       Reward: 31.0
87606:127678735       Q-min: 1.654       Q-max: 1.956       Lives: 2       Reward: 32.0
87606:127678793       Q-min: 1.934       Q-max: 1.967       Lives: 2       Reward: 33.0
87606:127678831       Q-min: 2.178       Q-max: 2.211       Lives: 2       Reward: 34.0
87606:127678865       Q-min: 2.051       Q-max: 2.127       Lives: 2       Reward: 35.0
87606:127678900       Q-min: 2.133       Q-max: 2.280       Lives: 2       Reward: 36.0
87606:127678933       Q-min: 2.284       Q-max: 2.330       Lives: 2       Reward: 37.0
87606:127678983       Q-min: 1.702       Q-max: 1.807       Lives: 2       Reward: 38.0
87606:127679049       Q-min: 1.585       Q-max: 1.716       Lives: 2       Reward: 42.0
87606:127679097       Q-min: -0.064       Q-max: 0.265       Lives: 1       Reward: 42.0
87606:127679155       Q-min: 1.777       Q-max: 1.864       Lives: 1       Reward: 46.0
87606:127679215       Q-min: 2.128       Q-max: 2.154       Lives: 1       Reward: 47.0
```

```
87606:127679261       Q-min: 2.095       Q-max: 2.386       Lives: 1       Reward: 51.0
87606:127679301       Q-min: 2.240       Q-max: 2.281       Lives: 1       Reward: 52.0
87606:127679336       Q-min: 2.182       Q-max: 2.221       Lives: 1       Reward: 53.0
87606:127679371       Q-min: 2.146       Q-max: 2.200       Lives: 1       Reward: 57.0
87606:127679403       Q-min: 2.166       Q-max: 2.235       Lives: 1       Reward: 58.0
87606:127679452       Q-min: 1.818       Q-max: 1.949       Lives: 1       Reward: 59.0
87606:127679521       Q-min: 1.925       Q-max: 2.035       Lives: 1       Reward: 60.0
87606:127679585       Q-min: 1.918       Q-max: 2.098       Lives: 1       Reward: 64.0
87606:127679662       Q-min: 1.814       Q-max: 2.746       Lives: 1       Reward: 68.0
87606:127679687       Q-min: 2.350       Q-max: 2.740       Lives: 1       Reward: 72.0
87606:127679702       Q-min: -0.177       Q-max: 0.151       Lives: 0       Reward: 72.0
87607:127679743       Q-min: 1.756       Q-max: 1.772       Lives: 5       Reward: 1.0
87607:127679792       Q-min: 1.627       Q-max: 1.654       Lives: 5       Reward: 2.0
87607:127679845       Q-min: 1.796       Q-max: 1.842       Lives: 5       Reward: 3.0
87607:127679882       Q-min: 1.859       Q-max: 1.878       Lives: 5       Reward: 4.0
87607:127679915       Q-min: 1.970       Q-max: 2.004       Lives: 5       Reward: 8.0
87607:127679949       Q-min: 1.986       Q-max: 2.019       Lives: 5       Reward: 9.0
87607:127679978       Q-min: 1.781       Q-max: 1.889       Lives: 5       Reward: 10.0
87607:127680025       Q-min: 1.672       Q-max: 1.699       Lives: 5       Reward: 11.0
87607:127680088       Q-min: 1.748       Q-max: 1.833       Lives: 5       Reward: 12.0
87607:127680154       Q-min: 1.716       Q-max: 1.752       Lives: 5       Reward: 13.0
87607:127680222       Q-min: 1.707       Q-max: 1.769       Lives: 5       Reward: 14.0
87607:127680274       Q-min: 1.999       Q-max: 2.047       Lives: 5       Reward: 15.0
87607:127680306       Q-min: 2.011       Q-max: 2.029       Lives: 5       Reward: 16.0
87607:127680328       Q-min: -0.113       Q-max: 0.253       Lives: 4       Reward: 16.0
87607:127680375       Q-min: 1.915       Q-max: 1.935       Lives: 4       Reward: 17.0
87607:127680421       Q-min: 1.981       Q-max: 2.003       Lives: 4       Reward: 18.0
87607:127680463       Q-min: 1.955       Q-max: 1.961       Lives: 4       Reward: 19.0
87607:127680500       Q-min: 2.005       Q-max: 2.065       Lives: 4       Reward: 20.0
87607:127680533       Q-min: 1.985       Q-max: 1.999       Lives: 4       Reward: 21.0
87607:127680565       Q-min: 2.096       Q-max: 2.145       Lives: 4       Reward: 22.0
87607:127680599       Q-min: 2.082       Q-max: 2.113       Lives: 4       Reward: 23.0
87607:127680649       Q-min: 1.753       Q-max: 1.813       Lives: 4       Reward: 24.0
87607:127680714       Q-min: 1.711       Q-max: 1.752       Lives: 4       Reward: 25.0
87607:127680781       Q-min: 1.694       Q-max: 1.736       Lives: 4       Reward: 26.0
87607:127680848       Q-min: 1.669       Q-max: 1.821       Lives: 4       Reward: 30.0
87607:127680899       Q-min: 2.096       Q-max: 2.222       Lives: 4       Reward: 34.0
87607:127680931       Q-min: 2.128       Q-max: 2.170       Lives: 4       Reward: 35.0
87607:127680961       Q-min: 2.139       Q-max: 2.193       Lives: 4       Reward: 36.0
87607:127680996       Q-min: 2.079       Q-max: 2.133       Lives: 4       Reward: 37.0
87607:127681031       Q-min: 2.045       Q-max: 2.180       Lives: 4       Reward: 38.0
87607:127681064       Q-min: 2.042       Q-max: 2.128       Lives: 4       Reward: 42.0
87607:127681098       Q-min: 2.222       Q-max: 2.232       Lives: 4       Reward: 43.0
87607:127681135       Q-min: 2.349       Q-max: 2.475       Lives: 4       Reward: 47.0
87607:127681156       Q-min: 2.286       Q-max: 2.440       Lives: 4       Reward: 48.0
87607:127681171       Q-min: 0.047       Q-max: 0.248       Lives: 3       Reward: 48.0
87607:127681217       Q-min: 2.070       Q-max: 2.095       Lives: 3       Reward: 49.0
87607:127681276       Q-min: 1.874       Q-max: 2.028       Lives: 3       Reward: 53.0
```

```
87607:127681334       Q-min: 2.101       Q-max: 2.147       Lives: 3       Reward: 54.0
87607:127681371       Q-min: 2.116       Q-max: 2.239       Lives: 3       Reward: 55.0
87607:127681408       Q-min: 2.468       Q-max: 2.547       Lives: 3       Reward: 59.0
87607:127681431       Q-min: 2.280       Q-max: 2.690       Lives: 3       Reward: 63.0
87607:127681453       Q-min: 2.054       Q-max: 2.563       Lives: 3       Reward: 67.0
87607:127681466       Q-min: -0.662       Q-max: 0.099       Lives: 2       Reward: 67.0
87607:127681513       Q-min: 2.119       Q-max: 2.174       Lives: 2       Reward: 71.0
87607:127681557       Q-min: 2.152       Q-max: 2.228       Lives: 2       Reward: 72.0
87607:127681615       Q-min: 2.401       Q-max: 2.595       Lives: 2       Reward: 76.0
87607:127681628       Q-min: 0.216       Q-max: 0.360       Lives: 1       Reward: 76.0
87607:127681673       Q-min: 2.201       Q-max: 2.312       Lives: 1       Reward: 80.0
87607:127681716       Q-min: 2.240       Q-max: 2.314       Lives: 1       Reward: 81.0
87607:127681758       Q-min: 2.284       Q-max: 2.382       Lives: 1       Reward: 82.0
87607:127681799       Q-min: 2.332       Q-max: 2.451       Lives: 1       Reward: 86.0
87607:127681835       Q-min: 2.389       Q-max: 2.679       Lives: 1       Reward: 90.0
87607:127681848       Q-min: -0.115       Q-max: 0.157       Lives: 0       Reward: 90.0
87608:127681894       Q-min: 1.754       Q-max: 1.766       Lives: 5       Reward: 1.0
87608:127681946       Q-min: 1.615       Q-max: 1.667       Lives: 5       Reward: 2.0
87608:127681999       Q-min: 1.888       Q-max: 1.900       Lives: 5       Reward: 3.0
87608:127682034       Q-min: 1.989       Q-max: 2.028       Lives: 5       Reward: 4.0
87608:127682066       Q-min: 1.922       Q-max: 1.962       Lives: 5       Reward: 5.0
87608:127682098       Q-min: 1.921       Q-max: 1.948       Lives: 5       Reward: 6.0
87608:127682132       Q-min: 1.779       Q-max: 1.816       Lives: 5       Reward: 7.0
87608:127682184       Q-min: 1.663       Q-max: 1.689       Lives: 5       Reward: 8.0
87608:127682253       Q-min: 1.671       Q-max: 1.748       Lives: 5       Reward: 9.0
87608:127682311       Q-min: 1.656       Q-max: 1.672       Lives: 5       Reward: 10.0
87608:127682374       Q-min: 1.611       Q-max: 1.666       Lives: 5       Reward: 11.0
87608:127682420       Q-min: 1.993       Q-max: 2.024       Lives: 5       Reward: 12.0
87608:127682452       Q-min: 1.965       Q-max: 2.005       Lives: 5       Reward: 13.0
87608:127682486       Q-min: 1.916       Q-max: 2.050       Lives: 5       Reward: 14.0
87608:127682517       Q-min: 1.985       Q-max: 2.014       Lives: 5       Reward: 15.0
87608:127682549       Q-min: 1.885       Q-max: 1.936       Lives: 5       Reward: 19.0
87608:127682585       Q-min: 1.965       Q-max: 1.990       Lives: 5       Reward: 23.0
87608:127682620       Q-min: 2.079       Q-max: 2.127       Lives: 5       Reward: 24.0
87608:127682653       Q-min: 2.049       Q-max: 2.104       Lives: 5       Reward: 25.0
87608:127682685       Q-min: 2.052       Q-max: 2.088       Lives: 5       Reward: 26.0
87608:127682717       Q-min: 2.025       Q-max: 2.052       Lives: 5       Reward: 27.0
87608:127682748       Q-min: 2.003       Q-max: 2.024       Lives: 5       Reward: 28.0
87608:127682782       Q-min: 2.108       Q-max: 2.209       Lives: 5       Reward: 32.0
87608:127682818       Q-min: 2.094       Q-max: 2.108       Lives: 5       Reward: 33.0
87608:127682849       Q-min: 2.079       Q-max: 2.142       Lives: 5       Reward: 34.0
87608:127682881       Q-min: 2.000       Q-max: 2.105       Lives: 5       Reward: 35.0
87608:127682916       Q-min: 2.170       Q-max: 2.207       Lives: 5       Reward: 36.0
87608:127682950       Q-min: 2.288       Q-max: 2.493       Lives: 5       Reward: 40.0
87608:127682972       Q-min: 2.439       Q-max: 2.481       Lives: 5       Reward: 41.0
87608:127682990       Q-min: 2.269       Q-max: 2.471       Lives: 5       Reward: 42.0
87608:127683012       Q-min: 2.143       Q-max: 2.405       Lives: 5       Reward: 46.0
87608:127683033       Q-min: 2.332       Q-max: 2.442       Lives: 5       Reward: 50.0
```

```
87608:127683056       Q-min: 2.458       Q-max: 2.539       Lives: 5       Reward: 54.0
87608:127683078       Q-min: 1.800       Q-max: 2.672       Lives: 5       Reward: 58.0
87608:127683092       Q-min: -0.059       Q-max: 0.069       Lives: 4       Reward: 58.0
87608:127683140       Q-min: 2.305       Q-max: 2.477       Lives: 4       Reward: 62.0
87608:127683162       Q-min: 2.416       Q-max: 2.487       Lives: 4       Reward: 63.0
87608:127683176       Q-min: -0.108       Q-max: 0.026       Lives: 3       Reward: 63.0
87608:127683224       Q-min: 2.500       Q-max: 2.744       Lives: 3       Reward: 70.0
87608:127683244       Q-min: 2.376       Q-max: 2.586       Lives: 3       Reward: 71.0
87608:127683258       Q-min: -0.027       Q-max: 0.220       Lives: 2       Reward: 71.0
87608:127683306       Q-min: 2.301       Q-max: 2.759       Lives: 2       Reward: 78.0
87608:127683330       Q-min: 2.334       Q-max: 2.649       Lives: 2       Reward: 82.0
87608:127683350       Q-min: 2.408       Q-max: 2.508       Lives: 2       Reward: 83.0
87608:127683362       Q-min: 0.001       Q-max: 0.376       Lives: 1       Reward: 83.0
87608:127683416       Q-min: 2.042       Q-max: 2.116       Lives: 1       Reward: 84.0
87608:127683469       Q-min: 2.351       Q-max: 2.478       Lives: 1       Reward: 85.0
87608:127683513       Q-min: 2.162       Q-max: 2.434       Lives: 1       Reward: 89.0
87608:127683553       Q-min: 2.296       Q-max: 2.536       Lives: 1       Reward: 93.0
87608:127683588       Q-min: 2.540       Q-max: 2.689       Lives: 1       Reward: 97.0
87608:127683622       Q-min: 2.363       Q-max: 2.576       Lives: 1       Reward: 98.0
87608:127683651       Q-min: 2.445       Q-max: 2.640       Lives: 1       Reward: 99.0
87608:127683703       Q-min: 2.028       Q-max: 2.240       Lives: 1       Reward: 103.0
87608:127683779       Q-min: 2.230       Q-max: 2.499       Lives: 1       Reward: 107.0
87608:127683858       Q-min: 2.058       Q-max: 3.091       Lives: 1       Reward: 111.0
87608:127683881       Q-min: 2.198       Q-max: 3.193       Lives: 1       Reward: 115.0
87608:127683903       Q-min: 2.595       Q-max: 3.421       Lives: 1       Reward: 122.0
87608:127683925       Q-min: 2.155       Q-max: 3.768       Lives: 1       Reward: 129.0
87608:127683954       Q-min: 2.611       Q-max: 5.506       Lives: 1       Reward: 136.0
87608:127683959       Q-min: 2.963       Q-max: 5.873       Lives: 1       Reward: 143.0
87608:127683965       Q-min: 2.843       Q-max: 5.788       Lives: 1       Reward: 150.0
87608:127683994       Q-min: 2.415       Q-max: 4.049       Lives: 1       Reward: 154.0
87608:127684008       Q-min: -0.035       Q-max: 0.309       Lives: 0       Reward: 154.0
87609:127684063       Q-min: 1.627       Q-max: 1.654       Lives: 5       Reward: 1.0
87609:127684114       Q-min: 1.841       Q-max: 1.865       Lives: 5       Reward: 2.0
87609:127684168       Q-min: 1.662       Q-max: 1.703       Lives: 5       Reward: 3.0
87609:127684216       Q-min: 2.002       Q-max: 2.020       Lives: 5       Reward: 4.0
87609:127684247       Q-min: 1.967       Q-max: 1.995       Lives: 5       Reward: 5.0
87609:127684279       Q-min: 1.965       Q-max: 1.986       Lives: 5       Reward: 6.0
87609:127684314       Q-min: 1.795       Q-max: 1.854       Lives: 5       Reward: 10.0
87609:127684360       Q-min: 1.695       Q-max: 1.724       Lives: 5       Reward: 11.0
87609:127684423       Q-min: 1.746       Q-max: 1.772       Lives: 5       Reward: 12.0
87609:127684484       Q-min: 1.606       Q-max: 1.772       Lives: 5       Reward: 13.0
87609:127684549       Q-min: 1.759       Q-max: 1.908       Lives: 5       Reward: 14.0
87609:127684595       Q-min: 2.024       Q-max: 2.086       Lives: 5       Reward: 15.0
87609:127684628       Q-min: 1.998       Q-max: 2.026       Lives: 5       Reward: 16.0
87609:127684661       Q-min: 1.952       Q-max: 2.014       Lives: 5       Reward: 17.0
87609:127684694       Q-min: 2.034       Q-max: 2.061       Lives: 5       Reward: 18.0
87609:127684726       Q-min: 2.015       Q-max: 2.051       Lives: 5       Reward: 19.0
87609:127684754       Q-min: 2.143       Q-max: 2.173       Lives: 5       Reward: 20.0
```

```
87609:127684788       Q-min: 1.963       Q-max: 2.037       Lives: 5       Reward: 21.0
87609:127684823       Q-min: 2.093       Q-max: 2.138       Lives: 5       Reward: 22.0
87609:127684846       Q-min: -0.186       Q-max: 0.232       Lives: 4       Reward: 22.0
87609:127684889       Q-min: 1.932       Q-max: 1.958       Lives: 4       Reward: 23.0
87609:127684930       Q-min: 2.099       Q-max: 2.142       Lives: 4       Reward: 24.0
87609:127684972       Q-min: 1.992       Q-max: 2.033       Lives: 4       Reward: 25.0
87609:127685009       Q-min: 2.075       Q-max: 2.135       Lives: 4       Reward: 29.0
87609:127685033       Q-min: 0.000       Q-max: 0.368       Lives: 3       Reward: 29.0
87609:127685077       Q-min: 2.295       Q-max: 2.514       Lives: 3       Reward: 33.0
87609:127685100       Q-min: 2.149       Q-max: 2.658       Lives: 3       Reward: 40.0
87609:127685123       Q-min: 1.812       Q-max: 3.201       Lives: 3       Reward: 47.0
87609:127685140       Q-min: -0.555       Q-max: 0.613       Lives: 2       Reward: 47.0
87609:127685185       Q-min: 2.128       Q-max: 2.161       Lives: 2       Reward: 48.0
87609:127685233       Q-min: 1.913       Q-max: 2.194       Lives: 2       Reward: 52.0
87609:127685280       Q-min: 2.126       Q-max: 2.274       Lives: 2       Reward: 53.0
87609:127685319       Q-min: 2.312       Q-max: 2.400       Lives: 2       Reward: 57.0
87609:127685357       Q-min: 2.606       Q-max: 2.748       Lives: 2       Reward: 61.0
87609:127685377       Q-min: 2.519       Q-max: 3.619       Lives: 2       Reward: 62.0
87609:127685389       Q-min: 0.083       Q-max: 0.184       Lives: 1       Reward: 62.0
87609:127685436       Q-min: 2.103       Q-max: 2.290       Lives: 1       Reward: 63.0
87609:127685494       Q-min: 1.586       Q-max: 3.198       Lives: 1       Reward: 70.0
87609:127685515       Q-min: 2.646       Q-max: 2.753       Lives: 1       Reward: 71.0
87609:127685529       Q-min: -0.298       Q-max: 0.219       Lives: 0       Reward: 71.0
87610:127685572       Q-min: 1.771       Q-max: 1.787       Lives: 5       Reward: 1.0
87610:127685612       Q-min: 1.795       Q-max: 1.833       Lives: 5       Reward: 2.0
87610:127685653       Q-min: 1.886       Q-max: 1.906       Lives: 5       Reward: 3.0
87610:127685684       Q-min: -0.097       Q-max: 0.119       Lives: 4       Reward: 3.0
87610:127685727       Q-min: 1.869       Q-max: 1.902       Lives: 4       Reward: 4.0
87610:127685773       Q-min: 1.935       Q-max: 1.952       Lives: 4       Reward: 8.0
87610:127685826       Q-min: 1.682       Q-max: 1.699       Lives: 4       Reward: 9.0
87610:127685876       Q-min: 1.935       Q-max: 1.961       Lives: 4       Reward: 10.0
87610:127685910       Q-min: 1.936       Q-max: 1.984       Lives: 4       Reward: 11.0
87610:127685941       Q-min: 1.960       Q-max: 1.993       Lives: 4       Reward: 12.0
87610:127685973       Q-min: 1.965       Q-max: 2.007       Lives: 4       Reward: 13.0
87610:127685996       Q-min: -0.274       Q-max: 0.102       Lives: 3       Reward: 13.0
87610:127686040       Q-min: 1.890       Q-max: 1.914       Lives: 3       Reward: 14.0
87610:127686082       Q-min: 1.975       Q-max: 2.052       Lives: 3       Reward: 15.0
87610:127686126       Q-min: 1.971       Q-max: 2.005       Lives: 3       Reward: 16.0
87610:127686165       Q-min: 2.055       Q-max: 2.075       Lives: 3       Reward: 17.0
87610:127686201       Q-min: 1.981       Q-max: 2.010       Lives: 3       Reward: 21.0
87610:127686236       Q-min: 1.987       Q-max: 2.014       Lives: 3       Reward: 22.0
87610:127686270       Q-min: 2.048       Q-max: 2.107       Lives: 3       Reward: 26.0
87610:127686319       Q-min: 1.779       Q-max: 1.806       Lives: 3       Reward: 27.0
87610:127686378       Q-min: 1.763       Q-max: 1.814       Lives: 3       Reward: 28.0
87610:127686441       Q-min: 1.784       Q-max: 1.854       Lives: 3       Reward: 29.0
87610:127686505       Q-min: 1.767       Q-max: 1.807       Lives: 3       Reward: 30.0
87610:127686552       Q-min: 2.130       Q-max: 2.151       Lives: 3       Reward: 31.0
87610:127686586       Q-min: 2.020       Q-max: 2.132       Lives: 3       Reward: 35.0
```

```
87610:127686619       Q-min: 2.229       Q-max: 2.302       Lives: 3       Reward: 39.0
87610:127686653       Q-min: 2.146       Q-max: 2.168       Lives: 3       Reward: 40.0
87610:127686688       Q-min: 2.089       Q-max: 2.103       Lives: 3       Reward: 41.0
87610:127686722       Q-min: 2.044       Q-max: 2.106       Lives: 3       Reward: 42.0
87610:127686758       Q-min: 2.272       Q-max: 2.601       Lives: 3       Reward: 46.0
87610:127686778       Q-min: 2.339       Q-max: 2.568       Lives: 3       Reward: 47.0
87610:127686793       Q-min: -0.078       Q-max: 0.390       Lives: 2       Reward: 47.0
87610:127686843       Q-min: 2.456       Q-max: 2.751       Lives: 2       Reward: 54.0
87610:127686864       Q-min: 2.580       Q-max: 2.685       Lives: 2       Reward: 55.0
87610:127686883       Q-min: 2.495       Q-max: 2.593       Lives: 2       Reward: 56.0
87610:127686902       Q-min: 2.405       Q-max: 2.642       Lives: 2       Reward: 57.0
87610:127686922       Q-min: 2.280       Q-max: 2.503       Lives: 2       Reward: 58.0
87610:127686941       Q-min: 2.468       Q-max: 2.574       Lives: 2       Reward: 62.0
87610:127686962       Q-min: 2.408       Q-max: 2.591       Lives: 2       Reward: 66.0
87610:127686985       Q-min: 2.310       Q-max: 2.628       Lives: 2       Reward: 70.0
87610:127687008       Q-min: 2.512       Q-max: 2.766       Lives: 2       Reward: 74.0
87610:127687030       Q-min: 2.183       Q-max: 2.847       Lives: 2       Reward: 78.0
87610:127687043       Q-min: 0.031       Q-max: 0.250       Lives: 1       Reward: 78.0
87610:127687082       Q-min: 2.263       Q-max: 2.332       Lives: 1       Reward: 79.0
87610:127687134       Q-min: 2.101       Q-max: 2.186       Lives: 1       Reward: 80.0
87610:127687203       Q-min: 2.399       Q-max: 2.755       Lives: 1       Reward: 84.0
87610:127687217       Q-min: 0.028       Q-max: 0.155       Lives: 0       Reward: 84.0
87611:127687262       Q-min: 1.731       Q-max: 1.745       Lives: 5       Reward: 1.0
87611:127687313       Q-min: 1.651       Q-max: 1.690       Lives: 5       Reward: 2.0
87611:127687367       Q-min: 1.864       Q-max: 1.887       Lives: 5       Reward: 3.0
87611:127687403       Q-min: 1.930       Q-max: 1.957       Lives: 5       Reward: 4.0
87611:127687433       Q-min: 1.943       Q-max: 1.969       Lives: 5       Reward: 5.0
87611:127687467       Q-min: 1.962       Q-max: 1.975       Lives: 5       Reward: 6.0
87611:127687500       Q-min: 1.687       Q-max: 1.752       Lives: 5       Reward: 7.0
87611:127687549       Q-min: 1.655       Q-max: 1.674       Lives: 5       Reward: 8.0
87611:127687613       Q-min: 1.662       Q-max: 1.678       Lives: 5       Reward: 9.0
87611:127687651       Q-min: -0.105       Q-max: 0.128       Lives: 4       Reward: 9.0
87611:127687705       Q-min: 1.583       Q-max: 1.631       Lives: 4       Reward: 10.0
87611:127687773       Q-min: 1.727       Q-max: 1.743       Lives: 4       Reward: 11.0
87611:127687828       Q-min: 1.937       Q-max: 1.957       Lives: 4       Reward: 12.0
87611:127687864       Q-min: 1.992       Q-max: 2.010       Lives: 4       Reward: 13.0
87611:127687897       Q-min: 1.985       Q-max: 2.009       Lives: 4       Reward: 14.0
87611:127687937       Q-min: 2.044       Q-max: 2.064       Lives: 4       Reward: 18.0
87611:127687970       Q-min: 2.033       Q-max: 2.052       Lives: 4       Reward: 19.0
87611:127688019       Q-min: 1.712       Q-max: 1.740       Lives: 4       Reward: 20.0
87611:127688085       Q-min: 1.721       Q-max: 1.762       Lives: 4       Reward: 21.0
87611:127688149       Q-min: 1.650       Q-max: 1.725       Lives: 4       Reward: 22.0
87611:127688214       Q-min: 1.663       Q-max: 1.756       Lives: 4       Reward: 23.0
87611:127688263       Q-min: 2.057       Q-max: 2.084       Lives: 4       Reward: 24.0
87611:127688299       Q-min: 1.769       Q-max: 1.903       Lives: 4       Reward: 28.0
87611:127688335       Q-min: 2.005       Q-max: 2.049       Lives: 4       Reward: 29.0
87611:127688367       Q-min: 2.091       Q-max: 2.151       Lives: 4       Reward: 30.0
87611:127688399       Q-min: 2.170       Q-max: 2.212       Lives: 4       Reward: 31.0
```

```
87611:127688431       Q-min: 2.025       Q-max: 2.069       Lives: 4       Reward: 32.0
87611:127688465       Q-min: 1.990       Q-max: 2.026       Lives: 4       Reward: 33.0
87611:127688500       Q-min: 2.126       Q-max: 2.194       Lives: 4       Reward: 37.0
87611:127688519       Q-min: 0.121       Q-max: 0.461       Lives: 3       Reward: 37.0
87611:127688552       Q-min: 0.028       Q-max: 0.190       Lives: 2       Reward: 37.0
87611:127688599       Q-min: 1.996       Q-max: 2.115       Lives: 2       Reward: 41.0
87611:127688647       Q-min: 2.269       Q-max: 2.482       Lives: 2       Reward: 45.0
87611:127688661       Q-min: -0.118       Q-max: 0.097       Lives: 1       Reward: 45.0
87611:127688708       Q-min: 2.254       Q-max: 2.884       Lives: 1       Reward: 52.0
87611:127688732       Q-min: 2.380       Q-max: 2.510       Lives: 1       Reward: 53.0
87611:127688751       Q-min: 2.509       Q-max: 2.535       Lives: 1       Reward: 54.0
87611:127688773       Q-min: 2.214       Q-max: 2.313       Lives: 1       Reward: 55.0
87611:127688784       Q-min: -0.137       Q-max: -0.064       Lives: 0       Reward: 55.0
87612:127688825       Q-min: 1.743       Q-max: 1.756       Lives: 5       Reward: 1.0
87612:127688868       Q-min: 1.769       Q-max: 1.795       Lives: 5       Reward: 2.0
87612:127688922       Q-min: 1.657       Q-max: 1.705       Lives: 5       Reward: 3.0
87612:127688973       Q-min: 1.955       Q-max: 1.978       Lives: 5       Reward: 4.0
87612:127689006       Q-min: 1.986       Q-max: 2.008       Lives: 5       Reward: 5.0
87612:127689035       Q-min: 1.966       Q-max: 1.974       Lives: 5       Reward: 6.0
87612:127689066       Q-min: 1.749       Q-max: 1.813       Lives: 5       Reward: 7.0
87612:127689111       Q-min: 1.692       Q-max: 1.708       Lives: 5       Reward: 8.0
87612:127689176       Q-min: 1.702       Q-max: 1.741       Lives: 5       Reward: 9.0
87612:127689241       Q-min: 1.659       Q-max: 1.676       Lives: 5       Reward: 10.0
87612:127689302       Q-min: 1.630       Q-max: 1.654       Lives: 5       Reward: 11.0
87612:127689355       Q-min: 1.973       Q-max: 2.007       Lives: 5       Reward: 12.0
87612:127689378       Q-min: -0.112       Q-max: 0.204       Lives: 4       Reward: 12.0
87612:127689419       Q-min: 1.939       Q-max: 1.980       Lives: 4       Reward: 13.0
87612:127689465       Q-min: 1.912       Q-max: 1.950       Lives: 4       Reward: 14.0
87612:127689522       Q-min: 1.673       Q-max: 1.718       Lives: 4       Reward: 15.0
87612:127689570       Q-min: 1.962       Q-max: 1.988       Lives: 4       Reward: 16.0
87612:127689603       Q-min: 2.014       Q-max: 2.075       Lives: 4       Reward: 20.0
87612:127689627       Q-min: -0.091       Q-max: 0.199       Lives: 3       Reward: 20.0
87612:127689682       Q-min: 1.711       Q-max: 1.726       Lives: 3       Reward: 21.0
87612:127689724       Q-min: -0.040       Q-max: 0.209       Lives: 2       Reward: 21.0
87612:127689767       Q-min: 1.848       Q-max: 1.880       Lives: 2       Reward: 22.0
87612:127689811       Q-min: 1.961       Q-max: 1.978       Lives: 2       Reward: 23.0
87612:127689853       Q-min: 1.906       Q-max: 1.925       Lives: 2       Reward: 24.0
87612:127689894       Q-min: 2.037       Q-max: 2.090       Lives: 2       Reward: 28.0
87612:127689933       Q-min: 2.121       Q-max: 2.210       Lives: 2       Reward: 32.0
87612:127689973       Q-min: 2.198       Q-max: 2.401       Lives: 2       Reward: 36.0
87612:127689995       Q-min: 2.288       Q-max: 2.436       Lives: 2       Reward: 37.0
87612:127690009       Q-min: -0.031       Q-max: 0.069       Lives: 1       Reward: 37.0
87612:127690057       Q-min: 2.013       Q-max: 2.063       Lives: 1       Reward: 38.0
87612:127690116       Q-min: 1.884       Q-max: 1.909       Lives: 1       Reward: 39.0
87612:127690174       Q-min: 2.133       Q-max: 2.171       Lives: 1       Reward: 40.0
87612:127690200       Q-min: -0.037       Q-max: 0.172       Lives: 0       Reward: 40.0
87613:127690252       Q-min: 1.621       Q-max: 1.693       Lives: 5       Reward: 1.0
87613:127690315       Q-min: 1.643       Q-max: 1.659       Lives: 5       Reward: 2.0
```

```
87613:127690366    Q-min: 1.864    Q-max: 1.898    Lives: 5    Reward: 3.0
87613:127690401    Q-min: 1.925    Q-max: 1.938    Lives: 5    Reward: 4.0
87613:127690435    Q-min: 1.933    Q-max: 1.949    Lives: 5    Reward: 5.0
87613:127690469    Q-min: 1.935    Q-max: 1.977    Lives: 5    Reward: 9.0
87613:127690504    Q-min: 2.179    Q-max: 2.305    Lives: 5    Reward: 13.0
87613:127690525    Q-min: 2.128    Q-max: 2.241    Lives: 5    Reward: 14.0
87613:127690545    Q-min: 1.717    Q-max: 2.213    Lives: 5    Reward: 15.0
87613:127690565    Q-min: 2.135    Q-max: 2.274    Lives: 5    Reward: 16.0
87613:127690585    Q-min: 2.200    Q-max: 2.348    Lives: 5    Reward: 17.0
87613:127690605    Q-min: 2.118    Q-max: 2.305    Lives: 5    Reward: 18.0
87613:127690627    Q-min: 2.129    Q-max: 2.312    Lives: 5    Reward: 19.0
87613:127690641    Q-min: 0.014    Q-max: 0.224    Lives: 4    Reward: 19.0
87613:127690695    Q-min: 1.714    Q-max: 1.727    Lives: 4    Reward: 20.0
87613:127690750    Q-min: 2.040    Q-max: 2.111    Lives: 4    Reward: 21.0
87613:127690792    Q-min: 2.043    Q-max: 2.105    Lives: 4    Reward: 22.0
87613:127690828    Q-min: 2.116    Q-max: 2.167    Lives: 4    Reward: 23.0
87613:127690860    Q-min: 2.157    Q-max: 2.172    Lives: 4    Reward: 24.0
87613:127690893    Q-min: 2.138    Q-max: 2.167    Lives: 4    Reward: 28.0
87613:127690915    Q-min: -0.427    Q-max: 0.563    Lives: 3    Reward: 28.0
87613:127690960    Q-min: 2.085    Q-max: 2.117    Lives: 3    Reward: 29.0
87613:127691006    Q-min: 2.026    Q-max: 2.062    Lives: 3    Reward: 30.0
87613:127691053    Q-min: 1.836    Q-max: 2.660    Lives: 3    Reward: 34.0
87613:127691075    Q-min: 2.278    Q-max: 2.366    Lives: 3    Reward: 38.0
87613:127691089    Q-min: -0.006    Q-max: 0.138    Lives: 2    Reward: 38.0
87613:127691129    Q-min: 1.988    Q-max: 2.011    Lives: 2    Reward: 39.0
87613:127691170    Q-min: 2.044    Q-max: 2.122    Lives: 2    Reward: 40.0
87613:127691224    Q-min: 1.812    Q-max: 1.864    Lives: 2    Reward: 41.0
87613:127691272    Q-min: 2.171    Q-max: 2.215    Lives: 2    Reward: 42.0
87613:127691301    Q-min: 2.344    Q-max: 2.418    Lives: 2    Reward: 43.0
87613:127691334    Q-min: 2.161    Q-max: 2.215    Lives: 2    Reward: 44.0
87613:127691372    Q-min: 2.097    Q-max: 2.232    Lives: 2    Reward: 48.0
87613:127691422    Q-min: 1.736    Q-max: 1.760    Lives: 2    Reward: 49.0
87613:127691494    Q-min: 1.720    Q-max: 1.880    Lives: 2    Reward: 53.0
87613:127691560    Q-min: 1.820    Q-max: 1.934    Lives: 2    Reward: 54.0
87613:127691626    Q-min: 1.752    Q-max: 2.086    Lives: 2    Reward: 58.0
87613:127691676    Q-min: 2.129    Q-max: 2.214    Lives: 2    Reward: 59.0
87613:127691710    Q-min: 2.299    Q-max: 2.406    Lives: 2    Reward: 63.0
87613:127691745    Q-min: 2.035    Q-max: 2.164    Lives: 2    Reward: 64.0
87613:127691785    Q-min: 2.293    Q-max: 2.472    Lives: 2    Reward: 71.0
87613:127691808    Q-min: 2.143    Q-max: 2.440    Lives: 2    Reward: 75.0
87613:127691829    Q-min: 1.921    Q-max: 2.489    Lives: 2    Reward: 79.0
87613:127691841    Q-min: 0.072    Q-max: 0.313    Lives: 1    Reward: 79.0
87613:127691888    Q-min: 2.163    Q-max: 2.211    Lives: 1    Reward: 83.0
87613:127691932    Q-min: 2.358    Q-max: 2.504    Lives: 1    Reward: 84.0
87613:127691978    Q-min: 2.047    Q-max: 2.285    Lives: 1    Reward: 85.0
87613:127692017    Q-min: 2.251    Q-max: 2.327    Lives: 1    Reward: 86.0
87613:127692055    Q-min: 2.361    Q-max: 2.563    Lives: 1    Reward: 93.0
87613:127692078    Q-min: 2.379    Q-max: 2.775    Lives: 1    Reward: 97.0
```

```
87613:127692101       Q-min: 2.345       Q-max: 3.082       Lives: 1       Reward: 104.0
87613:127692125       Q-min: 2.469       Q-max: 3.301       Lives: 1       Reward: 111.0
87613:127692151       Q-min: 2.195       Q-max: 3.053       Lives: 1       Reward: 115.0
87613:127692174       Q-min: 2.053       Q-max: 5.052       Lives: 1       Reward: 122.0
87613:127692196       Q-min: 3.218       Q-max: 3.777       Lives: 1       Reward: 129.0
87613:127692225       Q-min: 2.640       Q-max: 7.819       Lives: 1       Reward: 136.0
87613:127692231       Q-min: 3.148       Q-max: 6.144       Lives: 1       Reward: 143.0
87613:127692236       Q-min: 3.346       Q-max: 6.952       Lives: 1       Reward: 150.0
87613:127692241       Q-min: 3.525       Q-max: 6.480       Lives: 1       Reward: 157.0
87613:127692245       Q-min: 2.161       Q-max: 6.309       Lives: 1       Reward: 164.0
87613:127692249       Q-min: 3.869       Q-max: 6.309       Lives: 1       Reward: 171.0
87613:127692253       Q-min: 3.571       Q-max: 5.777       Lives: 1       Reward: 178.0
87613:127692258       Q-min: 3.166       Q-max: 5.918       Lives: 1       Reward: 185.0
87613:127692262       Q-min: 2.885       Q-max: 5.234       Lives: 1       Reward: 192.0
87613:127692270       Q-min: 2.900       Q-max: 3.831       Lives: 1       Reward: 193.0
87613:127692278       Q-min: 2.557       Q-max: 5.713       Lives: 1       Reward: 200.0
87613:127692284       Q-min: 4.380       Q-max: 4.935       Lives: 1       Reward: 207.0
87613:127692320       Q-min: 2.739       Q-max: 5.609       Lives: 1       Reward: 214.0
87613:127692327       Q-min: 3.526       Q-max: 5.042       Lives: 1       Reward: 218.0
87613:127692334       Q-min: 4.065       Q-max: 5.043       Lives: 1       Reward: 225.0
87613:127692339       Q-min: 1.078       Q-max: 4.988       Lives: 1       Reward: 232.0
87613:127692343       Q-min: 4.129       Q-max: 5.668       Lives: 1       Reward: 239.0
87613:127692348       Q-min: 3.897       Q-max: 6.170       Lives: 1       Reward: 246.0
87613:127692355       Q-min: 3.166       Q-max: 5.217       Lives: 1       Reward: 253.0
87613:127692395       Q-min: 2.545       Q-max: 3.698       Lives: 1       Reward: 260.0
87613:127692402       Q-min: 3.140       Q-max: 4.613       Lives: 1       Reward: 267.0
87613:127692408       Q-min: 2.794       Q-max: 4.494       Lives: 1       Reward: 274.0
87613:127692414       Q-min: 2.961       Q-max: 4.157       Lives: 1       Reward: 278.0
87613:127692420       Q-min: 2.604       Q-max: 4.554       Lives: 1       Reward: 285.0
87613:127692427       Q-min: 2.394       Q-max: 4.361       Lives: 1       Reward: 292.0
87613:127692465       Q-min: 2.877       Q-max: 4.571       Lives: 1       Reward: 299.0
87613:127692485       Q-min: -1.033       Q-max: 0.680       Lives: 0       Reward: 299.0
87614:127692527       Q-min: 1.766       Q-max: 1.789       Lives: 5       Reward: 1.0
87614:127692565       Q-min: 1.803       Q-max: 1.843       Lives: 5       Reward: 2.0
87614:127692591       Q-min: -0.225       Q-max: 0.154       Lives: 4       Reward: 2.0
87614:127692638       Q-min: 1.861       Q-max: 1.883       Lives: 4       Reward: 3.0
87614:127692677       Q-min: 1.971       Q-max: 1.990       Lives: 4       Reward: 4.0
87614:127692719       Q-min: 1.868       Q-max: 1.902       Lives: 4       Reward: 5.0
87614:127692757       Q-min: 1.970       Q-max: 2.008       Lives: 4       Reward: 6.0
87614:127692789       Q-min: 1.890       Q-max: 1.911       Lives: 4       Reward: 7.0
87614:127692823       Q-min: 1.920       Q-max: 1.938       Lives: 4       Reward: 8.0
87614:127692854       Q-min: 1.956       Q-max: 2.014       Lives: 4       Reward: 9.0
87614:127692900       Q-min: 1.622       Q-max: 1.698       Lives: 4       Reward: 10.0
87614:127692963       Q-min: 1.506       Q-max: 1.674       Lives: 4       Reward: 14.0
87614:127693028       Q-min: 1.711       Q-max: 1.739       Lives: 4       Reward: 15.0
87614:127693095       Q-min: 1.765       Q-max: 1.803       Lives: 4       Reward: 16.0
87614:127693141       Q-min: -0.065       Q-max: 0.344       Lives: 3       Reward: 16.0
87614:127693194       Q-min: 1.681       Q-max: 1.725       Lives: 3       Reward: 17.0
```

```
87614:127693248       Q-min: 1.985       Q-max: 1.995       Lives: 3       Reward: 18.0
87614:127693305       Q-min: 1.598       Q-max: 1.688       Lives: 3       Reward: 19.0
87614:127693358       Q-min: 1.962       Q-max: 1.997       Lives: 3       Reward: 20.0
87614:127693390       Q-min: 2.017       Q-max: 2.048       Lives: 3       Reward: 21.0
87614:127693423       Q-min: 2.144       Q-max: 2.350       Lives: 3       Reward: 25.0
87614:127693444       Q-min: 2.376       Q-max: 2.448       Lives: 3       Reward: 26.0
87614:127693462       Q-min: 2.332       Q-max: 2.458       Lives: 3       Reward: 27.0
87614:127693483       Q-min: 2.290       Q-max: 2.491       Lives: 3       Reward: 28.0
87614:127693504       Q-min: 2.354       Q-max: 2.425       Lives: 3       Reward: 29.0
87614:127693522       Q-min: 2.388       Q-max: 2.483       Lives: 3       Reward: 33.0
87614:127693541       Q-min: 2.270       Q-max: 2.347       Lives: 3       Reward: 34.0
87614:127693562       Q-min: 2.352       Q-max: 2.491       Lives: 3       Reward: 38.0
87614:127693585       Q-min: 2.124       Q-max: 2.533       Lives: 3       Reward: 42.0
87614:127693605       Q-min: 2.291       Q-max: 2.583       Lives: 3       Reward: 43.0
87614:127693625       Q-min: 2.335       Q-max: 2.473       Lives: 3       Reward: 44.0
87614:127693644       Q-min: 2.395       Q-max: 2.536       Lives: 3       Reward: 45.0
87614:127693657       Q-min: 0.026       Q-max: 0.412       Lives: 2       Reward: 45.0
87614:127693699       Q-min: 2.104       Q-max: 2.139       Lives: 2       Reward: 46.0
87614:127693754       Q-min: 1.830       Q-max: 1.937       Lives: 2       Reward: 47.0
87614:127693819       Q-min: 1.974       Q-max: 2.017       Lives: 2       Reward: 48.0
87614:127693870       Q-min: 2.307       Q-max: 2.335       Lives: 2       Reward: 49.0
87614:127693905       Q-min: 2.022       Q-max: 2.130       Lives: 2       Reward: 53.0
87614:127693942       Q-min: 2.223       Q-max: 2.292       Lives: 2       Reward: 54.0
87614:127693964       Q-min: -0.345       Q-max: 0.457       Lives: 1       Reward: 54.0
87614:127694008       Q-min: 2.152       Q-max: 2.202       Lives: 1       Reward: 58.0
87614:127694057       Q-min: 2.324       Q-max: 2.604       Lives: 1       Reward: 62.0
87614:127694079       Q-min: 2.336       Q-max: 2.557       Lives: 1       Reward: 66.0
87614:127694099       Q-min: 2.449       Q-max: 2.512       Lives: 1       Reward: 67.0
87614:127694120       Q-min: 2.384       Q-max: 2.646       Lives: 1       Reward: 71.0
87614:127694134       Q-min: 0.130       Q-max: 0.253       Lives: 0       Reward: 71.0
87615:127694189       Q-min: 1.683       Q-max: 1.693       Lives: 5       Reward: 1.0
87615:127694238       Q-min: 1.845       Q-max: 1.865       Lives: 5       Reward: 2.0
87615:127694287       Q-min: 1.685       Q-max: 1.715       Lives: 5       Reward: 3.0
87615:127694335       Q-min: 1.930       Q-max: 1.947       Lives: 5       Reward: 4.0
87615:127694365       Q-min: 1.954       Q-max: 1.982       Lives: 5       Reward: 5.0
87615:127694400       Q-min: 1.906       Q-max: 1.926       Lives: 5       Reward: 6.0
87615:127694420       Q-min: -0.096       Q-max: 0.063       Lives: 4       Reward: 6.0
87615:127694462       Q-min: 1.833       Q-max: 1.868       Lives: 4       Reward: 7.0
87615:127694502       Q-min: 1.910       Q-max: 1.920       Lives: 4       Reward: 8.0
87615:127694531       Q-min: -0.029       Q-max: 0.218       Lives: 3       Reward: 8.0
87615:127694582       Q-min: 1.631       Q-max: 1.664       Lives: 3       Reward: 9.0
87615:127694635       Q-min: 1.867       Q-max: 1.879       Lives: 3       Reward: 10.0
87615:127694683       Q-min: 1.914       Q-max: 1.928       Lives: 3       Reward: 11.0
87615:127694720       Q-min: 1.920       Q-max: 1.953       Lives: 3       Reward: 12.0
87615:127694753       Q-min: 1.966       Q-max: 2.038       Lives: 3       Reward: 16.0
87615:127694786       Q-min: 1.915       Q-max: 1.953       Lives: 3       Reward: 17.0
87615:127694818       Q-min: 2.045       Q-max: 2.117       Lives: 3       Reward: 21.0
87615:127694863       Q-min: 1.640       Q-max: 1.660       Lives: 3       Reward: 22.0
```

```
87615:127694936       Q-min: 1.704       Q-max: 1.790       Lives: 3       Reward: 23.0
87615:127695003       Q-min: 1.607       Q-max: 1.693       Lives: 3       Reward: 24.0
87615:127695066       Q-min: 1.545       Q-max: 1.706       Lives: 3       Reward: 25.0
87615:127695117       Q-min: 1.944       Q-max: 1.970       Lives: 3       Reward: 26.0
87615:127695148       Q-min: 2.001       Q-max: 2.020       Lives: 3       Reward: 30.0
87615:127695185       Q-min: 1.505       Q-max: 2.459       Lives: 3       Reward: 34.0
87615:127695206       Q-min: 2.292       Q-max: 2.447       Lives: 3       Reward: 35.0
87615:127695226       Q-min: 2.223       Q-max: 2.418       Lives: 3       Reward: 36.0
87615:127695245       Q-min: 2.313       Q-max: 2.373       Lives: 3       Reward: 37.0
87615:127695267       Q-min: 2.310       Q-max: 2.449       Lives: 3       Reward: 41.0
87615:127695283       Q-min: 0.034       Q-max: 0.248       Lives: 2       Reward: 41.0
87615:127695327       Q-min: 1.973       Q-max: 2.000       Lives: 2       Reward: 42.0
87615:127695368       Q-min: 2.111       Q-max: 2.145       Lives: 2       Reward: 43.0
87615:127695414       Q-min: 2.005       Q-max: 2.129       Lives: 2       Reward: 47.0
87615:127695454       Q-min: 2.254       Q-max: 2.286       Lives: 2       Reward: 48.0
87615:127695489       Q-min: 2.176       Q-max: 2.210       Lives: 2       Reward: 49.0
87615:127695521       Q-min: 2.159       Q-max: 2.215       Lives: 2       Reward: 50.0
87615:127695555       Q-min: 2.226       Q-max: 2.292       Lives: 2       Reward: 54.0
87615:127695612       Q-min: 1.677       Q-max: 1.804       Lives: 2       Reward: 55.0
87615:127695679       Q-min: 1.717       Q-max: 1.984       Lives: 2       Reward: 59.0
87615:127695750       Q-min: 1.788       Q-max: 1.857       Lives: 2       Reward: 60.0
87615:127695817       Q-min: 1.672       Q-max: 1.933       Lives: 2       Reward: 64.0
87615:127695868       Q-min: 2.129       Q-max: 2.172       Lives: 2       Reward: 65.0
87615:127695900       Q-min: 2.162       Q-max: 2.207       Lives: 2       Reward: 69.0
87615:127695934       Q-min: 2.146       Q-max: 2.353       Lives: 2       Reward: 70.0
87615:127695970       Q-min: 2.184       Q-max: 2.295       Lives: 2       Reward: 74.0
87615:127696006       Q-min: 2.072       Q-max: 2.672       Lives: 2       Reward: 78.0
87615:127696031       Q-min: 2.447       Q-max: 2.703       Lives: 2       Reward: 85.0
87615:127696059       Q-min: 2.400       Q-max: 2.714       Lives: 2       Reward: 92.0
87615:127696080       Q-min: 2.617       Q-max: 2.961       Lives: 2       Reward: 93.0
87615:127696101       Q-min: 2.716       Q-max: 3.000       Lives: 2       Reward: 97.0
87615:127696124       Q-min: 2.552       Q-max: 3.052       Lives: 2       Reward: 101.0
87615:127696147       Q-min: 2.177       Q-max: 3.415       Lives: 2       Reward: 108.0
87615:127696169       Q-min: 2.833       Q-max: 3.527       Lives: 2       Reward: 112.0
87615:127696199       Q-min: 1.329       Q-max: 7.426       Lives: 2       Reward: 119.0
87615:127696204       Q-min: 2.727       Q-max: 7.415       Lives: 2       Reward: 126.0
87615:127696208       Q-min: 4.507       Q-max: 7.164       Lives: 2       Reward: 133.0
87615:127696213       Q-min: 4.021       Q-max: 7.264       Lives: 2       Reward: 140.0
87615:127696218       Q-min: 4.524       Q-max: 6.419       Lives: 2       Reward: 147.0
87615:127696224       Q-min: 4.086       Q-max: 6.433       Lives: 2       Reward: 154.0
87615:127696229       Q-min: 3.476       Q-max: 6.955       Lives: 2       Reward: 161.0
87615:127696233       Q-min: 4.404       Q-max: 6.600       Lives: 2       Reward: 168.0
87615:127696237       Q-min: 4.124       Q-max: 6.776       Lives: 2       Reward: 175.0
87615:127696242       Q-min: 3.961       Q-max: 5.944       Lives: 2       Reward: 182.0
87615:127696247       Q-min: 4.544       Q-max: 5.938       Lives: 2       Reward: 189.0
87615:127696254       Q-min: 3.912       Q-max: 6.480       Lives: 2       Reward: 196.0
87615:127696262       Q-min: 4.059       Q-max: 5.634       Lives: 2       Reward: 203.0
87615:127696268       Q-min: 3.037       Q-max: 4.725       Lives: 2       Reward: 210.0
```

```
87615:127696274       Q-min: 2.643       Q-max: 5.178       Lives: 2       Reward: 217.0
87615:127696280       Q-min: 3.061       Q-max: 4.737       Lives: 2       Reward: 224.0
87615:127696286       Q-min: 3.398       Q-max: 4.795       Lives: 2       Reward: 231.0
87615:127696292       Q-min: 3.533       Q-max: 4.552       Lives: 2       Reward: 238.0
87615:127696298       Q-min: 3.064       Q-max: 5.131       Lives: 2       Reward: 245.0
87615:127696302       Q-min: 3.054       Q-max: 4.238       Lives: 2       Reward: 252.0
87615:127696309       Q-min: 3.495       Q-max: 4.537       Lives: 2       Reward: 259.0
87615:127696318       Q-min: 2.442       Q-max: 3.859       Lives: 2       Reward: 263.0
87615:127696325       Q-min: 2.675       Q-max: 4.446       Lives: 2       Reward: 267.0
87615:127696332       Q-min: 2.903       Q-max: 4.970       Lives: 2       Reward: 271.0
87615:127696339       Q-min: 3.041       Q-max: 5.599       Lives: 2       Reward: 278.0
87615:127696347       Q-min: 3.840       Q-max: 5.684       Lives: 2       Reward: 285.0
87615:127696354       Q-min: 3.338       Q-max: 5.494       Lives: 2       Reward: 292.0
87615:127696359       Q-min: 3.294       Q-max: 5.389       Lives: 2       Reward: 299.0
87615:127696366       Q-min: 2.294       Q-max: 3.083       Lives: 2       Reward: 300.0
87615:127696373       Q-min: 2.104       Q-max: 3.364       Lives: 2       Reward: 307.0
87615:127696379       Q-min: 1.393       Q-max: 4.202       Lives: 2       Reward: 314.0
87615:127696402       Q-min: 0.235       Q-max: 0.449       Lives: 1       Reward: 314.0
87615:127696467       Q-min: 1.152       Q-max: 2.538       Lives: 1       Reward: 318.0
87615:127696475       Q-min: 2.140       Q-max: 4.579       Lives: 1       Reward: 325.0
87615:127696481       Q-min: 1.566       Q-max: 2.914       Lives: 1       Reward: 332.0
87615:127696491       Q-min: 1.893       Q-max: 2.961       Lives: 1       Reward: 336.0
87615:127696499       Q-min: 2.189       Q-max: 3.113       Lives: 1       Reward: 340.0
87615:127696523       Q-min: 0.266       Q-max: 0.417       Lives: 0       Reward: 340.0
87616:127696566       Q-min: 1.755       Q-max: 1.768       Lives: 5       Reward: 1.0
87616:127696607       Q-min: 1.822       Q-max: 1.860       Lives: 5       Reward: 2.0
87616:127696650       Q-min: 1.905       Q-max: 1.942       Lives: 5       Reward: 3.0
87616:127696686       Q-min: 1.920       Q-max: 1.984       Lives: 5       Reward: 4.0
87616:127696717       Q-min: 1.984       Q-max: 2.012       Lives: 5       Reward: 5.0
87616:127696748       Q-min: 1.915       Q-max: 1.942       Lives: 5       Reward: 6.0
87616:127696778       Q-min: 1.698       Q-max: 1.758       Lives: 5       Reward: 7.0
87616:127696825       Q-min: 1.582       Q-max: 1.633       Lives: 5       Reward: 8.0
87616:127696887       Q-min: 1.685       Q-max: 1.710       Lives: 5       Reward: 9.0
87616:127696929       Q-min: -0.020       Q-max: 0.235       Lives: 4       Reward: 9.0
87616:127696986       Q-min: 1.568       Q-max: 1.667       Lives: 4       Reward: 13.0
87616:127697052       Q-min: 1.726       Q-max: 1.766       Lives: 4       Reward: 14.0
87616:127697114       Q-min: 1.681       Q-max: 1.714       Lives: 4       Reward: 15.0
87616:127697161       Q-min: 2.009       Q-max: 2.051       Lives: 4       Reward: 16.0
87616:127697180       Q-min: -0.006       Q-max: 0.308       Lives: 3       Reward: 16.0
87616:127697223       Q-min: 1.883       Q-max: 1.921       Lives: 3       Reward: 17.0
87616:127697264       Q-min: 1.967       Q-max: 1.993       Lives: 3       Reward: 18.0
87616:127697320       Q-min: 1.846       Q-max: 1.892       Lives: 3       Reward: 19.0
87616:127697370       Q-min: 2.026       Q-max: 2.101       Lives: 3       Reward: 20.0
87616:127697406       Q-min: 2.037       Q-max: 2.075       Lives: 3       Reward: 21.0
87616:127697436       Q-min: 2.030       Q-max: 2.055       Lives: 3       Reward: 22.0
87616:127697469       Q-min: 2.059       Q-max: 2.080       Lives: 3       Reward: 23.0
87616:127697521       Q-min: 1.619       Q-max: 1.701       Lives: 3       Reward: 24.0
87616:127697583       Q-min: 1.657       Q-max: 1.745       Lives: 3       Reward: 25.0
```

```
87616:127697651       Q-min: 1.690       Q-max: 1.752       Lives: 3       Reward: 26.0
87616:127697718       Q-min: 1.667       Q-max: 1.827       Lives: 3       Reward: 27.0
87616:127697766       Q-min: 2.042       Q-max: 2.062       Lives: 3       Reward: 31.0
87616:127697800       Q-min: 2.033       Q-max: 2.101       Lives: 3       Reward: 32.0
87616:127697830       Q-min: 2.066       Q-max: 2.106       Lives: 3       Reward: 33.0
87616:127697863       Q-min: 1.952       Q-max: 2.108       Lives: 3       Reward: 34.0
87616:127697898       Q-min: 2.262       Q-max: 2.323       Lives: 3       Reward: 38.0
87616:127697912       Q-min: 0.078       Q-max: 0.422       Lives: 2       Reward: 38.0
87616:127697968       Q-min: 1.802       Q-max: 1.856       Lives: 2       Reward: 39.0
87616:127698031       Q-min: 1.838       Q-max: 1.876       Lives: 2       Reward: 40.0
87616:127698099       Q-min: 1.856       Q-max: 1.898       Lives: 2       Reward: 41.0
87616:127698150       Q-min: 2.508       Q-max: 2.578       Lives: 2       Reward: 45.0
87616:127698172       Q-min: 2.409       Q-max: 2.576       Lives: 2       Reward: 49.0
87616:127698193       Q-min: 2.089       Q-max: 2.658       Lives: 2       Reward: 56.0
87616:127698216       Q-min: 2.436       Q-max: 2.506       Lives: 2       Reward: 57.0
87616:127698236       Q-min: 2.420       Q-max: 2.482       Lives: 2       Reward: 58.0
87616:127698257       Q-min: 2.457       Q-max: 2.508       Lives: 2       Reward: 62.0
87616:127698278       Q-min: 2.433       Q-max: 2.551       Lives: 2       Reward: 66.0
87616:127698291       Q-min: -0.079       Q-max: 0.126       Lives: 1       Reward: 66.0
87616:127698343       Q-min: 1.843       Q-max: 1.872       Lives: 1       Reward: 67.0
87616:127698399       Q-min: 2.090       Q-max: 2.351       Lives: 1       Reward: 71.0
87616:127698456       Q-min: 1.968       Q-max: 2.183       Lives: 1       Reward: 75.0
87616:127698512       Q-min: 2.150       Q-max: 2.390       Lives: 1       Reward: 79.0
87616:127698549       Q-min: 2.398       Q-max: 2.454       Lives: 1       Reward: 80.0
87616:127698584       Q-min: 2.203       Q-max: 2.419       Lives: 1       Reward: 84.0
87616:127698620       Q-min: 2.034       Q-max: 2.903       Lives: 1       Reward: 88.0
87616:127698641       Q-min: 2.441       Q-max: 2.569       Lives: 1       Reward: 89.0
87616:127698661       Q-min: 2.446       Q-max: 2.627       Lives: 1       Reward: 93.0
87616:127698684       Q-min: 2.499       Q-max: 2.649       Lives: 1       Reward: 94.0
87616:127698707       Q-min: 2.364       Q-max: 2.641       Lives: 1       Reward: 98.0
87616:127698721       Q-min: -0.372       Q-max: 0.212       Lives: 0       Reward: 98.0
87617:127698766       Q-min: 1.752       Q-max: 1.767       Lives: 5       Reward: 1.0
87617:127698808       Q-min: 1.814       Q-max: 1.833       Lives: 5       Reward: 2.0
87617:127698846       Q-min: 1.934       Q-max: 2.003       Lives: 5       Reward: 3.0
87617:127698884       Q-min: 1.993       Q-max: 2.013       Lives: 5       Reward: 4.0
87617:127698913       Q-min: 1.983       Q-max: 2.006       Lives: 5       Reward: 5.0
87617:127698943       Q-min: 1.895       Q-max: 1.934       Lives: 5       Reward: 6.0
87617:127698975       Q-min: 1.748       Q-max: 1.790       Lives: 5       Reward: 7.0
87617:127699023       Q-min: 1.579       Q-max: 1.655       Lives: 5       Reward: 8.0
87617:127699090       Q-min: 1.652       Q-max: 1.752       Lives: 5       Reward: 9.0
87617:127699159       Q-min: 1.698       Q-max: 1.752       Lives: 5       Reward: 10.0
87617:127699202       Q-min: -0.031       Q-max: 0.225       Lives: 4       Reward: 10.0
87617:127699246       Q-min: 1.889       Q-max: 1.922       Lives: 4       Reward: 11.0
87617:127699296       Q-min: 1.914       Q-max: 1.953       Lives: 4       Reward: 15.0
87617:127699353       Q-min: 1.757       Q-max: 1.781       Lives: 4       Reward: 16.0
87617:127699403       Q-min: 2.019       Q-max: 2.053       Lives: 4       Reward: 17.0
87617:127699434       Q-min: 2.157       Q-max: 2.220       Lives: 4       Reward: 18.0
87617:127699468       Q-min: 1.990       Q-max: 2.004       Lives: 4       Reward: 19.0
```

```
87617:127699501       Q-min: 1.990       Q-max: 2.028       Lives: 4       Reward: 20.0
87617:127699549       Q-min: 1.681       Q-max: 1.707       Lives: 4       Reward: 21.0
87617:127699614       Q-min: 1.751       Q-max: 1.785       Lives: 4       Reward: 22.0
87617:127699674       Q-min: 1.735       Q-max: 1.746       Lives: 4       Reward: 23.0
87617:127699736       Q-min: 1.770       Q-max: 1.788       Lives: 4       Reward: 24.0
87617:127699784       Q-min: 1.994       Q-max: 2.046       Lives: 4       Reward: 25.0
87617:127699816       Q-min: 2.013       Q-max: 2.036       Lives: 4       Reward: 26.0
87617:127699837       Q-min: -0.117       Q-max: 0.201       Lives: 3       Reward: 26.0
87617:127699880       Q-min: 1.936       Q-max: 1.952       Lives: 3       Reward: 27.0
87617:127699937       Q-min: 1.639       Q-max: 1.781       Lives: 3       Reward: 28.0
87617:127699992       Q-min: 2.026       Q-max: 2.058       Lives: 3       Reward: 29.0
87617:127700030       Q-min: 2.014       Q-max: 2.034       Lives: 3       Reward: 30.0
87617:127700066       Q-min: 2.029       Q-max: 2.075       Lives: 3       Reward: 31.0
87617:127700098       Q-min: 2.160       Q-max: 2.572       Lives: 3       Reward: 35.0
87617:127700118       Q-min: 2.057       Q-max: 2.423       Lives: 3       Reward: 39.0
87617:127700139       Q-min: 2.400       Q-max: 2.545       Lives: 3       Reward: 43.0
87617:127700158       Q-min: 2.066       Q-max: 2.565       Lives: 3       Reward: 50.0
87617:127700180       Q-min: 2.392       Q-max: 2.474       Lives: 3       Reward: 51.0
87617:127700200       Q-min: 2.233       Q-max: 2.483       Lives: 3       Reward: 52.0
87617:127700221       Q-min: 2.048       Q-max: 2.191       Lives: 3       Reward: 56.0
87617:127700245       Q-min: 2.330       Q-max: 2.450       Lives: 3       Reward: 60.0
87617:127700271       Q-min: 2.389       Q-max: 2.500       Lives: 3       Reward: 61.0
87617:127700293       Q-min: 2.307       Q-max: 2.508       Lives: 3       Reward: 65.0
87617:127700306       Q-min: -0.006       Q-max: 0.113       Lives: 2       Reward: 65.0
87617:127700351       Q-min: 1.970       Q-max: 2.111       Lives: 2       Reward: 69.0
87617:127700399       Q-min: 2.366       Q-max: 2.450       Lives: 2       Reward: 73.0
87617:127700422       Q-min: 2.140       Q-max: 2.488       Lives: 2       Reward: 74.0
87617:127700442       Q-min: 2.122       Q-max: 2.493       Lives: 2       Reward: 78.0
87617:127700464       Q-min: 2.243       Q-max: 2.550       Lives: 2       Reward: 82.0
87617:127700488       Q-min: 2.414       Q-max: 2.670       Lives: 2       Reward: 86.0
87617:127700510       Q-min: 2.252       Q-max: 2.626       Lives: 2       Reward: 93.0
87617:127700533       Q-min: 2.150       Q-max: 2.691       Lives: 2       Reward: 100.0
87617:127700556       Q-min: 1.972       Q-max: 2.779       Lives: 2       Reward: 101.0
87617:127700567       Q-min: -0.020       Q-max: 0.190       Lives: 1       Reward: 101.0
87617:127700613       Q-min: 2.301       Q-max: 2.671       Lives: 1       Reward: 105.0
87617:127700659       Q-min: 2.437       Q-max: 2.946       Lives: 1       Reward: 106.0
87617:127700709       Q-min: 0.855       Q-max: 2.276       Lives: 1       Reward: 113.0
87617:127700735       Q-min: 1.749       Q-max: 3.022       Lives: 1       Reward: 120.0
87617:127700761       Q-min: 2.380       Q-max: 3.517       Lives: 1       Reward: 127.0
87617:127700778       Q-min: -0.098       Q-max: 0.169       Lives: 0       Reward: 127.0
```

We can now print some statistics for the episode rewards, which vary greatly from one episode to the next.

```
In [25]: rewards = agent.episode_rewards
         print("Rewards for {0} episodes:".format(len(rewards)))
         print("- Min:   ", np.min(rewards))
```

```
    print("- Mean:  ", np.mean(rewards))
    print("- Max:   ", np.max(rewards))
    print("- Stdev: ", np.std(rewards))
```

```
Rewards for 30 episodes:
- Min:    40.0
- Mean:   145.166666667
- Max:    386.0
- Stdev:  105.131372842
```

We can also plot a histogram with the episode rewards.

```
In [19]: _ = plt.hist(rewards, bins=30)
```

```
---------------------------------------------------------------------------

NameError                                 Traceback (most recent call last)

<ipython-input-19-76d39b2da15f> in <module>()
----> 1 _ = plt.hist(rewards, bins=30)


NameError: name 'rewards' is not defined
```

2.16 Example States

We can plot examples of states from the game-environment and the Q-values that are estimated by the Neural Network.

This helper-function prints the Q-values for a given index in the replay-memory.

```
In [20]: def print_q_values(idx):
             """Print Q-values and actions from the replay-memory at the given index."""

             # Get the Q-values and action from the replay-memory.
             q_values = replay_memory.q_values[idx]
             action = replay_memory.actions[idx]

             print("Action:     Q-Value:")
             print("====================")

             # Print all the actions and their Q-values.
             for i, q_value in enumerate(q_values):
                 # Used to display which action was taken.
                 if i == action:
                     action_taken = "(Action Taken)"
                 else:
```

```
                action_taken = ""

            # Text-name of the action.
            action_name = agent.get_action_name(i)

            print("{0:12}{1:.3f} {2}".format(action_name, q_value,
                                             action_taken))

    # Newline.
    print()
```

This helper-function plots a state from the replay-memory and optionally prints the Q-values.

```
In [21]: def plot_state(idx, print_q=True):
             """Plot the state in the replay-memory with the given index."""

             # Get the state from the replay-memory.
             state = replay_memory.states[idx]

             # Create figure with a grid of sub-plots.
             fig, axes = plt.subplots(1, 2)

             # Plot the image from the game-environment.
             ax = axes.flat[0]
             ax.imshow(state[:, :, 0], vmin=0, vmax=255,
                       interpolation='lanczos', cmap='gray')

             # Plot the motion-trace.
             ax = axes.flat[1]
             ax.imshow(state[:, :, 1], vmin=0, vmax=255,
                       interpolation='lanczos', cmap='gray')

             # This is necessary if we show more than one plot in a single Notebook cell.
             plt.show()

             # Print the Q-values.
             if print_q:
                 print_q_values(idx=idx)
```

The replay-memory has room for 200k states but it is only partially full from the above call to `agent.run(num_episodes=1)`. This is how many states are actually used.

```
In [22]: num_used = replay_memory.num_used
         num_used
```

```
Out[22]: 2207
```

Get the Q-values from the replay-memory that are actually used.

```
In [23]: q_values = replay_memory.q_values[0:num_used, :]
```

For each state, calculate the min / max Q-values and their difference. This will be used to lookup interesting states in the following sections.

```
In [24]: q_values_min = q_values.min(axis=1)
         q_values_max = q_values.max(axis=1)
         q_values_dif = q_values_max - q_values_min
```

2.16.1 Example States: Highest Reward

This example shows the states surrounding the state with the highest reward.

During the training we limit the rewards to the range [-1, 1] so this basically just gets the first state that has a reward of 1.

```
In [25]: idx = np.argmax(replay_memory.rewards)
         idx
```

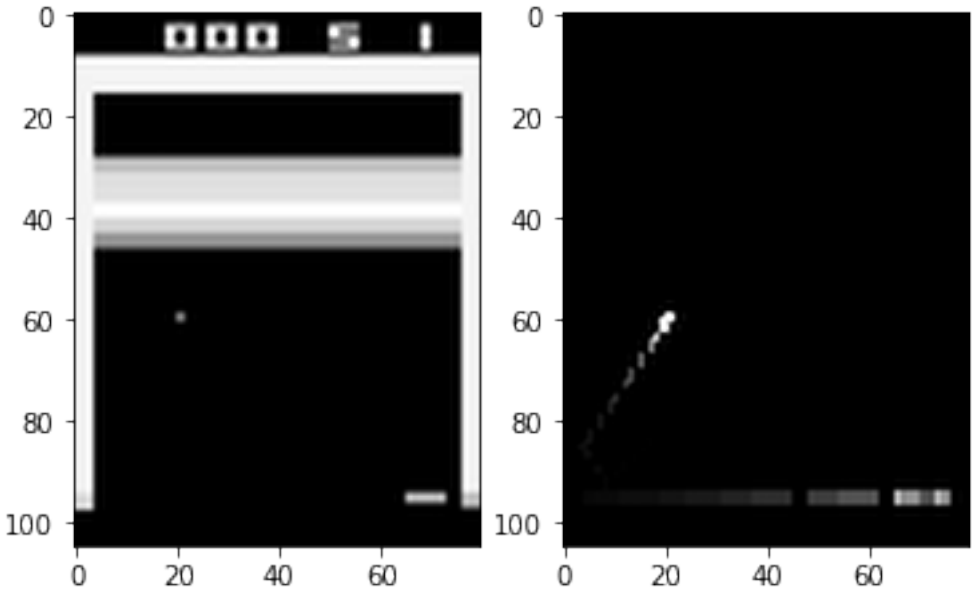
```
Out[25]: 40
```

This state is where the ball hits the wall so the agent scores a point.

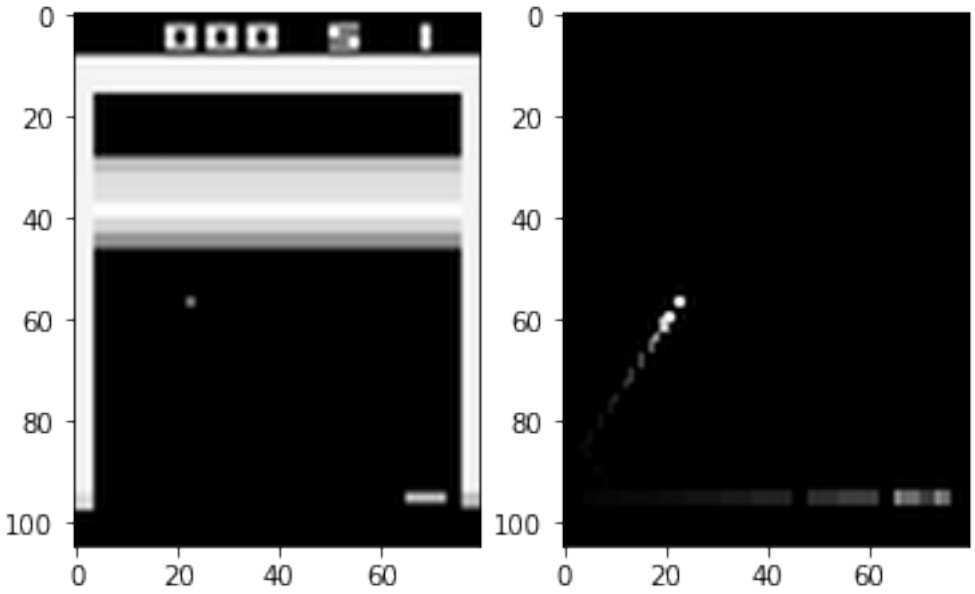
We can show the surrounding states leading up to and following this state. Note how the Q-values are very close for the different actions, because at this point it really does not matter what the agent does as the reward is already guaranteed. But note how the Q-values decrease significantly after the ball has hit the wall and a point has been scored.

Also note that the agent uses the Epsilon-greedy policy for taking actions, so there is a small probability that a random action is taken instead of the action with the highest Q-value.

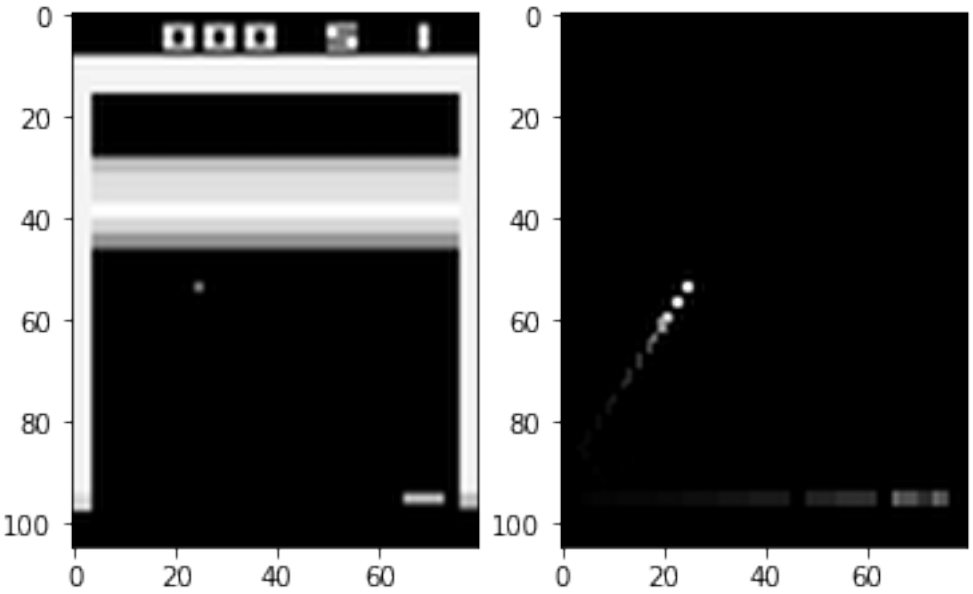
```
In [26]: for i in range(-5, 3):
             plot_state(idx=idx+i)
```



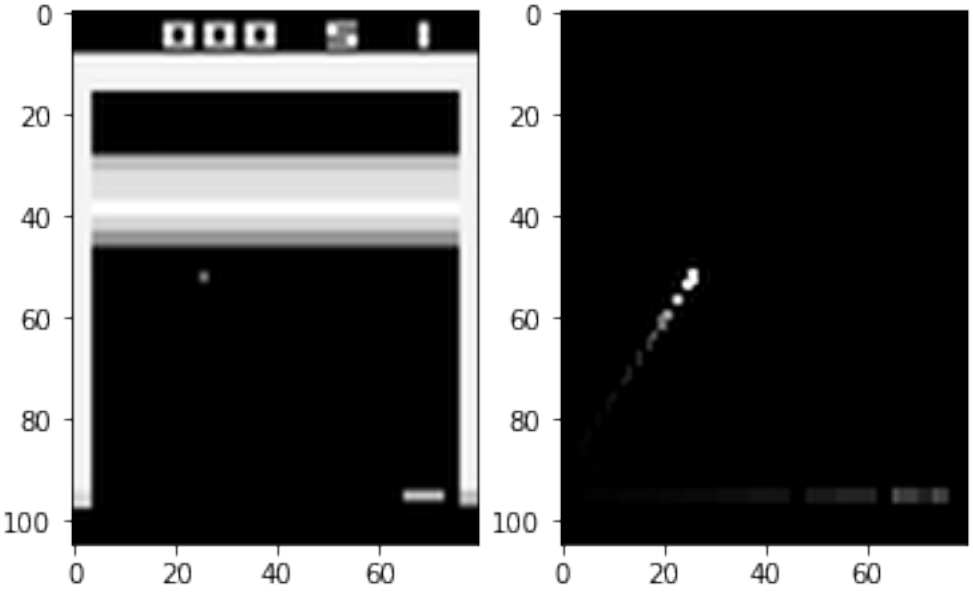
```
Action:     Q-Value:
====================
NOOP        1.201 (Action Taken)
FIRE        1.074
RIGHT       1.075
LEFT        1.185
```



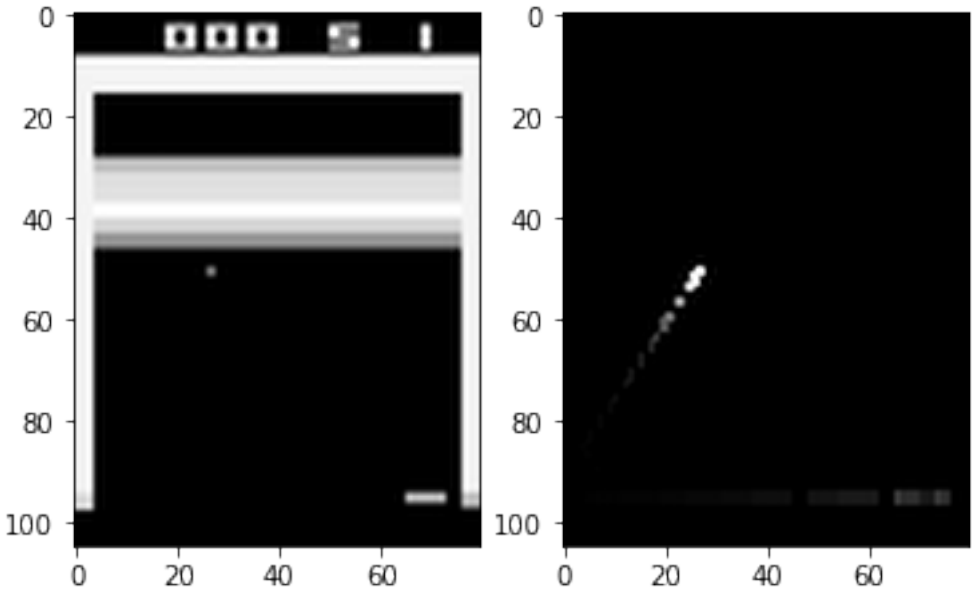

```
Action:     Q-Value:
====================
NOOP        1.282 (Action Taken)
FIRE        1.121
RIGHT       1.109
LEFT        1.234
```



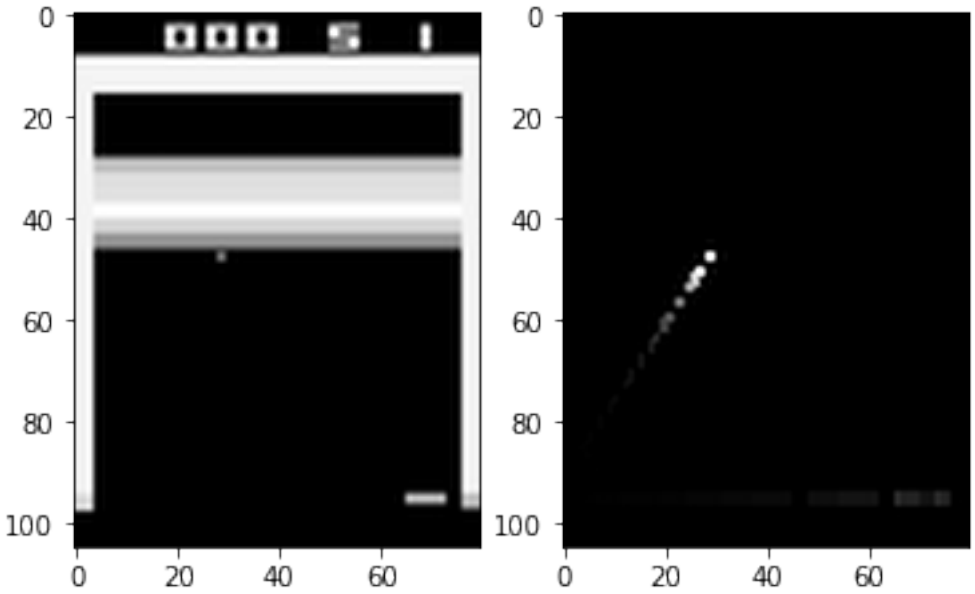

```
Action:     Q-Value:
====================
NOOP        1.326 (Action Taken)
FIRE        1.197
RIGHT       1.124
LEFT        1.191
```



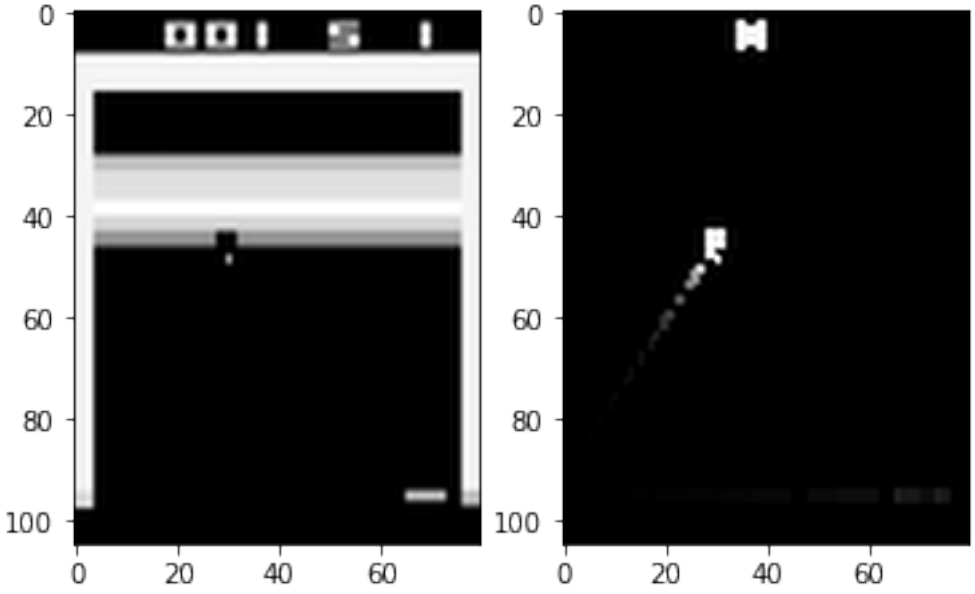
```
Action:     Q-Value:
====================
NOOP        1.352 (Action Taken)
FIRE        1.301
RIGHT       1.179
LEFT        1.235
```



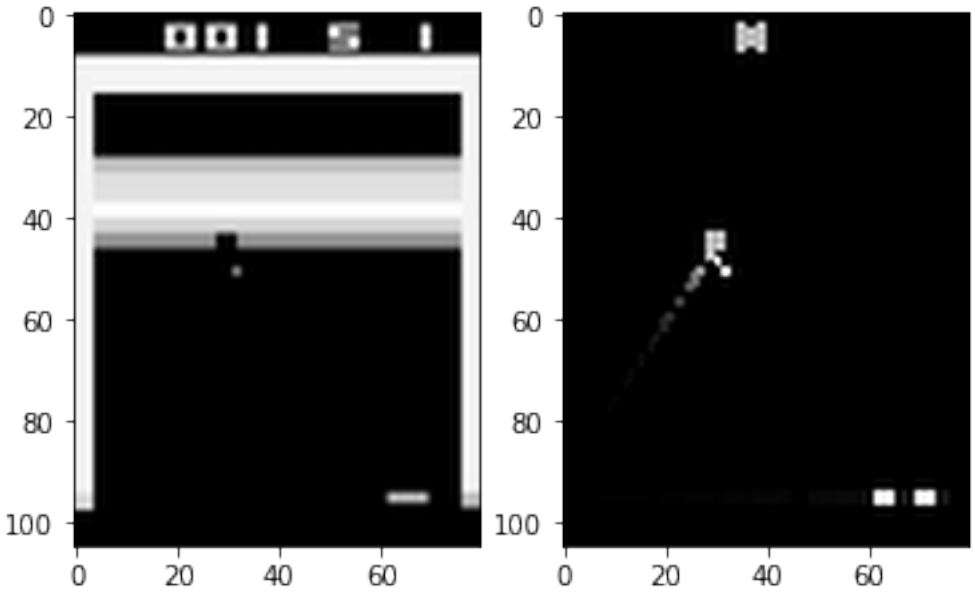

```
Action:     Q-Value:
====================
NOOP        1.353
FIRE        1.356 (Action Taken)
RIGHT       1.226
LEFT        1.331
```




```
Action:     Q-Value:
====================
NOOP        1.441 (Action Taken)
FIRE        1.403
RIGHT       1.238
LEFT        1.313
```



```
Action:     Q-Value:
====================
NOOP        0.363
FIRE        0.451
RIGHT       0.476
LEFT        0.483 (Action Taken)
```

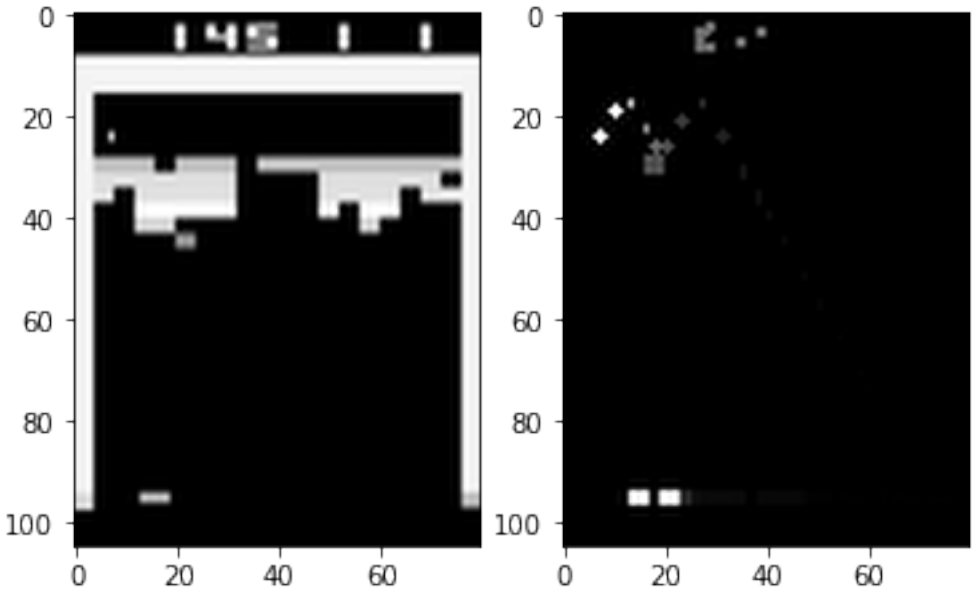



```
Action:     Q-Value:
====================
NOOP        0.420
FIRE        0.460
RIGHT       0.442
LEFT        0.477 (Action Taken)
```

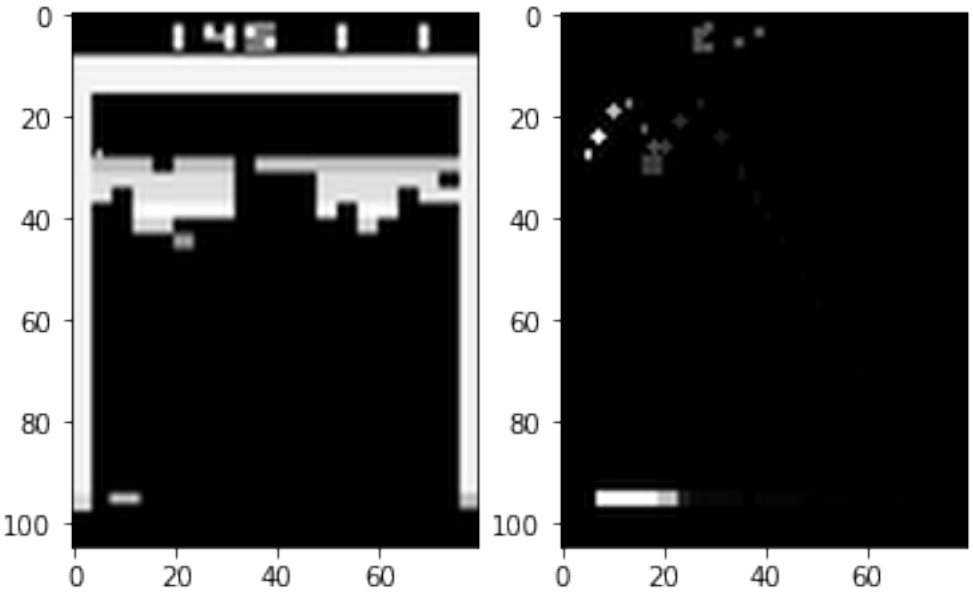
2.16.2 Example: Highest Q-Value

This example shows the states surrounding the one with the highest Q-values. This means that the agent has high expectation that several points will be scored in the following steps. Note that the Q-values decrease significantly after the points have been scored.

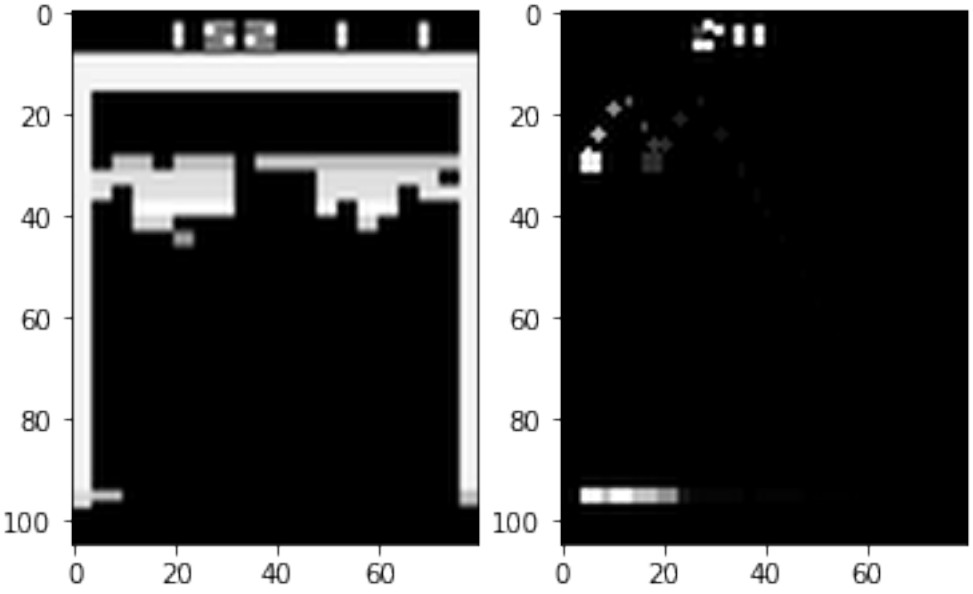
```
In [27]: idx = np.argmax(q_values_max)
         idx

Out[27]: 2107

In [28]: for i in range(0, 5):
             plot_state(idx=idx+i)
```



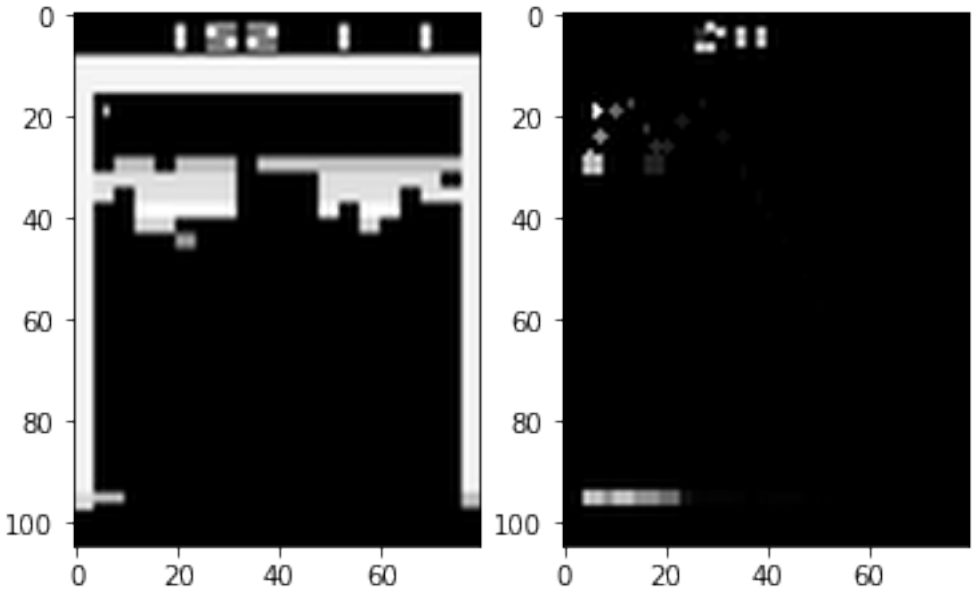

```
Action:     Q-Value:
====================
NOOP        3.216
FIRE        2.729
RIGHT       0.176
LEFT        6.751 (Action Taken)
```



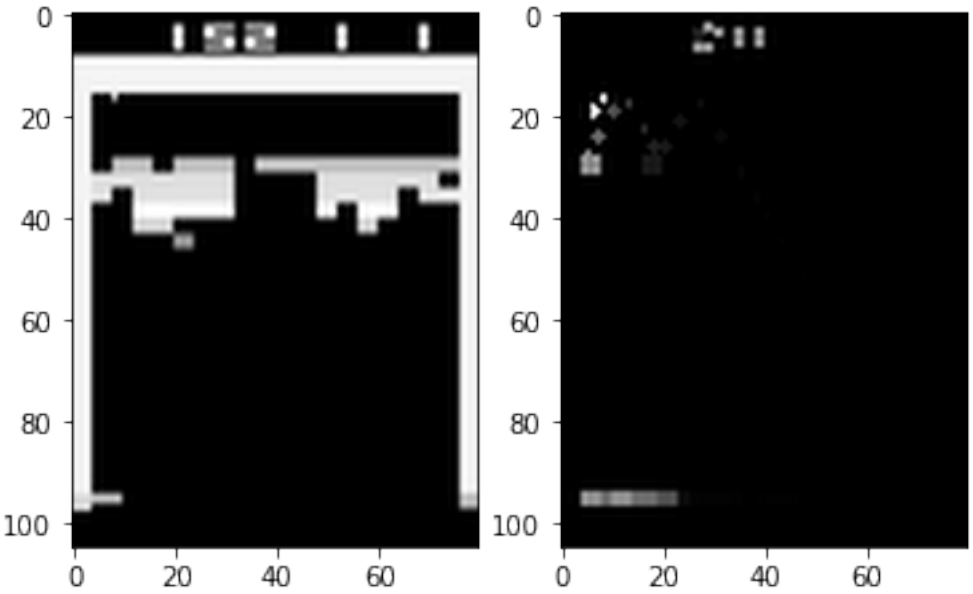

```
Action:     Q-Value:
====================
NOOP        2.236
FIRE        6.294 (Action Taken)
RIGHT       0.005
LEFT        3.473
```



```
Action:     Q-Value:
====================
NOOP        -0.016
FIRE        5.218 (Action Taken)
RIGHT       0.425
LEFT        2.046
```




```
Action:     Q-Value:
====================
NOOP        -0.195
FIRE        5.551 (Action Taken)
RIGHT       -0.107
LEFT        1.624
```

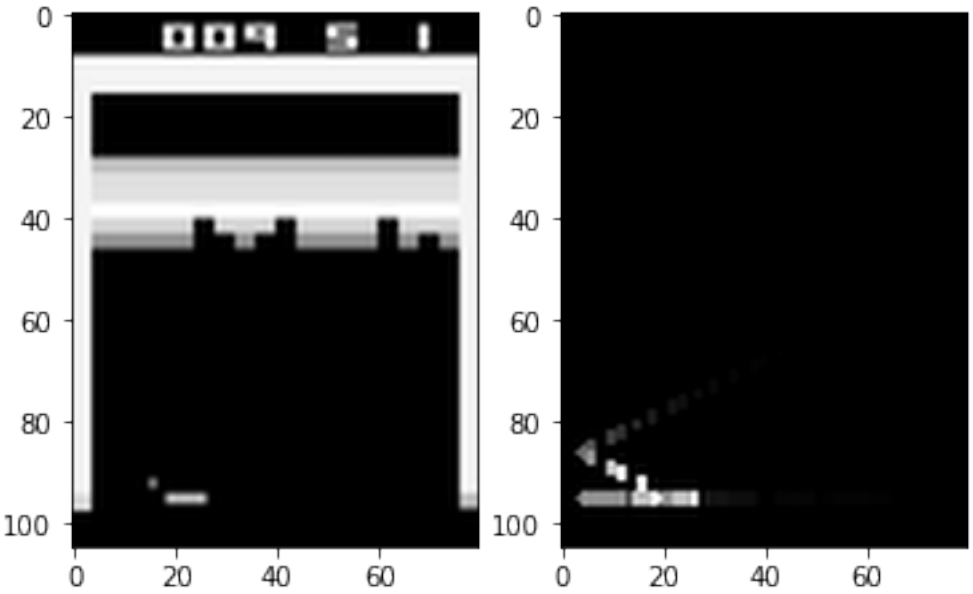



```
Action:     Q-Value:
====================
NOOP        0.800
FIRE        5.704 (Action Taken)
RIGHT       -0.312
LEFT        4.068
```

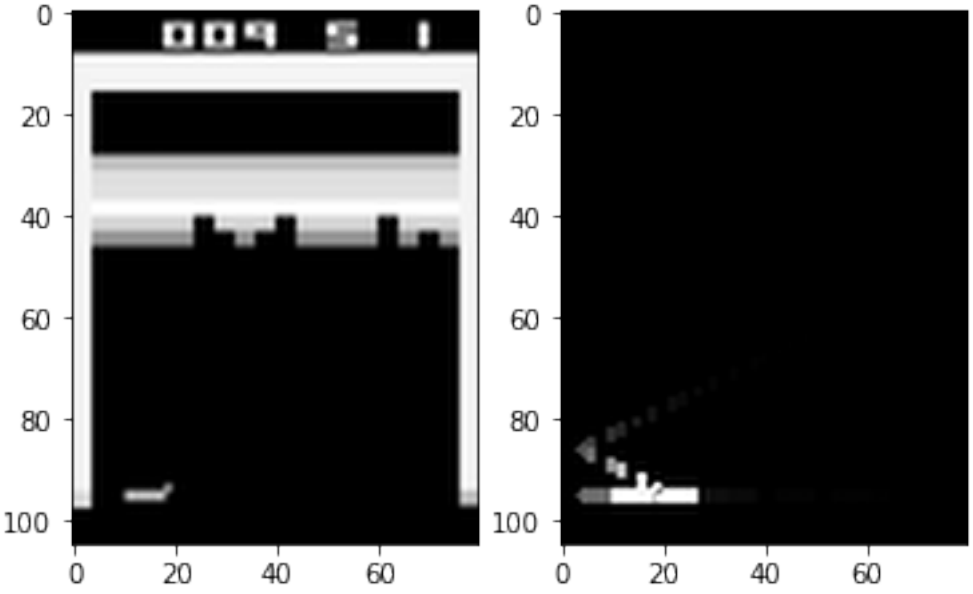
2.16.3 Example: Loss of Life

This example shows the states leading up to a loss of life for the agent.

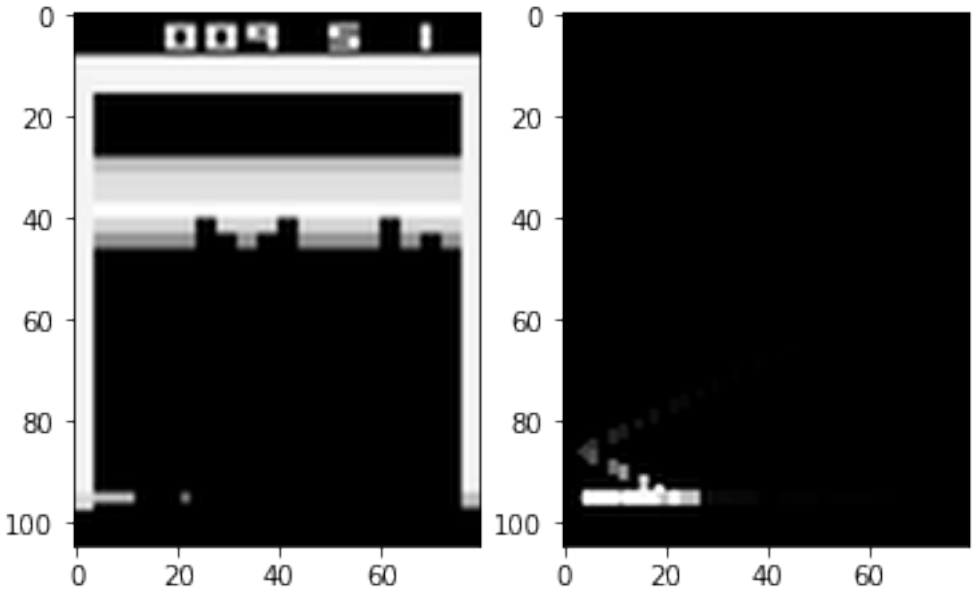
```
In [29]: idx = np.argmax(replay_memory.end_life)
         idx

Out[29]: 412

In [30]: for i in range(-10, 0):
             plot_state(idx=idx+i)
```



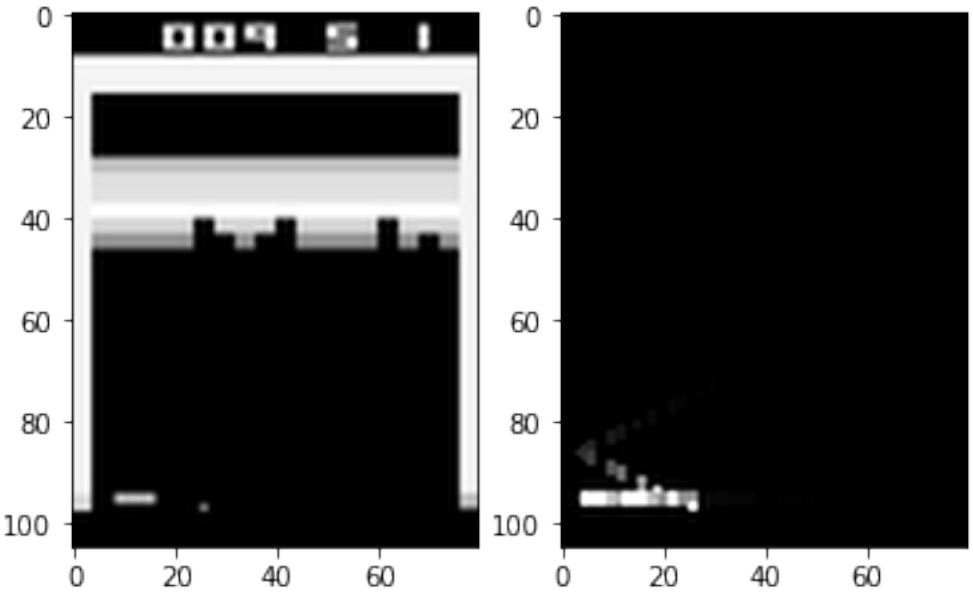

```
Action:     Q-Value:
====================
NOOP        0.531
FIRE        0.416
RIGHT       0.478
LEFT        0.547 (Action Taken)
```



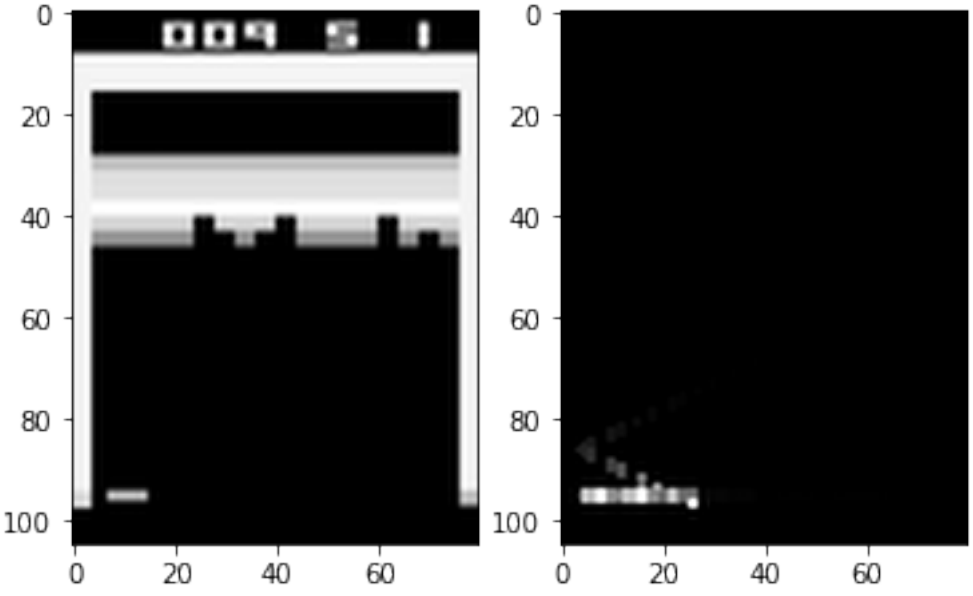
```
Action:     Q-Value:
====================
NOOP        0.479
FIRE        0.379
RIGHT       0.479
LEFT        0.497 (Action Taken)
```



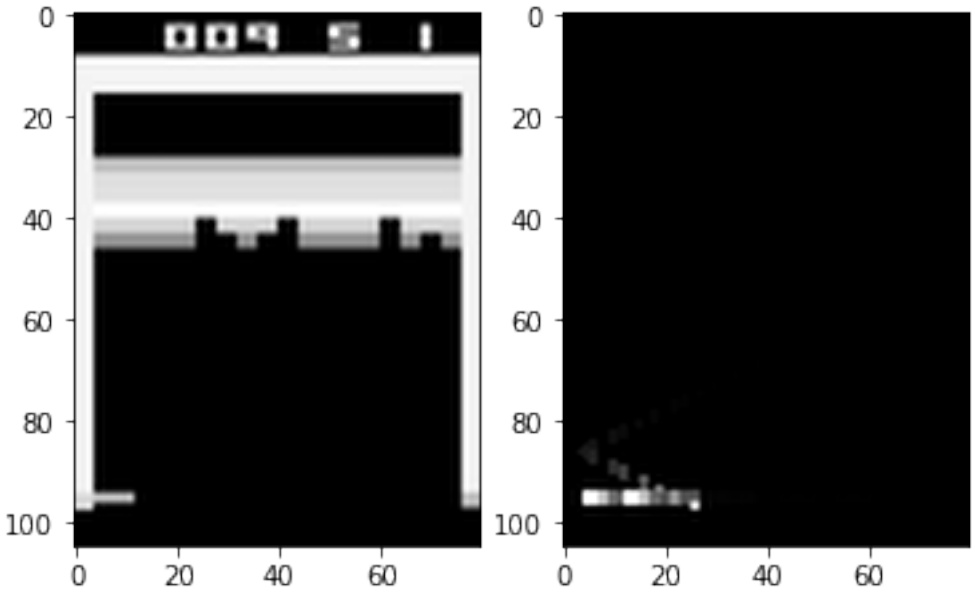

```
Action:     Q-Value:
====================
NOOP        0.346
FIRE        0.170
RIGHT       0.497 (Action Taken)
LEFT        0.262
```



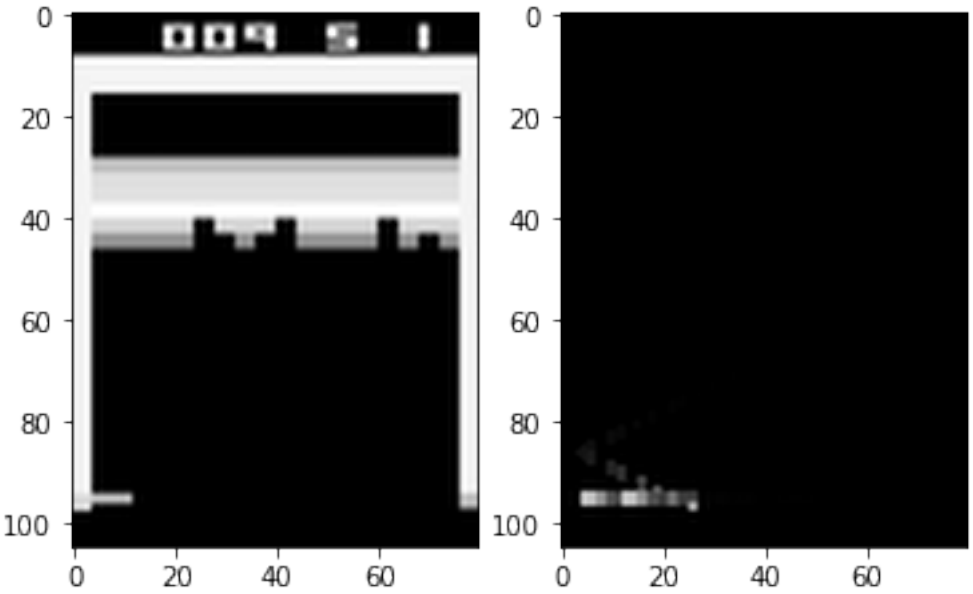

```
Action:     Q-Value:
====================
NOOP        0.147
FIRE        0.146
RIGHT       0.100
LEFT        0.151 (Action Taken)
```



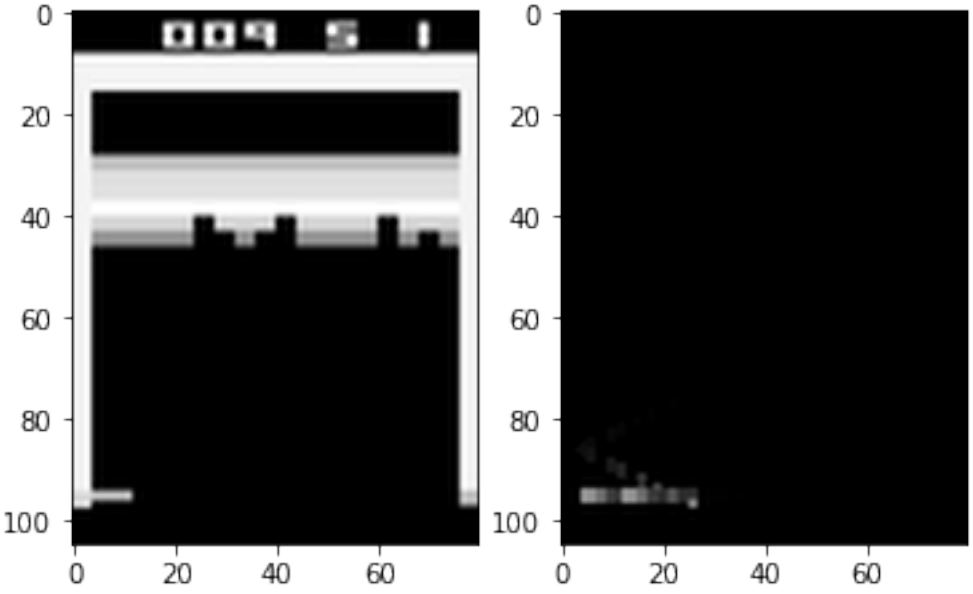
```
Action:     Q-Value:
====================
NOOP        0.181 (Action Taken)
FIRE        0.151
RIGHT       0.153
LEFT        0.174
```



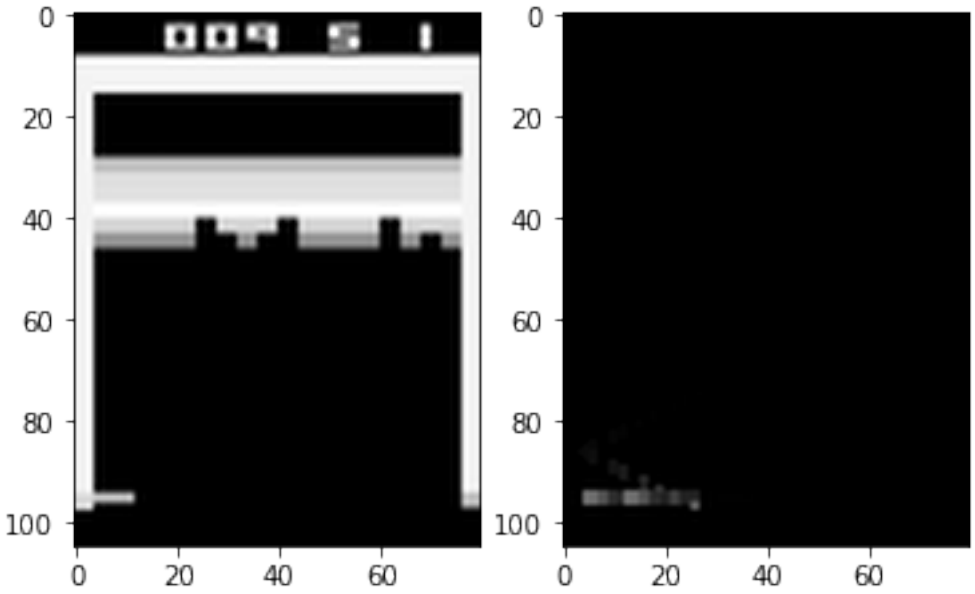

```
Action:     Q-Value:
====================
NOOP        0.155
FIRE        0.106
RIGHT       0.136
LEFT        0.179 (Action Taken)
```



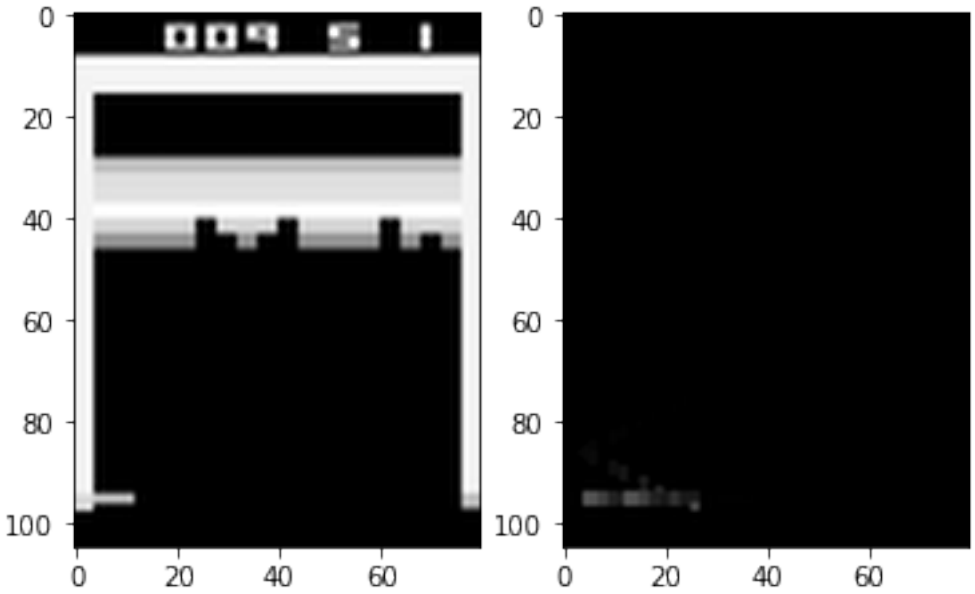

```
Action:     Q-Value:
====================
NOOP        0.119
FIRE        0.076
RIGHT       0.112
LEFT        0.137 (Action Taken)
```



```
Action:     Q-Value:
====================
NOOP        0.064
FIRE        0.031
RIGHT       0.060
LEFT        0.088 (Action Taken)
```




```
Action:     Q-Value:
====================
NOOP        0.031
FIRE        0.012
RIGHT       0.043
LEFT        0.064 (Action Taken)
```

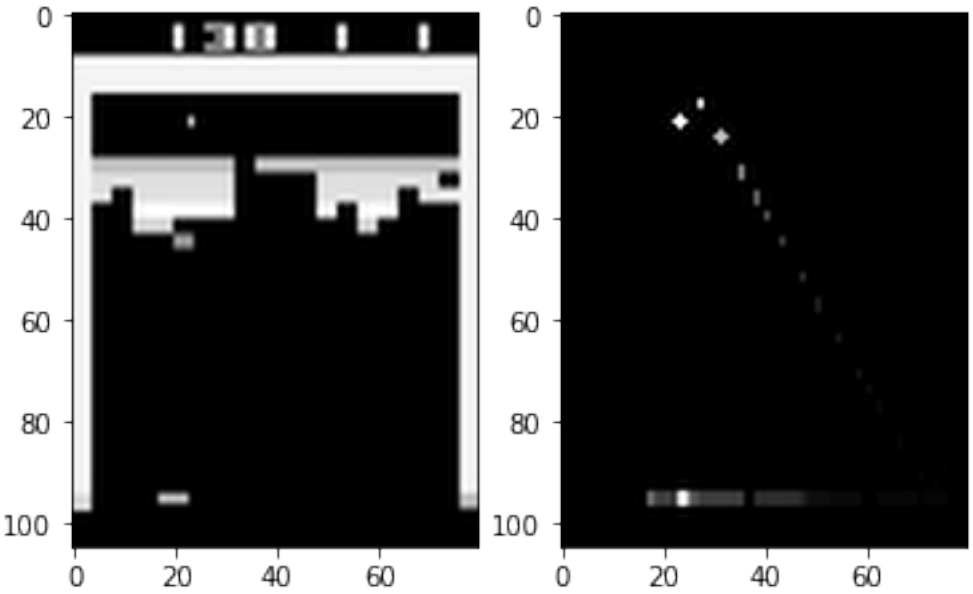



```
Action:     Q-Value:
====================
NOOP        0.042
FIRE        0.016
RIGHT       0.054
LEFT        0.082 (Action Taken)
```

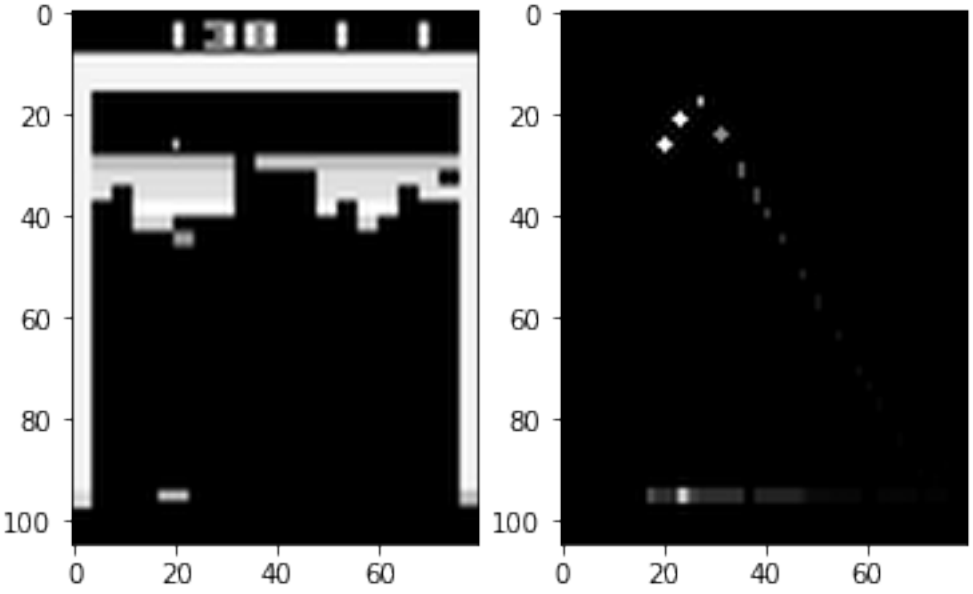
2.16.4 Example: Greatest Difference in Q-Values

This example shows the state where there is the greatest difference in Q-values, which means that the agent believes one action will be much more beneficial than another. But because the agent uses the Epsilon-greedy policy, it sometimes selects a random action instead.

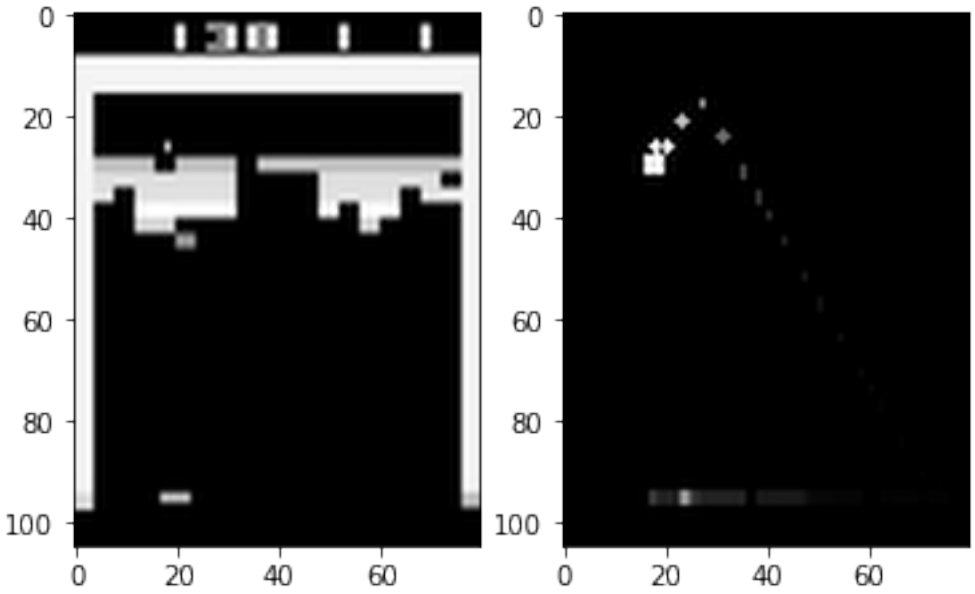
```
In [31]: idx = np.argmax(q_values_dif)
         idx

Out[31]: 2101

In [32]: for i in range(0, 5):
             plot_state(idx=idx+i)
```



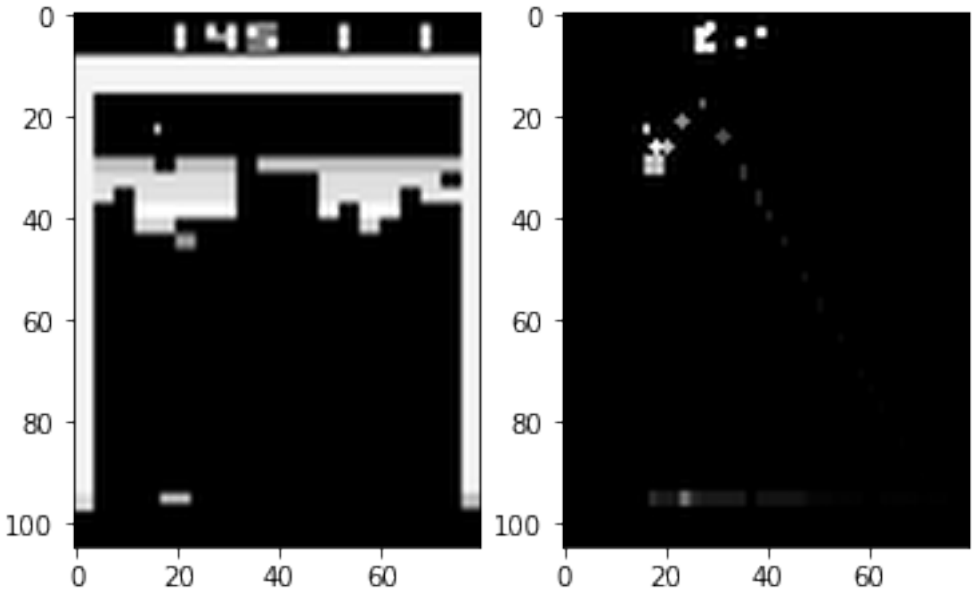

```
Action:     Q-Value:
====================
NOOP        6.126 (Action Taken)
FIRE        -2.656
RIGHT       -1.663
LEFT        0.403
```



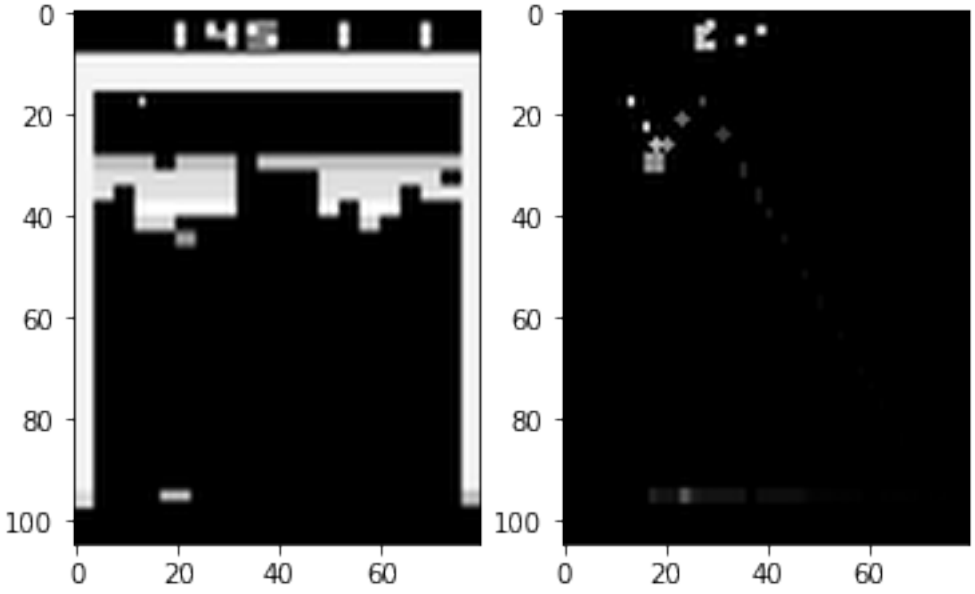
```
Action:     Q-Value:
====================
NOOP        6.730 (Action Taken)
FIRE        -2.017
RIGHT       -1.037
LEFT        0.698
```




```
Action:     Q-Value:
====================
NOOP        5.098 (Action Taken)
FIRE        -0.543
RIGHT       -0.231
LEFT        1.863
```




```
Action:     Q-Value:
====================
NOOP        -0.023
FIRE        5.527 (Action Taken)
RIGHT       0.325
LEFT        2.851
```



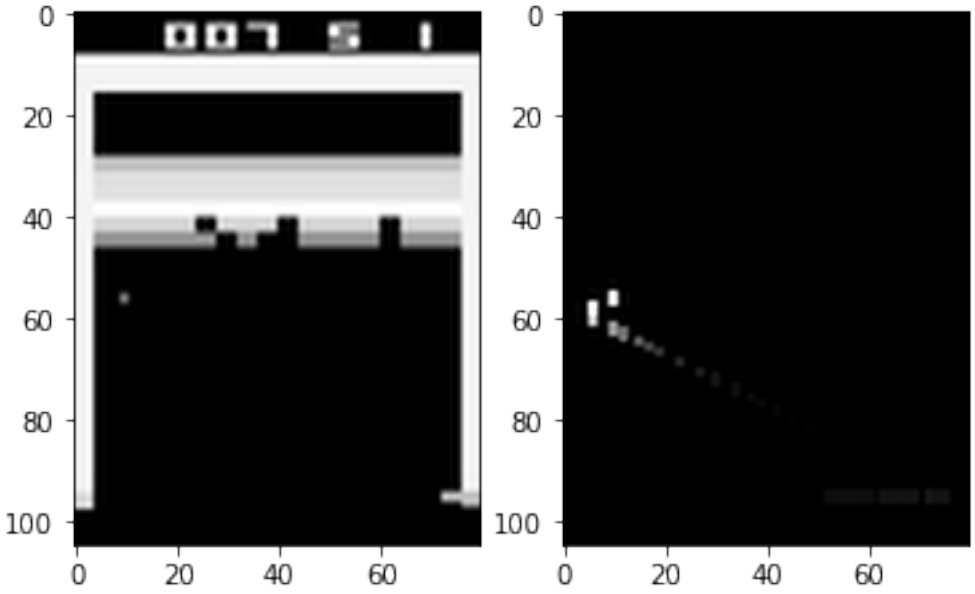
```
Action:     Q-Value:
====================
NOOP        0.269
FIRE        5.562 (Action Taken)
RIGHT       0.243
LEFT        3.503
```

2.16.5 Example: Smallest Difference in Q-Values

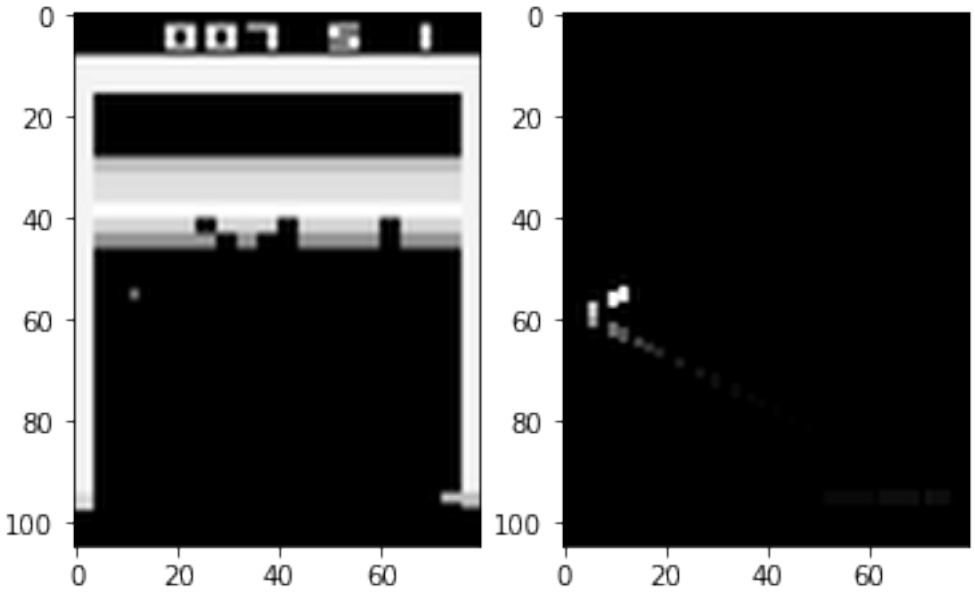
This example shows the state where there is the smallest difference in Q-values, which means that the agent believes it does not really matter which action it selects, as they all have roughly the same expectations for future rewards.

The Neural Network estimates these Q-values and they are not precise. The differences in Q-values may be so small that they fall within the error-range of the estimates.

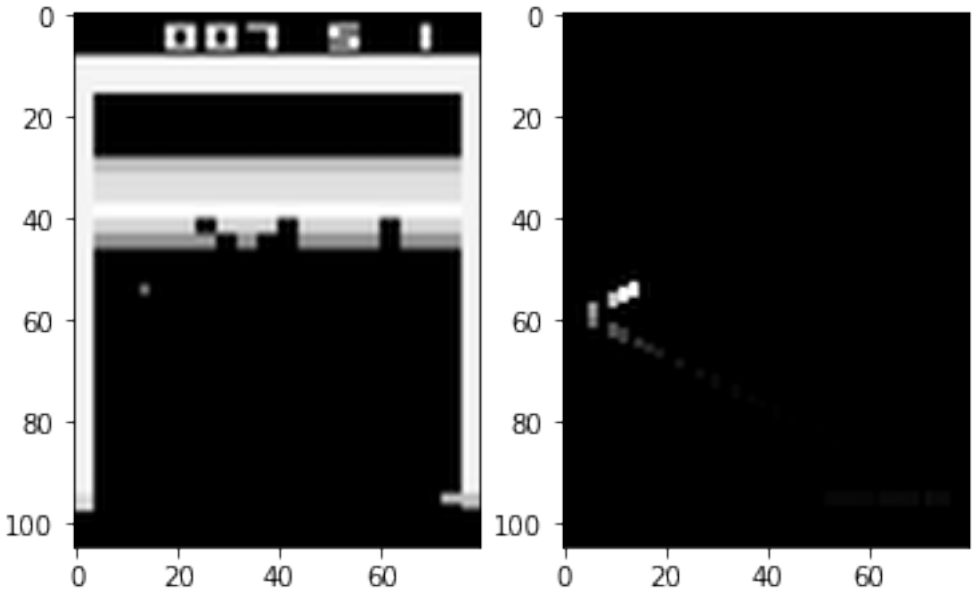
```
In [33]: idx = np.argmin(q_values_dif)
         idx

Out[33]: 304

In [34]: for i in range(0, 5):
             plot_state(idx=idx+i)
```



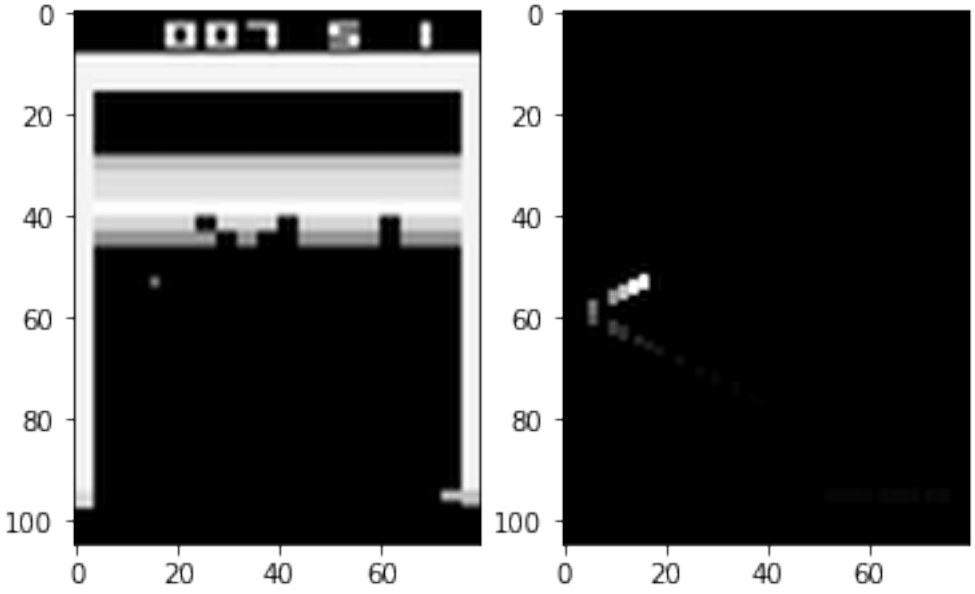
```
Action:     Q-Value:
====================
NOOP        1.139
FIRE        1.142
RIGHT       1.143 (Action Taken)
LEFT        1.140
```



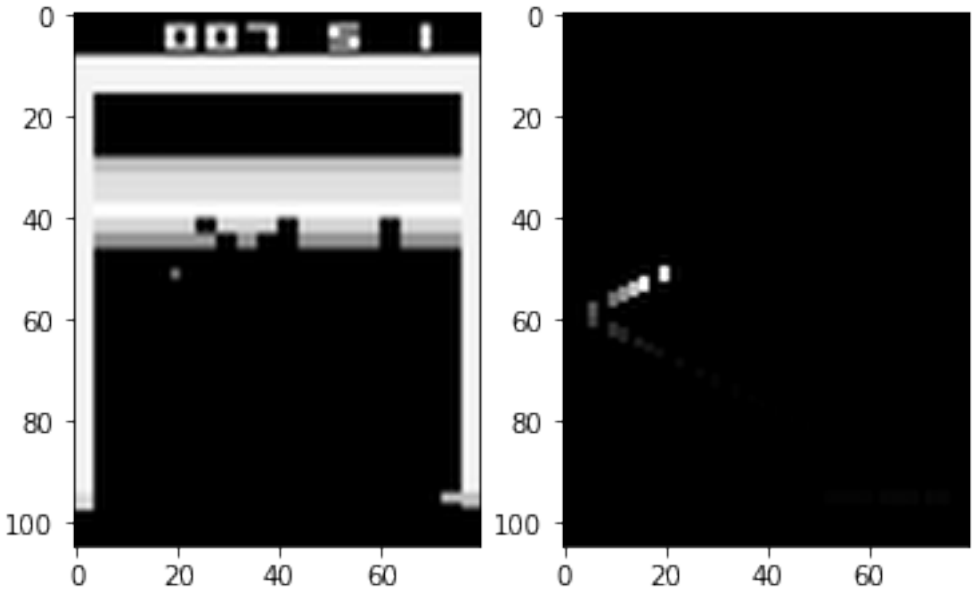

```
Action:     Q-Value:
====================
NOOP        1.147
FIRE        1.121
RIGHT       1.180 (Action Taken)
LEFT        1.179
```




```
Action:     Q-Value:
====================
NOOP        1.194
FIRE        1.104
RIGHT       1.230 (Action Taken)
LEFT        1.188
```



```
Action:     Q-Value:
====================
NOOP        1.124
FIRE        1.073
RIGHT       1.227 (Action Taken)
LEFT        1.090
```




```
Action:     Q-Value:
====================
NOOP        1.176
FIRE        1.165
RIGHT       1.275 (Action Taken)
LEFT        1.149
```

2.17 Output of Convolutional Layers

The outputs of the convolutional layers can be plotted so we can see how the images from the game-environment are being processed by the Neural Network.

This is the helper-function for plotting the output of the convolutional layer with the given name, when inputting the given state from the replay-memory.

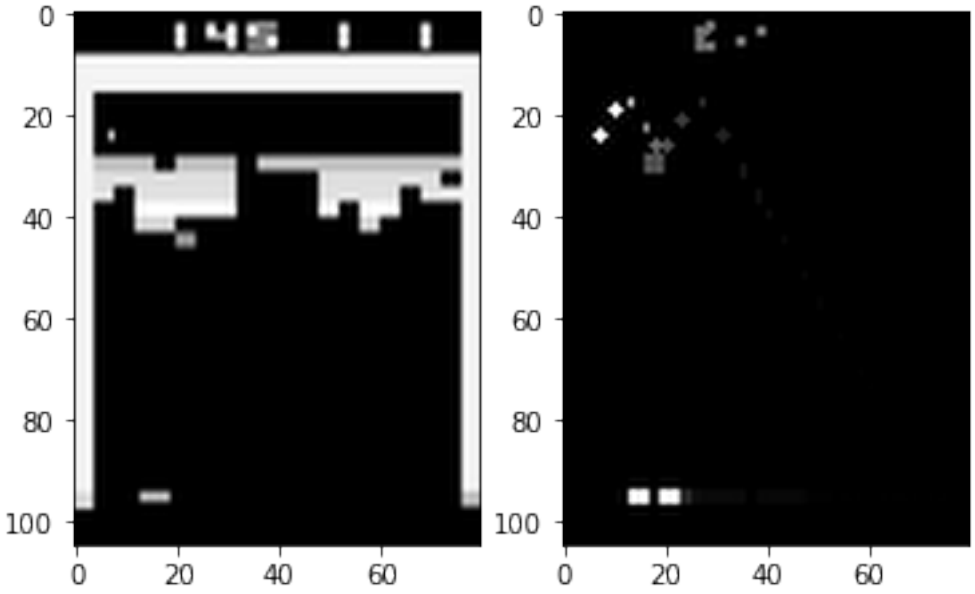
```
In [35]: def plot_layer_output(model, layer_name, state_index, inverse_cmap=False):
             """
             Plot the output of a convolutional layer.

             :param model: An instance of the NeuralNetwork-class.
             :param layer_name: Name of the convolutional layer.
             :param state_index: Index into the replay-memory for a state that
                                 will be input to the Neural Network.
             :param inverse_cmap: Boolean whether to inverse the color-map.
             """

             # Get the given state-array from the replay-memory.
             state = replay_memory.states[state_index]

             # Get the output tensor for the given layer inside the TensorFlow graph.
             # This is not the value-contents but merely a reference to the tensor.
             layer_tensor = model.get_layer_tensor(layer_name=layer_name)

             # Get the actual value of the tensor by feeding the state-data
             # to the TensorFlow graph and calculating the value of the tensor.
             values = model.get_tensor_value(tensor=layer_tensor, state=state)

             # Number of image channels output by the convolutional layer.
             num_images = values.shape[3]

             # Number of grid-cells to plot.
             # Rounded-up, square-root of the number of filters.
             num_grids = math.ceil(math.sqrt(num_images))

             # Create figure with a grid of sub-plots.
             fig, axes = plt.subplots(num_grids, num_grids, figsize=(10, 10))

             print("Dim. of each image:", values.shape)

             if inverse_cmap:
                 cmap = 'gray_r'
             else:
                 cmap = 'gray'

             # Plot the outputs of all the channels in the conv-layer.
             for i, ax in enumerate(axes.flat):
                 # Only plot the valid image-channels.
                 if i < num_images:
                     # Get the image for the i'th output channel.
                     img = values[0, :, :, i]

                     # Plot image.
                     ax.imshow(img, interpolation='nearest', cmap=cmap)
```

```
        # Remove ticks from the plot.
        ax.set_xticks([])
        ax.set_yticks([])

    # Ensure the plot is shown correctly with multiple plots
    # in a single Notebook cell.
    plt.show()
```

2.17.1 Game State

This is the state that is being input to the Neural Network. The image on the left is the last image from the game-environment. The image on the right is the processed motion-trace that shows the trajectories of objects in the game-environment.

```
In [36]: idx = np.argmax(q_values_max)
         plot_state(idx=idx, print_q=False)
```



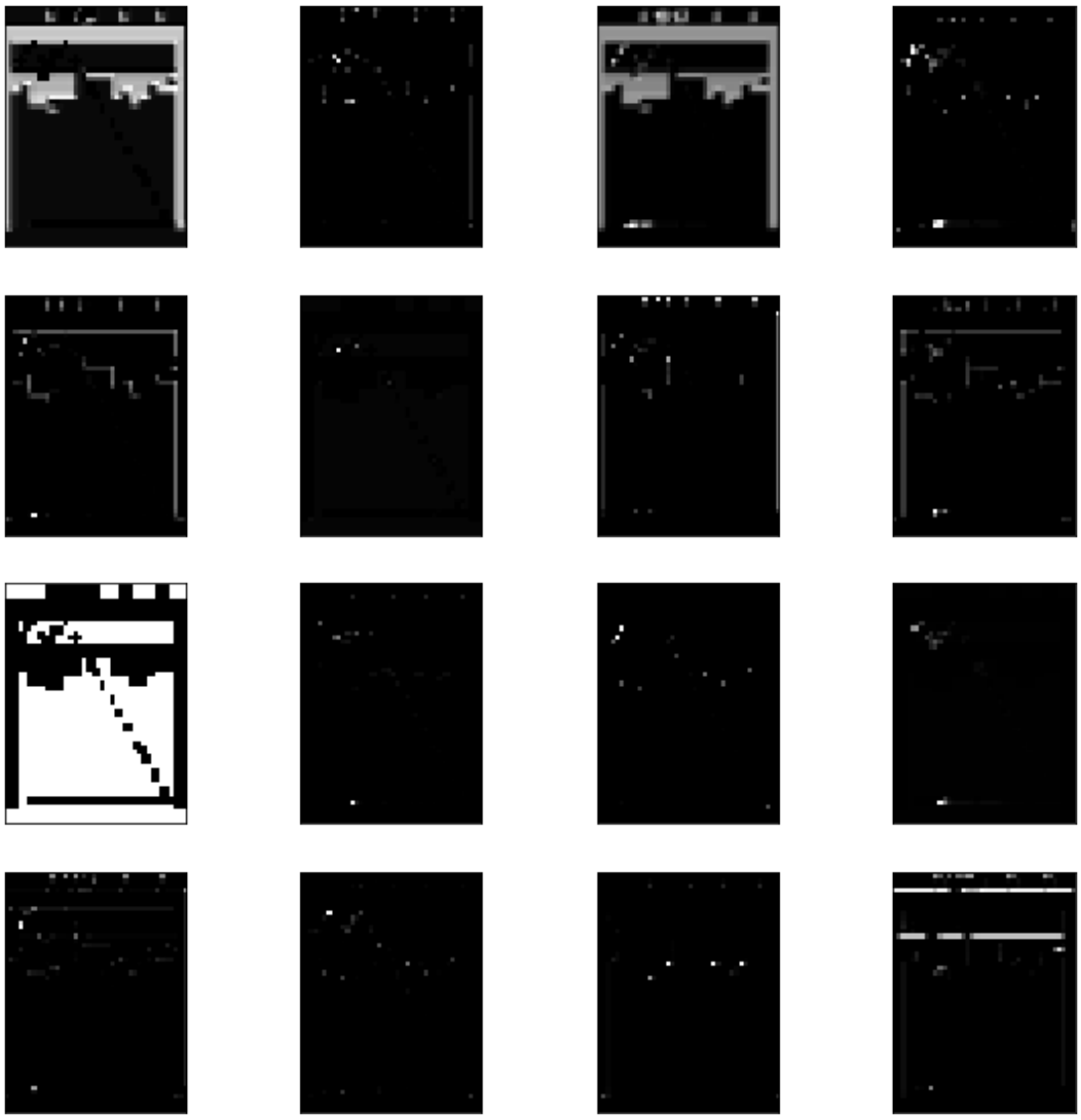
2.17.2 Output of Convolutional Layer 1

This shows the images that are output by the 1st convolutional layer, when inputting the above state to the Neural Network. There are 16 output channels of this convolutional layer.

Note that you can invert the colors by setting `inverse_cmap=True` in the parameters to this function.

```
In [37]: plot_layer_output(model=model, layer_name='layer_conv1', state_index=idx, inverse_cma
```

```
Dim. of each image: (1, 53, 40, 16)
```

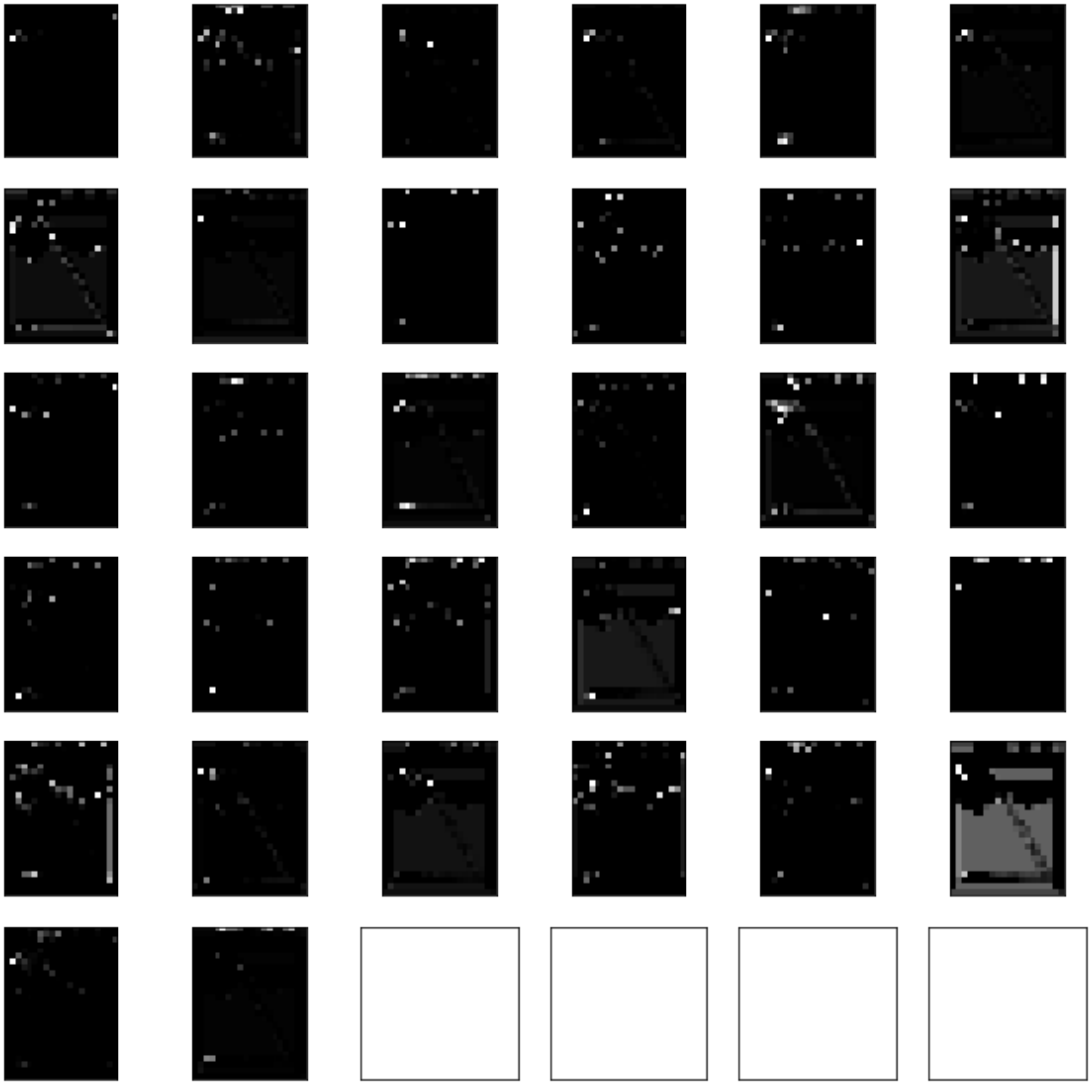


2.17.3 Output of Convolutional Layer 2

These are the images output by the 2nd convolutional layer, when inputting the above state to the Neural Network. There are 32 output channels of this convolutional layer.

```
In [38]: plot_layer_output(model=model, layer_name='layer_conv2', state_index=idx, inverse_cma
```

```
Dim. of each image: (1, 27, 20, 32)
```



2.17.4 Output of Convolutional Layer 3

These are the images output by the 3rd convolutional layer, when inputting the above state to the Neural Network. There are 64 output channels of this convolutional layer.

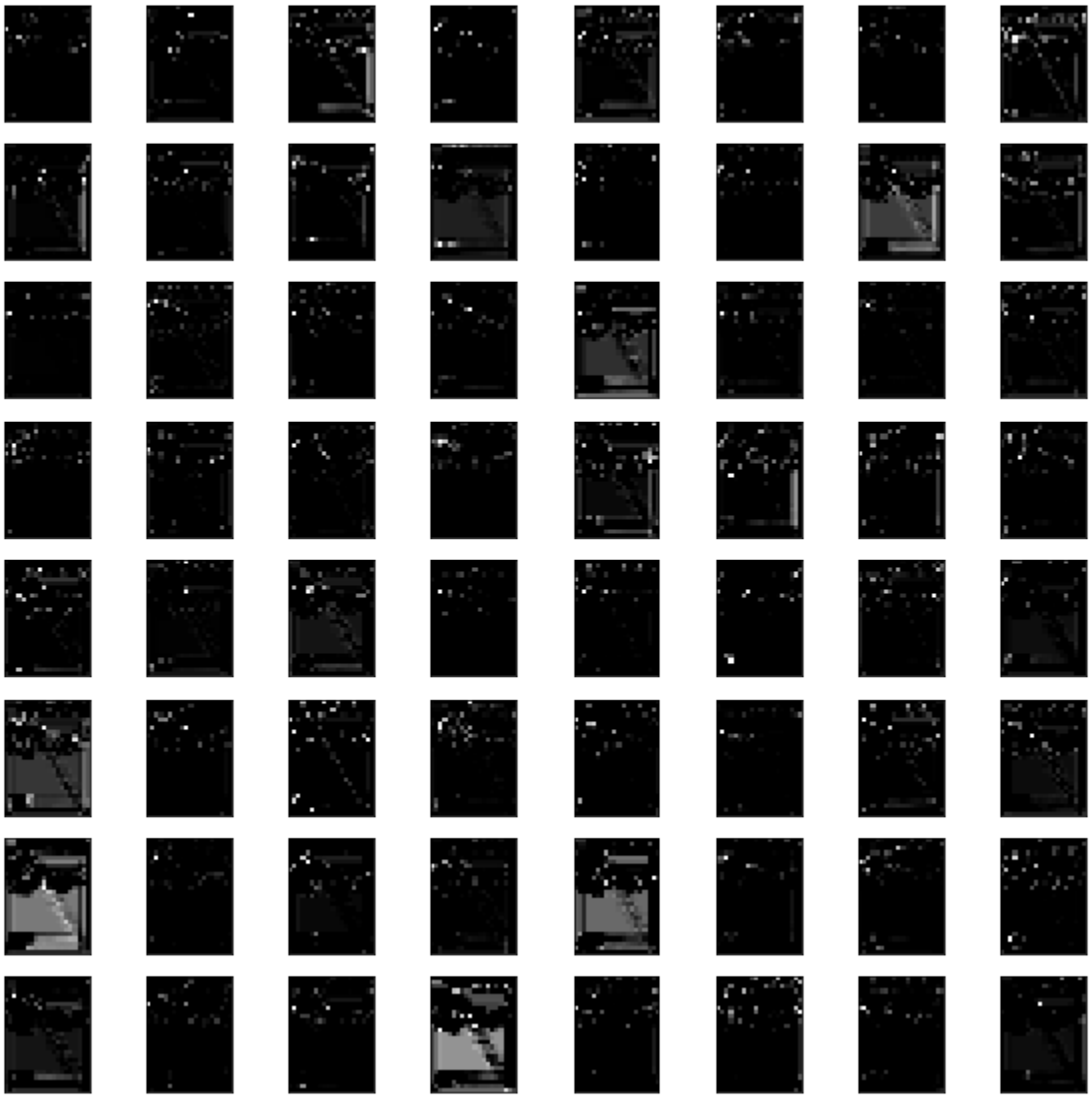
All these images are flattened to a one-dimensional array (or tensor) which is then used as the input to a fully-connected layer in the Neural Network.

During the training-process, the Neural Network has learnt what convolutional filters to apply to the images from the game-environment so as to produce these images, because they have proven to be useful when estimating Q-values.

Can you see what it is that the Neural Network has learned to detect in these images?

```
In [39]: plot_layer_output(model=model, layer_name='layer_conv3', state_index=idx, inverse_cma
```

```
Dim. of each image: (1, 27, 20, 64)
```



2.18 Weights for Convolutional Layers

We can also plot the weights of the convolutional layers in the Neural Network. These are the weights that are being optimized so as to improve the ability of the Neural Network to estimate Q-values. Tutorial #02 explains in greater detail what convolutional weights are. There are also weights for the fully-connected layers but they are not shown here.

This is the helper-function for plotting the weights of a convoluational layer.

```
In [40]: def plot_conv_weights(model, layer_name, input_channel=0):
             """
```

```
    Plot the weights for a convolutional layer.

    :param model: An instance of the NeuralNetwork-class.
    :param layer_name: Name of the convolutional layer.
    :param input_channel: Plot the weights for this input-channel.
    """

    # Get the variable for the weights of the given layer.
    # This is a reference to the variable inside TensorFlow,
    # not its actual value.
    weights_variable = model.get_weights_variable(layer_name=layer_name)

    # Retrieve the values of the weight-variable from TensorFlow.
    # The format of this 4-dim tensor is determined by the
    # TensorFlow API. See Tutorial #02 for more details.
    w = model.get_variable_value(variable=weights_variable)

    # Get the weights for the given input-channel.
    w_channel = w[:, :, input_channel, :]

    # Number of output-channels for the conv. layer.
    num_output_channels = w_channel.shape[2]

    # Get the lowest and highest values for the weights.
    # This is used to correct the colour intensity across
    # the images so they can be compared with each other.
    w_min = np.min(w_channel)
    w_max = np.max(w_channel)

    # This is used to center the colour intensity at zero.
    abs_max = max(abs(w_min), abs(w_max))

    # Print statistics for the weights.
    print("Min:  {0:.5f}, Max:   {1:.5f}".format(w_min, w_max))
    print("Mean: {0:.5f}, Stdev: {1:.5f}".format(w_channel.mean(),
                                                 w_channel.std()))

    # Number of grids to plot.
    # Rounded-up, square-root of the number of output-channels.
    num_grids = math.ceil(math.sqrt(num_output_channels))

    # Create figure with a grid of sub-plots.
    fig, axes = plt.subplots(num_grids, num_grids)

    # Plot all the filter-weights.
    for i, ax in enumerate(axes.flat):
        # Only plot the valid filter-weights.
        if i < num_output_channels:
```

```
            # Get the weights for the i'th filter of this input-channel.
            img = w_channel[:, :, i]

            # Plot image.
            ax.imshow(img, vmin=-abs_max, vmax=abs_max,
                      interpolation='nearest', cmap='seismic')

        # Remove ticks from the plot.
        ax.set_xticks([])
        ax.set_yticks([])

    # Ensure the plot is shown correctly with multiple plots
    # in a single Notebook cell.
    plt.show()
```

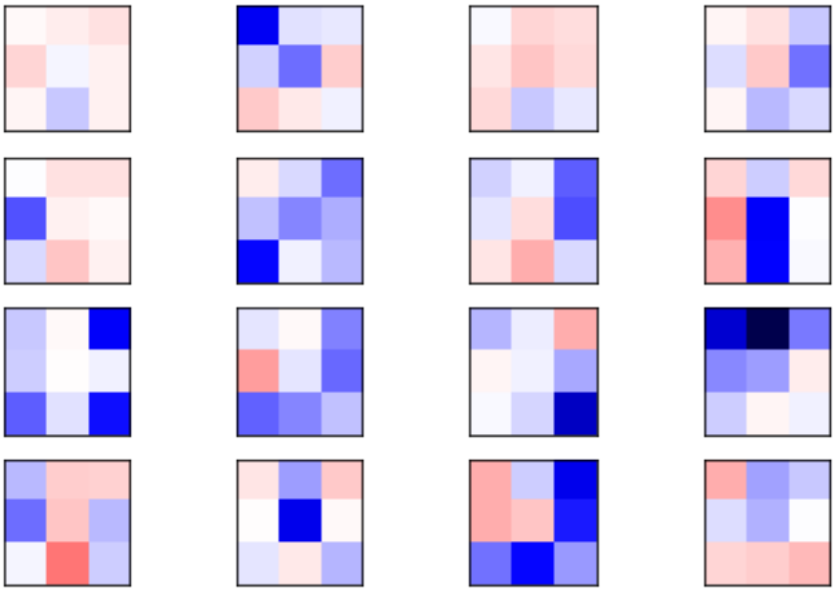
2.18.1 Weights for Convolutional Layer 1

These are the weights of the first convolutional layer of the Neural Network, with respect to the first input channel of the state. That is, these are the weights that are used on the image from the game-environment. Some basic statistics are also shown.

Note how the weights are more negative (blue) than positive (red). It is unclear why this happens as these weights are found through optimization. It is apparently beneficial for the following layers to have this processing with more negative weights in the first convolutional layer.

```
In [41]: plot_conv_weights(model=model, layer_name='layer_conv1', input_channel=0)
```

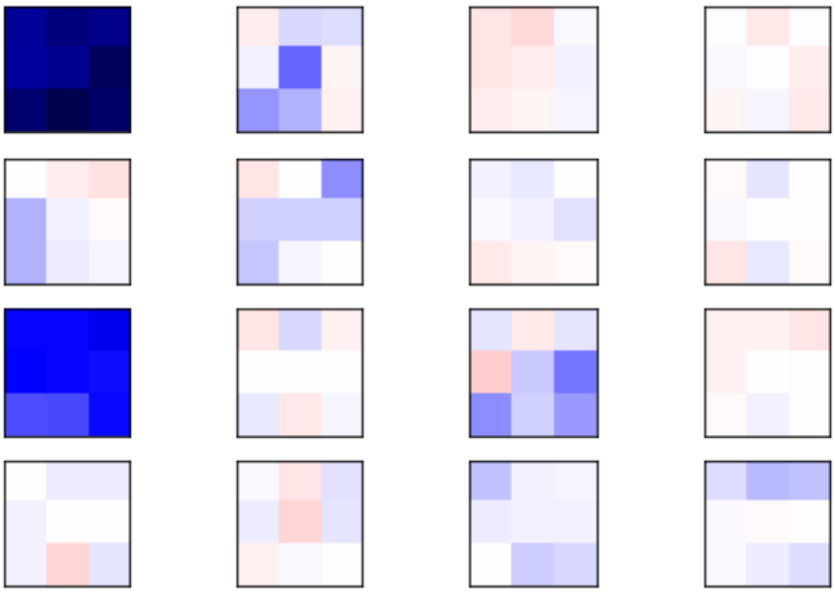
```
Min:  -0.40275, Max:   0.10862
Mean: -0.03399, Stdev: 0.08055
```



We can also plot the convolutional weights for the second input channel, that is, the motion-trace of the game-environment. Once again we see that the negative weights (blue) have a much greater magnitude than the positive weights (red).

```
In [42]: plot_conv_weights(model=model, layer_name='layer_conv1', input_channel=1)
```

```
Min:  -0.69614, Max:   0.06959
Mean: -0.07312, Stdev: 0.16320
```



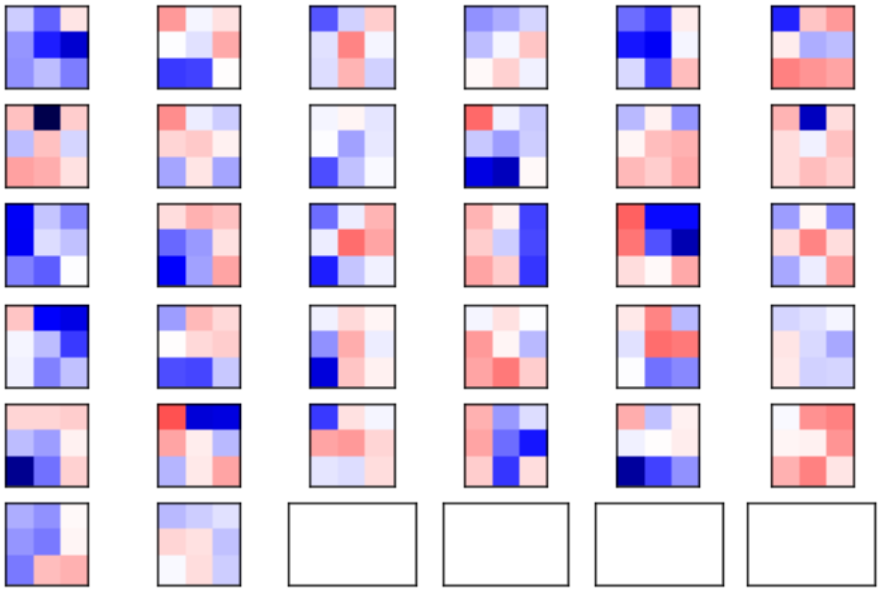
2.18.2 Weights for Convolutional Layer 2

These are the weights of the 2nd convolutional layer in the Neural Network. There are 16 input channels and 32 output channels of this layer. You can change the number for the input-channel to see the associated weights.

Note how the weights are more balanced between positive (red) and negative (blue) compared to the weights for the 1st convolutional layer above.

```
In [43]: plot_conv_weights(model=model, layer_name='layer_conv2', input_channel=0)
```

```
Min:  -0.50262, Max:   0.17076
Mean: -0.03282, Stdev: 0.11310
```



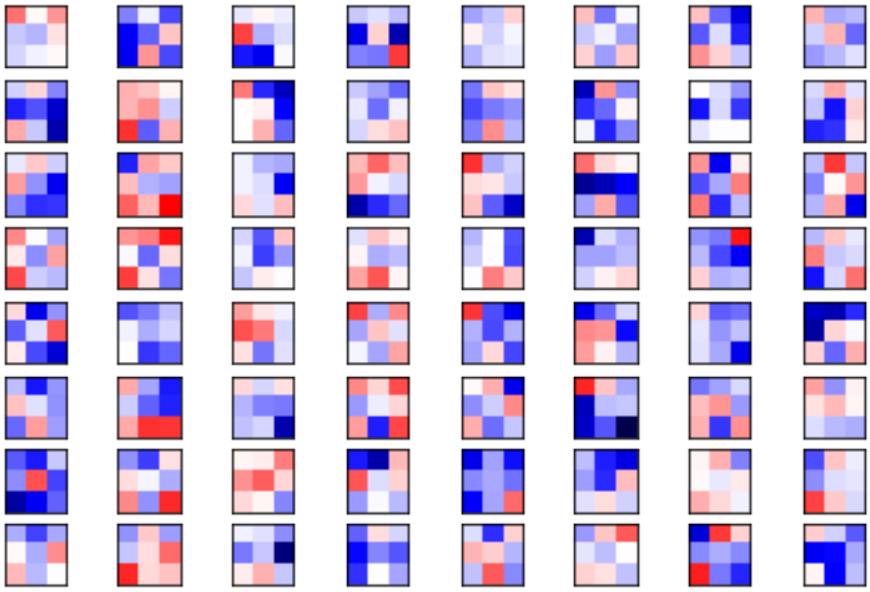
2.18.3 Weights for Convolutional Layer 3

These are the weights of the 3rd convolutional layer in the Neural Network. There are 32 input channels and 64 output channels of this layer. You can change the number for the input-channel to see the associated weights.

Note again how the weights are more balanced between positive (red) and negative (blue) compared to the weights for the 1st convolutional layer above.

```
In [44]: plot_conv_weights(model=model, layer_name='layer_conv3', input_channel=0)
```

```
Min:  -0.34386, Max:   0.17371
Mean: -0.03366, Stdev: 0.08802
```



2.19 Discussion

We trained an agent to play old Atari games quite well using Reinforcement Learning. Recent improvements to the training algorithm have improved the performance significantly. But is this true human-like intelligence? The answer is clearly NO!

Reinforcement Learning in its current form is a crude numerical algorithm for connecting visual images, actions, rewards and penalties when there is a time-lag between the signals. The learning is based on trial-and-error and cannot do logical reasoning like a human. The agent has no sense of "self" while a human has an understanding of what part of the game-environment it is controlling, so a human can reason logically like this: "(A) I control the paddle, and (B) I must avoid dying which happens when the ball flies past the paddle, so (C) I must move the paddle to hit the ball, and (D) this automatically scores points when the ball smashes bricks in the wall". A human would first learn these basic logical rules of the game - and then try and refine the eye-hand coordination to play the game better. Reinforcement Learning has no real comprehension of what is going on in the game and merely works on improving the eye-hand coordination until it gets lucky and does the right thing to score more points.

Furthermore, the training of the Reinforcement Learning algorithm required almost 150 hours of computation which played the game at high speeds. If the game was played at normal real-time speeds then it would have taken more than 1700 hours to train the agent, which is more than 70 days and nights.

Logical reasoning would allow for much faster learning than Reinforcement Learning, and it would be able to solve much more complicated problems than simple eye-hand coordination. I am skeptical if someone will be able to create true human-like intelligence from Reinforcement Learning algorithms.

Does that mean Reinforcement Learning is completely worthless? No, it has real-world applications that currently cannot be solved by other methods.

Another point of criticism is the use of Neural Networks. The majority of the research in Reinforcement Learning is actually spent on trying to stabilize the training of the Neural Network using various tricks. This is a waste of research time and strongly indicates that Neural Networks may not be a very good Machine Learning model compared to the human brain.

2.20 Exercises & Research Ideas

Below are suggestions for exercises and experiments that may help improve your skills with TensorFlow and Reinforcement Learning. Some of these ideas can easily be extended into full research problems that would help the community if you can solve them.

You should keep a log of your experiments, describing for each experiment the settings you tried and the results. You should also save the source-code and checkpoints / log-files.

It takes so much time to run these experiments, so please share your results with the rest of the community. Even if an experiment failed to produce anything useful, it will be helpful to others so they know not to redo the same experiment.

Thread on GitHub for discussing these experiments

You may want to backup this Notebook and the other files before making any changes.

You may find it helpful to add more command-line parameters to `reinforcement_learning.py` so you don't have to edit the source-code for testing other parameters.

- Change the epsilon-probability during testing to e.g. 0.001 or 0.05. Which gives the best results? Could you use this value during training? Why/not?
- Continue training the agent for the Breakout game using the downloaded checkpoint. Does the agent get better or worse the more you train it? Why? (You should run it in a terminal window as described above.)
- Try and change the game-environment to Space Invaders and re-run this Notebook. The checkpoint can be downloaded automatically. It was trained for about 150 hours, which is roughly the same as for Breakout, but note that it has processed far fewer states. The reason is that the hyper-parameters such as the learning-rate were tuned for Breakout. Can you make some kind of adaptive learning-rate that would work better for both Breakout and Space Invaders? What about the other hyper-parameters? What about other games?
- Try different architectures for the Neural Network. You will need to restart the training because the checkpoints cannot be reused for other architectures. You will need to train the agent for several days with each new architecture so as to properly assess its performance.
- The replay-memory throws away all data after optimization of the Neural Network. Can you make it reuse the data somehow? The ReplayMemory-class has the function `estimate_all_q_values()` which may be helpful.
- The reward is limited to -1 and 1 in the function `ReplayMemory.add()` so as to stabilize the training. This means the agent cannot distinguish between small and large rewards. Can you use batch normalization to fix this problem, so you can use the actual reward values?
- Can you improve the training by adding L2-regularization or dropout?
- Try using other optimizers for the Neural Network. Does it help with the training speed or stability?
- Let the agent take up to 30 random actions at the beginning of each new episode. This is used in some research papers to further randomize the game-environment, so the agent cannot memorize the first sequence of actions.

- Try and save the game at regular intervals. If the agent dies, then you can reload the last saved game. Would this help training the agent faster and better, because it does not need to play the game from the beginning?
- There are some invalid actions available to the agent in OpenAI Gym. Does it improve the training if you only allow the valid actions from the game-environment?
- Does the MotionTracer work for other games? Can you improve on the MotionTracer?
- Try and use the last 4 image-frames from the game instead of the MotionTracer.
- Try larger and smaller sizes for the replay memory.
- Try larger and smaller discount rates for updating the Q-values.
- If you look closely in the states and actions that are display above, you will note that the agent has sometimes taken actions that do not correspond to the movement of the paddle. For example, the action might be LEFT but the paddle has either not moved at all, or it has moved right instead. Is this a bug in the source-code for this tutorial, or is it a bug in OpenAI Gym, or is it a bug in the underlying Atari Learning Environment? Does it matter?

2.21 License (MIT)