Deep Q-Learning

In [2]:
```python
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-pyth
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all file


import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserv
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside o
```

```
/kaggle/input/mp5dataset-2/assignment5_materials/agent.py
/kaggle/input/mp5dataset-2/assignment5_materials/MP5_DQNLSTM_EC.ipynb
/kaggle/input/mp5dataset-2/assignment5_materials/memory.py
/kaggle/input/mp5dataset-2/assignment5_materials/model.py
/kaggle/input/mp5dataset-2/assignment5_materials/config.py
/kaggle/input/mp5dataset-2/assignment5_materials/agent_double.py
/kaggle/input/mp5dataset-2/assignment5_materials/utils.py
/kaggle/input/mp5dataset-2/assignment5_materials/MP5.ipynb
/kaggle/input/mp5dataset-2/assignment5_materials/agent_lstm.py
/kaggle/input/mp5dataset-2/assignment5_materials/.ipynb_checkpoints/MP5-checkpoint.ipynb
```

Install dependencies for AI gym to run properly (shouldn't take more than a minute). If running on google cloud or running locally, only need to run once. Colab may require installing everytime the vm shuts down.

In [3]:
```python
!pip3 install gym pyvirtualdisplay
!sudo apt-get install -y xvfb python-opengl ffmpeg
```

```
Requirement already satisfied: gym in /opt/conda/lib/python3.10/site-packages (0.26.2)
Collecting pyvirtualdisplay
  Downloading PyVirtualDisplay-3.0-py3-none-any.whl.metadata (943 bytes)
Requirement already satisfied: numpy>=1.18.0 in /opt/conda/lib/python3.10/site-packages
(from gym) (1.26.4)
Requirement already satisfied: cloudpickle>=1.2.0 in /opt/conda/lib/python3.10/site-pack
ages (from gym) (2.2.1)
Requirement already satisfied: gym-notices>=0.0.4 in /opt/conda/lib/python3.10/site-pack
ages (from gym) (0.0.8)
Downloading PyVirtualDisplay-3.0-py3-none-any.whl (15 kB)
Installing collected packages: pyvirtualdisplay
Successfully installed pyvirtualdisplay-3.0
Reading package lists... Done
Building dependency tree
Reading state information... Done
ffmpeg is already the newest version (7:4.2.7-0ubuntu0.1).
xvfb is already the newest version (2:1.20.13-1ubuntu1~20.04.17).
The following additional packages will be installed:
  freeglut3 libglu1-mesa libpython2-stdlib libpython2.7-minimal
  libpython2.7-stdlib python2 python2-minimal python2.7 python2.7-minimal
Suggested packages:
  python-tk python-numpy libgle3 python2-doc python2.7-doc binfmt-support
The following NEW packages will be installed:
```

```
  freeglut3 libglu1-mesa libpython2-stdlib libpython2.7-minimal
  libpython2.7-stdlib python-opengl python2 python2-minimal python2.7
  python2.7-minimal
0 upgraded, 10 newly installed, 0 to remove and 65 not upgraded.
Need to get 4540 kB of archives.
After this operation, 22.7 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu focal-updates/universe amd64 libpython2.7-minimal
amd64 2.7.18-1~20.04.4 [335 kB]
Get:2 http://archive.ubuntu.com/ubuntu focal-updates/universe amd64 python2.7-minimal am
d64 2.7.18-1~20.04.4 [1280 kB]
Get:3 http://archive.ubuntu.com/ubuntu focal/universe amd64 python2-minimal amd64 2.7.17
-2ubuntu4 [27.5 kB]
Get:4 http://archive.ubuntu.com/ubuntu focal-updates/universe amd64 libpython2.7-stdlib
amd64 2.7.18-1~20.04.4 [1887 kB]
Get:5 http://archive.ubuntu.com/ubuntu focal-updates/universe amd64 python2.7 amd64 2.7.
18-1~20.04.4 [248 kB]
Get:6 http://archive.ubuntu.com/ubuntu focal/universe amd64 libpython2-stdlib amd64 2.7.
17-2ubuntu4 [7072 B]
Get:7 http://archive.ubuntu.com/ubuntu focal/universe amd64 python2 amd64 2.7.17-2ubuntu
4 [26.5 kB]
Get:8 http://archive.ubuntu.com/ubuntu focal/universe amd64 freeglut3 amd64 2.8.1-3 [73.
6 kB]
Get:9 http://archive.ubuntu.com/ubuntu focal/main amd64 libglu1-mesa amd64 9.0.1-1build1
[168 kB]
Get:10 http://archive.ubuntu.com/ubuntu focal/universe amd64 python-opengl all 3.1.0+dfs
g-2build1 [486 kB]
Fetched 4540 kB in 1s (5655 kB/s)
Selecting previously unselected package libpython2.7-minimal:amd64.
(Reading database ... 113807 files and directories currently installed.)
Preparing to unpack .../0-libpython2.7-minimal_2.7.18-1~20.04.4_amd64.deb ...
Unpacking libpython2.7-minimal:amd64 (2.7.18-1~20.04.4) ...
Selecting previously unselected package python2.7-minimal.
Preparing to unpack .../1-python2.7-minimal_2.7.18-1~20.04.4_amd64.deb ...
Unpacking python2.7-minimal (2.7.18-1~20.04.4) ...
Selecting previously unselected package python2-minimal.
Preparing to unpack .../2-python2-minimal_2.7.17-2ubuntu4_amd64.deb ...
Unpacking python2-minimal (2.7.17-2ubuntu4) ...
Selecting previously unselected package libpython2.7-stdlib:amd64.
Preparing to unpack .../3-libpython2.7-stdlib_2.7.18-1~20.04.4_amd64.deb ...
Unpacking libpython2.7-stdlib:amd64 (2.7.18-1~20.04.4) ...
Selecting previously unselected package python2.7.
Preparing to unpack .../4-python2.7_2.7.18-1~20.04.4_amd64.deb ...
Unpacking python2.7 (2.7.18-1~20.04.4) ...
Selecting previously unselected package libpython2-stdlib:amd64.
Preparing to unpack .../5-libpython2-stdlib_2.7.17-2ubuntu4_amd64.deb ...
Unpacking libpython2-stdlib:amd64 (2.7.17-2ubuntu4) ...
Setting up libpython2.7-minimal:amd64 (2.7.18-1~20.04.4) ...
Setting up python2.7-minimal (2.7.18-1~20.04.4) ...
Linking and byte-compiling packages for runtime python2.7...
Setting up python2-minimal (2.7.17-2ubuntu4) ...
Selecting previously unselected package python2.
(Reading database ... 114554 files and directories currently installed.)
Preparing to unpack .../python2_2.7.17-2ubuntu4_amd64.deb ...
Unpacking python2 (2.7.17-2ubuntu4) ...
Selecting previously unselected package freeglut3:amd64.
Preparing to unpack .../freeglut3_2.8.1-3_amd64.deb ...
Unpacking freeglut3:amd64 (2.8.1-3) ...
Selecting previously unselected package libglu1-mesa:amd64.
Preparing to unpack .../libglu1-mesa_9.0.1-1build1_amd64.deb ...
Unpacking libglu1-mesa:amd64 (9.0.1-1build1) ...
Selecting previously unselected package python-opengl.
Preparing to unpack .../python-opengl_3.1.0+dfsg-2build1_all.deb ...
Unpacking python-opengl (3.1.0+dfsg-2build1) ...
Setting up freeglut3:amd64 (2.8.1-3) ...
Setting up libpython2.7-stdlib:amd64 (2.7.18-1~20.04.4) ...
Setting up libglu1-mesa:amd64 (9.0.1-1build1) ...
```

```
Setting up python2.7 (2.7.18-1~20.04.4) ...
Setting up libpython2-stdlib:amd64 (2.7.17-2ubuntu4) ...
Setting up python2 (2.7.17-2ubuntu4) ...
Setting up python-opengl (3.1.0+dfsg-2build1) ...
Processing triggers for man-db (2.9.1-1) ...
Processing triggers for mime-support (3.64ubuntu1) ...
Processing triggers for libc-bin (2.31-0ubuntu9.14) ...
```

In [4]:
```python
# !pip3 install gym==0.26.2
# !pip3 install gym-notices==0.0.8
!pip3 install --upgrade setuptools --user
!pip3 install ez_setup
!pip3 install gym[atari]
!pip3 install gym[accept-rom-license]
```

```
Requirement already satisfied: setuptools in /opt/conda/lib/python3.10/site-packages (6
9.0.3)
Collecting setuptools
  Downloading setuptools-69.5.1-py3-none-any.whl.metadata (6.2 kB)
Downloading setuptools-69.5.1-py3-none-any.whl (894 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 894.6/894.6 kB 16.2 MB/s eta 0:00:00a 0:00:01
Installing collected packages: setuptools
ERROR: pip's dependency resolver does not currently take into account all the packages t
hat are installed. This behaviour is the source of the following dependency conflicts.
tensorflow-decision-forests 1.8.1 requires wurlitzer, which is not installed.
tensorflow 2.15.0 requires keras<2.16,>=2.15.0, but you have keras 3.2.1 which is incomp
atible.
Successfully installed setuptools-69.5.1
Collecting ez_setup
  Downloading ez_setup-0.9.tar.gz (6.6 kB)
  Preparing metadata (setup.py) ... done
Building wheels for collected packages: ez_setup
  Building wheel for ez_setup (setup.py) ... done
  Created wheel for ez_setup: filename=ez_setup-0.9-py3-none-any.whl size=10995 sha256=1
9aa106977f5833610fb9d4d7feab78ba388eb4c242107fc9d7dbefd2b9aa324
  Stored in directory: /root/.cache/pip/wheels/7a/d6/77/8f495e85fb7df23d41c328b9ea3cf0d9
e83631b20bba479293
Successfully built ez_setup
Installing collected packages: ez_setup
Successfully installed ez_setup-0.9
Requirement already satisfied: gym[atari] in /opt/conda/lib/python3.10/site-packages (0.
26.2)
Requirement already satisfied: numpy>=1.18.0 in /opt/conda/lib/python3.10/site-packages
 (from gym[atari]) (1.26.4)
Requirement already satisfied: cloudpickle>=1.2.0 in /opt/conda/lib/python3.10/site-pack
ages (from gym[atari]) (2.2.1)
Requirement already satisfied: gym-notices>=0.0.4 in /opt/conda/lib/python3.10/site-pack
ages (from gym[atari]) (0.0.8)
Collecting ale-py~=0.8.0 (from gym[atari])
  Downloading ale_py-0.8.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.me
tadata (8.1 kB)
Requirement already satisfied: importlib-resources in /opt/conda/lib/python3.10/site-pac
kages (from ale-py~=0.8.0->gym[atari]) (6.1.1)
Requirement already satisfied: typing-extensions in /opt/conda/lib/python3.10/site-packa
ges (from ale-py~=0.8.0->gym[atari]) (4.9.0)
Downloading ale_py-0.8.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.7
 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.7/1.7 MB 25.5 MB/s eta 0:00:00a 0:00:01
Installing collected packages: ale-py
Successfully installed ale-py-0.8.1
Requirement already satisfied: gym[accept-rom-license] in /opt/conda/lib/python3.10/site
-packages (0.26.2)
Requirement already satisfied: numpy>=1.18.0 in /opt/conda/lib/python3.10/site-packages
 (from gym[accept-rom-license]) (1.26.4)
Requirement already satisfied: cloudpickle>=1.2.0 in /opt/conda/lib/python3.10/site-pack
ages (from gym[accept-rom-license]) (2.2.1)
```

```
Requirement already satisfied: gym-notices>=0.0.4 in /opt/conda/lib/python3.10/site-pack
ages (from gym[accept-rom-license]) (0.0.8)
Collecting autorom~=0.4.2 (from autorom[accept-rom-license]~=0.4.2; extra == "accept-rom
-license"->gym[accept-rom-license])
  Downloading AutoROM-0.4.2-py3-none-any.whl.metadata (2.8 kB)
Requirement already satisfied: click in /opt/conda/lib/python3.10/site-packages (from au
torom~=0.4.2->autorom[accept-rom-license]~=0.4.2; extra == "accept-rom-license"->gym[acc
ept-rom-license]) (8.1.7)
Requirement already satisfied: requests in /opt/conda/lib/python3.10/site-packages (from
 autorom~=0.4.2->autorom[accept-rom-license]~=0.4.2; extra == "accept-rom-license"->gym
[accept-rom-license]) (2.31.0)
Requirement already satisfied: tqdm in /opt/conda/lib/python3.10/site-packages (from aut
orom~=0.4.2->autorom[accept-rom-license]~=0.4.2; extra == "accept-rom-license"->gym[acce
pt-rom-license]) (4.66.1)
Collecting AutoROM.accept-rom-license (from autorom[accept-rom-license]~=0.4.2; extra ==
 "accept-rom-license"->gym[accept-rom-license])
  Downloading AutoROM.accept-rom-license-0.6.1.tar.gz (434 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 434.7/434.7 kB 9.3 MB/s eta 0:00:0000:01
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.10/sit
e-packages (from requests->autorom~=0.4.2->autorom[accept-rom-license]~=0.4.2; extra ==
 "accept-rom-license"->gym[accept-rom-license]) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages
 (from requests->autorom~=0.4.2->autorom[accept-rom-license]~=0.4.2; extra == "accept-ro
m-license"->gym[accept-rom-license]) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.10/site-pack
ages (from requests->autorom~=0.4.2->autorom[accept-rom-license]~=0.4.2; extra == "accep
t-rom-license"->gym[accept-rom-license]) (1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.10/site-pack
ages (from requests->autorom~=0.4.2->autorom[accept-rom-license]~=0.4.2; extra == "accep
t-rom-license"->gym[accept-rom-license]) (2024.2.2)
Downloading AutoROM-0.4.2-py3-none-any.whl (16 kB)
Building wheels for collected packages: AutoROM.accept-rom-license
  Building wheel for AutoROM.accept-rom-license (pyproject.toml) ... done
  Created wheel for AutoROM.accept-rom-license: filename=AutoROM.accept_rom_license-0.6.
1-py3-none-any.whl size=446659 sha256=25fbd1fd72b7a41412c9dee4d78292a405349dd66a437046c2
52858d8a53a862
  Stored in directory: /root/.cache/pip/wheels/6b/1b/ef/a43ff1a2f1736d5711faa1ba4c1f61be
1131b8899e6a057811
Successfully built AutoROM.accept-rom-license
Installing collected packages: AutoROM.accept-rom-license, autorom
Successfully installed AutoROM.accept-rom-license-0.6.1 autorom-0.4.2
```

For this assignment we will implement the Deep Q-Learning algorithm with Experience Replay as described in breakthrough paper **"Playing Atari with Deep Reinforcement Learning"**. We will train an agent to play the famous game of **Breakout**.

In [5]:
```python
%matplotlib inline
import sys
sys.path.insert(1, '/kaggle/input/mp5dataset-2/assignment5_materials')

import sys
import gym
import torch
import pylab
import random
import numpy as np
from collections import deque
from datetime import datetime
from copy import deepcopy
import torch.nn as nn
import torch.optim as optim
```

```
import torch.nn.functional as F
from torch.autograd import Variable
from utils import find_max_lives, check_live, get_frame, get_init_state
from model import DQN, DQN_LSTM
from config import *

import matplotlib.pyplot as plt
# %load_ext autoreload
# %autoreload 2
```

Understanding the environment

In the following cell, we initialize our game of **Breakout** and you can see how the environment looks like. For further documentation of the of the environment refer to https://gym.openai.com/envs.

In breakout, we will use 3 actions "fire", "left", and "right". "fire" is only used to reset the game when a life is lost, "left" moves the agent left and "right" moves the agent right.

In [6]:
```
env = gym.make('BreakoutDeterministic-v4')
state = env.reset()
```

```
A.L.E: Arcade Learning Environment (version 0.8.1+53f58b7)
[Powered by Stella]
```

In [7]:
```
number_lives = find_max_lives(env)
state_size = env.observation_space.shape
action_size = 3 #fire, left, and right
```

```
/opt/conda/lib/python3.10/site-packages/gym/utils/passive_env_checker.py:233: Deprecatio
nWarning: `np.bool8` is a deprecated alias for `np.bool_`.  (Deprecated NumPy 1.24)
  if not isinstance(terminated, (bool, np.bool8)):
```

Creating a DQN Agent

Here we create a DQN Agent. This agent is defined in the **agent.py**. The corresponding neural network is defined in the **model.py**. Once you've created a working DQN agent, use the code in agent.py to create a double DQN agent in **agent_double.py**. Set the flag "double_dqn" to True to train the double DQN agent.

Evaluation Reward : The average reward received in the past 100 episodes/games.

Frame : Number of frames processed in total.

Memory Size : The current size of the replay memory.

In [8]:
```
double_dqn = True # set to True if using double DQN agent

if double_dqn:
    from agent_double import Agent
else:
    from agent import Agent

agent = Agent(action_size)
evaluation_reward = deque(maxlen=evaluation_reward_length)
frame = 0
memory_size = 0
```

In this training loop, we do not render the screen because it slows down training signficantly. To watch the agent play the game, run the code in next section "Visualize Agent Performance"

```python
In [9]:   rewards, episodes = [], []
          best_eval_reward = 0
          EPISODES = 2500
          print(EPISODES)
          for e in range(EPISODES):
              print(e)
              done = False
              score = 0

              history = np.zeros([5, 84, 84], dtype=np.uint8)
              step = 0
              state, _ = env.reset()
              next_state = state
              life = number_lives

              get_init_state(history, state, HISTORY_SIZE)

              while not done:
                  step += 1
                  frame += 1

                  # Perform a fire action if ball is no longer on screen to continue onto next lif
                  if step > 1 and len(np.unique(next_state[:189] == state[:189])) < 2:
                      action = torch.tensor([[0]])
                      #.cuda()
                  else:
                      action = agent.get_action(np.float32(history[:4, :, :]) / 255.)
                  state = next_state
          #         print("reached")
                  next_state, reward, terminated, truncated, info = env.step(action + 1)
                  done = terminated or truncated

                  frame_next_state = get_frame(next_state)
                  history[4, :, :] = frame_next_state
                  terminal_state = check_live(life, info['lives'])

                  life = info['lives']
                  r = reward

                  # Store the transition in memory
          #         print("Type of action:", type(action))
          #         action = action.to(device)
                  agent.memory.push(deepcopy(frame_next_state), action.cpu(), r, terminal_state)

                  # Start training after random sample generation
                  if(frame >= train_frame): # You can set train_frame to a lower value while testi
                      agent.train_policy_net(frame)
                      # Update the target network only for Double DQN only
                      if double_dqn and (frame % update_target_network_frequency)== 0:
                          agent.update_target_net()
                  score += reward
                  history[:4, :, :] = history[1:, :, :]

                  if done:
                      evaluation_reward.append(score)
                      rewards.append(np.mean(evaluation_reward))
                      episodes.append(e)
                      pylab.plot(episodes, rewards, 'b')
                      pylab.xlabel('Episodes')
                      pylab.ylabel('Rewards')
                      pylab.title('Episodes vs Reward')
                      pylab.savefig("/kaggle/working/breakout_dqn.png") # save graph for training
          #             pylab.savefig("./save_graph/breakout_dqn.png")
```

```
            # every episode, plot the play time
            print("episode:", e, "  score:", score, "  memory length:",
                  len(agent.memory), "  epsilon:", agent.epsilon, "   steps:", step,
                  "   lr:", agent.optimizer.param_groups[0]['lr'], "    evaluation rewar

            # if the mean of scores of last 100 episode is bigger than 5 save model
            ### Change this save condition to whatever you prefer ###
#              if np.mean(evaluation_reward) > 5 and np.mean(evaluation_reward) > best_ev
#                  torch.save(agent.policy_net, "/kaggle/working/save_model/breakout_dqn.
#                  best_eval_reward = np.mean(evaluation_reward)
```

2500
0
episode: 0   score: 1.0   memory length: 168   epsilon: 1.0   steps: 168   lr: 0.0001
    evaluation reward: 1.0
1
episode: 1   score: 2.0   memory length: 386   epsilon: 1.0   steps: 218   lr: 0.0001
    evaluation reward: 1.5
2
episode: 2   score: 1.0   memory length: 555   epsilon: 1.0   steps: 169   lr: 0.0001
    evaluation reward: 1.3333333333333333
3
episode: 3   score: 1.0   memory length: 706   epsilon: 1.0   steps: 151   lr: 0.0001
    evaluation reward: 1.25
4
episode: 4   score: 1.0   memory length: 875   epsilon: 1.0   steps: 169   lr: 0.0001
    evaluation reward: 1.2
5
episode: 5   score: 0.0   memory length: 998   epsilon: 1.0   steps: 123   lr: 0.0001
    evaluation reward: 1.0
6
episode: 6   score: 6.0   memory length: 1374   epsilon: 1.0   steps: 376   lr: 0.0001
     evaluation reward: 1.7142857142857142
7
episode: 7   score: 1.0   memory length: 1543   epsilon: 1.0   steps: 169   lr: 0.0001
     evaluation reward: 1.625
8
episode: 8   score: 6.0   memory length: 1886   epsilon: 1.0   steps: 343   lr: 0.0001
     evaluation reward: 2.111111111111111
9
episode: 9   score: 1.0   memory length: 2054   epsilon: 1.0   steps: 168   lr: 0.0001
     evaluation reward: 2.0
10
episode: 10   score: 0.0   memory length: 2177   epsilon: 1.0   steps: 123   lr: 0.000
1   evaluation reward: 1.8181818181818181
11
episode: 11   score: 4.0   memory length: 2472   epsilon: 1.0   steps: 295   lr: 0.000
1   evaluation reward: 2.0
12
episode: 12   score: 0.0   memory length: 2595   epsilon: 1.0   steps: 123   lr: 0.000
1   evaluation reward: 1.8461538461538463
13
episode: 13   score: 3.0   memory length: 2806   epsilon: 1.0   steps: 211   lr: 0.000
1   evaluation reward: 1.9285714285714286
14
episode: 14   score: 2.0   memory length: 3007   epsilon: 1.0   steps: 201   lr: 0.000
1   evaluation reward: 1.9333333333333333
15
episode: 15   score: 0.0   memory length: 3129   epsilon: 1.0   steps: 122   lr: 0.000
1   evaluation reward: 1.8125
16
episode: 16   score: 2.0   memory length: 3326   epsilon: 1.0   steps: 197   lr: 0.000
1   evaluation reward: 1.8235294117647058
17
episode: 17   score: 0.0   memory length: 3448   epsilon: 1.0   steps: 122   lr: 0.000
1   evaluation reward: 1.7222222222222223
```

```
18
episode: 18    score: 2.0    memory length: 3665    epsilon: 1.0    steps: 217    lr: 0.000
1      evaluation reward: 1.736842105263158
19
episode: 19    score: 2.0    memory length: 3862    epsilon: 1.0    steps: 197    lr: 0.000
1      evaluation reward: 1.75
20
episode: 20    score: 3.0    memory length: 4087    epsilon: 1.0    steps: 225    lr: 0.000
1      evaluation reward: 1.8095238095238095
21
episode: 21    score: 1.0    memory length: 4256    epsilon: 1.0    steps: 169    lr: 0.000
1      evaluation reward: 1.7727272727272727
22
episode: 22    score: 3.0    memory length: 4502    epsilon: 1.0    steps: 246    lr: 0.000
1      evaluation reward: 1.826086956521739
23
episode: 23    score: 1.0    memory length: 4672    epsilon: 1.0    steps: 170    lr: 0.000
1      evaluation reward: 1.7916666666666667
24
episode: 24    score: 1.0    memory length: 4840    epsilon: 1.0    steps: 168    lr: 0.000
1      evaluation reward: 1.76
25
episode: 25    score: 1.0    memory length: 5008    epsilon: 1.0    steps: 168    lr: 0.000
1      evaluation reward: 1.7307692307692308
26
episode: 26    score: 1.0    memory length: 5177    epsilon: 1.0    steps: 169    lr: 0.000
1      evaluation reward: 1.7037037037037037
27
episode: 27    score: 1.0    memory length: 5328    epsilon: 1.0    steps: 151    lr: 0.000
1      evaluation reward: 1.6785714285714286
28
episode: 28    score: 0.0    memory length: 5451    epsilon: 1.0    steps: 123    lr: 0.000
1      evaluation reward: 1.6206896551724137
29
episode: 29    score: 1.0    memory length: 5602    epsilon: 1.0    steps: 151    lr: 0.000
1      evaluation reward: 1.6
30
episode: 30    score: 2.0    memory length: 5782    epsilon: 1.0    steps: 180    lr: 0.000
1      evaluation reward: 1.6129032258064515
31
episode: 31    score: 2.0    memory length: 6000    epsilon: 1.0    steps: 218    lr: 0.000
1      evaluation reward: 1.625
32
episode: 32    score: 0.0    memory length: 6123    epsilon: 1.0    steps: 123    lr: 0.000
1      evaluation reward: 1.5757575757575757
33
episode: 33    score: 0.0    memory length: 6246    epsilon: 1.0    steps: 123    lr: 0.000
1      evaluation reward: 1.5294117647058822
34
episode: 34    score: 1.0    memory length: 6415    epsilon: 1.0    steps: 169    lr: 0.000
1      evaluation reward: 1.5142857142857142
35
episode: 35    score: 0.0    memory length: 6537    epsilon: 1.0    steps: 122    lr: 0.000
1      evaluation reward: 1.4722222222222223
36
episode: 36    score: 2.0    memory length: 6735    epsilon: 1.0    steps: 198    lr: 0.000
1      evaluation reward: 1.4864864864864864
37
episode: 37    score: 4.0    memory length: 7009    epsilon: 1.0    steps: 274    lr: 0.000
1      evaluation reward: 1.5526315789473684
38
episode: 38    score: 1.0    memory length: 7178    epsilon: 1.0    steps: 169    lr: 0.000
1      evaluation reward: 1.5384615384615385
39
episode: 39    score: 1.0    memory length: 7347    epsilon: 1.0    steps: 169    lr: 0.000
1      evaluation reward: 1.525
```

```
40
episode: 40   score: 0.0   memory length: 7470   epsilon: 1.0   steps: 123   lr: 0.000
1     evaluation reward: 1.4878048780487805
41
episode: 41   score: 10.0   memory length: 7866   epsilon: 1.0   steps: 396   lr: 0.00
01     evaluation reward: 1.6904761904761905
42
episode: 42   score: 3.0   memory length: 8112   epsilon: 1.0   steps: 246   lr: 0.000
1     evaluation reward: 1.7209302325581395
43
episode: 43   score: 2.0   memory length: 8311   epsilon: 1.0   steps: 199   lr: 0.000
1     evaluation reward: 1.7272727272727273
44
episode: 44   score: 2.0   memory length: 8530   epsilon: 1.0   steps: 219   lr: 0.000
1     evaluation reward: 1.7333333333333334
45
episode: 45   score: 1.0   memory length: 8700   epsilon: 1.0   steps: 170   lr: 0.000
1     evaluation reward: 1.7173913043478262
46
episode: 46   score: 2.0   memory length: 8920   epsilon: 1.0   steps: 220   lr: 0.000
1     evaluation reward: 1.7234042553191489
47
episode: 47   score: 0.0   memory length: 9043   epsilon: 1.0   steps: 123   lr: 0.000
1     evaluation reward: 1.6875
48
episode: 48   score: 2.0   memory length: 9240   epsilon: 1.0   steps: 197   lr: 0.000
1     evaluation reward: 1.6938775510204083
49
episode: 49   score: 4.0   memory length: 9526   epsilon: 1.0   steps: 286   lr: 0.000
1     evaluation reward: 1.74
50
episode: 50   score: 0.0   memory length: 9648   epsilon: 1.0   steps: 122   lr: 0.000
1     evaluation reward: 1.7058823529411764
51
episode: 51   score: 1.0   memory length: 9816   epsilon: 1.0   steps: 168   lr: 0.000
1     evaluation reward: 1.6923076923076923
52
episode: 52   score: 0.0   memory length: 9938   epsilon: 1.0   steps: 122   lr: 0.000
1     evaluation reward: 1.6603773584905661
53
episode: 53   score: 1.0   memory length: 10106   epsilon: 1.0   steps: 168   lr: 0.00
01     evaluation reward: 1.6481481481481481
54
episode: 54   score: 0.0   memory length: 10228   epsilon: 1.0   steps: 122   lr: 0.00
01     evaluation reward: 1.6181818181818182
55
episode: 55   score: 3.0   memory length: 10476   epsilon: 1.0   steps: 248   lr: 0.00
01     evaluation reward: 1.6428571428571428
56
episode: 56   score: 2.0   memory length: 10673   epsilon: 1.0   steps: 197   lr: 0.00
01     evaluation reward: 1.6491228070175439
57
episode: 57   score: 0.0   memory length: 10795   epsilon: 1.0   steps: 122   lr: 0.00
01     evaluation reward: 1.6206896551724137
58
episode: 58   score: 1.0   memory length: 10964   epsilon: 1.0   steps: 169   lr: 0.00
01     evaluation reward: 1.6101694915254237
59
episode: 59   score: 5.0   memory length: 11305   epsilon: 1.0   steps: 341   lr: 0.00
01     evaluation reward: 1.6666666666666667
60
episode: 60   score: 3.0   memory length: 11569   epsilon: 1.0   steps: 264   lr: 0.00
01     evaluation reward: 1.6885245901639345
61
episode: 61   score: 1.0   memory length: 11741   epsilon: 1.0   steps: 172   lr: 0.00
01     evaluation reward: 1.6774193548387097
```

```
62
episode: 62    score: 2.0    memory length: 11959    epsilon: 1.0    steps: 218    lr: 0.00
01    evaluation reward: 1.6825396825396826
63
episode: 63    score: 1.0    memory length: 12109    epsilon: 1.0    steps: 150    lr: 0.00
01    evaluation reward: 1.671875
64
episode: 64    score: 4.0    memory length: 12405    epsilon: 1.0    steps: 296    lr: 0.00
01    evaluation reward: 1.7076923076923076
65
episode: 65    score: 4.0    memory length: 12683    epsilon: 1.0    steps: 278    lr: 0.00
01    evaluation reward: 1.7424242424242424
66
episode: 66    score: 6.0    memory length: 13077    epsilon: 1.0    steps: 394    lr: 0.00
01    evaluation reward: 1.8059701492537314
67
episode: 67    score: 1.0    memory length: 13246    epsilon: 1.0    steps: 169    lr: 0.00
01    evaluation reward: 1.7941176470588236
68
episode: 68    score: 3.0    memory length: 13516    epsilon: 1.0    steps: 270    lr: 0.00
01    evaluation reward: 1.8115942028985508
69
episode: 69    score: 1.0    memory length: 13666    epsilon: 1.0    steps: 150    lr: 0.00
01    evaluation reward: 1.8
70
episode: 70    score: 2.0    memory length: 13864    epsilon: 1.0    steps: 198    lr: 0.00
01    evaluation reward: 1.8028169014084507
71
episode: 71    score: 2.0    memory length: 14064    epsilon: 1.0    steps: 200    lr: 0.00
01    evaluation reward: 1.8055555555555556
72
episode: 72    score: 2.0    memory length: 14262    epsilon: 1.0    steps: 198    lr: 0.00
01    evaluation reward: 1.8082191780821917
73
episode: 73    score: 1.0    memory length: 14431    epsilon: 1.0    steps: 169    lr: 0.00
01    evaluation reward: 1.7972972972972974
74
episode: 74    score: 0.0    memory length: 14553    epsilon: 1.0    steps: 122    lr: 0.00
01    evaluation reward: 1.7733333333333334
75
episode: 75    score: 0.0    memory length: 14676    epsilon: 1.0    steps: 123    lr: 0.00
01    evaluation reward: 1.75
76
episode: 76    score: 0.0    memory length: 14799    epsilon: 1.0    steps: 123    lr: 0.00
01    evaluation reward: 1.7272727272727273
77
episode: 77    score: 4.0    memory length: 15096    epsilon: 1.0    steps: 297    lr: 0.00
01    evaluation reward: 1.7564102564102564
78
episode: 78    score: 2.0    memory length: 15314    epsilon: 1.0    steps: 218    lr: 0.00
01    evaluation reward: 1.759493670886076
79
episode: 79    score: 5.0    memory length: 15613    epsilon: 1.0    steps: 299    lr: 0.00
01    evaluation reward: 1.8
80
episode: 80    score: 2.0    memory length: 15828    epsilon: 1.0    steps: 215    lr: 0.00
01    evaluation reward: 1.8024691358024691
81
episode: 81    score: 2.0    memory length: 16026    epsilon: 1.0    steps: 198    lr: 0.00
01    evaluation reward: 1.8048780487804879
82
episode: 82    score: 3.0    memory length: 16292    epsilon: 1.0    steps: 266    lr: 0.00
01    evaluation reward: 1.819277108433735
83
episode: 83    score: 1.0    memory length: 16463    epsilon: 1.0    steps: 171    lr: 0.00
01    evaluation reward: 1.8095238095238095
```

```
84
episode: 84    score: 2.0    memory length: 16685    epsilon: 1.0    steps: 222    lr: 0.00
01      evaluation reward: 1.811764705882353
85
episode: 85    score: 0.0    memory length: 16807    epsilon: 1.0    steps: 122    lr: 0.00
01      evaluation reward: 1.7906976744186047
86
episode: 86    score: 2.0    memory length: 17005    epsilon: 1.0    steps: 198    lr: 0.00
01      evaluation reward: 1.793103448275862
87
episode: 87    score: 1.0    memory length: 17175    epsilon: 1.0    steps: 170    lr: 0.00
01      evaluation reward: 1.7840909090909092
88
episode: 88    score: 3.0    memory length: 17401    epsilon: 1.0    steps: 226    lr: 0.00
01      evaluation reward: 1.797752808988764
89
episode: 89    score: 1.0    memory length: 17552    epsilon: 1.0    steps: 151    lr: 0.00
01      evaluation reward: 1.788888888888889
90
episode: 90    score: 1.0    memory length: 17723    epsilon: 1.0    steps: 171    lr: 0.00
01      evaluation reward: 1.7802197802197801
91
episode: 91    score: 0.0    memory length: 17845    epsilon: 1.0    steps: 122    lr: 0.00
01      evaluation reward: 1.7608695652173914
92
episode: 92    score: 0.0    memory length: 17968    epsilon: 1.0    steps: 123    lr: 0.00
01      evaluation reward: 1.7419354838709677
93
episode: 93    score: 0.0    memory length: 18091    epsilon: 1.0    steps: 123    lr: 0.00
01      evaluation reward: 1.7234042553191489
94
episode: 94    score: 1.0    memory length: 18242    epsilon: 1.0    steps: 151    lr: 0.00
01      evaluation reward: 1.7157894736842105
95
episode: 95    score: 0.0    memory length: 18365    epsilon: 1.0    steps: 123    lr: 0.00
01      evaluation reward: 1.6979166666666667
96
episode: 96    score: 0.0    memory length: 18488    epsilon: 1.0    steps: 123    lr: 0.00
01      evaluation reward: 1.6804123711340206
97
episode: 97    score: 2.0    memory length: 18686    epsilon: 1.0    steps: 198    lr: 0.00
01      evaluation reward: 1.683673469387755
98
episode: 98    score: 4.0    memory length: 18981    epsilon: 1.0    steps: 295    lr: 0.00
01      evaluation reward: 1.707070707070707
99
episode: 99    score: 0.0    memory length: 19104    epsilon: 1.0    steps: 123    lr: 0.00
01      evaluation reward: 1.69
100
episode: 100    score: 0.0    memory length: 19226    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.68
101
episode: 101    score: 1.0    memory length: 19395    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.67
102
episode: 102    score: 1.0    memory length: 19545    epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.67
103
episode: 103    score: 4.0    memory length: 19838    epsilon: 1.0    steps: 293    lr: 0.0
001      evaluation reward: 1.7
104
episode: 104    score: 1.0    memory length: 19988    epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.7
105
episode: 105    score: 2.0    memory length: 20170    epsilon: 1.0    steps: 182    lr: 0.0
001      evaluation reward: 1.72
```

```
106
episode: 106   score: 2.0   memory length: 20371   epsilon: 1.0    steps: 201    lr: 0.0
001     evaluation reward: 1.68
107
episode: 107   score: 1.0   memory length: 20540   epsilon: 1.0    steps: 169    lr: 0.0
001     evaluation reward: 1.68
108
episode: 108   score: 0.0   memory length: 20662   epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.62
109
episode: 109   score: 1.0   memory length: 20833   epsilon: 1.0    steps: 171    lr: 0.0
001     evaluation reward: 1.62
110
episode: 110   score: 1.0   memory length: 21002   epsilon: 1.0    steps: 169    lr: 0.0
001     evaluation reward: 1.63
111
episode: 111   score: 5.0   memory length: 21310   epsilon: 1.0    steps: 308    lr: 0.0
001     evaluation reward: 1.64
112
episode: 112   score: 0.0   memory length: 21432   epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.64
113
episode: 113   score: 3.0   memory length: 21660   epsilon: 1.0    steps: 228    lr: 0.0
001     evaluation reward: 1.64
114
episode: 114   score: 1.0   memory length: 21812   epsilon: 1.0    steps: 152    lr: 0.0
001     evaluation reward: 1.63
115
episode: 115   score: 3.0   memory length: 22078   epsilon: 1.0    steps: 266    lr: 0.0
001     evaluation reward: 1.66
116
episode: 116   score: 0.0   memory length: 22201   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.64
117
episode: 117   score: 1.0   memory length: 22370   epsilon: 1.0    steps: 169    lr: 0.0
001     evaluation reward: 1.65
118
episode: 118   score: 0.0   memory length: 22493   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.63
119
episode: 119   score: 0.0   memory length: 22615   epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.61
120
episode: 120   score: 7.0   memory length: 22987   epsilon: 1.0    steps: 372    lr: 0.0
001     evaluation reward: 1.65
121
episode: 121   score: 0.0   memory length: 23110   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.64
122
episode: 122   score: 3.0   memory length: 23357   epsilon: 1.0    steps: 247    lr: 0.0
001     evaluation reward: 1.64
123
episode: 123   score: 0.0   memory length: 23480   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.63
124
episode: 124   score: 2.0   memory length: 23698   epsilon: 1.0    steps: 218    lr: 0.0
001     evaluation reward: 1.64
125
episode: 125   score: 3.0   memory length: 23946   epsilon: 1.0    steps: 248    lr: 0.0
001     evaluation reward: 1.66
126
episode: 126   score: 2.0   memory length: 24144   epsilon: 1.0    steps: 198    lr: 0.0
001     evaluation reward: 1.67
127
episode: 127   score: 3.0   memory length: 24372   epsilon: 1.0    steps: 228    lr: 0.0
001     evaluation reward: 1.69
```

```
128
episode: 128    score: 2.0    memory length: 24570    epsilon: 1.0    steps: 198    lr: 0.0
001     evaluation reward: 1.71
129
episode: 129    score: 2.0    memory length: 24786    epsilon: 1.0    steps: 216    lr: 0.0
001     evaluation reward: 1.72
130
episode: 130    score: 2.0    memory length: 24983    epsilon: 1.0    steps: 197    lr: 0.0
001     evaluation reward: 1.72
131
episode: 131    score: 2.0    memory length: 25198    epsilon: 1.0    steps: 215    lr: 0.0
001     evaluation reward: 1.72
132
episode: 132    score: 3.0    memory length: 25431    epsilon: 1.0    steps: 233    lr: 0.0
001     evaluation reward: 1.75
133
episode: 133    score: 1.0    memory length: 25584    epsilon: 1.0    steps: 153    lr: 0.0
001     evaluation reward: 1.76
134
episode: 134    score: 0.0    memory length: 25707    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.75
135
episode: 135    score: 0.0    memory length: 25829    epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.75
136
episode: 136    score: 1.0    memory length: 25980    epsilon: 1.0    steps: 151    lr: 0.0
001     evaluation reward: 1.74
137
episode: 137    score: 2.0    memory length: 26162    epsilon: 1.0    steps: 182    lr: 0.0
001     evaluation reward: 1.72
138
episode: 138    score: 2.0    memory length: 26349    epsilon: 1.0    steps: 187    lr: 0.0
001     evaluation reward: 1.73
139
episode: 139    score: 1.0    memory length: 26518    epsilon: 1.0    steps: 169    lr: 0.0
001     evaluation reward: 1.73
140
episode: 140    score: 0.0    memory length: 26641    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.73
141
episode: 141    score: 2.0    memory length: 26839    epsilon: 1.0    steps: 198    lr: 0.0
001     evaluation reward: 1.65
142
episode: 142    score: 2.0    memory length: 27057    epsilon: 1.0    steps: 218    lr: 0.0
001     evaluation reward: 1.64
143
episode: 143    score: 2.0    memory length: 27255    epsilon: 1.0    steps: 198    lr: 0.0
001     evaluation reward: 1.64
144
episode: 144    score: 1.0    memory length: 27406    epsilon: 1.0    steps: 151    lr: 0.0
001     evaluation reward: 1.63
145
episode: 145    score: 5.0    memory length: 27751    epsilon: 1.0    steps: 345    lr: 0.0
001     evaluation reward: 1.67
146
episode: 146    score: 1.0    memory length: 27919    epsilon: 1.0    steps: 168    lr: 0.0
001     evaluation reward: 1.66
147
episode: 147    score: 3.0    memory length: 28149    epsilon: 1.0    steps: 230    lr: 0.0
001     evaluation reward: 1.69
148
episode: 148    score: 2.0    memory length: 28330    epsilon: 1.0    steps: 181    lr: 0.0
001     evaluation reward: 1.69
149
episode: 149    score: 0.0    memory length: 28452    epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.65
```

```
150
episode: 150    score: 2.0    memory length: 28650    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.67
151
episode: 151    score: 2.0    memory length: 28870    epsilon: 1.0    steps: 220    lr: 0.0
001      evaluation reward: 1.68
152
episode: 152    score: 3.0    memory length: 29137    epsilon: 1.0    steps: 267    lr: 0.0
001      evaluation reward: 1.71
153
episode: 153    score: 5.0    memory length: 29485    epsilon: 1.0    steps: 348    lr: 0.0
001      evaluation reward: 1.75
154
episode: 154    score: 0.0    memory length: 29608    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.75
155
episode: 155    score: 1.0    memory length: 29777    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.73
156
episode: 156    score: 5.0    memory length: 30113    epsilon: 1.0    steps: 336    lr: 0.0
001      evaluation reward: 1.76
157
episode: 157    score: 0.0    memory length: 30235    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.76
158
episode: 158    score: 1.0    memory length: 30404    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.76
159
episode: 159    score: 2.0    memory length: 30605    epsilon: 1.0    steps: 201    lr: 0.0
001      evaluation reward: 1.73
160
episode: 160    score: 0.0    memory length: 30728    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.7
161
episode: 161    score: 2.0    memory length: 30928    epsilon: 1.0    steps: 200    lr: 0.0
001      evaluation reward: 1.71
162
episode: 162    score: 1.0    memory length: 31097    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.7
163
episode: 163    score: 0.0    memory length: 31219    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.69
164
episode: 164    score: 1.0    memory length: 31387    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.66
165
episode: 165    score: 3.0    memory length: 31633    epsilon: 1.0    steps: 246    lr: 0.0
001      evaluation reward: 1.65
166
episode: 166    score: 4.0    memory length: 31930    epsilon: 1.0    steps: 297    lr: 0.0
001      evaluation reward: 1.63
167
episode: 167    score: 1.0    memory length: 32099    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.63
168
episode: 168    score: 2.0    memory length: 32296    epsilon: 1.0    steps: 197    lr: 0.0
001      evaluation reward: 1.62
169
episode: 169    score: 2.0    memory length: 32494    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.63
170
episode: 170    score: 1.0    memory length: 32665    epsilon: 1.0    steps: 171    lr: 0.0
001      evaluation reward: 1.62
171
episode: 171    score: 2.0    memory length: 32884    epsilon: 1.0    steps: 219    lr: 0.0
001      evaluation reward: 1.62
```

```
172
episode: 172    score: 4.0    memory length: 33181    epsilon: 1.0    steps: 297    lr: 0.0
001      evaluation reward: 1.64
173
episode: 173    score: 1.0    memory length: 33350    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.64
174
episode: 174    score: 0.0    memory length: 33472    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.64
175
episode: 175    score: 2.0    memory length: 33690    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.66
176
episode: 176    score: 1.0    memory length: 33841    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.67
177
episode: 177    score: 0.0    memory length: 33963    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.63
178
episode: 178    score: 0.0    memory length: 34085    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.61
179
episode: 179    score: 0.0    memory length: 34207    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.56
180
episode: 180    score: 3.0    memory length: 34453    epsilon: 1.0    steps: 246    lr: 0.0
001      evaluation reward: 1.57
181
episode: 181    score: 0.0    memory length: 34576    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.55
182
episode: 182    score: 2.0    memory length: 34774    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.54
183
episode: 183    score: 0.0    memory length: 34897    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.53
184
episode: 184    score: 0.0    memory length: 35020    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.51
185
episode: 185    score: 2.0    memory length: 35241    epsilon: 1.0    steps: 221    lr: 0.0
001      evaluation reward: 1.53
186
episode: 186    score: 2.0    memory length: 35459    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.53
187
episode: 187    score: 2.0    memory length: 35659    epsilon: 1.0    steps: 200    lr: 0.0
001      evaluation reward: 1.54
188
episode: 188    score: 3.0    memory length: 35909    epsilon: 1.0    steps: 250    lr: 0.0
001      evaluation reward: 1.54
189
episode: 189    score: 0.0    memory length: 36032    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.53
190
episode: 190    score: 0.0    memory length: 36155    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.52
191
episode: 191    score: 0.0    memory length: 36278    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.52
192
episode: 192    score: 6.0    memory length: 36595    epsilon: 1.0    steps: 317    lr: 0.0
001      evaluation reward: 1.58
193
episode: 193    score: 1.0    memory length: 36763    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.59
```

```
194
episode: 194    score: 3.0    memory length: 36996    epsilon: 1.0    steps: 233    lr: 0.0
001      evaluation reward: 1.61
195
episode: 195    score: 2.0    memory length: 37196    epsilon: 1.0    steps: 200    lr: 0.0
001      evaluation reward: 1.63
196
episode: 196    score: 0.0    memory length: 37319    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.63
197
episode: 197    score: 0.0    memory length: 37441    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.61
198
episode: 198    score: 2.0    memory length: 37621    epsilon: 1.0    steps: 180    lr: 0.0
001      evaluation reward: 1.59
199
episode: 199    score: 2.0    memory length: 37819    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.61
200
episode: 200    score: 2.0    memory length: 38034    epsilon: 1.0    steps: 215    lr: 0.0
001      evaluation reward: 1.63
201
episode: 201    score: 0.0    memory length: 38157    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.62
202
episode: 202    score: 1.0    memory length: 38309    epsilon: 1.0    steps: 152    lr: 0.0
001      evaluation reward: 1.62
203
episode: 203    score: 0.0    memory length: 38432    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.58
204
episode: 204    score: 0.0    memory length: 38554    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.57
205
episode: 205    score: 2.0    memory length: 38754    epsilon: 1.0    steps: 200    lr: 0.0
001      evaluation reward: 1.57
206
episode: 206    score: 0.0    memory length: 38877    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.55
207
episode: 207    score: 2.0    memory length: 39095    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.56
208
episode: 208    score: 2.0    memory length: 39293    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.58
209
episode: 209    score: 4.0    memory length: 39571    epsilon: 1.0    steps: 278    lr: 0.0
001      evaluation reward: 1.61
210
episode: 210    score: 1.0    memory length: 39742    epsilon: 1.0    steps: 171    lr: 0.0
001      evaluation reward: 1.61
211
episode: 211    score: 1.0    memory length: 39911    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.57
212
episode: 212    score: 0.0    memory length: 40034    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.57
213
episode: 213    score: 1.0    memory length: 40186    epsilon: 1.0    steps: 152    lr: 0.0
001      evaluation reward: 1.55
214
episode: 214    score: 2.0    memory length: 40404    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.56
215
episode: 215    score: 0.0    memory length: 40527    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.53
```

```
216
episode: 216    score: 1.0    memory length: 40696    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.54
217
episode: 217    score: 1.0    memory length: 40865    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.54
218
episode: 218    score: 1.0    memory length: 41016    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.55
219
episode: 219    score: 2.0    memory length: 41214    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.57
220
episode: 220    score: 0.0    memory length: 41337    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.5
221
episode: 221    score: 0.0    memory length: 41459    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.5
222
episode: 222    score: 5.0    memory length: 41771    epsilon: 1.0    steps: 312    lr: 0.0
001      evaluation reward: 1.52
223
episode: 223    score: 0.0    memory length: 41894    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.52
224
episode: 224    score: 0.0    memory length: 42017    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.5
225
episode: 225    score: 0.0    memory length: 42140    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.47
226
episode: 226    score: 5.0    memory length: 42448    epsilon: 1.0    steps: 308    lr: 0.0
001      evaluation reward: 1.5
227
episode: 227    score: 2.0    memory length: 42667    epsilon: 1.0    steps: 219    lr: 0.0
001      evaluation reward: 1.49
228
episode: 228    score: 2.0    memory length: 42865    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.49
229
episode: 229    score: 2.0    memory length: 43083    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.49
230
episode: 230    score: 0.0    memory length: 43205    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.47
231
episode: 231    score: 1.0    memory length: 43356    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.46
232
episode: 232    score: 0.0    memory length: 43479    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.43
233
episode: 233    score: 0.0    memory length: 43601    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.42
234
episode: 234    score: 2.0    memory length: 43816    epsilon: 1.0    steps: 215    lr: 0.0
001      evaluation reward: 1.44
235
episode: 235    score: 4.0    memory length: 44083    epsilon: 1.0    steps: 267    lr: 0.0
001      evaluation reward: 1.48
236
episode: 236    score: 2.0    memory length: 44281    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.49
237
episode: 237    score: 1.0    memory length: 44453    epsilon: 1.0    steps: 172    lr: 0.0
001      evaluation reward: 1.48
```

```
238
episode: 238    score: 0.0    memory length: 44576    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.46
239
episode: 239    score: 0.0    memory length: 44698    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.45
240
episode: 240    score: 2.0    memory length: 44896    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.47
241
episode: 241    score: 1.0    memory length: 45046    epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.46
242
episode: 242    score: 0.0    memory length: 45168    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.44
243
episode: 243    score: 1.0    memory length: 45319    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.43
244
episode: 244    score: 2.0    memory length: 45536    epsilon: 1.0    steps: 217    lr: 0.0
001      evaluation reward: 1.44
245
episode: 245    score: 3.0    memory length: 45764    epsilon: 1.0    steps: 228    lr: 0.0
001      evaluation reward: 1.42
246
episode: 246    score: 0.0    memory length: 45886    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.41
247
episode: 247    score: 4.0    memory length: 46178    epsilon: 1.0    steps: 292    lr: 0.0
001      evaluation reward: 1.42
248
episode: 248    score: 3.0    memory length: 46444    epsilon: 1.0    steps: 266    lr: 0.0
001      evaluation reward: 1.43
249
episode: 249    score: 1.0    memory length: 46615    epsilon: 1.0    steps: 171    lr: 0.0
001      evaluation reward: 1.44
250
episode: 250    score: 2.0    memory length: 46833    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.44
251
episode: 251    score: 1.0    memory length: 47002    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.43
252
episode: 252    score: 1.0    memory length: 47170    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.41
253
episode: 253    score: 1.0    memory length: 47322    epsilon: 1.0    steps: 152    lr: 0.0
001      evaluation reward: 1.37
254
episode: 254    score: 2.0    memory length: 47539    epsilon: 1.0    steps: 217    lr: 0.0
001      evaluation reward: 1.39
255
episode: 255    score: 0.0    memory length: 47662    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.38
256
episode: 256    score: 2.0    memory length: 47884    epsilon: 1.0    steps: 222    lr: 0.0
001      evaluation reward: 1.35
257
episode: 257    score: 0.0    memory length: 48007    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.35
258
episode: 258    score: 1.0    memory length: 48175    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.35
259
episode: 259    score: 1.0    memory length: 48346    epsilon: 1.0    steps: 171    lr: 0.0
001      evaluation reward: 1.34
```

```
260
episode: 260   score: 8.0   memory length: 48664   epsilon: 1.0     steps: 318    lr: 0.0
001     evaluation reward: 1.42
261
episode: 261   score: 1.0   memory length: 48836   epsilon: 1.0     steps: 172    lr: 0.0
001     evaluation reward: 1.41
262
episode: 262   score: 2.0   memory length: 49015   epsilon: 1.0     steps: 179    lr: 0.0
001     evaluation reward: 1.42
263
episode: 263   score: 4.0   memory length: 49291   epsilon: 1.0     steps: 276    lr: 0.0
001     evaluation reward: 1.46
264
episode: 264   score: 0.0   memory length: 49413   epsilon: 1.0     steps: 122    lr: 0.0
001     evaluation reward: 1.45
265
episode: 265   score: 2.0   memory length: 49610   epsilon: 1.0     steps: 197    lr: 0.0
001     evaluation reward: 1.44
266
episode: 266   score: 2.0   memory length: 49809   epsilon: 1.0     steps: 199    lr: 0.0
001     evaluation reward: 1.42
267
episode: 267   score: 0.0   memory length: 49932   epsilon: 1.0     steps: 123    lr: 0.0
001     evaluation reward: 1.41
268
episode: 268   score: 1.0   memory length: 50102   epsilon: 1.0     steps: 170    lr: 0.0
001     evaluation reward: 1.4
269
episode: 269   score: 0.0   memory length: 50225   epsilon: 1.0     steps: 123    lr: 0.0
001     evaluation reward: 1.38
270
episode: 270   score: 1.0   memory length: 50393   epsilon: 1.0     steps: 168    lr: 0.0
001     evaluation reward: 1.38
271
episode: 271   score: 2.0   memory length: 50594   epsilon: 1.0     steps: 201    lr: 0.0
001     evaluation reward: 1.38
272
episode: 272   score: 0.0   memory length: 50717   epsilon: 1.0     steps: 123    lr: 0.0
001     evaluation reward: 1.34
273
episode: 273   score: 0.0   memory length: 50839   epsilon: 1.0     steps: 122    lr: 0.0
001     evaluation reward: 1.33
274
episode: 274   score: 0.0   memory length: 50962   epsilon: 1.0     steps: 123    lr: 0.0
001     evaluation reward: 1.33
275
episode: 275   score: 2.0   memory length: 51159   epsilon: 1.0     steps: 197    lr: 0.0
001     evaluation reward: 1.33
276
episode: 276   score: 0.0   memory length: 51282   epsilon: 1.0     steps: 123    lr: 0.0
001     evaluation reward: 1.32
277
episode: 277   score: 0.0   memory length: 51405   epsilon: 1.0     steps: 123    lr: 0.0
001     evaluation reward: 1.32
278
episode: 278   score: 5.0   memory length: 51747   epsilon: 1.0     steps: 342    lr: 0.0
001     evaluation reward: 1.37
279
episode: 279   score: 4.0   memory length: 52039   epsilon: 1.0     steps: 292    lr: 0.0
001     evaluation reward: 1.41
280
episode: 280   score: 0.0   memory length: 52161   epsilon: 1.0     steps: 122    lr: 0.0
001     evaluation reward: 1.38
281
episode: 281   score: 2.0   memory length: 52379   epsilon: 1.0     steps: 218    lr: 0.0
001     evaluation reward: 1.4
```

```
282
episode: 282    score: 0.0    memory length: 52501    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.38
283
episode: 283    score: 2.0    memory length: 52719    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.4
284
episode: 284    score: 1.0    memory length: 52888    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.41
285
episode: 285    score: 1.0    memory length: 53039    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.4
286
episode: 286    score: 1.0    memory length: 53208    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.39
287
episode: 287    score: 2.0    memory length: 53426    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.39
288
episode: 288    score: 0.0    memory length: 53549    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.36
289
episode: 289    score: 0.0    memory length: 53672    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.36
290
episode: 290    score: 0.0    memory length: 53795    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.36
291
episode: 291    score: 0.0    memory length: 53918    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.36
292
episode: 292    score: 0.0    memory length: 54040    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.3
293
episode: 293    score: 2.0    memory length: 54260    epsilon: 1.0    steps: 220    lr: 0.0
001      evaluation reward: 1.31
294
episode: 294    score: 2.0    memory length: 54457    epsilon: 1.0    steps: 197    lr: 0.0
001      evaluation reward: 1.3
295
episode: 295    score: 3.0    memory length: 54720    epsilon: 1.0    steps: 263    lr: 0.0
001      evaluation reward: 1.31
296
episode: 296    score: 0.0    memory length: 54843    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.31
297
episode: 297    score: 0.0    memory length: 54965    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.31
298
episode: 298    score: 1.0    memory length: 55136    epsilon: 1.0    steps: 171    lr: 0.0
001      evaluation reward: 1.3
299
episode: 299    score: 0.0    memory length: 55258    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.28
300
episode: 300    score: 3.0    memory length: 55524    epsilon: 1.0    steps: 266    lr: 0.0
001      evaluation reward: 1.29
301
episode: 301    score: 2.0    memory length: 55721    epsilon: 1.0    steps: 197    lr: 0.0
001      evaluation reward: 1.31
302
episode: 302    score: 0.0    memory length: 55844    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.3
303
episode: 303    score: 5.0    memory length: 56187    epsilon: 1.0    steps: 343    lr: 0.0
001      evaluation reward: 1.35
```

```
304
episode: 304    score: 2.0    memory length: 56384    epsilon: 1.0    steps: 197    lr: 0.0
001      evaluation reward: 1.37
305
episode: 305    score: 0.0    memory length: 56507    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.35
306
episode: 306    score: 1.0    memory length: 56658    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.36
307
episode: 307    score: 2.0    memory length: 56873    epsilon: 1.0    steps: 215    lr: 0.0
001      evaluation reward: 1.36
308
episode: 308    score: 1.0    memory length: 57044    epsilon: 1.0    steps: 171    lr: 0.0
001      evaluation reward: 1.35
309
episode: 309    score: 1.0    memory length: 57195    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.32
310
episode: 310    score: 0.0    memory length: 57318    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.31
311
episode: 311    score: 2.0    memory length: 57498    epsilon: 1.0    steps: 180    lr: 0.0
001      evaluation reward: 1.32
312
episode: 312    score: 0.0    memory length: 57621    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.32
313
episode: 313    score: 1.0    memory length: 57790    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.32
314
episode: 314    score: 2.0    memory length: 58006    epsilon: 1.0    steps: 216    lr: 0.0
001      evaluation reward: 1.32
315
episode: 315    score: 1.0    memory length: 58157    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.33
316
episode: 316    score: 1.0    memory length: 58329    epsilon: 1.0    steps: 172    lr: 0.0
001      evaluation reward: 1.33
317
episode: 317    score: 3.0    memory length: 58556    epsilon: 1.0    steps: 227    lr: 0.0
001      evaluation reward: 1.35
318
episode: 318    score: 0.0    memory length: 58678    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.34
319
episode: 319    score: 2.0    memory length: 58876    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.34
320
episode: 320    score: 2.0    memory length: 59074    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.36
321
episode: 321    score: 0.0    memory length: 59197    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.36
322
episode: 322    score: 1.0    memory length: 59365    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.32
323
episode: 323    score: 2.0    memory length: 59564    epsilon: 1.0    steps: 199    lr: 0.0
001      evaluation reward: 1.34
324
episode: 324    score: 2.0    memory length: 59764    epsilon: 1.0    steps: 200    lr: 0.0
001      evaluation reward: 1.36
325
episode: 325    score: 2.0    memory length: 59962    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.38
```

```
326
episode: 326    score: 0.0    memory length: 60085    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.33
327
episode: 327    score: 0.0    memory length: 60208    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.31
328
episode: 328    score: 2.0    memory length: 60426    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.31
329
episode: 329    score: 2.0    memory length: 60642    epsilon: 1.0    steps: 216    lr: 0.0
001      evaluation reward: 1.31
330
episode: 330    score: 3.0    memory length: 60868    epsilon: 1.0    steps: 226    lr: 0.0
001      evaluation reward: 1.34
331
episode: 331    score: 0.0    memory length: 60991    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.33
332
episode: 332    score: 1.0    memory length: 61142    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.34
333
episode: 333    score: 2.0    memory length: 61358    epsilon: 1.0    steps: 216    lr: 0.0
001      evaluation reward: 1.36
334
episode: 334    score: 3.0    memory length: 61605    epsilon: 1.0    steps: 247    lr: 0.0
001      evaluation reward: 1.37
335
episode: 335    score: 2.0    memory length: 61823    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.35
336
episode: 336    score: 0.0    memory length: 61946    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.33
337
episode: 337    score: 0.0    memory length: 62069    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.32
338
episode: 338    score: 0.0    memory length: 62192    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.32
339
episode: 339    score: 2.0    memory length: 62410    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.34
340
episode: 340    score: 0.0    memory length: 62533    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.32
341
episode: 341    score: 2.0    memory length: 62752    epsilon: 1.0    steps: 219    lr: 0.0
001      evaluation reward: 1.33
342
episode: 342    score: 3.0    memory length: 62979    epsilon: 1.0    steps: 227    lr: 0.0
001      evaluation reward: 1.36
343
episode: 343    score: 0.0    memory length: 63101    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.35
344
episode: 344    score: 1.0    memory length: 63270    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.34
345
episode: 345    score: 1.0    memory length: 63420    epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.32
346
episode: 346    score: 1.0    memory length: 63570    epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.33
347
episode: 347    score: 0.0    memory length: 63692    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.29
```

```
348
episode: 348    score: 1.0    memory length: 63843    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.27
349
episode: 349    score: 2.0    memory length: 64062    epsilon: 1.0    steps: 219    lr: 0.0
001      evaluation reward: 1.28
350
episode: 350    score: 5.0    memory length: 64365    epsilon: 1.0    steps: 303    lr: 0.0
001      evaluation reward: 1.31
351
episode: 351    score: 0.0    memory length: 64487    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.3
352
episode: 352    score: 3.0    memory length: 64697    epsilon: 1.0    steps: 210    lr: 0.0
001      evaluation reward: 1.32
353
episode: 353    score: 4.0    memory length: 64987    epsilon: 1.0    steps: 290    lr: 0.0
001      evaluation reward: 1.35
354
episode: 354    score: 1.0    memory length: 65156    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.34
355
episode: 355    score: 2.0    memory length: 65361    epsilon: 1.0    steps: 205    lr: 0.0
001      evaluation reward: 1.36
356
episode: 356    score: 0.0    memory length: 65483    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.34
357
episode: 357    score: 3.0    memory length: 65751    epsilon: 1.0    steps: 268    lr: 0.0
001      evaluation reward: 1.37
358
episode: 358    score: 1.0    memory length: 65919    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.37
359
episode: 359    score: 0.0    memory length: 66042    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.36
360
episode: 360    score: 0.0    memory length: 66165    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.28
361
episode: 361    score: 2.0    memory length: 66380    epsilon: 1.0    steps: 215    lr: 0.0
001      evaluation reward: 1.29
362
episode: 362    score: 0.0    memory length: 66503    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.27
363
episode: 363    score: 0.0    memory length: 66625    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.23
364
episode: 364    score: 1.0    memory length: 66776    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.24
365
episode: 365    score: 2.0    memory length: 66957    epsilon: 1.0    steps: 181    lr: 0.0
001      evaluation reward: 1.24
366
episode: 366    score: 1.0    memory length: 67125    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.23
367
episode: 367    score: 0.0    memory length: 67248    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.23
368
episode: 368    score: 1.0    memory length: 67418    epsilon: 1.0    steps: 170    lr: 0.0
001      evaluation reward: 1.23
369
episode: 369    score: 4.0    memory length: 67701    epsilon: 1.0    steps: 283    lr: 0.0
001      evaluation reward: 1.27
```

```
370
episode: 370    score: 2.0    memory length: 67898    epsilon: 1.0    steps: 197    lr: 0.0
001      evaluation reward: 1.28
371
episode: 371    score: 1.0    memory length: 68068    epsilon: 1.0    steps: 170    lr: 0.0
001      evaluation reward: 1.27
372
episode: 372    score: 0.0    memory length: 68191    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.27
373
episode: 373    score: 3.0    memory length: 68417    epsilon: 1.0    steps: 226    lr: 0.0
001      evaluation reward: 1.3
374
episode: 374    score: 2.0    memory length: 68632    epsilon: 1.0    steps: 215    lr: 0.0
001      evaluation reward: 1.32
375
episode: 375    score: 1.0    memory length: 68800    epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.31
376
episode: 376    score: 1.0    memory length: 68951    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.32
377
episode: 377    score: 1.0    memory length: 69121    epsilon: 1.0    steps: 170    lr: 0.0
001      evaluation reward: 1.33
378
episode: 378    score: 5.0    memory length: 69459    epsilon: 1.0    steps: 338    lr: 0.0
001      evaluation reward: 1.33
379
episode: 379    score: 2.0    memory length: 69677    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.31
380
episode: 380    score: 2.0    memory length: 69897    epsilon: 1.0    steps: 220    lr: 0.0
001      evaluation reward: 1.33
381
episode: 381    score: 0.0    memory length: 70020    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.31
382
episode: 382    score: 2.0    memory length: 70238    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.33
383
episode: 383    score: 1.0    memory length: 70407    epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.32
384
episode: 384    score: 1.0    memory length: 70557    epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.32
385
episode: 385    score: 2.0    memory length: 70755    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.33
386
episode: 386    score: 0.0    memory length: 70878    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.32
387
episode: 387    score: 0.0    memory length: 71000    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.3
388
episode: 388    score: 3.0    memory length: 71264    epsilon: 1.0    steps: 264    lr: 0.0
001      evaluation reward: 1.33
389
episode: 389    score: 0.0    memory length: 71386    epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.33
390
episode: 390    score: 2.0    memory length: 71584    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.35
391
episode: 391    score: 3.0    memory length: 71849    epsilon: 1.0    steps: 265    lr: 0.0
001      evaluation reward: 1.38
```

```
392
episode: 392   score: 0.0   memory length: 71972   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.38
393
episode: 393   score: 5.0   memory length: 72294   epsilon: 1.0    steps: 322    lr: 0.0
001      evaluation reward: 1.41
394
episode: 394   score: 1.0   memory length: 72445   epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.4
395
episode: 395   score: 1.0   memory length: 72613   epsilon: 1.0    steps: 168    lr: 0.0
001      evaluation reward: 1.38
396
episode: 396   score: 2.0   memory length: 72811   epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.4
397
episode: 397   score: 0.0   memory length: 72934   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.4
398
episode: 398   score: 5.0   memory length: 73239   epsilon: 1.0    steps: 305    lr: 0.0
001      evaluation reward: 1.44
399
episode: 399   score: 5.0   memory length: 73566   epsilon: 1.0    steps: 327    lr: 0.0
001      evaluation reward: 1.49
400
episode: 400   score: 2.0   memory length: 73747   epsilon: 1.0    steps: 181    lr: 0.0
001      evaluation reward: 1.48
401
episode: 401   score: 0.0   memory length: 73869   epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.46
402
episode: 402   score: 0.0   memory length: 73992   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.46
403
episode: 403   score: 1.0   memory length: 74161   epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.42
404
episode: 404   score: 0.0   memory length: 74284   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.4
405
episode: 405   score: 4.0   memory length: 74559   epsilon: 1.0    steps: 275    lr: 0.0
001      evaluation reward: 1.44
406
episode: 406   score: 0.0   memory length: 74681   epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.43
407
episode: 407   score: 3.0   memory length: 74908   epsilon: 1.0    steps: 227    lr: 0.0
001      evaluation reward: 1.44
408
episode: 408   score: 0.0   memory length: 75031   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.43
409
episode: 409   score: 1.0   memory length: 75200   epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.43
410
episode: 410   score: 2.0   memory length: 75398   epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.45
411
episode: 411   score: 1.0   memory length: 75567   epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.44
412
episode: 412   score: 2.0   memory length: 75784   epsilon: 1.0    steps: 217    lr: 0.0
001      evaluation reward: 1.46
413
episode: 413   score: 4.0   memory length: 76080   epsilon: 1.0    steps: 296    lr: 0.0
001      evaluation reward: 1.49
```

```
414
episode: 414    score: 3.0    memory length: 76308    epsilon: 1.0    steps: 228    lr: 0.0
001     evaluation reward: 1.5
415
episode: 415    score: 0.0    memory length: 76430    epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.49
416
episode: 416    score: 0.0    memory length: 76553    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.48
417
episode: 417    score: 2.0    memory length: 76753    epsilon: 1.0    steps: 200    lr: 0.0
001     evaluation reward: 1.47
418
episode: 418    score: 2.0    memory length: 76974    epsilon: 1.0    steps: 221    lr: 0.0
001     evaluation reward: 1.49
419
episode: 419    score: 0.0    memory length: 77096    epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.47
420
episode: 420    score: 2.0    memory length: 77294    epsilon: 1.0    steps: 198    lr: 0.0
001     evaluation reward: 1.47
421
episode: 421    score: 0.0    memory length: 77417    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.47
422
episode: 422    score: 0.0    memory length: 77540    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.46
423
episode: 423    score: 2.0    memory length: 77756    epsilon: 1.0    steps: 216    lr: 0.0
001     evaluation reward: 1.46
424
episode: 424    score: 1.0    memory length: 77928    epsilon: 1.0    steps: 172    lr: 0.0
001     evaluation reward: 1.45
425
episode: 425    score: 3.0    memory length: 78173    epsilon: 1.0    steps: 245    lr: 0.0
001     evaluation reward: 1.46
426
episode: 426    score: 1.0    memory length: 78345    epsilon: 1.0    steps: 172    lr: 0.0
001     evaluation reward: 1.47
427
episode: 427    score: 2.0    memory length: 78544    epsilon: 1.0    steps: 199    lr: 0.0
001     evaluation reward: 1.49
428
episode: 428    score: 2.0    memory length: 78728    epsilon: 1.0    steps: 184    lr: 0.0
001     evaluation reward: 1.49
429
episode: 429    score: 0.0    memory length: 78851    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.47
430
episode: 430    score: 0.0    memory length: 78974    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.44
431
episode: 431    score: 2.0    memory length: 79172    epsilon: 1.0    steps: 198    lr: 0.0
001     evaluation reward: 1.46
432
episode: 432    score: 1.0    memory length: 79325    epsilon: 1.0    steps: 153    lr: 0.0
001     evaluation reward: 1.46
433
episode: 433    score: 1.0    memory length: 79476    epsilon: 1.0    steps: 151    lr: 0.0
001     evaluation reward: 1.45
434
episode: 434    score: 3.0    memory length: 79724    epsilon: 1.0    steps: 248    lr: 0.0
001     evaluation reward: 1.45
435
episode: 435    score: 2.0    memory length: 79943    epsilon: 1.0    steps: 219    lr: 0.0
001     evaluation reward: 1.45
```

```
436
episode: 436   score: 2.0   memory length: 80140   epsilon: 1.0    steps: 197    lr: 0.0
001     evaluation reward: 1.47
437
episode: 437   score: 0.0   memory length: 80263   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.47
438
episode: 438   score: 0.0   memory length: 80385   epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.47
439
episode: 439   score: 0.0   memory length: 80508   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.45
440
episode: 440   score: 1.0   memory length: 80679   epsilon: 1.0    steps: 171    lr: 0.0
001     evaluation reward: 1.46
441
episode: 441   score: 0.0   memory length: 80802   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.44
442
episode: 442   score: 2.0   memory length: 81021   epsilon: 1.0    steps: 219    lr: 0.0
001     evaluation reward: 1.43
443
episode: 443   score: 0.0   memory length: 81144   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.43
444
episode: 444   score: 5.0   memory length: 81450   epsilon: 1.0    steps: 306    lr: 0.0
001     evaluation reward: 1.47
445
episode: 445   score: 4.0   memory length: 81745   epsilon: 1.0    steps: 295    lr: 0.0
001     evaluation reward: 1.5
446
episode: 446   score: 3.0   memory length: 81991   epsilon: 1.0    steps: 246    lr: 0.0
001     evaluation reward: 1.52
447
episode: 447   score: 0.0   memory length: 82114   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.52
448
episode: 448   score: 1.0   memory length: 82264   epsilon: 1.0    steps: 150    lr: 0.0
001     evaluation reward: 1.52
449
episode: 449   score: 5.0   memory length: 82606   epsilon: 1.0    steps: 342    lr: 0.0
001     evaluation reward: 1.55
450
episode: 450   score: 2.0   memory length: 82824   epsilon: 1.0    steps: 218    lr: 0.0
001     evaluation reward: 1.52
451
episode: 451   score: 0.0   memory length: 82947   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.52
452
episode: 452   score: 2.0   memory length: 83164   epsilon: 1.0    steps: 217    lr: 0.0
001     evaluation reward: 1.51
453
episode: 453   score: 2.0   memory length: 83387   epsilon: 1.0    steps: 223    lr: 0.0
001     evaluation reward: 1.49
454
episode: 454   score: 3.0   memory length: 83614   epsilon: 1.0    steps: 227    lr: 0.0
001     evaluation reward: 1.51
455
episode: 455   score: 0.0   memory length: 83737   epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.49
456
episode: 456   score: 1.0   memory length: 83906   epsilon: 1.0    steps: 169    lr: 0.0
001     evaluation reward: 1.5
457
episode: 457   score: 1.0   memory length: 84057   epsilon: 1.0    steps: 151    lr: 0.0
001     evaluation reward: 1.48
```

```
458
episode: 458   score: 2.0   memory length: 84274   epsilon: 1.0     steps: 217     lr: 0.0
001     evaluation reward: 1.49
459
episode: 459   score: 0.0   memory length: 84396   epsilon: 1.0     steps: 122     lr: 0.0
001     evaluation reward: 1.49
460
episode: 460   score: 0.0   memory length: 84519   epsilon: 1.0     steps: 123     lr: 0.0
001     evaluation reward: 1.49
461
episode: 461   score: 2.0   memory length: 84717   epsilon: 1.0     steps: 198     lr: 0.0
001     evaluation reward: 1.49
462
episode: 462   score: 1.0   memory length: 84887   epsilon: 1.0     steps: 170     lr: 0.0
001     evaluation reward: 1.5
463
episode: 463   score: 0.0   memory length: 85010   epsilon: 1.0     steps: 123     lr: 0.0
001     evaluation reward: 1.5
464
episode: 464   score: 1.0   memory length: 85161   epsilon: 1.0     steps: 151     lr: 0.0
001     evaluation reward: 1.5
465
episode: 465   score: 0.0   memory length: 85284   epsilon: 1.0     steps: 123     lr: 0.0
001     evaluation reward: 1.48
466
episode: 466   score: 5.0   memory length: 85619   epsilon: 1.0     steps: 335     lr: 0.0
001     evaluation reward: 1.52
467
episode: 467   score: 1.0   memory length: 85789   epsilon: 1.0     steps: 170     lr: 0.0
001     evaluation reward: 1.53
468
episode: 468   score: 1.0   memory length: 85940   epsilon: 1.0     steps: 151     lr: 0.0
001     evaluation reward: 1.53
469
episode: 469   score: 4.0   memory length: 86204   epsilon: 1.0     steps: 264     lr: 0.0
001     evaluation reward: 1.53
470
episode: 470   score: 2.0   memory length: 86401   epsilon: 1.0     steps: 197     lr: 0.0
001     evaluation reward: 1.53
471
episode: 471   score: 1.0   memory length: 86571   epsilon: 1.0     steps: 170     lr: 0.0
001     evaluation reward: 1.53
472
episode: 472   score: 4.0   memory length: 86887   epsilon: 1.0     steps: 316     lr: 0.0
001     evaluation reward: 1.57
473
episode: 473   score: 4.0   memory length: 87206   epsilon: 1.0     steps: 319     lr: 0.0
001     evaluation reward: 1.58
474
episode: 474   score: 2.0   memory length: 87404   epsilon: 1.0     steps: 198     lr: 0.0
001     evaluation reward: 1.58
475
episode: 475   score: 2.0   memory length: 87622   epsilon: 1.0     steps: 218     lr: 0.0
001     evaluation reward: 1.59
476
episode: 476   score: 7.0   memory length: 88035   epsilon: 1.0     steps: 413     lr: 0.0
001     evaluation reward: 1.65
477
episode: 477   score: 3.0   memory length: 88260   epsilon: 1.0     steps: 225     lr: 0.0
001     evaluation reward: 1.67
478
episode: 478   score: 1.0   memory length: 88429   epsilon: 1.0     steps: 169     lr: 0.0
001     evaluation reward: 1.63
479
episode: 479   score: 3.0   memory length: 88681   epsilon: 1.0     steps: 252     lr: 0.0
001     evaluation reward: 1.64
```

```
480
episode: 480   score: 3.0   memory length: 88947   epsilon: 1.0    steps: 266    lr: 0.0
001      evaluation reward: 1.65
481
episode: 481   score: 2.0   memory length: 89145   epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.67
482
episode: 482   score: 3.0   memory length: 89371   epsilon: 1.0    steps: 226    lr: 0.0
001      evaluation reward: 1.68
483
episode: 483   score: 2.0   memory length: 89590   epsilon: 1.0    steps: 219    lr: 0.0
001      evaluation reward: 1.69
484
episode: 484   score: 0.0   memory length: 89713   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.68
485
episode: 485   score: 0.0   memory length: 89836   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.66
486
episode: 486   score: 2.0   memory length: 90055   epsilon: 1.0    steps: 219    lr: 0.0
001      evaluation reward: 1.68
487
episode: 487   score: 1.0   memory length: 90205   epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.69
488
episode: 488   score: 0.0   memory length: 90328   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.66
489
episode: 489   score: 1.0   memory length: 90497   epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.67
490
episode: 490   score: 0.0   memory length: 90620   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.65
491
episode: 491   score: 0.0   memory length: 90743   epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.62
492
episode: 492   score: 1.0   memory length: 90912   epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.63
493
episode: 493   score: 0.0   memory length: 91034   epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.58
494
episode: 494   score: 0.0   memory length: 91156   epsilon: 1.0    steps: 122    lr: 0.0
001      evaluation reward: 1.57
495
episode: 495   score: 1.0   memory length: 91325   epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.57
496
episode: 496   score: 1.0   memory length: 91494   epsilon: 1.0    steps: 169    lr: 0.0
001      evaluation reward: 1.56
497
episode: 497   score: 2.0   memory length: 91691   epsilon: 1.0    steps: 197    lr: 0.0
001      evaluation reward: 1.58
498
episode: 498   score: 2.0   memory length: 91909   epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.55
499
episode: 499   score: 2.0   memory length: 92107   epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.52
500
episode: 500   score: 4.0   memory length: 92401   epsilon: 1.0    steps: 294    lr: 0.0
001      evaluation reward: 1.54
501
episode: 501   score: 2.0   memory length: 92599   epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.56
```

```
502
episode: 502    score: 3.0    memory length: 92848    epsilon: 1.0    steps: 249    lr: 0.0
001      evaluation reward: 1.59
503
episode: 503    score: 3.0    memory length: 93098    epsilon: 1.0    steps: 250    lr: 0.0
001      evaluation reward: 1.61
504
episode: 504    score: 2.0    memory length: 93279    epsilon: 1.0    steps: 181    lr: 0.0
001      evaluation reward: 1.63
505
episode: 505    score: 3.0    memory length: 93506    epsilon: 1.0    steps: 227    lr: 0.0
001      evaluation reward: 1.62
506
episode: 506    score: 3.0    memory length: 93774    epsilon: 1.0    steps: 268    lr: 0.0
001      evaluation reward: 1.65
507
episode: 507    score: 2.0    memory length: 93991    epsilon: 1.0    steps: 217    lr: 0.0
001      evaluation reward: 1.64
508
episode: 508    score: 1.0    memory length: 94142    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.65
509
episode: 509    score: 2.0    memory length: 94340    epsilon: 1.0    steps: 198    lr: 0.0
001      evaluation reward: 1.66
510
episode: 510    score: 1.0    memory length: 94491    epsilon: 1.0    steps: 151    lr: 0.0
001      evaluation reward: 1.65
511
episode: 511    score: 1.0    memory length: 94661    epsilon: 1.0    steps: 170    lr: 0.0
001      evaluation reward: 1.65
512
episode: 512    score: 3.0    memory length: 94889    epsilon: 1.0    steps: 228    lr: 0.0
001      evaluation reward: 1.66
513
episode: 513    score: 0.0    memory length: 95012    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.62
514
episode: 514    score: 2.0    memory length: 95215    epsilon: 1.0    steps: 203    lr: 0.0
001      evaluation reward: 1.61
515
episode: 515    score: 2.0    memory length: 95434    epsilon: 1.0    steps: 219    lr: 0.0
001      evaluation reward: 1.63
516
episode: 516    score: 2.0    memory length: 95652    epsilon: 1.0    steps: 218    lr: 0.0
001      evaluation reward: 1.65
517
episode: 517    score: 3.0    memory length: 95899    epsilon: 1.0    steps: 247    lr: 0.0
001      evaluation reward: 1.66
518
episode: 518    score: 1.0    memory length: 96049    epsilon: 1.0    steps: 150    lr: 0.0
001      evaluation reward: 1.65
519
episode: 519    score: 0.0    memory length: 96172    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.65
520
episode: 520    score: 0.0    memory length: 96295    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.63
521
episode: 521    score: 0.0    memory length: 96418    epsilon: 1.0    steps: 123    lr: 0.0
001      evaluation reward: 1.63
522
episode: 522    score: 2.0    memory length: 96598    epsilon: 1.0    steps: 180    lr: 0.0
001      evaluation reward: 1.65
523
episode: 523    score: 1.0    memory length: 96768    epsilon: 1.0    steps: 170    lr: 0.0
001      evaluation reward: 1.64
```

```
524
episode: 524    score: 0.0    memory length: 96891    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.63
525
episode: 525    score: 0.0    memory length: 97014    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.6
526
episode: 526    score: 2.0    memory length: 97216    epsilon: 1.0    steps: 202    lr: 0.0
001     evaluation reward: 1.61
527
episode: 527    score: 1.0    memory length: 97385    epsilon: 1.0    steps: 169    lr: 0.0
001     evaluation reward: 1.6
528
episode: 528    score: 2.0    memory length: 97583    epsilon: 1.0    steps: 198    lr: 0.0
001     evaluation reward: 1.6
529
episode: 529    score: 3.0    memory length: 97828    epsilon: 1.0    steps: 245    lr: 0.0
001     evaluation reward: 1.63
530
episode: 530    score: 1.0    memory length: 97998    epsilon: 1.0    steps: 170    lr: 0.0
001     evaluation reward: 1.64
531
episode: 531    score: 1.0    memory length: 98148    epsilon: 1.0    steps: 150    lr: 0.0
001     evaluation reward: 1.63
532
episode: 532    score: 0.0    memory length: 98270    epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.62
533
episode: 533    score: 0.0    memory length: 98393    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.61
534
episode: 534    score: 5.0    memory length: 98735    epsilon: 1.0    steps: 342    lr: 0.0
001     evaluation reward: 1.63
535
episode: 535    score: 0.0    memory length: 98857    epsilon: 1.0    steps: 122    lr: 0.0
001     evaluation reward: 1.61
536
episode: 536    score: 2.0    memory length: 99054    epsilon: 1.0    steps: 197    lr: 0.0
001     evaluation reward: 1.61
537
episode: 537    score: 2.0    memory length: 99255    epsilon: 1.0    steps: 201    lr: 0.0
001     evaluation reward: 1.63
538
episode: 538    score: 0.0    memory length: 99378    epsilon: 1.0    steps: 123    lr: 0.0
001     evaluation reward: 1.63
539
episode: 539    score: 3.0    memory length: 99627    epsilon: 1.0    steps: 249    lr: 0.0
001     evaluation reward: 1.66
540
episode: 540    score: 3.0    memory length: 99895    epsilon: 1.0    steps: 268    lr: 0.0
001     evaluation reward: 1.68
541
episode: 541    score: 2.0    memory length: 100095    epsilon: 0.9998099200000041    step
s: 200    lr: 0.0001     evaluation reward: 1.7
542
episode: 542    score: 0.0    memory length: 100218    epsilon: 0.9995663800000094    step
s: 123    lr: 0.0001     evaluation reward: 1.68
543
episode: 543    score: 8.0    memory length: 100644    epsilon: 0.9987229000000277    step
s: 426    lr: 0.0001     evaluation reward: 1.76
544
episode: 544    score: 1.0    memory length: 100812    epsilon: 0.998390260000035    steps:
 168    lr: 0.0001     evaluation reward: 1.72
545
episode: 545    score: 1.0    memory length: 100981    epsilon: 0.9980556400000422    step
s: 169    lr: 0.0001     evaluation reward: 1.69
```

```
546
episode: 546   score: 2.0   memory length: 101178   epsilon: 0.9976655800000507      step
s: 197    lr: 0.0001     evaluation reward: 1.68
547
episode: 547   score: 0.0   memory length: 101301   epsilon: 0.997422040000056      steps:
 123    lr: 0.0001     evaluation reward: 1.68
548
episode: 548   score: 3.0   memory length: 101548   epsilon: 0.9969329800000666      step
s: 247    lr: 0.0001     evaluation reward: 1.7
549
episode: 549   score: 2.0   memory length: 101745   epsilon: 0.996542920000075      steps:
 197    lr: 0.0001     evaluation reward: 1.67
550
episode: 550   score: 0.0   memory length: 101868   epsilon: 0.9962993800000803      step
s: 123    lr: 0.0001     evaluation reward: 1.65
551
episode: 551   score: 3.0   memory length: 102113   epsilon: 0.9958142800000909      step
s: 245    lr: 0.0001     evaluation reward: 1.68
552
episode: 552   score: 6.0   memory length: 102491   epsilon: 0.9950658400001071      step
s: 378    lr: 0.0001     evaluation reward: 1.72
553
episode: 553   score: 5.0   memory length: 102783   epsilon: 0.9944876800001197      step
s: 292    lr: 0.0001     evaluation reward: 1.75
554
episode: 554   score: 0.0   memory length: 102906   epsilon: 0.994244140000125      steps:
 123    lr: 0.0001     evaluation reward: 1.72
555
episode: 555   score: 5.0   memory length: 103185   epsilon: 0.993691720000137      steps:
 279    lr: 0.0001     evaluation reward: 1.77
556
episode: 556   score: 3.0   memory length: 103431   epsilon: 0.9932046400001475      step
s: 246    lr: 0.0001     evaluation reward: 1.79
557
episode: 557   score: 2.0   memory length: 103629   epsilon: 0.992812600000156      steps:
 198    lr: 0.0001     evaluation reward: 1.8
558
episode: 558   score: 2.0   memory length: 103847   epsilon: 0.9923809600001654      step
s: 218    lr: 0.0001     evaluation reward: 1.8
559
episode: 559   score: 1.0   memory length: 104015   epsilon: 0.9920483200001726      step
s: 168    lr: 0.0001     evaluation reward: 1.81
560
episode: 560   score: 0.0   memory length: 104138   epsilon: 0.9918047800001779      step
s: 123    lr: 0.0001     evaluation reward: 1.81
561
episode: 561   score: 3.0   memory length: 104385   epsilon: 0.9913157200001885      step
s: 247    lr: 0.0001     evaluation reward: 1.82
562
episode: 562   score: 1.0   memory length: 104536   epsilon: 0.991016740000195      steps:
 151    lr: 0.0001     evaluation reward: 1.82
563
episode: 563   score: 3.0   memory length: 104763   epsilon: 0.9905672800002048      step
s: 227    lr: 0.0001     evaluation reward: 1.85
564
episode: 564   score: 0.0   memory length: 104886   epsilon: 0.9903237400002101      step
s: 123    lr: 0.0001     evaluation reward: 1.84
565
episode: 565   score: 0.0   memory length: 105008   epsilon: 0.9900821800002153      step
s: 122    lr: 0.0001     evaluation reward: 1.84
566
episode: 566   score: 1.0   memory length: 105177   epsilon: 0.9897475600002226      step
s: 169    lr: 0.0001     evaluation reward: 1.8
567
episode: 567   score: 0.0   memory length: 105300   epsilon: 0.9895040200002279      step
s: 123    lr: 0.0001     evaluation reward: 1.79
```

```
568
episode: 568   score: 3.0   memory length: 105564   epsilon: 0.9889813000002392     step
s: 264    lr: 0.0001     evaluation reward: 1.81
569
episode: 569   score: 2.0   memory length: 105781   epsilon: 0.9885516400002485     step
s: 217    lr: 0.0001     evaluation reward: 1.79
570
episode: 570   score: 3.0   memory length: 106028   epsilon: 0.9880625800002591     step
s: 247    lr: 0.0001     evaluation reward: 1.8
571
episode: 571   score: 5.0   memory length: 106371   epsilon: 0.9873834400002739     step
s: 343    lr: 0.0001     evaluation reward: 1.84
572
episode: 572   score: 0.0   memory length: 106493   epsilon: 0.9871418800002791     step
s: 122    lr: 0.0001     evaluation reward: 1.8
573
episode: 573   score: 2.0   memory length: 106710   epsilon: 0.9867122200002885     step
s: 217    lr: 0.0001     evaluation reward: 1.78
574
episode: 574   score: 0.0   memory length: 106832   epsilon: 0.9864706600002937     step
s: 122    lr: 0.0001     evaluation reward: 1.76
575
episode: 575   score: 1.0   memory length: 106983   epsilon: 0.9861716800003002     step
s: 151    lr: 0.0001     evaluation reward: 1.75
576
episode: 576   score: 3.0   memory length: 107250   epsilon: 0.9856430200003117     step
s: 267    lr: 0.0001     evaluation reward: 1.71
577
episode: 577   score: 0.0   memory length: 107373   epsilon: 0.985399480000317     steps:
 123    lr: 0.0001     evaluation reward: 1.68
578
episode: 578   score: 0.0   memory length: 107495   epsilon: 0.9851579200003222     step
s: 122    lr: 0.0001     evaluation reward: 1.67
579
episode: 579   score: 5.0   memory length: 107837   epsilon: 0.9844807600003369     step
s: 342    lr: 0.0001     evaluation reward: 1.69
580
episode: 580   score: 0.0   memory length: 107960   epsilon: 0.9842372200003422     step
s: 123    lr: 0.0001     evaluation reward: 1.66
581
episode: 581   score: 3.0   memory length: 108227   epsilon: 0.9837085600003537     step
s: 267    lr: 0.0001     evaluation reward: 1.67
582
episode: 582   score: 0.0   memory length: 108350   epsilon: 0.983465020000359     steps:
 123    lr: 0.0001     evaluation reward: 1.64
583
episode: 583   score: 0.0   memory length: 108473   epsilon: 0.9832214800003642     step
s: 123    lr: 0.0001     evaluation reward: 1.62
584
episode: 584   score: 1.0   memory length: 108623   epsilon: 0.9829244800003707     step
s: 150    lr: 0.0001     evaluation reward: 1.63
585
episode: 585   score: 0.0   memory length: 108746   epsilon: 0.982680940000376     steps:
 123    lr: 0.0001     evaluation reward: 1.63
586
episode: 586   score: 1.0   memory length: 108914   epsilon: 0.9823483000003832     step
s: 168    lr: 0.0001     evaluation reward: 1.62
587
episode: 587   score: 3.0   memory length: 109129   epsilon: 0.9819226000003924     step
s: 215    lr: 0.0001     evaluation reward: 1.64
588
episode: 588   score: 0.0   memory length: 109252   epsilon: 0.9816790600003977     step
s: 123    lr: 0.0001     evaluation reward: 1.64
589
episode: 589   score: 1.0   memory length: 109402   epsilon: 0.9813820600004042     step
s: 150    lr: 0.0001     evaluation reward: 1.64
```

```
590
episode: 590   score: 1.0   memory length: 109571   epsilon: 0.9810474400004114     step
s: 169    lr: 0.0001     evaluation reward: 1.65
591
episode: 591   score: 0.0   memory length: 109693   epsilon: 0.9808058800004167     step
s: 122    lr: 0.0001     evaluation reward: 1.65
592
episode: 592   score: 3.0   memory length: 109922   epsilon: 0.9803524600004265     step
s: 229    lr: 0.0001     evaluation reward: 1.67
593
episode: 593   score: 0.0   memory length: 110044   epsilon: 0.9801109000004318     step
s: 122    lr: 0.0001     evaluation reward: 1.67
594
episode: 594   score: 0.0   memory length: 110167   epsilon: 0.9798673600004371     step
s: 123    lr: 0.0001     evaluation reward: 1.67
595
episode: 595   score: 1.0   memory length: 110318   epsilon: 0.9795683800004436     step
s: 151    lr: 0.0001     evaluation reward: 1.67
596
episode: 596   score: 2.0   memory length: 110535   epsilon: 0.9791387200004529     step
s: 217    lr: 0.0001     evaluation reward: 1.68
597
episode: 597   score: 2.0   memory length: 110735   epsilon: 0.9787427200004615     step
s: 200    lr: 0.0001     evaluation reward: 1.68
598
episode: 598   score: 4.0   memory length: 111051   epsilon: 0.9781170400004751     step
s: 316    lr: 0.0001     evaluation reward: 1.7
599
episode: 599   score: 0.0   memory length: 111174   epsilon: 0.9778735000004803     step
s: 123    lr: 0.0001     evaluation reward: 1.68
600
episode: 600   score: 0.0   memory length: 111297   epsilon: 0.9776299600004856     step
s: 123    lr: 0.0001     evaluation reward: 1.64
601
episode: 601   score: 0.0   memory length: 111420   epsilon: 0.9773864200004909     step
s: 123    lr: 0.0001     evaluation reward: 1.62
602
episode: 602   score: 2.0   memory length: 111617   epsilon: 0.9769963600004994     step
s: 197    lr: 0.0001     evaluation reward: 1.61
603
episode: 603   score: 2.0   memory length: 111834   epsilon: 0.9765667000005087     step
s: 217    lr: 0.0001     evaluation reward: 1.6
604
episode: 604   score: 0.0   memory length: 111956   epsilon: 0.976325140000514     steps:
 122    lr: 0.0001     evaluation reward: 1.58
605
episode: 605   score: 0.0   memory length: 112078   epsilon: 0.9760835800005192     step
s: 122    lr: 0.0001     evaluation reward: 1.55
606
episode: 606   score: 1.0   memory length: 112228   epsilon: 0.9757865800005256     step
s: 150    lr: 0.0001     evaluation reward: 1.53
607
episode: 607   score: 5.0   memory length: 112535   epsilon: 0.9751787200005388     step
s: 307    lr: 0.0001     evaluation reward: 1.56
608
episode: 608   score: 2.0   memory length: 112732   epsilon: 0.9747886600005473     step
s: 197    lr: 0.0001     evaluation reward: 1.57
609
episode: 609   score: 1.0   memory length: 112903   epsilon: 0.9744500800005547     step
s: 171    lr: 0.0001     evaluation reward: 1.56
610
episode: 610   score: 2.0   memory length: 113101   epsilon: 0.9740580400005632     step
s: 198    lr: 0.0001     evaluation reward: 1.57
611
episode: 611   score: 0.0   memory length: 113223   epsilon: 0.9738164800005684     step
s: 122    lr: 0.0001     evaluation reward: 1.56
```

```
612
episode: 612   score: 1.0   memory length: 113393   epsilon: 0.9734798800005757    step
s: 170    lr: 0.0001    evaluation reward: 1.54
613
episode: 613   score: 2.0   memory length: 113591   epsilon: 0.9730878400005842    step
s: 198    lr: 0.0001    evaluation reward: 1.56
614
episode: 614   score: 2.0   memory length: 113773   epsilon: 0.9727274800005921    step
s: 182    lr: 0.0001    evaluation reward: 1.56
615
episode: 615   score: 0.0   memory length: 113895   epsilon: 0.9724859200005973    step
s: 122    lr: 0.0001    evaluation reward: 1.54
616
episode: 616   score: 3.0   memory length: 114140   epsilon: 0.9720008200006078    step
s: 245    lr: 0.0001    evaluation reward: 1.55
617
episode: 617   score: 1.0   memory length: 114292   epsilon: 0.9716998600006144    step
s: 152    lr: 0.0001    evaluation reward: 1.53
618
episode: 618   score: 1.0   memory length: 114442   epsilon: 0.9714028600006208    step
s: 150    lr: 0.0001    evaluation reward: 1.53
619
episode: 619   score: 0.0   memory length: 114565   epsilon: 0.9711593200006261    step
s: 123    lr: 0.0001    evaluation reward: 1.53
620
episode: 620   score: 1.0   memory length: 114736   epsilon: 0.9708207400006335    step
s: 171    lr: 0.0001    evaluation reward: 1.54
621
episode: 621   score: 0.0   memory length: 114858   epsilon: 0.9705791800006387    step
s: 122    lr: 0.0001    evaluation reward: 1.54
622
episode: 622   score: 0.0   memory length: 114981   epsilon: 0.970335640000644    steps:
 123    lr: 0.0001    evaluation reward: 1.52
623
episode: 623   score: 0.0   memory length: 115104   epsilon: 0.9700921000006493    step
s: 123    lr: 0.0001    evaluation reward: 1.51
624
episode: 624   score: 1.0   memory length: 115255   epsilon: 0.9697931200006558    step
s: 151    lr: 0.0001    evaluation reward: 1.52
625
episode: 625   score: 1.0   memory length: 115405   epsilon: 0.9694961200006622    step
s: 150    lr: 0.0001    evaluation reward: 1.53
626
episode: 626   score: 1.0   memory length: 115576   epsilon: 0.9691575400006696    step
s: 171    lr: 0.0001    evaluation reward: 1.52
627
episode: 627   score: 2.0   memory length: 115774   epsilon: 0.9687655000006781    step
s: 198    lr: 0.0001    evaluation reward: 1.53
628
episode: 628   score: 1.0   memory length: 115942   epsilon: 0.9684328600006853    step
s: 168    lr: 0.0001    evaluation reward: 1.52
629
episode: 629   score: 1.0   memory length: 116093   epsilon: 0.9681338800006918    step
s: 151    lr: 0.0001    evaluation reward: 1.5
630
episode: 630   score: 0.0   memory length: 116216   epsilon: 0.9678903400006971    step
s: 123    lr: 0.0001    evaluation reward: 1.49
631
episode: 631   score: 2.0   memory length: 116414   epsilon: 0.9674983000007056    step
s: 198    lr: 0.0001    evaluation reward: 1.5
632
episode: 632   score: 1.0   memory length: 116564   epsilon: 0.967201300000712    steps:
 150    lr: 0.0001    evaluation reward: 1.51
633
episode: 633   score: 2.0   memory length: 116762   epsilon: 0.9668092600007205    step
s: 198    lr: 0.0001    evaluation reward: 1.53
```

```
634
episode: 634   score: 0.0   memory length: 116884   epsilon: 0.9665677000007258    step
s: 122    lr: 0.0001     evaluation reward: 1.48
635
episode: 635   score: 0.0   memory length: 117007   epsilon: 0.9663241600007311    step
s: 123    lr: 0.0001     evaluation reward: 1.48
636
episode: 636   score: 2.0   memory length: 117223   epsilon: 0.9658964800007404    step
s: 216    lr: 0.0001     evaluation reward: 1.48
637
episode: 637   score: 0.0   memory length: 117346   epsilon: 0.9656529400007456    step
s: 123    lr: 0.0001     evaluation reward: 1.46
638
episode: 638   score: 2.0   memory length: 117562   epsilon: 0.9652252600007549    step
s: 216    lr: 0.0001     evaluation reward: 1.48
639
episode: 639   score: 0.0   memory length: 117685   epsilon: 0.9649817200007602    step
s: 123    lr: 0.0001     evaluation reward: 1.45
640
episode: 640   score: 3.0   memory length: 117952   epsilon: 0.9644530600007717    step
s: 267    lr: 0.0001     evaluation reward: 1.45
641
episode: 641   score: 3.0   memory length: 118197   epsilon: 0.9639679600007822    step
s: 245    lr: 0.0001     evaluation reward: 1.46
642
episode: 642   score: 1.0   memory length: 118365   epsilon: 0.9636353200007894    step
s: 168    lr: 0.0001     evaluation reward: 1.47
643
episode: 643   score: 0.0   memory length: 118487   epsilon: 0.9633937600007947    step
s: 122    lr: 0.0001     evaluation reward: 1.39
644
episode: 644   score: 0.0   memory length: 118610   epsilon: 0.9631502200008    steps: 1
23    lr: 0.0001     evaluation reward: 1.38
645
episode: 645   score: 5.0   memory length: 118913   epsilon: 0.962550280000813    steps:
 303    lr: 0.0001     evaluation reward: 1.42
646
episode: 646   score: 1.0   memory length: 119066   epsilon: 0.9622473400008196    step
s: 153    lr: 0.0001     evaluation reward: 1.41
647
episode: 647   score: 1.0   memory length: 119235   epsilon: 0.9619127200008268    step
s: 169    lr: 0.0001     evaluation reward: 1.42
648
episode: 648   score: 1.0   memory length: 119404   epsilon: 0.9615781000008341    step
s: 169    lr: 0.0001     evaluation reward: 1.4
649
episode: 649   score: 0.0   memory length: 119527   epsilon: 0.9613345600008394    step
s: 123    lr: 0.0001     evaluation reward: 1.38
650
episode: 650   score: 3.0   memory length: 119772   epsilon: 0.9608494600008499    step
s: 245    lr: 0.0001     evaluation reward: 1.41
651
episode: 651   score: 4.0   memory length: 120087   epsilon: 0.9602257600008635    step
s: 315    lr: 0.0001     evaluation reward: 1.42
652
episode: 652   score: 3.0   memory length: 120353   epsilon: 0.9596990800008749    step
s: 266    lr: 0.0001     evaluation reward: 1.39
653
episode: 653   score: 1.0   memory length: 120504   epsilon: 0.9594001000008814    step
s: 151    lr: 0.0001     evaluation reward: 1.35
654
episode: 654   score: 0.0   memory length: 120627   epsilon: 0.9591565600008867    step
s: 123    lr: 0.0001     evaluation reward: 1.35
655
episode: 655   score: 5.0   memory length: 120952   epsilon: 0.9585130600009006    step
s: 325    lr: 0.0001     evaluation reward: 1.35
```

656
episode: 656   score: 3.0   memory length: 121200   epsilon: 0.9580220200009113   step
s: 248   lr: 0.0001   evaluation reward: 1.35
657
episode: 657   score: 0.0   memory length: 121323   epsilon: 0.9577784800009166   step
s: 123   lr: 0.0001   evaluation reward: 1.33
658
episode: 658   score: 9.0   memory length: 121741   epsilon: 0.9569508400009346   step
s: 418   lr: 0.0001   evaluation reward: 1.4
659
episode: 659   score: 0.0   memory length: 121864   epsilon: 0.9567073000009398   step
s: 123   lr: 0.0001   evaluation reward: 1.39
660
episode: 660   score: 2.0   memory length: 122044   epsilon: 0.9563509000009476   step
s: 180   lr: 0.0001   evaluation reward: 1.41
661
episode: 661   score: 3.0   memory length: 122293   epsilon: 0.9558578800009583   step
s: 249   lr: 0.0001   evaluation reward: 1.41
662
episode: 662   score: 0.0   memory length: 122416   epsilon: 0.9556143400009636   step
s: 123   lr: 0.0001   evaluation reward: 1.4
663
episode: 663   score: 2.0   memory length: 122613   epsilon: 0.955224280000972   steps:
 197   lr: 0.0001   evaluation reward: 1.39
664
episode: 664   score: 4.0   memory length: 122892   epsilon: 0.954671860000984   steps:
 279   lr: 0.0001   evaluation reward: 1.43
665
episode: 665   score: 1.0   memory length: 123062   epsilon: 0.9543352600009913   step
s: 170   lr: 0.0001   evaluation reward: 1.44
666
episode: 666   score: 0.0   memory length: 123185   epsilon: 0.9540917200009966   step
s: 123   lr: 0.0001   evaluation reward: 1.43
667
episode: 667   score: 0.0   memory length: 123307   epsilon: 0.9538501600010019   step
s: 122   lr: 0.0001   evaluation reward: 1.43
668
episode: 668   score: 2.0   memory length: 123507   epsilon: 0.9534541600010105   step
s: 200   lr: 0.0001   evaluation reward: 1.42
669
episode: 669   score: 0.0   memory length: 123629   epsilon: 0.9532126000010157   step
s: 122   lr: 0.0001   evaluation reward: 1.4
670
episode: 670   score: 3.0   memory length: 123896   epsilon: 0.9526839400010272   step
s: 267   lr: 0.0001   evaluation reward: 1.4
671
episode: 671   score: 3.0   memory length: 124125   epsilon: 0.952230520001037   steps:
 229   lr: 0.0001   evaluation reward: 1.38
672
episode: 672   score: 0.0   memory length: 124248   epsilon: 0.9519869800010423   step
s: 123   lr: 0.0001   evaluation reward: 1.38
673
episode: 673   score: 2.0   memory length: 124428   epsilon: 0.95163058000105   steps:
 180   lr: 0.0001   evaluation reward: 1.38
674
episode: 674   score: 0.0   memory length: 124551   epsilon: 0.9513870400010553   step
s: 123   lr: 0.0001   evaluation reward: 1.38
675
episode: 675   score: 0.0   memory length: 124674   epsilon: 0.9511435000010606   step
s: 123   lr: 0.0001   evaluation reward: 1.37
676
episode: 676   score: 2.0   memory length: 124872   epsilon: 0.9507514600010691   step
s: 198   lr: 0.0001   evaluation reward: 1.36
677
episode: 677   score: 4.0   memory length: 125168   epsilon: 0.9501653800010819   step
s: 296   lr: 0.0001   evaluation reward: 1.4

```
678
episode: 678   score: 1.0   memory length: 125319   epsilon: 0.9498664000010884     step
s: 151    lr: 0.0001    evaluation reward: 1.41
679
episode: 679   score: 2.0   memory length: 125499   epsilon: 0.9495100000010961     step
s: 180    lr: 0.0001    evaluation reward: 1.38
680
episode: 680   score: 2.0   memory length: 125716   epsilon: 0.9490803400011054     step
s: 217    lr: 0.0001    evaluation reward: 1.4
681
episode: 681   score: 1.0   memory length: 125866   epsilon: 0.9487833400011119     step
s: 150    lr: 0.0001    evaluation reward: 1.38
682
episode: 682   score: 3.0   memory length: 126091   epsilon: 0.9483378400011215     step
s: 225    lr: 0.0001    evaluation reward: 1.41
683
episode: 683   score: 0.0   memory length: 126214   epsilon: 0.9480943000011268     step
s: 123    lr: 0.0001    evaluation reward: 1.41
684
episode: 684   score: 4.0   memory length: 126509   epsilon: 0.9475102000011395     step
s: 295    lr: 0.0001    evaluation reward: 1.44
685
episode: 685   score: 0.0   memory length: 126632   epsilon: 0.9472666600011448     step
s: 123    lr: 0.0001    evaluation reward: 1.44
686
episode: 686   score: 3.0   memory length: 126876   epsilon: 0.9467835400011553     step
s: 244    lr: 0.0001    evaluation reward: 1.46
687
episode: 687   score: 1.0   memory length: 127046   epsilon: 0.9464469400011626     step
s: 170    lr: 0.0001    evaluation reward: 1.44
688
episode: 688   score: 0.0   memory length: 127169   epsilon: 0.9462034000011679     step
s: 123    lr: 0.0001    evaluation reward: 1.44
689
episode: 689   score: 4.0   memory length: 127448   epsilon: 0.9456509800011799     step
s: 279    lr: 0.0001    evaluation reward: 1.47
690
episode: 690   score: 5.0   memory length: 127782   epsilon: 0.9449896600011942     step
s: 334    lr: 0.0001    evaluation reward: 1.51
691
episode: 691   score: 1.0   memory length: 127954   epsilon: 0.9446491000012016     step
s: 172    lr: 0.0001    evaluation reward: 1.52
692
episode: 692   score: 1.0   memory length: 128106   epsilon: 0.9443481400012081     step
s: 152    lr: 0.0001    evaluation reward: 1.5
693
episode: 693   score: 0.0   memory length: 128229   epsilon: 0.9441046000012134     step
s: 123    lr: 0.0001    evaluation reward: 1.5
694
episode: 694   score: 0.0   memory length: 128351   epsilon: 0.9438630400012187     step
s: 122    lr: 0.0001    evaluation reward: 1.5
695
episode: 695   score: 2.0   memory length: 128568   epsilon: 0.943433380001228     steps:
 217    lr: 0.0001     evaluation reward: 1.51
696
episode: 696   score: 1.0   memory length: 128737   epsilon: 0.9430987600012353     step
s: 169    lr: 0.0001    evaluation reward: 1.5
697
episode: 697   score: 1.0   memory length: 128888   epsilon: 0.9427997800012418     step
s: 151    lr: 0.0001    evaluation reward: 1.49
698
episode: 698   score: 3.0   memory length: 129114   epsilon: 0.9423523000012515     step
s: 226    lr: 0.0001    evaluation reward: 1.48
699
episode: 699   score: 2.0   memory length: 129312   epsilon: 0.94196026000126     steps:
 198    lr: 0.0001     evaluation reward: 1.5
```

```
700
episode: 700   score: 2.0   memory length: 129510   epsilon: 0.9415682200012685      step
s: 198    lr: 0.0001     evaluation reward: 1.52
701
episode: 701   score: 1.0   memory length: 129661   epsilon: 0.941269240001275     steps:
 151    lr: 0.0001     evaluation reward: 1.53
702
episode: 702   score: 3.0   memory length: 129909   epsilon: 0.9407782000012856      step
s: 248    lr: 0.0001     evaluation reward: 1.54
703
episode: 703   score: 3.0   memory length: 130134   epsilon: 0.94033270000012953      step
s: 225    lr: 0.0001     evaluation reward: 1.55
704
episode: 704   score: 1.0   memory length: 130285   epsilon: 0.9400337200013018      step
s: 151    lr: 0.0001     evaluation reward: 1.56
705
episode: 705   score: 0.0   memory length: 130408   epsilon: 0.9397901800013071      step
s: 123    lr: 0.0001     evaluation reward: 1.56
706
episode: 706   score: 2.0   memory length: 130588   epsilon: 0.9394337800013148      step
s: 180    lr: 0.0001     evaluation reward: 1.57
707
episode: 707   score: 2.0   memory length: 130789   epsilon: 0.9390358000013235      step
s: 201    lr: 0.0001     evaluation reward: 1.54
708
episode: 708   score: 1.0   memory length: 130959   epsilon: 0.9386992000013308      step
s: 170    lr: 0.0001     evaluation reward: 1.53
709
episode: 709   score: 2.0   memory length: 131177   epsilon: 0.9382675600013402      step
s: 218    lr: 0.0001     evaluation reward: 1.54
710
episode: 710   score: 0.0   memory length: 131299   epsilon: 0.9380260000013454      step
s: 122    lr: 0.0001     evaluation reward: 1.52
711
episode: 711   score: 2.0   memory length: 131518   epsilon: 0.9375923800013548      step
s: 219    lr: 0.0001     evaluation reward: 1.54
712
episode: 712   score: 1.0   memory length: 131690   epsilon: 0.9372518200013622      step
s: 172    lr: 0.0001     evaluation reward: 1.54
713
episode: 713   score: 5.0   memory length: 132015   epsilon: 0.9366083200013762      step
s: 325    lr: 0.0001     evaluation reward: 1.57
714
episode: 714   score: 2.0   memory length: 132213   epsilon: 0.9362162800013847      step
s: 198    lr: 0.0001     evaluation reward: 1.57
715
episode: 715   score: 0.0   memory length: 132336   epsilon: 0.93597274000139     steps:
 123    lr: 0.0001     evaluation reward: 1.57
716
episode: 716   score: 2.0   memory length: 132518   epsilon: 0.9356123800013978      step
s: 182    lr: 0.0001     evaluation reward: 1.56
717
episode: 717   score: 1.0   memory length: 132687   epsilon: 0.9352777600014051      step
s: 169    lr: 0.0001     evaluation reward: 1.56
718
episode: 718   score: 2.0   memory length: 132869   epsilon: 0.9349174000014129      step
s: 182    lr: 0.0001     evaluation reward: 1.57
719
episode: 719   score: 1.0   memory length: 133041   epsilon: 0.9345768400014203      step
s: 172    lr: 0.0001     evaluation reward: 1.58
720
episode: 720   score: 3.0   memory length: 133286   epsilon: 0.9340917400014308      step
s: 245    lr: 0.0001     evaluation reward: 1.6
721
episode: 721   score: 3.0   memory length: 133555   epsilon: 0.9335591200014424      step
s: 269    lr: 0.0001     evaluation reward: 1.63
```

```
722
episode: 722   score: 0.0   memory length: 133678   epsilon: 0.9333155800014477    step
s: 123    lr: 0.0001    evaluation reward: 1.63
723
episode: 723   score: 0.0   memory length: 133801   epsilon: 0.9330720400014529    step
s: 123    lr: 0.0001    evaluation reward: 1.63
724
episode: 724   score: 1.0   memory length: 133952   epsilon: 0.9327730600014594    step
s: 151    lr: 0.0001    evaluation reward: 1.63
725
episode: 725   score: 1.0   memory length: 134103   epsilon: 0.9324740800014659    step
s: 151    lr: 0.0001    evaluation reward: 1.63
726
episode: 726   score: 0.0   memory length: 134225   epsilon: 0.9322325200014712    step
s: 122    lr: 0.0001    evaluation reward: 1.62
727
episode: 727   score: 0.0   memory length: 134348   epsilon: 0.9319889800014765    step
s: 123    lr: 0.0001    evaluation reward: 1.6
728
episode: 728   score: 3.0   memory length: 134599   epsilon: 0.9314920000014872    step
s: 251    lr: 0.0001    evaluation reward: 1.62
729
episode: 729   score: 0.0   memory length: 134722   epsilon: 0.9312484600014925    step
s: 123    lr: 0.0001    evaluation reward: 1.61
730
episode: 730   score: 3.0   memory length: 134968   epsilon: 0.9307613800015031    step
s: 246    lr: 0.0001    evaluation reward: 1.64
731
episode: 731   score: 0.0   memory length: 135090   epsilon: 0.9305198200015083    step
s: 122    lr: 0.0001    evaluation reward: 1.62
732
episode: 732   score: 2.0   memory length: 135308   epsilon: 0.9300881800015177    step
s: 218    lr: 0.0001    evaluation reward: 1.63
733
episode: 733   score: 0.0   memory length: 135431   epsilon: 0.929844640001523    steps:
 123    lr: 0.0001    evaluation reward: 1.61
734
episode: 734   score: 1.0   memory length: 135600   epsilon: 0.9295100200015303    step
s: 169    lr: 0.0001    evaluation reward: 1.62
735
episode: 735   score: 1.0   memory length: 135750   epsilon: 0.9292130200015367    step
s: 150    lr: 0.0001    evaluation reward: 1.63
736
episode: 736   score: 0.0   memory length: 135873   epsilon: 0.928969480001542    steps:
 123    lr: 0.0001    evaluation reward: 1.61
737
episode: 737   score: 3.0   memory length: 136120   epsilon: 0.9284804200015526    step
s: 247    lr: 0.0001    evaluation reward: 1.64
738
episode: 738   score: 0.0   memory length: 136243   epsilon: 0.9282368800015579    step
s: 123    lr: 0.0001    evaluation reward: 1.62
739
episode: 739   score: 0.0   memory length: 136365   epsilon: 0.9279953200015632    step
s: 122    lr: 0.0001    evaluation reward: 1.62
740
episode: 740   score: 1.0   memory length: 136517   epsilon: 0.9276943600015697    step
s: 152    lr: 0.0001    evaluation reward: 1.6
741
episode: 741   score: 2.0   memory length: 136717   epsilon: 0.9272983600015783    step
s: 200    lr: 0.0001    evaluation reward: 1.59
742
episode: 742   score: 4.0   memory length: 136995   epsilon: 0.9267479200015902    step
s: 278    lr: 0.0001    evaluation reward: 1.62
743
episode: 743   score: 2.0   memory length: 137193   epsilon: 0.9263558800015987    step
s: 198    lr: 0.0001    evaluation reward: 1.64
```

```
744
episode: 744    score: 6.0    memory length: 137548    epsilon: 0.925652980001614      steps:
 355    lr: 0.0001      evaluation reward: 1.7
745
episode: 745    score: 3.0    memory length: 137792    epsilon: 0.9251698600016245      step
s: 244    lr: 0.0001      evaluation reward: 1.68
746
episode: 746    score: 1.0    memory length: 137960    epsilon: 0.9248372200016317      step
s: 168    lr: 0.0001      evaluation reward: 1.68
747
episode: 747    score: 2.0    memory length: 138158    epsilon: 0.9244451800016402      step
s: 198    lr: 0.0001      evaluation reward: 1.69
748
episode: 748    score: 4.0    memory length: 138426    epsilon: 0.9239145400016517      step
s: 268    lr: 0.0001      evaluation reward: 1.72
749
episode: 749    score: 2.0    memory length: 138643    epsilon: 0.9234848800016611      step
s: 217    lr: 0.0001      evaluation reward: 1.74
750
episode: 750    score: 0.0    memory length: 138765    epsilon: 0.9232433200016663      step
s: 122    lr: 0.0001      evaluation reward: 1.71
751
episode: 751    score: 1.0    memory length: 138915    epsilon: 0.9229463200016728      step
s: 150    lr: 0.0001      evaluation reward: 1.68
752
episode: 752    score: 0.0    memory length: 139038    epsilon: 0.922702780001678      steps:
 123    lr: 0.0001      evaluation reward: 1.65
753
episode: 753    score: 1.0    memory length: 139189    epsilon: 0.9224038000016845      step
s: 151    lr: 0.0001      evaluation reward: 1.65
754
episode: 754    score: 2.0    memory length: 139407    epsilon: 0.9219721600016939      step
s: 218    lr: 0.0001      evaluation reward: 1.67
755
episode: 755    score: 3.0    memory length: 139654    epsilon: 0.9214831000017045      step
s: 247    lr: 0.0001      evaluation reward: 1.65
756
episode: 756    score: 1.0    memory length: 139826    epsilon: 0.9211425400017119      step
s: 172    lr: 0.0001      evaluation reward: 1.63
757
episode: 757    score: 0.0    memory length: 139949    epsilon: 0.9208990000017172      step
s: 123    lr: 0.0001      evaluation reward: 1.63
758
episode: 758    score: 4.0    memory length: 140247    epsilon: 0.92030896000173      steps:
 298    lr: 0.0001      evaluation reward: 1.58
759
episode: 759    score: 1.0    memory length: 140397    epsilon: 0.9200119600017365      step
s: 150    lr: 0.0001      evaluation reward: 1.59
760
episode: 760    score: 4.0    memory length: 140674    epsilon: 0.9194635000017484      step
s: 277    lr: 0.0001      evaluation reward: 1.61
761
episode: 761    score: 1.0    memory length: 140844    epsilon: 0.9191269000017557      step
s: 170    lr: 0.0001      evaluation reward: 1.59
762
episode: 762    score: 0.0    memory length: 140966    epsilon: 0.9188853400017609      step
s: 122    lr: 0.0001      evaluation reward: 1.59
763
episode: 763    score: 1.0    memory length: 141135    epsilon: 0.9185507200017682      step
s: 169    lr: 0.0001      evaluation reward: 1.58
764
episode: 764    score: 1.0    memory length: 141304    epsilon: 0.9182161000017754      step
s: 169    lr: 0.0001      evaluation reward: 1.55
765
episode: 765    score: 0.0    memory length: 141426    epsilon: 0.9179745400017807      step
s: 122    lr: 0.0001      evaluation reward: 1.54
```

```
766
episode: 766   score: 2.0   memory length: 141623   epsilon: 0.9175844800017892    step
s: 197    lr: 0.0001    evaluation reward: 1.56
767
episode: 767   score: 2.0   memory length: 141821   epsilon: 0.9171924400017977    step
s: 198    lr: 0.0001    evaluation reward: 1.58
768
episode: 768   score: 1.0   memory length: 141992   epsilon: 0.916853860001805    steps:
 171    lr: 0.0001    evaluation reward: 1.57
769
episode: 769   score: 2.0   memory length: 142190   epsilon: 0.9164618200018135    step
s: 198    lr: 0.0001    evaluation reward: 1.59
770
episode: 770   score: 2.0   memory length: 142388   epsilon: 0.916069780001822    steps:
 198    lr: 0.0001    evaluation reward: 1.58
771
episode: 771   score: 3.0   memory length: 142634   epsilon: 0.9155827000018326    step
s: 246    lr: 0.0001    evaluation reward: 1.58
772
episode: 772   score: 3.0   memory length: 142860   epsilon: 0.9151352200018423    step
s: 226    lr: 0.0001    evaluation reward: 1.61
773
episode: 773   score: 1.0   memory length: 143028   epsilon: 0.9148025800018496    step
s: 168    lr: 0.0001    evaluation reward: 1.6
774
episode: 774   score: 1.0   memory length: 143197   epsilon: 0.9144679600018568    step
s: 169    lr: 0.0001    evaluation reward: 1.61
775
episode: 775   score: 1.0   memory length: 143366   epsilon: 0.9141333400018641    step
s: 169    lr: 0.0001    evaluation reward: 1.62
776
episode: 776   score: 2.0   memory length: 143563   epsilon: 0.9137432800018725    step
s: 197    lr: 0.0001    evaluation reward: 1.62
777
episode: 777   score: 3.0   memory length: 143806   epsilon: 0.913262140001883    steps:
 243    lr: 0.0001    evaluation reward: 1.61
778
episode: 778   score: 2.0   memory length: 144004   epsilon: 0.9128701000018915    step
s: 198    lr: 0.0001    evaluation reward: 1.62
779
episode: 779   score: 1.0   memory length: 144155   epsilon: 0.912571120001898    steps:
 151    lr: 0.0001    evaluation reward: 1.61
780
episode: 780   score: 1.0   memory length: 144325   epsilon: 0.9122345200019053    step
s: 170    lr: 0.0001    evaluation reward: 1.6
781
episode: 781   score: 1.0   memory length: 144475   epsilon: 0.9119375200019118    step
s: 150    lr: 0.0001    evaluation reward: 1.6
782
episode: 782   score: 2.0   memory length: 144673   epsilon: 0.9115454800019203    step
s: 198    lr: 0.0001    evaluation reward: 1.59
783
episode: 783   score: 5.0   memory length: 144981   epsilon: 0.9109356400019335    step
s: 308    lr: 0.0001    evaluation reward: 1.64
784
episode: 784   score: 2.0   memory length: 145180   epsilon: 0.910541620001942    steps:
 199    lr: 0.0001    evaluation reward: 1.62
785
episode: 785   score: 2.0   memory length: 145377   epsilon: 0.9101515600019505    step
s: 197    lr: 0.0001    evaluation reward: 1.64
786
episode: 786   score: 6.0   memory length: 145720   epsilon: 0.9094724200019653    step
s: 343    lr: 0.0001    evaluation reward: 1.67
787
episode: 787   score: 1.0   memory length: 145871   epsilon: 0.9091734400019718    step
s: 151    lr: 0.0001    evaluation reward: 1.67
```

```
788
episode: 788    score: 0.0    memory length: 145993    epsilon: 0.908931880001977    steps:
 122    lr: 0.0001    evaluation reward: 1.67
789
episode: 789    score: 3.0    memory length: 146241    epsilon: 0.9084408400019877    step
s: 248    lr: 0.0001    evaluation reward: 1.66
790
episode: 790    score: 1.0    memory length: 146392    epsilon: 0.9081418600019942    step
s: 151    lr: 0.0001    evaluation reward: 1.62
791
episode: 791    score: 4.0    memory length: 146691    epsilon: 0.907549840002007    steps:
 299    lr: 0.0001    evaluation reward: 1.65
792
episode: 792    score: 2.0    memory length: 146891    epsilon: 0.9071538400020156    step
s: 200    lr: 0.0001    evaluation reward: 1.66
793
episode: 793    score: 3.0    memory length: 147135    epsilon: 0.9066707200020261    step
s: 244    lr: 0.0001    evaluation reward: 1.69
794
episode: 794    score: 0.0    memory length: 147258    epsilon: 0.9064271800020314    step
s: 123    lr: 0.0001    evaluation reward: 1.69
795
episode: 795    score: 2.0    memory length: 147461    epsilon: 0.9060252400020401    step
s: 203    lr: 0.0001    evaluation reward: 1.69
796
episode: 796    score: 3.0    memory length: 147712    epsilon: 0.9055282600020509    step
s: 251    lr: 0.0001    evaluation reward: 1.71
797
episode: 797    score: 4.0    memory length: 147986    epsilon: 0.9049857400020627    step
s: 274    lr: 0.0001    evaluation reward: 1.74
798
episode: 798    score: 1.0    memory length: 148157    epsilon: 0.90464716000207    steps:
 171    lr: 0.0001    evaluation reward: 1.72
799
episode: 799    score: 2.0    memory length: 148375    epsilon: 0.9042155200020794    step
s: 218    lr: 0.0001    evaluation reward: 1.72
800
episode: 800    score: 4.0    memory length: 148675    epsilon: 0.9036215200020923    step
s: 300    lr: 0.0001    evaluation reward: 1.74
801
episode: 801    score: 1.0    memory length: 148844    epsilon: 0.9032869000020995    step
s: 169    lr: 0.0001    evaluation reward: 1.74
802
episode: 802    score: 1.0    memory length: 148995    epsilon: 0.902987920002106    steps:
 151    lr: 0.0001    evaluation reward: 1.72
803
episode: 803    score: 3.0    memory length: 149262    epsilon: 0.9024592600021175    step
s: 267    lr: 0.0001    evaluation reward: 1.72
804
episode: 804    score: 3.0    memory length: 149491    epsilon: 0.9020058400021274    step
s: 229    lr: 0.0001    evaluation reward: 1.74
805
episode: 805    score: 1.0    memory length: 149641    epsilon: 0.9017088400021338    step
s: 150    lr: 0.0001    evaluation reward: 1.75
806
episode: 806    score: 2.0    memory length: 149858    epsilon: 0.9012791800021431    step
s: 217    lr: 0.0001    evaluation reward: 1.75
807
episode: 807    score: 1.0    memory length: 150027    epsilon: 0.9009445600021504    step
s: 169    lr: 0.0001    evaluation reward: 1.74
808
episode: 808    score: 3.0    memory length: 150273    epsilon: 0.900457480002161    steps:
 246    lr: 0.0001    evaluation reward: 1.76
809
episode: 809    score: 1.0    memory length: 150423    epsilon: 0.9001604800021674    step
s: 150    lr: 0.0001    evaluation reward: 1.75
```

```
810
episode: 810   score: 2.0   memory length: 150638   epsilon: 0.8997347800021767   step
s: 215   lr: 0.0001   evaluation reward: 1.77
811
episode: 811   score: 2.0   memory length: 150861   epsilon: 0.8992932400021862   step
s: 223   lr: 0.0001   evaluation reward: 1.77
812
episode: 812   score: 3.0   memory length: 151108   epsilon: 0.8988041800021969   step
s: 247   lr: 0.0001   evaluation reward: 1.79
813
episode: 813   score: 2.0   memory length: 151306   epsilon: 0.8984121400022054   step
s: 198   lr: 0.0001   evaluation reward: 1.76
814
episode: 814   score: 1.0   memory length: 151457   epsilon: 0.8981131600022119   step
s: 151   lr: 0.0001   evaluation reward: 1.75
815
episode: 815   score: 6.0   memory length: 151831   epsilon: 0.8973726400022279   step
s: 374   lr: 0.0001   evaluation reward: 1.81
816
episode: 816   score: 1.0   memory length: 151981   epsilon: 0.8970756400022344   step
s: 150   lr: 0.0001   evaluation reward: 1.8
817
episode: 817   score: 2.0   memory length: 152179   epsilon: 0.8966836000022429   step
s: 198   lr: 0.0001   evaluation reward: 1.81
818
episode: 818   score: 1.0   memory length: 152330   epsilon: 0.8963846200022494   step
s: 151   lr: 0.0001   evaluation reward: 1.8
819
episode: 819   score: 1.0   memory length: 152480   epsilon: 0.8960876200022558   step
s: 150   lr: 0.0001   evaluation reward: 1.8
820
episode: 820   score: 4.0   memory length: 152757   epsilon: 0.8955391600022677   step
s: 277   lr: 0.0001   evaluation reward: 1.81
821
episode: 821   score: 5.0   memory length: 153062   epsilon: 0.8949352600022809   step
s: 305   lr: 0.0001   evaluation reward: 1.83
822
episode: 822   score: 2.0   memory length: 153260   epsilon: 0.8945432200022894   step
s: 198   lr: 0.0001   evaluation reward: 1.85
823
episode: 823   score: 4.0   memory length: 153557   epsilon: 0.8939551600023021   step
s: 297   lr: 0.0001   evaluation reward: 1.89
824
episode: 824   score: 1.0   memory length: 153707   epsilon: 0.8936581600023086   step
s: 150   lr: 0.0001   evaluation reward: 1.89
825
episode: 825   score: 0.0   memory length: 153830   epsilon: 0.8934146200023139   step
s: 123   lr: 0.0001   evaluation reward: 1.88
826
episode: 826   score: 1.0   memory length: 153982   epsilon: 0.8931136600023204   step
s: 152   lr: 0.0001   evaluation reward: 1.89
827
episode: 827   score: 7.0   memory length: 154351   epsilon: 0.8923830400023363   step
s: 369   lr: 0.0001   evaluation reward: 1.96
828
episode: 828   score: 1.0   memory length: 154522   epsilon: 0.8920444600023436   step
s: 171   lr: 0.0001   evaluation reward: 1.94
829
episode: 829   score: 0.0   memory length: 154644   epsilon: 0.8918029000023489   step
s: 122   lr: 0.0001   evaluation reward: 1.94
830
episode: 830   score: 5.0   memory length: 154939   epsilon: 0.8912188000023615   step
s: 295   lr: 0.0001   evaluation reward: 1.96
831
episode: 831   score: 3.0   memory length: 155165   epsilon: 0.8907713200023712   step
s: 226   lr: 0.0001   evaluation reward: 1.99
```

```
832
episode: 832    score: 2.0    memory length: 155362    epsilon: 0.8903812600023797    step
s: 197    lr: 0.0001    evaluation reward: 1.99
833
episode: 833    score: 1.0    memory length: 155531    epsilon: 0.890046640002387    steps:
 169    lr: 0.0001    evaluation reward: 2.0
834
episode: 834    score: 4.0    memory length: 155826    epsilon: 0.8894625400023997    step
s: 295    lr: 0.0001    evaluation reward: 2.03
835
episode: 835    score: 5.0    memory length: 156139    epsilon: 0.8888428000024131    step
s: 313    lr: 0.0001    evaluation reward: 2.07
836
episode: 836    score: 3.0    memory length: 156365    epsilon: 0.8883953200024228    step
s: 226    lr: 0.0001    evaluation reward: 2.1
837
episode: 837    score: 4.0    memory length: 156643    epsilon: 0.8878448800024348    step
s: 278    lr: 0.0001    evaluation reward: 2.11
838
episode: 838    score: 2.0    memory length: 156841    epsilon: 0.8874528400024433    step
s: 198    lr: 0.0001    evaluation reward: 2.13
839
episode: 839    score: 4.0    memory length: 157116    epsilon: 0.8869083400024551    step
s: 275    lr: 0.0001    evaluation reward: 2.17
840
episode: 840    score: 0.0    memory length: 157239    epsilon: 0.8866648000024604    step
s: 123    lr: 0.0001    evaluation reward: 2.16
841
episode: 841    score: 9.0    memory length: 157615    epsilon: 0.8859203200024766    step
s: 376    lr: 0.0001    evaluation reward: 2.23
842
episode: 842    score: 0.0    memory length: 157737    epsilon: 0.8856787600024818    step
s: 122    lr: 0.0001    evaluation reward: 2.19
843
episode: 843    score: 1.0    memory length: 157905    epsilon: 0.885346120002489    steps:
 168    lr: 0.0001    evaluation reward: 2.18
844
episode: 844    score: 2.0    memory length: 158126    epsilon: 0.8849085400024985    step
s: 221    lr: 0.0001    evaluation reward: 2.14
845
episode: 845    score: 8.0    memory length: 158534    epsilon: 0.8841007000025161    step
s: 408    lr: 0.0001    evaluation reward: 2.19
846
episode: 846    score: 3.0    memory length: 158800    epsilon: 0.8835740200025275    step
s: 266    lr: 0.0001    evaluation reward: 2.21
847
episode: 847    score: 0.0    memory length: 158922    epsilon: 0.8833324600025327    step
s: 122    lr: 0.0001    evaluation reward: 2.19
848
episode: 848    score: 2.0    memory length: 159120    epsilon: 0.8829404200025412    step
s: 198    lr: 0.0001    evaluation reward: 2.17
849
episode: 849    score: 4.0    memory length: 159416    epsilon: 0.882354340002554    steps:
 296    lr: 0.0001    evaluation reward: 2.19
850
episode: 850    score: 2.0    memory length: 159634    epsilon: 0.8819227000025633    step
s: 218    lr: 0.0001    evaluation reward: 2.21
851
episode: 851    score: 1.0    memory length: 159803    epsilon: 0.8815880800025706    step
s: 169    lr: 0.0001    evaluation reward: 2.21
852
episode: 852    score: 2.0    memory length: 160021    epsilon: 0.88115644000258    steps:
 218    lr: 0.0001    evaluation reward: 2.23
853
episode: 853    score: 2.0    memory length: 160218    epsilon: 0.8807663800025884    step
s: 197    lr: 0.0001    evaluation reward: 2.24
```

854
episode: 854   score: 1.0   memory length: 160390   epsilon: 0.8804258200025958   step
s: 172   lr: 0.0001   evaluation reward: 2.23
855
episode: 855   score: 1.0   memory length: 160541   epsilon: 0.8801268400026023   step
s: 151   lr: 0.0001   evaluation reward: 2.21
856
episode: 856   score: 3.0   memory length: 160766   epsilon: 0.879681340002612   steps:
 225   lr: 0.0001   evaluation reward: 2.23
857
episode: 857   score: 2.0   memory length: 160967   epsilon: 0.8792833600026206   step
s: 201   lr: 0.0001   evaluation reward: 2.25
858
episode: 858   score: 2.0   memory length: 161149   epsilon: 0.8789230000026285   step
s: 182   lr: 0.0001   evaluation reward: 2.23
859
episode: 859   score: 6.0   memory length: 161515   epsilon: 0.8781983200026442   step
s: 366   lr: 0.0001   evaluation reward: 2.28
860
episode: 860   score: 3.0   memory length: 161742   epsilon: 0.877748860002654   steps:
 227   lr: 0.0001   evaluation reward: 2.27
861
episode: 861   score: 4.0   memory length: 162020   epsilon: 0.8771984200026659   step
s: 278   lr: 0.0001   evaluation reward: 2.3
862
episode: 862   score: 2.0   memory length: 162238   epsilon: 0.8767667800026753   step
s: 218   lr: 0.0001   evaluation reward: 2.32
863
episode: 863   score: 2.0   memory length: 162457   epsilon: 0.8763331600026847   step
s: 219   lr: 0.0001   evaluation reward: 2.33
864
episode: 864   score: 5.0   memory length: 162763   epsilon: 0.8757272800026978   step
s: 306   lr: 0.0001   evaluation reward: 2.37
865
episode: 865   score: 1.0   memory length: 162913   epsilon: 0.8754302800027043   step
s: 150   lr: 0.0001   evaluation reward: 2.38
866
episode: 866   score: 3.0   memory length: 163159   epsilon: 0.8749432000027149   step
s: 246   lr: 0.0001   evaluation reward: 2.39
867
episode: 867   score: 2.0   memory length: 163357   epsilon: 0.8745511600027234   step
s: 198   lr: 0.0001   evaluation reward: 2.39
868
episode: 868   score: 2.0   memory length: 163555   epsilon: 0.8741591200027319   step
s: 198   lr: 0.0001   evaluation reward: 2.4
869
episode: 869   score: 4.0   memory length: 163828   epsilon: 0.8736185800027436   step
s: 273   lr: 0.0001   evaluation reward: 2.42
870
episode: 870   score: 0.0   memory length: 163951   epsilon: 0.8733750400027489   step
s: 123   lr: 0.0001   evaluation reward: 2.4
871
episode: 871   score: 0.0   memory length: 164073   epsilon: 0.8731334800027541   step
s: 122   lr: 0.0001   evaluation reward: 2.37
872
episode: 872   score: 2.0   memory length: 164293   epsilon: 0.8726978800027636   step
s: 220   lr: 0.0001   evaluation reward: 2.36
873
episode: 873   score: 2.0   memory length: 164490   epsilon: 0.8723078200027721   step
s: 197   lr: 0.0001   evaluation reward: 2.37
874
episode: 874   score: 2.0   memory length: 164687   epsilon: 0.8719177600027805   step
s: 197   lr: 0.0001   evaluation reward: 2.38
875
episode: 875   score: 6.0   memory length: 165021   epsilon: 0.8712564400027949   step
s: 334   lr: 0.0001   evaluation reward: 2.43

```
876
episode: 876   score: 2.0   memory length: 165223   epsilon: 0.8708564800028036    step
s: 202    lr: 0.0001     evaluation reward: 2.43
877
episode: 877   score: 3.0   memory length: 165488   epsilon: 0.870331780002815    steps:
 265    lr: 0.0001     evaluation reward: 2.43
878
episode: 878   score: 4.0   memory length: 165783   epsilon: 0.8697476800028277    step
s: 295    lr: 0.0001     evaluation reward: 2.45
879
episode: 879   score: 0.0   memory length: 165906   epsilon: 0.8695041400028329    step
s: 123    lr: 0.0001     evaluation reward: 2.44
880
episode: 880   score: 3.0   memory length: 166152   epsilon: 0.8690170600028435    step
s: 246    lr: 0.0001     evaluation reward: 2.46
881
episode: 881   score: 0.0   memory length: 166275   epsilon: 0.8687735200028488    step
s: 123    lr: 0.0001     evaluation reward: 2.45
882
episode: 882   score: 4.0   memory length: 166552   epsilon: 0.8682250600028607    step
s: 277    lr: 0.0001     evaluation reward: 2.47
883
episode: 883   score: 3.0   memory length: 166802   epsilon: 0.8677300600028715    step
s: 250    lr: 0.0001     evaluation reward: 2.45
884
episode: 884   score: 1.0   memory length: 166952   epsilon: 0.8674330600028779    step
s: 150    lr: 0.0001     evaluation reward: 2.44
885
episode: 885   score: 5.0   memory length: 167260   epsilon: 0.8668232200028911    step
s: 308    lr: 0.0001     evaluation reward: 2.47
886
episode: 886   score: 5.0   memory length: 167586   epsilon: 0.8661777400029051    step
s: 326    lr: 0.0001     evaluation reward: 2.46
887
episode: 887   score: 2.0   memory length: 167765   epsilon: 0.8658233200029128    step
s: 179    lr: 0.0001     evaluation reward: 2.47
888
episode: 888   score: 3.0   memory length: 167991   epsilon: 0.8653758400029226    step
s: 226    lr: 0.0001     evaluation reward: 2.5
889
episode: 889   score: 2.0   memory length: 168207   epsilon: 0.8649481600029318    step
s: 216    lr: 0.0001     evaluation reward: 2.49
890
episode: 890   score: 2.0   memory length: 168407   epsilon: 0.8645521600029404    step
s: 200    lr: 0.0001     evaluation reward: 2.5
891
episode: 891   score: 3.0   memory length: 168619   epsilon: 0.8641324000029496    step
s: 212    lr: 0.0001     evaluation reward: 2.49
892
episode: 892   score: 1.0   memory length: 168770   epsilon: 0.863833420002956    steps:
 151    lr: 0.0001     evaluation reward: 2.48
893
episode: 893   score: 2.0   memory length: 168968   epsilon: 0.8634413800029646    step
s: 198    lr: 0.0001     evaluation reward: 2.47
894
episode: 894   score: 2.0   memory length: 169166   epsilon: 0.8630493400029731    step
s: 198    lr: 0.0001     evaluation reward: 2.49
895
episode: 895   score: 4.0   memory length: 169431   epsilon: 0.8625246400029845    step
s: 265    lr: 0.0001     evaluation reward: 2.51
896
episode: 896   score: 0.0   memory length: 169554   epsilon: 0.8622811000029897    step
s: 123    lr: 0.0001     evaluation reward: 2.48
897
episode: 897   score: 2.0   memory length: 169757   epsilon: 0.8618791600029985    step
s: 203    lr: 0.0001     evaluation reward: 2.46
```

```
898
episode: 898   score: 2.0   memory length: 169976   epsilon: 0.8614455400030079     step
s: 219    lr: 0.0001     evaluation reward: 2.47
899
episode: 899   score: 2.0   memory length: 170194   epsilon: 0.8610139000030173     step
s: 218    lr: 0.0001     evaluation reward: 2.47
900
episode: 900   score: 6.0   memory length: 170571   epsilon: 0.8602674400030335     step
s: 377    lr: 0.0001     evaluation reward: 2.49
901
episode: 901   score: 2.0   memory length: 170770   epsilon: 0.859873420003042    steps:
 199    lr: 0.0001     evaluation reward: 2.5
902
episode: 902   score: 3.0   memory length: 171020   epsilon: 0.8593784200030528     step
s: 250    lr: 0.0001     evaluation reward: 2.52
903
episode: 903   score: 2.0   memory length: 171238   epsilon: 0.8589467800030621     step
s: 218    lr: 0.0001     evaluation reward: 2.51
904
episode: 904   score: 3.0   memory length: 171465   epsilon: 0.8584973200030719     step
s: 227    lr: 0.0001     evaluation reward: 2.51
905
episode: 905   score: 4.0   memory length: 171741   epsilon: 0.8579508400030837     step
s: 276    lr: 0.0001     evaluation reward: 2.54
906
episode: 906   score: 3.0   memory length: 171987   epsilon: 0.8574637600030943     step
s: 246    lr: 0.0001     evaluation reward: 2.55
907
episode: 907   score: 3.0   memory length: 172201   epsilon: 0.8570400400031035     step
s: 214    lr: 0.0001     evaluation reward: 2.57
908
episode: 908   score: 1.0   memory length: 172351   epsilon: 0.85674304000311    steps:
 150    lr: 0.0001     evaluation reward: 2.55
909
episode: 909   score: 2.0   memory length: 172548   epsilon: 0.8563529800031184     step
s: 197    lr: 0.0001     evaluation reward: 2.56
910
episode: 910   score: 2.0   memory length: 172745   epsilon: 0.8559629200031269     step
s: 197    lr: 0.0001     evaluation reward: 2.56
911
episode: 911   score: 1.0   memory length: 172897   epsilon: 0.8556619600031334     step
s: 152    lr: 0.0001     evaluation reward: 2.55
912
episode: 912   score: 3.0   memory length: 173128   epsilon: 0.8552045800031434     step
s: 231    lr: 0.0001     evaluation reward: 2.55
913
episode: 913   score: 1.0   memory length: 173299   epsilon: 0.8548660000031507     step
s: 171    lr: 0.0001     evaluation reward: 2.54
914
episode: 914   score: 1.0   memory length: 173469   epsilon: 0.854529400003158    steps:
 170    lr: 0.0001     evaluation reward: 2.54
915
episode: 915   score: 0.0   memory length: 173592   epsilon: 0.8542858600031633     step
s: 123    lr: 0.0001     evaluation reward: 2.48
916
episode: 916   score: 2.0   memory length: 173810   epsilon: 0.8538542200031727     step
s: 218    lr: 0.0001     evaluation reward: 2.49
917
episode: 917   score: 2.0   memory length: 174008   epsilon: 0.8534621800031812     step
s: 198    lr: 0.0001     evaluation reward: 2.49
918
episode: 918   score: 2.0   memory length: 174187   epsilon: 0.8531077600031889     step
s: 179    lr: 0.0001     evaluation reward: 2.5
919
episode: 919   score: 1.0   memory length: 174356   epsilon: 0.8527731400031962     step
s: 169    lr: 0.0001     evaluation reward: 2.5
```

```
920
episode: 920   score: 3.0   memory length: 174625   epsilon: 0.8522405200032077     step
s: 269    lr: 0.0001     evaluation reward: 2.49
921
episode: 921   score: 3.0   memory length: 174851   epsilon: 0.8517930400032174     step
s: 226    lr: 0.0001     evaluation reward: 2.47
922
episode: 922   score: 5.0   memory length: 175173   epsilon: 0.8511554800032313     step
s: 322    lr: 0.0001     evaluation reward: 2.5
923
episode: 923   score: 3.0   memory length: 175403   epsilon: 0.8507000800032412     step
s: 230    lr: 0.0001     evaluation reward: 2.49
924
episode: 924   score: 4.0   memory length: 175719   epsilon: 0.8500744000032547     step
s: 316    lr: 0.0001     evaluation reward: 2.52
925
episode: 925   score: 2.0   memory length: 175919   epsilon: 0.8496784000032633     step
s: 200    lr: 0.0001     evaluation reward: 2.54
926
episode: 926   score: 4.0   memory length: 176195   epsilon: 0.8491319200032752     step
s: 276    lr: 0.0001     evaluation reward: 2.57
927
episode: 927   score: 4.0   memory length: 176495   epsilon: 0.8485379200032881     step
s: 300    lr: 0.0001     evaluation reward: 2.54
928
episode: 928   score: 3.0   memory length: 176728   epsilon: 0.8480765800032981     step
s: 233    lr: 0.0001     evaluation reward: 2.56
929
episode: 929   score: 1.0   memory length: 176897   epsilon: 0.8477419600033054     step
s: 169    lr: 0.0001     evaluation reward: 2.57
930
episode: 930   score: 3.0   memory length: 177144   epsilon: 0.847252900003316     steps:
 247    lr: 0.0001     evaluation reward: 2.55
931
episode: 931   score: 1.0   memory length: 177295   epsilon: 0.8469539200033225     step
s: 151    lr: 0.0001     evaluation reward: 2.53
932
episode: 932   score: 4.0   memory length: 177591   epsilon: 0.8463678400033352     step
s: 296    lr: 0.0001     evaluation reward: 2.55
933
episode: 933   score: 2.0   memory length: 177788   epsilon: 0.8459777800033437     step
s: 197    lr: 0.0001     evaluation reward: 2.56
934
episode: 934   score: 4.0   memory length: 178107   epsilon: 0.8453461600033574     step
s: 319    lr: 0.0001     evaluation reward: 2.56
935
episode: 935   score: 0.0   memory length: 178229   epsilon: 0.8451046000033626     step
s: 122    lr: 0.0001     evaluation reward: 2.51
936
episode: 936   score: 1.0   memory length: 178380   epsilon: 0.8448056200033691     step
s: 151    lr: 0.0001     evaluation reward: 2.49
937
episode: 937   score: 0.0   memory length: 178503   epsilon: 0.8445620800033744     step
s: 123    lr: 0.0001     evaluation reward: 2.45
938
episode: 938   score: 0.0   memory length: 178626   epsilon: 0.8443185400033797     step
s: 123    lr: 0.0001     evaluation reward: 2.43
939
episode: 939   score: 0.0   memory length: 178749   epsilon: 0.844075000003385     steps:
 123    lr: 0.0001     evaluation reward: 2.39
940
episode: 940   score: 5.0   memory length: 179065   epsilon: 0.8434493200033986     step
s: 316    lr: 0.0001     evaluation reward: 2.44
941
episode: 941   score: 2.0   memory length: 179265   epsilon: 0.8430533200034072     step
s: 200    lr: 0.0001     evaluation reward: 2.37
```

```
942
episode: 942   score: 1.0   memory length: 179433   epsilon: 0.8427206800034144   step
s: 168   lr: 0.0001     evaluation reward: 2.38
943
episode: 943   score: 2.0   memory length: 179652   epsilon: 0.8422870600034238   step
s: 219   lr: 0.0001     evaluation reward: 2.39
944
episode: 944   score: 1.0   memory length: 179802   epsilon: 0.8419900600034302   step
s: 150   lr: 0.0001     evaluation reward: 2.38
945
episode: 945   score: 2.0   memory length: 180020   epsilon: 0.8415584200034396   step
s: 218   lr: 0.0001     evaluation reward: 2.32
946
episode: 946   score: 2.0   memory length: 180220   epsilon: 0.8411624200034482   step
s: 200   lr: 0.0001     evaluation reward: 2.31
947
episode: 947   score: 2.0   memory length: 180419   epsilon: 0.8407684000034568   step
s: 199   lr: 0.0001     evaluation reward: 2.33
948
episode: 948   score: 1.0   memory length: 180569   epsilon: 0.8404714000034632   step
s: 150   lr: 0.0001     evaluation reward: 2.32
949
episode: 949   score: 4.0   memory length: 180881   epsilon: 0.8398536400034766   step
s: 312   lr: 0.0001     evaluation reward: 2.32
950
episode: 950   score: 3.0   memory length: 181126   epsilon: 0.8393685400034872   step
s: 245   lr: 0.0001     evaluation reward: 2.33
951
episode: 951   score: 0.0   memory length: 181248   epsilon: 0.8391269800034924   step
s: 122   lr: 0.0001     evaluation reward: 2.32
952
episode: 952   score: 0.0   memory length: 181370   epsilon: 0.8388854200034976   step
s: 122   lr: 0.0001     evaluation reward: 2.3
953
episode: 953   score: 1.0   memory length: 181522   epsilon: 0.8385844600035042   step
s: 152   lr: 0.0001     evaluation reward: 2.29
954
episode: 954   score: 5.0   memory length: 181846   epsilon: 0.8379429400035181   step
s: 324   lr: 0.0001     evaluation reward: 2.33
955
episode: 955   score: 3.0   memory length: 182073   epsilon: 0.8374934800035279   step
s: 227   lr: 0.0001     evaluation reward: 2.35
956
episode: 956   score: 1.0   memory length: 182245   epsilon: 0.8371529200035353   step
s: 172   lr: 0.0001     evaluation reward: 2.33
957
episode: 957   score: 6.0   memory length: 182625   epsilon: 0.8364005200035516   step
s: 380   lr: 0.0001     evaluation reward: 2.37
958
episode: 958   score: 5.0   memory length: 182933   epsilon: 0.8357906800035648   step
s: 308   lr: 0.0001     evaluation reward: 2.4
959
episode: 959   score: 2.0   memory length: 183133   epsilon: 0.8353946800035734   step
s: 200   lr: 0.0001     evaluation reward: 2.36
960
episode: 960   score: 5.0   memory length: 183452   epsilon: 0.8347630600035871   step
s: 319   lr: 0.0001     evaluation reward: 2.38
961
episode: 961   score: 0.0   memory length: 183574   epsilon: 0.8345215000035924   step
s: 122   lr: 0.0001     evaluation reward: 2.34
962
episode: 962   score: 1.0   memory length: 183725   epsilon: 0.8342225200035989   step
s: 151   lr: 0.0001     evaluation reward: 2.33
963
episode: 963   score: 3.0   memory length: 183952   epsilon: 0.8337730600036086   step
s: 227   lr: 0.0001     evaluation reward: 2.34
```

```
964
episode: 964    score: 3.0    memory length: 184199    epsilon: 0.8332840000036192    step
s: 247    lr: 0.0001    evaluation reward: 2.32
965
episode: 965    score: 6.0    memory length: 184560    epsilon: 0.8325692200036348    step
s: 361    lr: 0.0001    evaluation reward: 2.37
966
episode: 966    score: 5.0    memory length: 184870    epsilon: 0.8319554200036481    step
s: 310    lr: 0.0001    evaluation reward: 2.39
967
episode: 967    score: 6.0    memory length: 185248    epsilon: 0.8312069800036643    step
s: 378    lr: 0.0001    evaluation reward: 2.43
968
episode: 968    score: 3.0    memory length: 185475    epsilon: 0.8307575200036741    step
s: 227    lr: 0.0001    evaluation reward: 2.44
969
episode: 969    score: 3.0    memory length: 185723    epsilon: 0.8302664800036847    step
s: 248    lr: 0.0001    evaluation reward: 2.43
970
episode: 970    score: 2.0    memory length: 185904    epsilon: 0.8299081000036925    step
s: 181    lr: 0.0001    evaluation reward: 2.45
971
episode: 971    score: 1.0    memory length: 186073    epsilon: 0.8295734800036998    step
s: 169    lr: 0.0001    evaluation reward: 2.46
972
episode: 972    score: 3.0    memory length: 186299    epsilon: 0.8291260000037095    step
s: 226    lr: 0.0001    evaluation reward: 2.47
973
episode: 973    score: 1.0    memory length: 186450    epsilon: 0.828827020003716    steps:
 151    lr: 0.0001    evaluation reward: 2.46
974
episode: 974    score: 2.0    memory length: 186631    epsilon: 0.8284686400037238    step
s: 181    lr: 0.0001    evaluation reward: 2.46
975
episode: 975    score: 1.0    memory length: 186783    epsilon: 0.8281676800037303    step
s: 152    lr: 0.0001    evaluation reward: 2.41
976
episode: 976    score: 6.0    memory length: 187131    epsilon: 0.8274786400037453    step
s: 348    lr: 0.0001    evaluation reward: 2.45
977
episode: 977    score: 3.0    memory length: 187380    epsilon: 0.826985620003756    steps:
 249    lr: 0.0001    evaluation reward: 2.45
978
episode: 978    score: 4.0    memory length: 187656    epsilon: 0.8264391400037678    step
s: 276    lr: 0.0001    evaluation reward: 2.45
979
episode: 979    score: 0.0    memory length: 187779    epsilon: 0.8261956000037731    step
s: 123    lr: 0.0001    evaluation reward: 2.45
980
episode: 980    score: 4.0    memory length: 188021    epsilon: 0.8257164400037835    step
s: 242    lr: 0.0001    evaluation reward: 2.46
981
episode: 981    score: 4.0    memory length: 188317    epsilon: 0.8251303600037962    step
s: 296    lr: 0.0001    evaluation reward: 2.5
982
episode: 982    score: 2.0    memory length: 188496    epsilon: 0.8247759400038039    step
s: 179    lr: 0.0001    evaluation reward: 2.48
983
episode: 983    score: 2.0    memory length: 188676    epsilon: 0.8244195400038117    step
s: 180    lr: 0.0001    evaluation reward: 2.47
984
episode: 984    score: 4.0    memory length: 188952    epsilon: 0.8238730600038235    step
s: 276    lr: 0.0001    evaluation reward: 2.5
985
episode: 985    score: 5.0    memory length: 189257    epsilon: 0.8232691600038367    step
s: 305    lr: 0.0001    evaluation reward: 2.5
```

```
986
episode: 986   score: 7.0   memory length: 189667   epsilon: 0.8224573600038543     step
s: 410    lr: 0.0001    evaluation reward: 2.52
987
episode: 987   score: 8.0   memory length: 190120   epsilon: 0.8215604200038737     step
s: 453    lr: 0.0001    evaluation reward: 2.58
988
episode: 988   score: 0.0   memory length: 190243   epsilon: 0.821316880003879    steps:
 123    lr: 0.0001    evaluation reward: 2.55
989
episode: 989   score: 0.0   memory length: 190366   epsilon: 0.8210733400038843     step
s: 123    lr: 0.0001    evaluation reward: 2.53
990
episode: 990   score: 3.0   memory length: 190613   epsilon: 0.8205842800038949     step
s: 247    lr: 0.0001    evaluation reward: 2.54
991
episode: 991   score: 2.0   memory length: 190793   epsilon: 0.8202278800039027     step
s: 180    lr: 0.0001    evaluation reward: 2.53
992
episode: 992   score: 3.0   memory length: 191003   epsilon: 0.8198120800039117     step
s: 210    lr: 0.0001    evaluation reward: 2.55
993
episode: 993   score: 2.0   memory length: 191183   epsilon: 0.8194556800039194     step
s: 180    lr: 0.0001    evaluation reward: 2.55
994
episode: 994   score: 2.0   memory length: 191399   epsilon: 0.8190280000039287     step
s: 216    lr: 0.0001    evaluation reward: 2.55
995
episode: 995   score: 1.0   memory length: 191568   epsilon: 0.818693380003936    steps:
 169    lr: 0.0001    evaluation reward: 2.52
996
episode: 996   score: 3.0   memory length: 191816   epsilon: 0.8182023400039466     step
s: 248    lr: 0.0001    evaluation reward: 2.55
997
episode: 997   score: 0.0   memory length: 191938   epsilon: 0.8179607800039519     step
s: 122    lr: 0.0001    evaluation reward: 2.53
998
episode: 998   score: 1.0   memory length: 192089   epsilon: 0.8176618000039584     step
s: 151    lr: 0.0001    evaluation reward: 2.52
999
episode: 999   score: 3.0   memory length: 192299   epsilon: 0.8172460000039674     step
s: 210    lr: 0.0001    evaluation reward: 2.53
1000
episode: 1000   score: 4.0   memory length: 192592   epsilon: 0.81666586000398    steps:
 293    lr: 0.0001    evaluation reward: 2.51
1001
episode: 1001   score: 1.0   memory length: 192743   epsilon: 0.8163668800039865     step
s: 151    lr: 0.0001    evaluation reward: 2.5
1002
episode: 1002   score: 2.0   memory length: 192941   epsilon: 0.815974840003995     step
s: 198    lr: 0.0001    evaluation reward: 2.49
1003
episode: 1003   score: 2.0   memory length: 193160   epsilon: 0.8155412200040044     step
s: 219    lr: 0.0001    evaluation reward: 2.49
1004
episode: 1004   score: 5.0   memory length: 193469   epsilon: 0.8149294000040177     step
s: 309    lr: 0.0001    evaluation reward: 2.51
1005
episode: 1005   score: 4.0   memory length: 193784   epsilon: 0.8143057000040312     step
s: 315    lr: 0.0001    evaluation reward: 2.51
1006
episode: 1006   score: 2.0   memory length: 193984   epsilon: 0.8139097000040398     step
s: 200    lr: 0.0001    evaluation reward: 2.5
1007
episode: 1007   score: 3.0   memory length: 194232   epsilon: 0.8134186600040505     step
s: 248    lr: 0.0001    evaluation reward: 2.5
```

```
1008
episode: 1008   score: 0.0   memory length: 194354   epsilon: 0.8131771000040557   step
s: 122   lr: 0.0001   evaluation reward: 2.49
1009
episode: 1009   score: 1.0   memory length: 194523   epsilon: 0.812842480004063   step
s: 169   lr: 0.0001   evaluation reward: 2.48
1010
episode: 1010   score: 2.0   memory length: 194721   epsilon: 0.8124504400040715   step
s: 198   lr: 0.0001   evaluation reward: 2.48
1011
episode: 1011   score: 3.0   memory length: 194964   epsilon: 0.811969300004082   step
s: 243   lr: 0.0001   evaluation reward: 2.5
1012
episode: 1012   score: 4.0   memory length: 195221   epsilon: 0.811460440004093   step
s: 257   lr: 0.0001   evaluation reward: 2.51
1013
episode: 1013   score: 3.0   memory length: 195449   epsilon: 0.8110090000041028   step
s: 228   lr: 0.0001   evaluation reward: 2.53
1014
episode: 1014   score: 2.0   memory length: 195630   epsilon: 0.8106506200041106   step
s: 181   lr: 0.0001   evaluation reward: 2.54
1015
episode: 1015   score: 2.0   memory length: 195849   epsilon: 0.81021700000412   steps:
 219   lr: 0.0001   evaluation reward: 2.56
1016
episode: 1016   score: 3.0   memory length: 196097   epsilon: 0.8097259600041307   step
s: 248   lr: 0.0001   evaluation reward: 2.57
1017
episode: 1017   score: 2.0   memory length: 196295   epsilon: 0.8093339200041392   step
s: 198   lr: 0.0001   evaluation reward: 2.57
1018
episode: 1018   score: 4.0   memory length: 196611   epsilon: 0.8087082400041528   step
s: 316   lr: 0.0001   evaluation reward: 2.59
1019
episode: 1019   score: 2.0   memory length: 196808   epsilon: 0.8083181800041612   step
s: 197   lr: 0.0001   evaluation reward: 2.6
1020
episode: 1020   score: 1.0   memory length: 196976   epsilon: 0.8079855400041684   step
s: 168   lr: 0.0001   evaluation reward: 2.58
1021
episode: 1021   score: 2.0   memory length: 197175   epsilon: 0.807591520004177   step
s: 199   lr: 0.0001   evaluation reward: 2.57
1022
episode: 1022   score: 4.0   memory length: 197469   epsilon: 0.8070094000041896   step
s: 294   lr: 0.0001   evaluation reward: 2.56
1023
episode: 1023   score: 3.0   memory length: 197718   epsilon: 0.8065163800042003   step
s: 249   lr: 0.0001   evaluation reward: 2.56
1024
episode: 1024   score: 2.0   memory length: 197916   epsilon: 0.8061243400042089   step
s: 198   lr: 0.0001   evaluation reward: 2.54
1025
episode: 1025   score: 1.0   memory length: 198067   epsilon: 0.8058253600042153   step
s: 151   lr: 0.0001   evaluation reward: 2.53
1026
episode: 1026   score: 0.0   memory length: 198190   epsilon: 0.8055818200042206   step
s: 123   lr: 0.0001   evaluation reward: 2.49
1027
episode: 1027   score: 3.0   memory length: 198440   epsilon: 0.8050868200042314   step
s: 250   lr: 0.0001   evaluation reward: 2.48
1028
episode: 1028   score: 2.0   memory length: 198661   epsilon: 0.8046492400042409   step
s: 221   lr: 0.0001   evaluation reward: 2.47
1029
episode: 1029   score: 4.0   memory length: 198936   epsilon: 0.8041047400042527   step
s: 275   lr: 0.0001   evaluation reward: 2.5
```

```
1030
episode: 1030    score: 3.0    memory length: 199162    epsilon: 0.8036572600042624    step
s: 226    lr: 0.0001    evaluation reward: 2.5
1031
episode: 1031    score: 1.0    memory length: 199331    epsilon: 0.8033226400042697    step
s: 169    lr: 0.0001    evaluation reward: 2.5
1032
episode: 1032    score: 4.0    memory length: 199606    epsilon: 0.8027781400042815    step
s: 275    lr: 0.0001    evaluation reward: 2.5
1033
episode: 1033    score: 1.0    memory length: 199775    epsilon: 0.8024435200042888    step
s: 169    lr: 0.0001    evaluation reward: 2.49
1034
episode: 1034    score: 3.0    memory length: 200020    epsilon: 0.8019584200042993    step
s: 245    lr: 4e-05    evaluation reward: 2.48
1035
episode: 1035    score: 5.0    memory length: 200344    epsilon: 0.8013169000043132    step
s: 324    lr: 4e-05    evaluation reward: 2.53
1036
episode: 1036    score: 7.0    memory length: 200741    epsilon: 0.8005308400043303    step
s: 397    lr: 4e-05    evaluation reward: 2.59
1037
episode: 1037    score: 2.0    memory length: 200940    epsilon: 0.8001368200043388    step
s: 199    lr: 4e-05    evaluation reward: 2.61
1038
episode: 1038    score: 3.0    memory length: 201169    epsilon: 0.7996834000043487    step
s: 229    lr: 4e-05    evaluation reward: 2.64
1039
episode: 1039    score: 4.0    memory length: 201446    epsilon: 0.7991349400043606    step
s: 277    lr: 4e-05    evaluation reward: 2.68
1040
episode: 1040    score: 3.0    memory length: 201690    epsilon: 0.7986518200043711    step
s: 244    lr: 4e-05    evaluation reward: 2.66
1041
episode: 1041    score: 4.0    memory length: 201986    epsilon: 0.7980657400043838    step
s: 296    lr: 4e-05    evaluation reward: 2.68
1042
episode: 1042    score: 6.0    memory length: 202361    epsilon: 0.7973232400043999    step
s: 375    lr: 4e-05    evaluation reward: 2.73
1043
episode: 1043    score: 3.0    memory length: 202587    epsilon: 0.7968757600044096    step
s: 226    lr: 4e-05    evaluation reward: 2.74
1044
episode: 1044    score: 0.0    memory length: 202710    epsilon: 0.7966322200044149    step
s: 123    lr: 4e-05    evaluation reward: 2.73
1045
episode: 1045    score: 3.0    memory length: 202977    epsilon: 0.7961035600044264    step
s: 267    lr: 4e-05    evaluation reward: 2.74
1046
episode: 1046    score: 3.0    memory length: 203224    epsilon: 0.795614500004437    step
s: 247    lr: 4e-05    evaluation reward: 2.75
1047
episode: 1047    score: 5.0    memory length: 203529    epsilon: 0.7950106000044501    step
s: 305    lr: 4e-05    evaluation reward: 2.78
1048
episode: 1048    score: 5.0    memory length: 203852    epsilon: 0.794371060004464    step
s: 323    lr: 4e-05    evaluation reward: 2.82
1049
episode: 1049    score: 0.0    memory length: 203974    epsilon: 0.7941295000044692    step
s: 122    lr: 4e-05    evaluation reward: 2.78
1050
episode: 1050    score: 1.0    memory length: 204125    epsilon: 0.7938305200044757    step
s: 151    lr: 4e-05    evaluation reward: 2.76
1051
episode: 1051    score: 0.0    memory length: 204248    epsilon: 0.793586980004481    step
s: 123    lr: 4e-05    evaluation reward: 2.76
```

```
1052
episode: 1052    score: 5.0    memory length: 204574    epsilon: 0.792941500004495      step
s: 326    lr: 4e-05     evaluation reward: 2.81
1053
episode: 1053    score: 4.0    memory length: 204891    epsilon: 0.7923138400045087     step
s: 317    lr: 4e-05     evaluation reward: 2.84
1054
episode: 1054    score: 1.0    memory length: 205060    epsilon: 0.7919792200045159     step
s: 169    lr: 4e-05     evaluation reward: 2.8
1055
episode: 1055    score: 1.0    memory length: 205211    epsilon: 0.7916802400045224     step
s: 151    lr: 4e-05     evaluation reward: 2.78
1056
episode: 1056    score: 5.0    memory length: 205553    epsilon: 0.7910030800045371     step
s: 342    lr: 4e-05     evaluation reward: 2.82
1057
episode: 1057    score: 8.0    memory length: 205996    epsilon: 0.7901259400045562     step
s: 443    lr: 4e-05     evaluation reward: 2.84
1058
episode: 1058    score: 4.0    memory length: 206291    epsilon: 0.7895418400045688     step
s: 295    lr: 4e-05     evaluation reward: 2.83
1059
episode: 1059    score: 1.0    memory length: 206461    epsilon: 0.7892052400045761     step
s: 170    lr: 4e-05     evaluation reward: 2.82
1060
episode: 1060    score: 5.0    memory length: 206769    epsilon: 0.7885954000045894     step
s: 308    lr: 4e-05     evaluation reward: 2.82
1061
episode: 1061    score: 3.0    memory length: 206999    epsilon: 0.7881400000045993     step
s: 230    lr: 4e-05     evaluation reward: 2.85
1062
episode: 1062    score: 5.0    memory length: 207325    epsilon: 0.7874945200046133     step
s: 326    lr: 4e-05     evaluation reward: 2.89
1063
episode: 1063    score: 3.0    memory length: 207572    epsilon: 0.7870054600046239     step
s: 247    lr: 4e-05     evaluation reward: 2.89
1064
episode: 1064    score: 3.0    memory length: 207786    epsilon: 0.7865817400046331     step
s: 214    lr: 4e-05     evaluation reward: 2.89
1065
episode: 1065    score: 4.0    memory length: 208043    epsilon: 0.7860728800046441     step
s: 257    lr: 4e-05     evaluation reward: 2.87
1066
episode: 1066    score: 3.0    memory length: 208289    epsilon: 0.7855858000046547     step
s: 246    lr: 4e-05     evaluation reward: 2.85
1067
episode: 1067    score: 4.0    memory length: 208564    epsilon: 0.7850413000046665     step
s: 275    lr: 4e-05     evaluation reward: 2.83
1068
episode: 1068    score: 1.0    memory length: 208714    epsilon: 0.784744300004673     step
s: 150    lr: 4e-05     evaluation reward: 2.81
1069
episode: 1069    score: 3.0    memory length: 208942    epsilon: 0.7842928600046828     step
s: 228    lr: 4e-05     evaluation reward: 2.81
1070
episode: 1070    score: 1.0    memory length: 209093    epsilon: 0.7839938800046893     step
s: 151    lr: 4e-05     evaluation reward: 2.8
1071
episode: 1071    score: 4.0    memory length: 209394    epsilon: 0.7833979000047022     step
s: 301    lr: 4e-05     evaluation reward: 2.83
1072
episode: 1072    score: 3.0    memory length: 209639    epsilon: 0.7829128000047128     step
s: 245    lr: 4e-05     evaluation reward: 2.83
1073
episode: 1073    score: 3.0    memory length: 209867    epsilon: 0.7824613600047226     step
s: 228    lr: 4e-05     evaluation reward: 2.85
```

```
1074
episode: 1074    score: 5.0    memory length: 210183    epsilon: 0.7818356800047361    step
s: 316    lr: 4e-05    evaluation reward: 2.88
1075
episode: 1075    score: 3.0    memory length: 210415    epsilon: 0.7813763200047461    step
s: 232    lr: 4e-05    evaluation reward: 2.9
1076
episode: 1076    score: 2.0    memory length: 210597    epsilon: 0.7810159600047539    step
s: 182    lr: 4e-05    evaluation reward: 2.86
1077
episode: 1077    score: 5.0    memory length: 210905    epsilon: 0.7804061200047672    step
s: 308    lr: 4e-05    evaluation reward: 2.88
1078
episode: 1078    score: 6.0    memory length: 211223    epsilon: 0.7797764800047808    step
s: 318    lr: 4e-05    evaluation reward: 2.9
1079
episode: 1079    score: 6.0    memory length: 211616    epsilon: 0.7789983400047977    step
s: 393    lr: 4e-05    evaluation reward: 2.96
1080
episode: 1080    score: 4.0    memory length: 211893    epsilon: 0.7784498800048096    step
s: 277    lr: 4e-05    evaluation reward: 2.96
1081
episode: 1081    score: 3.0    memory length: 212118    epsilon: 0.7780043800048193    step
s: 225    lr: 4e-05    evaluation reward: 2.95
1082
episode: 1082    score: 7.0    memory length: 212511    epsilon: 0.7772262400048362    step
s: 393    lr: 4e-05    evaluation reward: 3.0
1083
episode: 1083    score: 1.0    memory length: 212662    epsilon: 0.7769272600048427    step
s: 151    lr: 4e-05    evaluation reward: 2.99
1084
episode: 1084    score: 2.0    memory length: 212843    epsilon: 0.7765688800048505    step
s: 181    lr: 4e-05    evaluation reward: 2.97
1085
episode: 1085    score: 4.0    memory length: 213085    epsilon: 0.7760897200048609    step
s: 242    lr: 4e-05    evaluation reward: 2.96
1086
episode: 1086    score: 3.0    memory length: 213333    epsilon: 0.7755986800048715    step
s: 248    lr: 4e-05    evaluation reward: 2.92
1087
episode: 1087    score: 5.0    memory length: 213669    epsilon: 0.774933400004886    step
s: 336    lr: 4e-05    evaluation reward: 2.89
1088
episode: 1088    score: 5.0    memory length: 213994    epsilon: 0.7742899000049    steps:
 325    lr: 4e-05    evaluation reward: 2.94
1089
episode: 1089    score: 3.0    memory length: 214260    epsilon: 0.7737632200049114    step
s: 266    lr: 4e-05    evaluation reward: 2.97
1090
episode: 1090    score: 3.0    memory length: 214506    epsilon: 0.773276140004922    step
s: 246    lr: 4e-05    evaluation reward: 2.97
1091
episode: 1091    score: 5.0    memory length: 214832    epsilon: 0.772630660004936    step
s: 326    lr: 4e-05    evaluation reward: 3.0
1092
episode: 1092    score: 2.0    memory length: 215030    epsilon: 0.7722386200049445    step
s: 198    lr: 4e-05    evaluation reward: 2.99
1093
episode: 1093    score: 7.0    memory length: 215474    epsilon: 0.7713595000049636    step
s: 444    lr: 4e-05    evaluation reward: 3.04
1094
episode: 1094    score: 5.0    memory length: 215778    epsilon: 0.7707575800049766    step
s: 304    lr: 4e-05    evaluation reward: 3.07
1095
episode: 1095    score: 4.0    memory length: 216034    epsilon: 0.7702507000049876    step
s: 256    lr: 4e-05    evaluation reward: 3.1
```

```
1096
episode: 1096    score: 6.0    memory length: 216381    epsilon: 0.7695636400050025    step
s: 347    lr: 4e-05    evaluation reward: 3.13
1097
episode: 1097    score: 4.0    memory length: 216666    epsilon: 0.7689993400050148    step
s: 285    lr: 4e-05    evaluation reward: 3.17
1098
episode: 1098    score: 2.0    memory length: 216848    epsilon: 0.7686389800050226    step
s: 182    lr: 4e-05    evaluation reward: 3.18
1099
episode: 1099    score: 0.0    memory length: 216971    epsilon: 0.7683954400050279    step
s: 123    lr: 4e-05    evaluation reward: 3.15
1100
episode: 1100    score: 4.0    memory length: 217269    epsilon: 0.7678054000050407    step
s: 298    lr: 4e-05    evaluation reward: 3.15
1101
episode: 1101    score: 1.0    memory length: 217420    epsilon: 0.7675064200050472    step
s: 151    lr: 4e-05    evaluation reward: 3.15
1102
episode: 1102    score: 4.0    memory length: 217664    epsilon: 0.7670233000050577    step
s: 244    lr: 4e-05    evaluation reward: 3.17
1103
episode: 1103    score: 5.0    memory length: 217954    epsilon: 0.7664491000050702    step
s: 290    lr: 4e-05    evaluation reward: 3.2
1104
episode: 1104    score: 4.0    memory length: 218231    epsilon: 0.7659006400050821    step
s: 277    lr: 4e-05    evaluation reward: 3.19
1105
episode: 1105    score: 2.0    memory length: 218447    epsilon: 0.7654729600050914    step
s: 216    lr: 4e-05    evaluation reward: 3.17
1106
episode: 1106    score: 1.0    memory length: 218597    epsilon: 0.7651759600050978    step
s: 150    lr: 4e-05    evaluation reward: 3.16
1107
episode: 1107    score: 4.0    memory length: 218855    epsilon: 0.7646651200051089    step
s: 258    lr: 4e-05    evaluation reward: 3.17
1108
episode: 1108    score: 2.0    memory length: 219052    epsilon: 0.7642750600051174    step
s: 197    lr: 4e-05    evaluation reward: 3.19
1109
episode: 1109    score: 2.0    memory length: 219233    epsilon: 0.7639166800051251    step
s: 181    lr: 4e-05    evaluation reward: 3.2
1110
episode: 1110    score: 5.0    memory length: 219577    epsilon: 0.7632355600051399    step
s: 344    lr: 4e-05    evaluation reward: 3.23
1111
episode: 1111    score: 2.0    memory length: 219758    epsilon: 0.7628771800051477    step
s: 181    lr: 4e-05    evaluation reward: 3.22
1112
episode: 1112    score: 3.0    memory length: 220005    epsilon: 0.7623881200051583    step
s: 247    lr: 4e-05    evaluation reward: 3.21
1113
episode: 1113    score: 6.0    memory length: 220380    epsilon: 0.7616456200051744    step
s: 375    lr: 4e-05    evaluation reward: 3.24
1114
episode: 1114    score: 1.0    memory length: 220530    epsilon: 0.7613486200051809    step
s: 150    lr: 4e-05    evaluation reward: 3.23
1115
episode: 1115    score: 12.0    memory length: 221006    epsilon: 0.7604061400052013    ste
ps: 476    lr: 4e-05    evaluation reward: 3.33
1116
episode: 1116    score: 5.0    memory length: 221350    epsilon: 0.7597250200052161    step
s: 344    lr: 4e-05    evaluation reward: 3.35
1117
episode: 1117    score: 5.0    memory length: 221625    epsilon: 0.759180520005228    step
s: 275    lr: 4e-05    evaluation reward: 3.38
```

```
1118
episode: 1118    score: 3.0    memory length: 221870    epsilon: 0.7586954200052385    step
s: 245    lr: 4e-05    evaluation reward: 3.37
1119
episode: 1119    score: 6.0    memory length: 222263    epsilon: 0.7579172800052554    step
s: 393    lr: 4e-05    evaluation reward: 3.41
1120
episode: 1120    score: 3.0    memory length: 222515    epsilon: 0.7574183200052662    step
s: 252    lr: 4e-05    evaluation reward: 3.43
1121
episode: 1121    score: 0.0    memory length: 222637    epsilon: 0.7571767600052715    step
s: 122    lr: 4e-05    evaluation reward: 3.41
1122
episode: 1122    score: 3.0    memory length: 222866    epsilon: 0.7567233400052813    step
s: 229    lr: 4e-05    evaluation reward: 3.4
1123
episode: 1123    score: 2.0    memory length: 223050    epsilon: 0.7563590200052892    step
s: 184    lr: 4e-05    evaluation reward: 3.39
1124
episode: 1124    score: 7.0    memory length: 223462    epsilon: 0.7555432600053069    step
s: 412    lr: 4e-05    evaluation reward: 3.44
1125
episode: 1125    score: 9.0    memory length: 223895    epsilon: 0.7546859200053255    step
s: 433    lr: 4e-05    evaluation reward: 3.52
1126
episode: 1126    score: 2.0    memory length: 224095    epsilon: 0.7542899200053341    step
s: 200    lr: 4e-05    evaluation reward: 3.54
1127
episode: 1127    score: 3.0    memory length: 224325    epsilon: 0.753834520005344    step
s: 230    lr: 4e-05    evaluation reward: 3.54
1128
episode: 1128    score: 5.0    memory length: 224653    epsilon: 0.7531850800053581    step
s: 328    lr: 4e-05    evaluation reward: 3.57
1129
episode: 1129    score: 8.0    memory length: 225129    epsilon: 0.7522426000053786    step
s: 476    lr: 4e-05    evaluation reward: 3.61
1130
episode: 1130    score: 2.0    memory length: 225327    epsilon: 0.7518505600053871    step
s: 198    lr: 4e-05    evaluation reward: 3.6
1131
episode: 1131    score: 6.0    memory length: 225718    epsilon: 0.7510763800054039    step
s: 391    lr: 4e-05    evaluation reward: 3.65
1132
episode: 1132    score: 1.0    memory length: 225869    epsilon: 0.7507774000054104    step
s: 151    lr: 4e-05    evaluation reward: 3.62
1133
episode: 1133    score: 6.0    memory length: 226205    epsilon: 0.7501121200054248    step
s: 336    lr: 4e-05    evaluation reward: 3.67
1134
episode: 1134    score: 1.0    memory length: 226375    epsilon: 0.7497755200054321    step
s: 170    lr: 4e-05    evaluation reward: 3.65
1135
episode: 1135    score: 3.0    memory length: 226601    epsilon: 0.7493280400054418    step
s: 226    lr: 4e-05    evaluation reward: 3.63
1136
episode: 1136    score: 4.0    memory length: 226859    epsilon: 0.7488172000054529    step
s: 258    lr: 4e-05    evaluation reward: 3.6
1137
episode: 1137    score: 4.0    memory length: 227136    epsilon: 0.7482687400054648    step
s: 277    lr: 4e-05    evaluation reward: 3.62
1138
episode: 1138    score: 1.0    memory length: 227286    epsilon: 0.7479717400054713    step
s: 150    lr: 4e-05    evaluation reward: 3.6
1139
episode: 1139    score: 3.0    memory length: 227512    epsilon: 0.747524260005481    step
s: 226    lr: 4e-05    evaluation reward: 3.59
```

```
1140
episode: 1140    score: 6.0    memory length: 227870    epsilon: 0.7468154200054964    step
s: 358    lr: 4e-05    evaluation reward: 3.62
1141
episode: 1141    score: 5.0    memory length: 228171    epsilon: 0.7462194400055093    step
s: 301    lr: 4e-05    evaluation reward: 3.63
1142
episode: 1142    score: 4.0    memory length: 228469    epsilon: 0.7456294000055221    step
s: 298    lr: 4e-05    evaluation reward: 3.61
1143
episode: 1143    score: 3.0    memory length: 228717    epsilon: 0.7451383600055328    step
s: 248    lr: 4e-05    evaluation reward: 3.61
1144
episode: 1144    score: 4.0    memory length: 228996    epsilon: 0.7445859400055448    step
s: 279    lr: 4e-05    evaluation reward: 3.65
1145
episode: 1145    score: 4.0    memory length: 229296    epsilon: 0.7439919400055577    step
s: 300    lr: 4e-05    evaluation reward: 3.66
1146
episode: 1146    score: 5.0    memory length: 229622    epsilon: 0.7433464600055717    step
s: 326    lr: 4e-05    evaluation reward: 3.68
1147
episode: 1147    score: 0.0    memory length: 229745    epsilon: 0.743102920005577    step
s: 123    lr: 4e-05    evaluation reward: 3.63
1148
episode: 1148    score: 3.0    memory length: 229971    epsilon: 0.7426554400055867    step
s: 226    lr: 4e-05    evaluation reward: 3.61
1149
episode: 1149    score: 3.0    memory length: 230221    epsilon: 0.7421604400055974    step
s: 250    lr: 4e-05    evaluation reward: 3.64
1150
episode: 1150    score: 6.0    memory length: 230594    epsilon: 0.7414219000056135    step
s: 373    lr: 4e-05    evaluation reward: 3.69
1151
episode: 1151    score: 6.0    memory length: 230923    epsilon: 0.7407704800056276    step
s: 329    lr: 4e-05    evaluation reward: 3.75
1152
episode: 1152    score: 3.0    memory length: 231132    epsilon: 0.7403566600056366    step
s: 209    lr: 4e-05    evaluation reward: 3.73
1153
episode: 1153    score: 15.0    memory length: 231668    epsilon: 0.7392953800056596    ste
ps: 536    lr: 4e-05    evaluation reward: 3.84
1154
episode: 1154    score: 5.0    memory length: 231994    epsilon: 0.7386499000056737    step
s: 326    lr: 4e-05    evaluation reward: 3.88
1155
episode: 1155    score: 4.0    memory length: 232253    epsilon: 0.7381370800056848    step
s: 259    lr: 4e-05    evaluation reward: 3.91
1156
episode: 1156    score: 5.0    memory length: 232556    epsilon: 0.7375371400056978    step
s: 303    lr: 4e-05    evaluation reward: 3.91
1157
episode: 1157    score: 4.0    memory length: 232815    epsilon: 0.7370243200057089    step
s: 259    lr: 4e-05    evaluation reward: 3.87
1158
episode: 1158    score: 3.0    memory length: 233064    epsilon: 0.7365313000057196    step
s: 249    lr: 4e-05    evaluation reward: 3.86
1159
episode: 1159    score: 0.0    memory length: 233187    epsilon: 0.7362877600057249    step
s: 123    lr: 4e-05    evaluation reward: 3.85
1160
episode: 1160    score: 6.0    memory length: 233582    epsilon: 0.7355056600057419    step
s: 395    lr: 4e-05    evaluation reward: 3.86
1161
episode: 1161    score: 0.0    memory length: 233705    epsilon: 0.7352621200057472    step
s: 123    lr: 4e-05    evaluation reward: 3.83
```

```
1162
episode: 1162   score: 2.0   memory length: 233925   epsilon: 0.7348265200057567   step
s: 220    lr: 4e-05    evaluation reward: 3.8
1163
episode: 1163   score: 4.0   memory length: 234168   epsilon: 0.7343453800057671   step
s: 243    lr: 4e-05    evaluation reward: 3.81
1164
episode: 1164   score: 3.0   memory length: 234414   epsilon: 0.7338583000057777   step
s: 246    lr: 4e-05    evaluation reward: 3.81
1165
episode: 1165   score: 0.0   memory length: 234537   epsilon: 0.7333614760005783   step
s: 123    lr: 4e-05    evaluation reward: 3.77
1166
episode: 1166   score: 2.0   memory length: 234753   epsilon: 0.7331870800057922   step
s: 216    lr: 4e-05    evaluation reward: 3.76
1167
episode: 1167   score: 6.0   memory length: 235107   epsilon: 0.7324861600058075   step
s: 354    lr: 4e-05    evaluation reward: 3.78
1168
episode: 1168   score: 3.0   memory length: 235351   epsilon: 0.732003040005818    step
s: 244    lr: 4e-05    evaluation reward: 3.8
1169
episode: 1169   score: 4.0   memory length: 235608   epsilon: 0.731494180005829    step
s: 257    lr: 4e-05    evaluation reward: 3.81
1170
episode: 1170   score: 4.0   memory length: 235865   epsilon: 0.73098532000584    steps:
 257    lr: 4e-05    evaluation reward: 3.84
1171
episode: 1171   score: 6.0   memory length: 236257   epsilon: 0.7302091600058569   step
s: 392    lr: 4e-05    evaluation reward: 3.86
1172
episode: 1172   score: 5.0   memory length: 236599   epsilon: 0.7295320000058716   step
s: 342    lr: 4e-05    evaluation reward: 3.88
1173
episode: 1173   score: 2.0   memory length: 236797   epsilon: 0.7291399600058801   step
s: 198    lr: 4e-05    evaluation reward: 3.87
1174
episode: 1174   score: 2.0   memory length: 237014   epsilon: 0.7287103000058894   step
s: 217    lr: 4e-05    evaluation reward: 3.84
1175
episode: 1175   score: 3.0   memory length: 237260   epsilon: 0.7282232200059    steps:
 246    lr: 4e-05    evaluation reward: 3.84
1176
episode: 1176   score: 6.0   memory length: 237597   epsilon: 0.7275559600059145   step
s: 337    lr: 4e-05    evaluation reward: 3.88
1177
episode: 1177   score: 5.0   memory length: 237872   epsilon: 0.7270114600059263   step
s: 275    lr: 4e-05    evaluation reward: 3.88
1178
episode: 1178   score: 4.0   memory length: 238170   epsilon: 0.7264214200059391   step
s: 298    lr: 4e-05    evaluation reward: 3.86
1179
episode: 1179   score: 4.0   memory length: 238444   epsilon: 0.7258789000059509   step
s: 274    lr: 4e-05    evaluation reward: 3.84
1180
episode: 1180   score: 6.0   memory length: 238830   epsilon: 0.7251146200059675   step
s: 386    lr: 4e-05    evaluation reward: 3.86
1181
episode: 1181   score: 7.0   memory length: 239206   epsilon: 0.7243701400059837   step
s: 376    lr: 4e-05    evaluation reward: 3.9
1182
episode: 1182   score: 9.0   memory length: 239686   epsilon: 0.7234197400060043   step
s: 480    lr: 4e-05    evaluation reward: 3.92
1183
episode: 1183   score: 3.0   memory length: 239915   epsilon: 0.7229663200060141   step
s: 229    lr: 4e-05    evaluation reward: 3.94
```

```
1184
episode: 1184    score: 5.0    memory length: 240218    epsilon: 0.7223663800060272    step
s: 303    lr: 4e-05    evaluation reward: 3.97
1185
episode: 1185    score: 6.0    memory length: 240562    epsilon: 0.7216852600060419    step
s: 344    lr: 4e-05    evaluation reward: 3.99
1186
episode: 1186    score: 5.0    memory length: 240854    epsilon: 0.7211071000060545    step
s: 292    lr: 4e-05    evaluation reward: 4.01
1187
episode: 1187    score: 3.0    memory length: 241100    epsilon: 0.7206200200060651    step
s: 246    lr: 4e-05    evaluation reward: 3.99
1188
episode: 1188    score: 3.0    memory length: 241345    epsilon: 0.7201349200060756    step
s: 245    lr: 4e-05    evaluation reward: 3.97
1189
episode: 1189    score: 3.0    memory length: 241592    epsilon: 0.7196458600060862    step
s: 247    lr: 4e-05    evaluation reward: 3.97
1190
episode: 1190    score: 5.0    memory length: 241893    epsilon: 0.7190498800060992    step
s: 301    lr: 4e-05    evaluation reward: 3.99
1191
episode: 1191    score: 7.0    memory length: 242293    epsilon: 0.7182578800061163    step
s: 400    lr: 4e-05    evaluation reward: 4.01
1192
episode: 1192    score: 2.0    memory length: 242491    epsilon: 0.7178658400061249    step
s: 198    lr: 4e-05    evaluation reward: 4.01
1193
episode: 1193    score: 2.0    memory length: 242710    epsilon: 0.7174322200061343    step
s: 219    lr: 4e-05    evaluation reward: 3.96
1194
episode: 1194    score: 1.0    memory length: 242860    epsilon: 0.7171352200061407    step
s: 150    lr: 4e-05    evaluation reward: 3.92
1195
episode: 1195    score: 3.0    memory length: 243087    epsilon: 0.7166857600061505    step
s: 227    lr: 4e-05    evaluation reward: 3.91
1196
episode: 1196    score: 5.0    memory length: 243395    epsilon: 0.7160759200061637    step
s: 308    lr: 4e-05    evaluation reward: 3.9
1197
episode: 1197    score: 2.0    memory length: 243593    epsilon: 0.7156838800061722    step
s: 198    lr: 4e-05    evaluation reward: 3.88
1198
episode: 1198    score: 1.0    memory length: 243743    epsilon: 0.7153868800061787    step
s: 150    lr: 4e-05    evaluation reward: 3.87
1199
episode: 1199    score: 5.0    memory length: 244041    epsilon: 0.7147968400061915    step
s: 298    lr: 4e-05    evaluation reward: 3.92
1200
episode: 1200    score: 5.0    memory length: 244349    epsilon: 0.7141870000062047    step
s: 308    lr: 4e-05    evaluation reward: 3.93
1201
episode: 1201    score: 2.0    memory length: 244531    epsilon: 0.7138266400062125    step
s: 182    lr: 4e-05    evaluation reward: 3.94
1202
episode: 1202    score: 11.0    memory length: 245113    epsilon: 0.7126742800062376    ste
ps: 582    lr: 4e-05    evaluation reward: 4.01
1203
episode: 1203    score: 8.0    memory length: 245544    epsilon: 0.7118209000062561    step
s: 431    lr: 4e-05    evaluation reward: 4.04
1204
episode: 1204    score: 0.0    memory length: 245667    epsilon: 0.7115773600062614    step
s: 123    lr: 4e-05    evaluation reward: 4.0
1205
episode: 1205    score: 3.0    memory length: 245899    epsilon: 0.7111180000062713    step
s: 232    lr: 4e-05    evaluation reward: 4.01
```

```
1206
episode: 1206    score: 7.0    memory length: 246266    epsilon: 0.7103913400062871    step
s: 367    lr: 4e-05    evaluation reward: 4.07
1207
episode: 1207    score: 3.0    memory length: 246516    epsilon: 0.7098963400062979    step
s: 250    lr: 4e-05    evaluation reward: 4.06
1208
episode: 1208    score: 8.0    memory length: 246866    epsilon: 0.7092033400063129    step
s: 350    lr: 4e-05    evaluation reward: 4.12
1209
episode: 1209    score: 4.0    memory length: 247126    epsilon: 0.7086885400063241    step
s: 260    lr: 4e-05    evaluation reward: 4.14
1210
episode: 1210    score: 3.0    memory length: 247357    epsilon: 0.708231160006334    step
s: 231    lr: 4e-05    evaluation reward: 4.12
1211
episode: 1211    score: 6.0    memory length: 247720    epsilon: 0.7075124200063496    step
s: 363    lr: 4e-05    evaluation reward: 4.16
1212
episode: 1212    score: 4.0    memory length: 247996    epsilon: 0.7069659400063615    step
s: 276    lr: 4e-05    evaluation reward: 4.17
1213
episode: 1213    score: 6.0    memory length: 248341    epsilon: 0.7062828400063763    step
s: 345    lr: 4e-05    evaluation reward: 4.17
1214
episode: 1214    score: 6.0    memory length: 248690    epsilon: 0.7055918200063913    step
s: 349    lr: 4e-05    evaluation reward: 4.22
1215
episode: 1215    score: 3.0    memory length: 248898    epsilon: 0.7051799800064003    step
s: 208    lr: 4e-05    evaluation reward: 4.13
1216
episode: 1216    score: 7.0    memory length: 249217    epsilon: 0.704548360006414    step
s: 319    lr: 4e-05    evaluation reward: 4.15
1217
episode: 1217    score: 5.0    memory length: 249562    epsilon: 0.7038652600064288    step
s: 345    lr: 4e-05    evaluation reward: 4.15
1218
episode: 1218    score: 6.0    memory length: 249914    epsilon: 0.7031683000064439    step
s: 352    lr: 4e-05    evaluation reward: 4.18
1219
episode: 1219    score: 3.0    memory length: 250164    epsilon: 0.7026733000064547    step
s: 250    lr: 4e-05    evaluation reward: 4.15
1220
episode: 1220    score: 3.0    memory length: 250415    epsilon: 0.7021763200064655    step
s: 251    lr: 4e-05    evaluation reward: 4.15
1221
episode: 1221    score: 5.0    memory length: 250721    epsilon: 0.7015704400064786    step
s: 306    lr: 4e-05    evaluation reward: 4.2
1222
episode: 1222    score: 7.0    memory length: 251109    epsilon: 0.7008022000064953    step
s: 388    lr: 4e-05    evaluation reward: 4.24
1223
episode: 1223    score: 8.0    memory length: 251558    epsilon: 0.6999131800065146    step
s: 449    lr: 4e-05    evaluation reward: 4.3
1224
episode: 1224    score: 8.0    memory length: 251838    epsilon: 0.6993587800065266    step
s: 280    lr: 4e-05    evaluation reward: 4.31
1225
episode: 1225    score: 3.0    memory length: 252067    epsilon: 0.6989053600065365    step
s: 229    lr: 4e-05    evaluation reward: 4.25
1226
episode: 1226    score: 2.0    memory length: 252284    epsilon: 0.6984757000065458    step
s: 217    lr: 4e-05    evaluation reward: 4.25
1227
episode: 1227    score: 7.0    memory length: 252634    epsilon: 0.6977827000065608    step
s: 350    lr: 4e-05    evaluation reward: 4.29
```

```
1228
episode: 1228    score: 4.0    memory length: 252900    epsilon: 0.6972560200065723    step
s: 266    lr: 4e-05    evaluation reward: 4.28
1229
episode: 1229    score: 4.0    memory length: 253176    epsilon: 0.6967095400065841    step
s: 276    lr: 4e-05    evaluation reward: 4.24
1230
episode: 1230    score: 5.0    memory length: 253500    epsilon: 0.6960680200065981    step
s: 324    lr: 4e-05    evaluation reward: 4.27
1231
episode: 1231    score: 4.0    memory length: 253776    epsilon: 0.6955215400066099    step
s: 276    lr: 4e-05    evaluation reward: 4.25
1232
episode: 1232    score: 5.0    memory length: 254100    epsilon: 0.6948800200066239    step
s: 324    lr: 4e-05    evaluation reward: 4.29
1233
episode: 1233    score: 5.0    memory length: 254389    epsilon: 0.6943078000066363    step
s: 289    lr: 4e-05    evaluation reward: 4.28
1234
episode: 1234    score: 5.0    memory length: 254739    epsilon: 0.6936148000066513    step
s: 350    lr: 4e-05    evaluation reward: 4.32
1235
episode: 1235    score: 4.0    memory length: 255020    epsilon: 0.6930584200066634    step
s: 281    lr: 4e-05    evaluation reward: 4.33
1236
episode: 1236    score: 4.0    memory length: 255265    epsilon: 0.6925733200066739    step
s: 245    lr: 4e-05    evaluation reward: 4.33
1237
episode: 1237    score: 7.0    memory length: 255668    epsilon: 0.6917753800066913    step
s: 403    lr: 4e-05    evaluation reward: 4.36
1238
episode: 1238    score: 4.0    memory length: 255920    epsilon: 0.6912764200067021    step
s: 252    lr: 4e-05    evaluation reward: 4.39
1239
episode: 1239    score: 4.0    memory length: 256218    epsilon: 0.6906863800067149    step
s: 298    lr: 4e-05    evaluation reward: 4.4
1240
episode: 1240    score: 2.0    memory length: 256400    epsilon: 0.6903260200067227    step
s: 182    lr: 4e-05    evaluation reward: 4.36
1241
episode: 1241    score: 4.0    memory length: 256651    epsilon: 0.6898290400067335    step
s: 251    lr: 4e-05    evaluation reward: 4.35
1242
episode: 1242    score: 3.0    memory length: 256881    epsilon: 0.6893736400067434    step
s: 230    lr: 4e-05    evaluation reward: 4.34
1243
episode: 1243    score: 4.0    memory length: 257179    epsilon: 0.6887836000067562    step
s: 298    lr: 4e-05    evaluation reward: 4.35
1244
episode: 1244    score: 7.0    memory length: 257564    epsilon: 0.6880213000067728    step
s: 385    lr: 4e-05    evaluation reward: 4.38
1245
episode: 1245    score: 2.0    memory length: 257762    epsilon: 0.6876292600067813    step
s: 198    lr: 4e-05    evaluation reward: 4.36
1246
episode: 1246    score: 1.0    memory length: 257932    epsilon: 0.6872926600067886    step
s: 170    lr: 4e-05    evaluation reward: 4.32
1247
episode: 1247    score: 6.0    memory length: 258287    epsilon: 0.6865897600068038    step
s: 355    lr: 4e-05    evaluation reward: 4.38
1248
episode: 1248    score: 3.0    memory length: 258517    epsilon: 0.6861343600068137    step
s: 230    lr: 4e-05    evaluation reward: 4.38
1249
episode: 1249    score: 6.0    memory length: 258873    epsilon: 0.685429480006829    step
s: 356    lr: 4e-05    evaluation reward: 4.41
```

```
1250
episode: 1250   score: 2.0   memory length: 259070   epsilon: 0.6850394200068375    step
s: 197    lr: 4e-05     evaluation reward: 4.37
1251
episode: 1251   score: 3.0   memory length: 259317   epsilon: 0.6845503600068481    step
s: 247    lr: 4e-05     evaluation reward: 4.34
1252
episode: 1252   score: 7.0   memory length: 259710   epsilon: 0.683772220006865    step
s: 393    lr: 4e-05     evaluation reward: 4.38
1253
episode: 1253   score: 1.0   memory length: 259861   epsilon: 0.6834732400068715    step
s: 151    lr: 4e-05     evaluation reward: 4.24
1254
episode: 1254   score: 4.0   memory length: 260139   epsilon: 0.6829228000068834    step
s: 278    lr: 4e-05     evaluation reward: 4.23
1255
episode: 1255   score: 7.0   memory length: 260567   epsilon: 0.6820753600069018    step
s: 428    lr: 4e-05     evaluation reward: 4.26
1256
episode: 1256   score: 3.0   memory length: 260794   epsilon: 0.6816259000069116    step
s: 227    lr: 4e-05     evaluation reward: 4.24
1257
episode: 1257   score: 5.0   memory length: 261118   epsilon: 0.6809843800069255    step
s: 324    lr: 4e-05     evaluation reward: 4.25
1258
episode: 1258   score: 10.0   memory length: 261625   epsilon: 0.6799805200069473    ste
ps: 507    lr: 4e-05     evaluation reward: 4.32
1259
episode: 1259   score: 8.0   memory length: 262024   epsilon: 0.6791905000069645    step
s: 399    lr: 4e-05     evaluation reward: 4.4
1260
episode: 1260   score: 3.0   memory length: 262250   epsilon: 0.6787430200069742    step
s: 226    lr: 4e-05     evaluation reward: 4.37
1261
episode: 1261   score: 4.0   memory length: 262545   epsilon: 0.6781589200069869    step
s: 295    lr: 4e-05     evaluation reward: 4.41
1262
episode: 1262   score: 2.0   memory length: 262742   epsilon: 0.6777688600069953    step
s: 197    lr: 4e-05     evaluation reward: 4.41
1263
episode: 1263   score: 3.0   memory length: 262968   epsilon: 0.677321380007005    step
s: 226    lr: 4e-05     evaluation reward: 4.4
1264
episode: 1264   score: 1.0   memory length: 263119   epsilon: 0.6770224000070115    step
s: 151    lr: 4e-05     evaluation reward: 4.38
1265
episode: 1265   score: 6.0   memory length: 263457   epsilon: 0.676353160007026    step
s: 338    lr: 4e-05     evaluation reward: 4.44
1266
episode: 1266   score: 3.0   memory length: 263686   epsilon: 0.6758997400070359    step
s: 229    lr: 4e-05     evaluation reward: 4.45
1267
episode: 1267   score: 10.0   memory length: 264136   epsilon: 0.6750087400070552    ste
ps: 450    lr: 4e-05     evaluation reward: 4.49
1268
episode: 1268   score: 2.0   memory length: 264334   epsilon: 0.6746167000070638    step
s: 198    lr: 4e-05     evaluation reward: 4.48
1269
episode: 1269   score: 2.0   memory length: 264531   epsilon: 0.6742266400070722    step
s: 197    lr: 4e-05     evaluation reward: 4.46
1270
episode: 1270   score: 8.0   memory length: 265026   epsilon: 0.6732465400070935    step
s: 495    lr: 4e-05     evaluation reward: 4.5
1271
episode: 1271   score: 5.0   memory length: 265316   epsilon: 0.672672340007106    step
s: 290    lr: 4e-05     evaluation reward: 4.49
```

```
1272
episode: 1272   score: 3.0   memory length: 265529   epsilon: 0.6722506000071151     step
s: 213   lr: 4e-05     evaluation reward: 4.47
1273
episode: 1273   score: 1.0   memory length: 265698   epsilon: 0.6719159800071224     step
s: 169   lr: 4e-05     evaluation reward: 4.46
1274
episode: 1274   score: 3.0   memory length: 265929   epsilon: 0.6714586000071323     step
s: 231   lr: 4e-05     evaluation reward: 4.47
1275
episode: 1275   score: 1.0   memory length: 266080   epsilon: 0.6711596200071388     step
s: 151   lr: 4e-05     evaluation reward: 4.45
1276
episode: 1276   score: 3.0   memory length: 266310   epsilon: 0.6707042200071487     step
s: 230   lr: 4e-05     evaluation reward: 4.42
1277
episode: 1277   score: 5.0   memory length: 266635   epsilon: 0.6700607200071627     step
s: 325   lr: 4e-05     evaluation reward: 4.42
1278
episode: 1278   score: 2.0   memory length: 266856   epsilon: 0.6696231400071722     step
s: 221   lr: 4e-05     evaluation reward: 4.4
1279
episode: 1279   score: 1.0   memory length: 267007   epsilon: 0.6693241600071786     step
s: 151   lr: 4e-05     evaluation reward: 4.37
1280
episode: 1280   score: 6.0   memory length: 267402   epsilon: 0.6685420600071956     step
s: 395   lr: 4e-05     evaluation reward: 4.37
1281
episode: 1281   score: 3.0   memory length: 267612   epsilon: 0.6681262600072047     step
s: 210   lr: 4e-05     evaluation reward: 4.33
1282
episode: 1282   score: 7.0   memory length: 268001   epsilon: 0.6673560400072214     step
s: 389   lr: 4e-05     evaluation reward: 4.31
1283
episode: 1283   score: 6.0   memory length: 268375   epsilon: 0.6666155200072374     step
s: 374   lr: 4e-05     evaluation reward: 4.34
1284
episode: 1284   score: 5.0   memory length: 268681   epsilon: 0.6660096400072506     step
s: 306   lr: 4e-05     evaluation reward: 4.34
1285
episode: 1285   score: 6.0   memory length: 269039   epsilon: 0.665300800007266     step
s: 358   lr: 4e-05     evaluation reward: 4.34
1286
episode: 1286   score: 6.0   memory length: 269395   epsilon: 0.6645959200072813     step
s: 356   lr: 4e-05     evaluation reward: 4.35
1287
episode: 1287   score: 4.0   memory length: 269671   epsilon: 0.6640494400072932     step
s: 276   lr: 4e-05     evaluation reward: 4.36
1288
episode: 1288   score: 3.0   memory length: 269901   epsilon: 0.663594040007303     step
s: 230   lr: 4e-05     evaluation reward: 4.36
1289
episode: 1289   score: 6.0   memory length: 270237   epsilon: 0.6629287600073175     step
s: 336   lr: 4e-05     evaluation reward: 4.39
1290
episode: 1290   score: 9.0   memory length: 270700   epsilon: 0.6620120200073374     step
s: 463   lr: 4e-05     evaluation reward: 4.43
1291
episode: 1291   score: 4.0   memory length: 270995   epsilon: 0.6614279200073501     step
s: 295   lr: 4e-05     evaluation reward: 4.4
1292
episode: 1292   score: 2.0   memory length: 271176   epsilon: 0.6610695400073578     step
s: 181   lr: 4e-05     evaluation reward: 4.4
1293
episode: 1293   score: 3.0   memory length: 271408   epsilon: 0.6606101800073678     step
s: 232   lr: 4e-05     evaluation reward: 4.41
```

```
1294
episode: 1294    score: 9.0    memory length: 271846    epsilon: 0.6597429400073866    step
s: 438    lr: 4e-05    evaluation reward: 4.49
1295
episode: 1295    score: 5.0    memory length: 272135    epsilon: 0.6591707200073991    step
s: 289    lr: 4e-05    evaluation reward: 4.51
1296
episode: 1296    score: 5.0    memory length: 272440    epsilon: 0.6585668200074122    step
s: 305    lr: 4e-05    evaluation reward: 4.51
1297
episode: 1297    score: 5.0    memory length: 272786    epsilon: 0.657881740007427     step
s: 346    lr: 4e-05    evaluation reward: 4.54
1298
episode: 1298    score: 3.0    memory length: 272996    epsilon: 0.6574659400074361    step
s: 210    lr: 4e-05    evaluation reward: 4.56
1299
episode: 1299    score: 3.0    memory length: 273224    epsilon: 0.6570145000074459    step
s: 228    lr: 4e-05    evaluation reward: 4.54
1300
episode: 1300    score: 7.0    memory length: 273598    epsilon: 0.656273980007462     step
s: 374    lr: 4e-05    evaluation reward: 4.56
1301
episode: 1301    score: 8.0    memory length: 274046    epsilon: 0.6553869400074812    step
s: 448    lr: 4e-05    evaluation reward: 4.62
1302
episode: 1302    score: 1.0    memory length: 274197    epsilon: 0.6550879600074877    step
s: 151    lr: 4e-05    evaluation reward: 4.52
1303
episode: 1303    score: 5.0    memory length: 274523    epsilon: 0.6544424800075017    step
s: 326    lr: 4e-05    evaluation reward: 4.49
1304
episode: 1304    score: 7.0    memory length: 274908    epsilon: 0.6536801800075183    step
s: 385    lr: 4e-05    evaluation reward: 4.56
1305
episode: 1305    score: 6.0    memory length: 275256    epsilon: 0.6529911400075332    step
s: 348    lr: 4e-05    evaluation reward: 4.59
1306
episode: 1306    score: 4.0    memory length: 275534    epsilon: 0.6524407000075452    step
s: 278    lr: 4e-05    evaluation reward: 4.56
1307
episode: 1307    score: 5.0    memory length: 275822    epsilon: 0.6518704600075576    step
s: 288    lr: 4e-05    evaluation reward: 4.58
1308
episode: 1308    score: 4.0    memory length: 276099    epsilon: 0.6513220000075695    step
s: 277    lr: 4e-05    evaluation reward: 4.54
1309
episode: 1309    score: 5.0    memory length: 276386    epsilon: 0.6507537400075818    step
s: 287    lr: 4e-05    evaluation reward: 4.55
1310
episode: 1310    score: 2.0    memory length: 276583    epsilon: 0.6503636800075903    step
s: 197    lr: 4e-05    evaluation reward: 4.54
1311
episode: 1311    score: 7.0    memory length: 277001    epsilon: 0.6495360400076082    step
s: 418    lr: 4e-05    evaluation reward: 4.55
1312
episode: 1312    score: 7.0    memory length: 277425    epsilon: 0.6486965200076265    step
s: 424    lr: 4e-05    evaluation reward: 4.58
1313
episode: 1313    score: 4.0    memory length: 277705    epsilon: 0.6481421200076385    step
s: 280    lr: 4e-05    evaluation reward: 4.56
1314
episode: 1314    score: 7.0    memory length: 278049    epsilon: 0.6474610000076533    step
s: 344    lr: 4e-05    evaluation reward: 4.57
1315
episode: 1315    score: 5.0    memory length: 278360    epsilon: 0.6468452200076666    step
s: 311    lr: 4e-05    evaluation reward: 4.59
```

```
1316
episode: 1316   score: 8.0   memory length: 278790   epsilon: 0.6459938200076851   step
s: 430   lr: 4e-05     evaluation reward: 4.6
1317
episode: 1317   score: 6.0   memory length: 279148   epsilon: 0.6452849800077005   step
s: 358   lr: 4e-05     evaluation reward: 4.61
1318
episode: 1318   score: 6.0   memory length: 279504   epsilon: 0.6445801000077158   step
s: 356   lr: 4e-05     evaluation reward: 4.61
1319
episode: 1319   score: 4.0   memory length: 279821   epsilon: 0.6439524400077294   step
s: 317   lr: 4e-05     evaluation reward: 4.62
1320
episode: 1320   score: 4.0   memory length: 280098   epsilon: 0.6434039800077413   step
s: 277   lr: 4e-05     evaluation reward: 4.63
1321
episode: 1321   score: 4.0   memory length: 280376   epsilon: 0.6428535400077533   step
s: 278   lr: 4e-05     evaluation reward: 4.62
1322
episode: 1322   score: 9.0   memory length: 280831   epsilon: 0.6419526400077729   step
s: 455   lr: 4e-05     evaluation reward: 4.64
1323
episode: 1323   score: 5.0   memory length: 281158   epsilon: 0.6413051800077869   step
s: 327   lr: 4e-05     evaluation reward: 4.61
1324
episode: 1324   score: 4.0   memory length: 281416   epsilon: 0.640794340007798    step
s: 258   lr: 4e-05     evaluation reward: 4.57
1325
episode: 1325   score: 3.0   memory length: 281664   epsilon: 0.6403033000078087   step
s: 248   lr: 4e-05     evaluation reward: 4.57
1326
episode: 1326   score: 4.0   memory length: 281984   epsilon: 0.6396697000078224   step
s: 320   lr: 4e-05     evaluation reward: 4.59
1327
episode: 1327   score: 4.0   memory length: 282243   epsilon: 0.6391568800078335   step
s: 259   lr: 4e-05     evaluation reward: 4.56
1328
episode: 1328   score: 5.0   memory length: 282560   epsilon: 0.6385292200078472   step
s: 317   lr: 4e-05     evaluation reward: 4.57
1329
episode: 1329   score: 4.0   memory length: 282820   epsilon: 0.6380144200078584   step
s: 260   lr: 4e-05     evaluation reward: 4.57
1330
episode: 1330   score: 5.0   memory length: 283130   epsilon: 0.6374006200078717   step
s: 310   lr: 4e-05     evaluation reward: 4.57
1331
episode: 1331   score: 5.0   memory length: 283438   epsilon: 0.6367907800078849   step
s: 308   lr: 4e-05     evaluation reward: 4.58
1332
episode: 1332   score: 3.0   memory length: 283664   epsilon: 0.6363433000078946   step
s: 226   lr: 4e-05     evaluation reward: 4.56
1333
episode: 1333   score: 3.0   memory length: 283893   epsilon: 0.6358898800079045   step
s: 229   lr: 4e-05     evaluation reward: 4.54
1334
episode: 1334   score: 8.0   memory length: 284210   epsilon: 0.6352622200079181   step
s: 317   lr: 4e-05     evaluation reward: 4.57
1335
episode: 1335   score: 7.0   memory length: 284578   epsilon: 0.6345335800079339   step
s: 368   lr: 4e-05     evaluation reward: 4.6
1336
episode: 1336   score: 3.0   memory length: 284824   epsilon: 0.6340465000079445   step
s: 246   lr: 4e-05     evaluation reward: 4.59
1337
episode: 1337   score: 4.0   memory length: 285099   epsilon: 0.6335020000079563   step
s: 275   lr: 4e-05     evaluation reward: 4.56
```

```
1338
episode: 1338   score: 4.0   memory length: 285378   epsilon: 0.6329495800079683     step
s: 279    lr: 4e-05     evaluation reward: 4.56
1339
episode: 1339   score: 8.0   memory length: 285803   epsilon: 0.6321080800079866     step
s: 425    lr: 4e-05     evaluation reward: 4.6
1340
episode: 1340   score: 3.0   memory length: 286053   epsilon: 0.6316130800079973     step
s: 250    lr: 4e-05     evaluation reward: 4.61
1341
episode: 1341   score: 7.0   memory length: 286474   epsilon: 0.6307795000080154     step
s: 421    lr: 4e-05     evaluation reward: 4.64
1342
episode: 1342   score: 6.0   memory length: 286804   epsilon: 0.6301261000080296     step
s: 330    lr: 4e-05     evaluation reward: 4.67
1343
episode: 1343   score: 5.0   memory length: 287147   epsilon: 0.6294469600080443     step
s: 343    lr: 4e-05     evaluation reward: 4.68
1344
episode: 1344   score: 4.0   memory length: 287428   epsilon: 0.6288905800080564     step
s: 281    lr: 4e-05     evaluation reward: 4.65
1345
episode: 1345   score: 15.0   memory length: 288006   epsilon: 0.6277461400080813     ste
ps: 578    lr: 4e-05     evaluation reward: 4.78
1346
episode: 1346   score: 4.0   memory length: 288282   epsilon: 0.6271996600080931     step
s: 276    lr: 4e-05     evaluation reward: 4.81
1347
episode: 1347   score: 2.0   memory length: 288501   epsilon: 0.6267660400081025     step
s: 219    lr: 4e-05     evaluation reward: 4.77
1348
episode: 1348   score: 6.0   memory length: 288839   epsilon: 0.6260968000081171     step
s: 338    lr: 4e-05     evaluation reward: 4.8
1349
episode: 1349   score: 2.0   memory length: 289054   epsilon: 0.6256711000081263     step
s: 215    lr: 4e-05     evaluation reward: 4.76
1350
episode: 1350   score: 3.0   memory length: 289302   epsilon: 0.625180060008137     step
s: 248    lr: 4e-05     evaluation reward: 4.77
1351
episode: 1351   score: 3.0   memory length: 289532   epsilon: 0.6247246600081469     step
s: 230    lr: 4e-05     evaluation reward: 4.77
1352
episode: 1352   score: 8.0   memory length: 289954   epsilon: 0.623889100008165     step
s: 422    lr: 4e-05     evaluation reward: 4.78
1353
episode: 1353   score: 4.0   memory length: 290232   epsilon: 0.623338660008177     step
s: 278    lr: 4e-05     evaluation reward: 4.81
1354
episode: 1354   score: 3.0   memory length: 290461   epsilon: 0.6228852400081868     step
s: 229    lr: 4e-05     evaluation reward: 4.8
1355
episode: 1355   score: 4.0   memory length: 290777   epsilon: 0.6222595600082004     step
s: 316    lr: 4e-05     evaluation reward: 4.77
1356
episode: 1356   score: 6.0   memory length: 291133   epsilon: 0.6215546800082157     step
s: 356    lr: 4e-05     evaluation reward: 4.8
1357
episode: 1357   score: 5.0   memory length: 291443   epsilon: 0.620940880008229     step
s: 310    lr: 4e-05     evaluation reward: 4.8
1358
episode: 1358   score: 4.0   memory length: 291719   epsilon: 0.6203944000082409     step
s: 276    lr: 4e-05     evaluation reward: 4.74
1359
episode: 1359   score: 2.0   memory length: 291937   epsilon: 0.6199627600082502     step
s: 218    lr: 4e-05     evaluation reward: 4.68
```

```
1360
episode: 1360    score: 6.0    memory length: 292282    epsilon: 0.6192796600082651    step
s: 345    lr: 4e-05    evaluation reward: 4.71
1361
episode: 1361    score: 4.0    memory length: 292575    epsilon: 0.6186995200082777    step
s: 293    lr: 4e-05    evaluation reward: 4.71
1362
episode: 1362    score: 5.0    memory length: 292919    epsilon: 0.6180184000082924    step
s: 344    lr: 4e-05    evaluation reward: 4.74
1363
episode: 1363    score: 5.0    memory length: 293212    epsilon: 0.617438260008305    step
s: 293    lr: 4e-05    evaluation reward: 4.76
1364
episode: 1364    score: 5.0    memory length: 293531    epsilon: 0.6168066400083188    step
s: 319    lr: 4e-05    evaluation reward: 4.8
1365
episode: 1365    score: 5.0    memory length: 293878    epsilon: 0.6161195800083337    step
s: 347    lr: 4e-05    evaluation reward: 4.79
1366
episode: 1366    score: 6.0    memory length: 294243    epsilon: 0.6153968800083494    step
s: 365    lr: 4e-05    evaluation reward: 4.82
1367
episode: 1367    score: 4.0    memory length: 294515    epsilon: 0.614858320008361    step
s: 272    lr: 4e-05    evaluation reward: 4.76
1368
episode: 1368    score: 3.0    memory length: 294746    epsilon: 0.614400940008371    step
s: 231    lr: 4e-05    evaluation reward: 4.77
1369
episode: 1369    score: 8.0    memory length: 295149    epsilon: 0.6136030000083883    step
s: 403    lr: 4e-05    evaluation reward: 4.83
1370
episode: 1370    score: 3.0    memory length: 295378    epsilon: 0.6131495800083981    step
s: 229    lr: 4e-05    evaluation reward: 4.78
1371
episode: 1371    score: 7.0    memory length: 295797    epsilon: 0.6123199600084162    step
s: 419    lr: 4e-05    evaluation reward: 4.8
1372
episode: 1372    score: 3.0    memory length: 296008    epsilon: 0.6119021800084252    step
s: 211    lr: 4e-05    evaluation reward: 4.8
1373
episode: 1373    score: 8.0    memory length: 296435    epsilon: 0.6110567200084436    step
s: 427    lr: 4e-05    evaluation reward: 4.87
1374
episode: 1374    score: 2.0    memory length: 296633    epsilon: 0.6106646800084521    step
s: 198    lr: 4e-05    evaluation reward: 4.86
1375
episode: 1375    score: 11.0    memory length: 297034    epsilon: 0.6098707000084693    ste
ps: 401    lr: 4e-05    evaluation reward: 4.96
1376
episode: 1376    score: 8.0    memory length: 297473    epsilon: 0.6090014800084882    step
s: 439    lr: 4e-05    evaluation reward: 5.01
1377
episode: 1377    score: 5.0    memory length: 297797    epsilon: 0.6083599600085021    step
s: 324    lr: 4e-05    evaluation reward: 5.01
1378
episode: 1378    score: 9.0    memory length: 298278    epsilon: 0.6074075800085228    step
s: 481    lr: 4e-05    evaluation reward: 5.08
1379
episode: 1379    score: 2.0    memory length: 298478    epsilon: 0.6070115800085314    step
s: 200    lr: 4e-05    evaluation reward: 5.09
1380
episode: 1380    score: 9.0    memory length: 298979    epsilon: 0.6060196000085529    step
s: 501    lr: 4e-05    evaluation reward: 5.12
1381
episode: 1381    score: 6.0    memory length: 299316    epsilon: 0.6053523400085674    step
s: 337    lr: 4e-05    evaluation reward: 5.15
```

```
1382
episode: 1382   score: 14.0   memory length: 299866   epsilon: 0.604263340008591   step
s: 550    lr: 4e-05      evaluation reward: 5.22
1383
episode: 1383   score: 3.0   memory length: 300095   epsilon: 0.6038099200086009   step
s: 229    lr: 1.6000000000000003e-05     evaluation reward: 5.19
1384
episode: 1384   score: 7.0   memory length: 300469   epsilon: 0.603069400008617   step
s: 374    lr: 1.6000000000000003e-05     evaluation reward: 5.21
1385
episode: 1385   score: 4.0   memory length: 300711   epsilon: 0.6025902400086274   step
s: 242    lr: 1.6000000000000003e-05     evaluation reward: 5.19
1386
episode: 1386   score: 6.0   memory length: 301039   epsilon: 0.6019408000086415   step
s: 328    lr: 1.6000000000000003e-05     evaluation reward: 5.19
1387
episode: 1387   score: 7.0   memory length: 301450   epsilon: 0.6011270200086591   step
s: 411    lr: 1.6000000000000003e-05     evaluation reward: 5.22
1388
episode: 1388   score: 7.0   memory length: 301800   epsilon: 0.6004340200086742   step
s: 350    lr: 1.6000000000000003e-05     evaluation reward: 5.26
1389
episode: 1389   score: 7.0   memory length: 302163   epsilon: 0.5997152800086898   step
s: 363    lr: 1.6000000000000003e-05     evaluation reward: 5.27
1390
episode: 1390   score: 7.0   memory length: 302557   epsilon: 0.5989351600087067   step
s: 394    lr: 1.6000000000000003e-05     evaluation reward: 5.25
1391
episode: 1391   score: 5.0   memory length: 302866   epsilon: 0.59832334000872   steps:
 309    lr: 1.6000000000000003e-05     evaluation reward: 5.26
1392
episode: 1392   score: 5.0   memory length: 303158   epsilon: 0.5977451800087326   step
s: 292    lr: 1.6000000000000003e-05     evaluation reward: 5.29
1393
episode: 1393   score: 3.0   memory length: 303371   epsilon: 0.5973234400087417   step
s: 213    lr: 1.6000000000000003e-05     evaluation reward: 5.29
1394
episode: 1394   score: 2.0   memory length: 303570   epsilon: 0.5969294200087503   step
s: 199    lr: 1.6000000000000003e-05     evaluation reward: 5.22
1395
episode: 1395   score: 6.0   memory length: 303892   epsilon: 0.5962918600087641   step
s: 322    lr: 1.6000000000000003e-05     evaluation reward: 5.23
1396
episode: 1396   score: 7.0   memory length: 304353   epsilon: 0.5953790800087839   step
s: 461    lr: 1.6000000000000003e-05     evaluation reward: 5.25
1397
episode: 1397   score: 3.0   memory length: 304566   epsilon: 0.5949573400087931   step
s: 213    lr: 1.6000000000000003e-05     evaluation reward: 5.23
1398
episode: 1398   score: 4.0   memory length: 304843   epsilon: 0.594408880008805   step
s: 277    lr: 1.6000000000000003e-05     evaluation reward: 5.24
1399
episode: 1399   score: 6.0   memory length: 305196   epsilon: 0.5937099400088202   step
s: 353    lr: 1.6000000000000003e-05     evaluation reward: 5.27
1400
episode: 1400   score: 5.0   memory length: 305521   epsilon: 0.5930664400088341   step
s: 325    lr: 1.6000000000000003e-05     evaluation reward: 5.25
1401
episode: 1401   score: 4.0   memory length: 305783   epsilon: 0.5925476800088454   step
s: 262    lr: 1.6000000000000003e-05     evaluation reward: 5.21
1402
episode: 1402   score: 1.0   memory length: 305933   epsilon: 0.5922506800088518   step
s: 150    lr: 1.6000000000000003e-05     evaluation reward: 5.21
1403
episode: 1403   score: 3.0   memory length: 306144   epsilon: 0.5918329000088609   step
s: 211    lr: 1.6000000000000003e-05     evaluation reward: 5.19
```

```
1404
episode: 1404   score: 5.0   memory length: 306456   epsilon: 0.5912151400088743    step
s: 312   lr: 1.6000000000000003e-05    evaluation reward: 5.17
1405
episode: 1405   score: 3.0   memory length: 306666   epsilon: 0.5907993400088833    step
s: 210   lr: 1.6000000000000003e-05    evaluation reward: 5.14
1406
episode: 1406   score: 1.0   memory length: 306817   epsilon: 0.5905003600088898    step
s: 151   lr: 1.6000000000000003e-05    evaluation reward: 5.11
1407
episode: 1407   score: 6.0   memory length: 307167   epsilon: 0.5898073600089049    step
s: 350   lr: 1.6000000000000003e-05    evaluation reward: 5.12
1408
episode: 1408   score: 6.0   memory length: 307522   epsilon: 0.5891044600089201    step
s: 355   lr: 1.6000000000000003e-05    evaluation reward: 5.14
1409
episode: 1409   score: 5.0   memory length: 307826   epsilon: 0.5885025400089332    step
s: 304   lr: 1.6000000000000003e-05    evaluation reward: 5.14
1410
episode: 1410   score: 5.0   memory length: 308128   epsilon: 0.5879045800089462    step
s: 302   lr: 1.6000000000000003e-05    evaluation reward: 5.17
1411
episode: 1411   score: 4.0   memory length: 308427   epsilon: 0.587312560008959    step
s: 299   lr: 1.6000000000000003e-05    evaluation reward: 5.14
1412
episode: 1412   score: 3.0   memory length: 308675   epsilon: 0.5868215200089697    step
s: 248   lr: 1.6000000000000003e-05    evaluation reward: 5.1
1413
episode: 1413   score: 7.0   memory length: 308921   epsilon: 0.5863344400089803    step
s: 246   lr: 1.6000000000000003e-05    evaluation reward: 5.13
1414
episode: 1414   score: 6.0   memory length: 309247   epsilon: 0.5856889600089943    step
s: 326   lr: 1.6000000000000003e-05    evaluation reward: 5.12
1415
episode: 1415   score: 2.0   memory length: 309444   epsilon: 0.5852989000090028    step
s: 197   lr: 1.6000000000000003e-05    evaluation reward: 5.09
1416
episode: 1416   score: 2.0   memory length: 309625   epsilon: 0.5849405200090105    step
s: 181   lr: 1.6000000000000003e-05    evaluation reward: 5.03
1417
episode: 1417   score: 4.0   memory length: 309882   epsilon: 0.5844316600090216    step
s: 257   lr: 1.6000000000000003e-05    evaluation reward: 5.01
1418
episode: 1418   score: 6.0   memory length: 310253   epsilon: 0.5836970800090375    step
s: 371   lr: 1.6000000000000003e-05    evaluation reward: 5.01
1419
episode: 1419   score: 7.0   memory length: 310620   epsilon: 0.5829704200090533    step
s: 367   lr: 1.6000000000000003e-05    evaluation reward: 5.04
1420
episode: 1420   score: 5.0   memory length: 310965   epsilon: 0.5822873200090681    step
s: 345   lr: 1.6000000000000003e-05    evaluation reward: 5.05
1421
episode: 1421   score: 6.0   memory length: 311303   epsilon: 0.5816180800090827    step
s: 338   lr: 1.6000000000000003e-05    evaluation reward: 5.07
1422
episode: 1422   score: 6.0   memory length: 311656   epsilon: 0.5809191400090978    step
s: 353   lr: 1.6000000000000003e-05    evaluation reward: 5.04
1423
episode: 1423   score: 3.0   memory length: 311884   epsilon: 0.5804677000091076    step
s: 228   lr: 1.6000000000000003e-05    evaluation reward: 5.02
1424
episode: 1424   score: 7.0   memory length: 312280   epsilon: 0.5796836200091247    step
s: 396   lr: 1.6000000000000003e-05    evaluation reward: 5.05
1425
episode: 1425   score: 4.0   memory length: 312558   epsilon: 0.5791331800091366    step
s: 278   lr: 1.6000000000000003e-05    evaluation reward: 5.06
```

```
1426
episode: 1426    score: 7.0    memory length: 312936    epsilon: 0.5783847400091529    step
s: 378    lr: 1.6000000000000003e-05    evaluation reward: 5.09
1427
episode: 1427    score: 5.0    memory length: 313236    epsilon: 0.5777907400091657    step
s: 300    lr: 1.6000000000000003e-05    evaluation reward: 5.1
1428
episode: 1428    score: 9.0    memory length: 313753    epsilon: 0.576767080009188    step
s: 517    lr: 1.6000000000000003e-05    evaluation reward: 5.14
1429
episode: 1429    score: 5.0    memory length: 314032    epsilon: 0.5762146600092    steps:
 279    lr: 1.6000000000000003e-05    evaluation reward: 5.15
1430
episode: 1430    score: 4.0    memory length: 314308    epsilon: 0.5756681800092118    step
s: 276    lr: 1.6000000000000003e-05    evaluation reward: 5.14
1431
episode: 1431    score: 5.0    memory length: 314633    epsilon: 0.5750246800092258    step
s: 325    lr: 1.6000000000000003e-05    evaluation reward: 5.14
1432
episode: 1432    score: 7.0    memory length: 315055    epsilon: 0.5741891200092439    step
s: 422    lr: 1.6000000000000003e-05    evaluation reward: 5.18
1433
episode: 1433    score: 5.0    memory length: 315380    epsilon: 0.5735456200092579    step
s: 325    lr: 1.6000000000000003e-05    evaluation reward: 5.2
1434
episode: 1434    score: 8.0    memory length: 315781    epsilon: 0.5727516400092751    step
s: 401    lr: 1.6000000000000003e-05    evaluation reward: 5.2
1435
episode: 1435    score: 4.0    memory length: 316038    epsilon: 0.5722427800092862    step
s: 257    lr: 1.6000000000000003e-05    evaluation reward: 5.17
1436
episode: 1436    score: 6.0    memory length: 316359    epsilon: 0.5716072000093    steps:
 321    lr: 1.6000000000000003e-05    evaluation reward: 5.2
1437
episode: 1437    score: 3.0    memory length: 316572    epsilon: 0.5711854600093091    step
s: 213    lr: 1.6000000000000003e-05    evaluation reward: 5.19
1438
episode: 1438    score: 4.0    memory length: 316853    epsilon: 0.5706290800093212    step
s: 281    lr: 1.6000000000000003e-05    evaluation reward: 5.19
1439
episode: 1439    score: 2.0    memory length: 317053    epsilon: 0.5702330800093298    step
s: 200    lr: 1.6000000000000003e-05    evaluation reward: 5.13
1440
episode: 1440    score: 5.0    memory length: 317362    epsilon: 0.5696212600093431    step
s: 309    lr: 1.6000000000000003e-05    evaluation reward: 5.15
1441
episode: 1441    score: 0.0    memory length: 317485    epsilon: 0.5693777200093484    step
s: 123    lr: 1.6000000000000003e-05    evaluation reward: 5.08
1442
episode: 1442    score: 8.0    memory length: 317920    epsilon: 0.5685164200093671    step
s: 435    lr: 1.6000000000000003e-05    evaluation reward: 5.1
1443
episode: 1443    score: 6.0    memory length: 318275    epsilon: 0.5678135200093823    step
s: 355    lr: 1.6000000000000003e-05    evaluation reward: 5.11
1444
episode: 1444    score: 10.0    memory length: 318776    epsilon: 0.5668215400094039    ste
ps: 501    lr: 1.6000000000000003e-05    evaluation reward: 5.17
1445
episode: 1445    score: 2.0    memory length: 318974    epsilon: 0.5664295000094124    step
s: 198    lr: 1.6000000000000003e-05    evaluation reward: 5.04
1446
episode: 1446    score: 5.0    memory length: 319317    epsilon: 0.5657503600094271    step
s: 343    lr: 1.6000000000000003e-05    evaluation reward: 5.05
1447
episode: 1447    score: 6.0    memory length: 319714    epsilon: 0.5649643000094442    step
s: 397    lr: 1.6000000000000003e-05    evaluation reward: 5.09
```

```
1448
episode: 1448   score: 9.0   memory length: 320221   epsilon: 0.563960440009466     step
s: 507   lr: 1.6000000000000003e-05     evaluation reward: 5.12
1449
episode: 1449   score: 1.0   memory length: 320390   epsilon: 0.56362582000094733    step
s: 169   lr: 1.6000000000000003e-05     evaluation reward: 5.11
1450
episode: 1450   score: 7.0   memory length: 320758   epsilon: 0.5628971800094891     step
s: 368   lr: 1.6000000000000003e-05     evaluation reward: 5.15
1451
episode: 1451   score: 8.0   memory length: 321194   epsilon: 0.5620339000095078     step
s: 436   lr: 1.6000000000000003e-05     evaluation reward: 5.2
1452
episode: 1452   score: 3.0   memory length: 321407   epsilon: 0.561612160009517     step
s: 213   lr: 1.6000000000000003e-05     evaluation reward: 5.15
1453
episode: 1453   score: 8.0   memory length: 321792   epsilon: 0.5608498600095335     step
s: 385   lr: 1.6000000000000003e-05     evaluation reward: 5.19
1454
episode: 1454   score: 4.0   memory length: 322087   epsilon: 0.56026576000954    62     step
s: 295   lr: 1.6000000000000003e-05     evaluation reward: 5.2
1455
episode: 1455   score: 4.0   memory length: 322368   epsilon: 0.5597093800095583     step
s: 281   lr: 1.6000000000000003e-05     evaluation reward: 5.2
1456
episode: 1456   score: 6.0   memory length: 322739   epsilon: 0.5589748000095742     step
s: 371   lr: 1.6000000000000003e-05     evaluation reward: 5.2
1457
episode: 1457   score: 7.0   memory length: 323149   epsilon: 0.5581630000095918     step
s: 410   lr: 1.6000000000000003e-05     evaluation reward: 5.22
1458
episode: 1458   score: 5.0   memory length: 323494   epsilon: 0.5574799000096067     step
s: 345   lr: 1.6000000000000003e-05     evaluation reward: 5.23
1459
episode: 1459   score: 6.0   memory length: 323829   epsilon: 0.5568166000096211     step
s: 335   lr: 1.6000000000000003e-05     evaluation reward: 5.27
1460
episode: 1460   score: 4.0   memory length: 324124   epsilon: 0.5562325000096338     step
s: 295   lr: 1.6000000000000003e-05     evaluation reward: 5.25
1461
episode: 1461   score: 7.0   memory length: 324547   epsilon: 0.5553949600096519     step
s: 423   lr: 1.6000000000000003e-05     evaluation reward: 5.28
1462
episode: 1462   score: 4.0   memory length: 324807   epsilon: 0.5548801600096631     step
s: 260   lr: 1.6000000000000003e-05     evaluation reward: 5.27
1463
episode: 1463   score: 2.0   memory length: 324989   epsilon: 0.5545198000096709     step
s: 182   lr: 1.6000000000000003e-05     evaluation reward: 5.24
1464
episode: 1464   score: 5.0   memory length: 325295   epsilon: 0.5539139200096841     step
s: 306   lr: 1.6000000000000003e-05     evaluation reward: 5.24
1465
episode: 1465   score: 3.0   memory length: 325523   epsilon: 0.5534624800096939     step
s: 228   lr: 1.6000000000000003e-05     evaluation reward: 5.22
1466
episode: 1466   score: 6.0   memory length: 325881   epsilon: 0.5527536400097093     step
s: 358   lr: 1.6000000000000003e-05     evaluation reward: 5.22
1467
episode: 1467   score: 5.0   memory length: 326204   epsilon: 0.5521141000097232     step
s: 323   lr: 1.6000000000000003e-05     evaluation reward: 5.23
1468
episode: 1468   score: 5.0   memory length: 326530   epsilon: 0.5514686200097372     step
s: 326   lr: 1.6000000000000003e-05     evaluation reward: 5.25
1469
episode: 1469   score: 5.0   memory length: 326834   epsilon: 0.5508667000097502     step
s: 304   lr: 1.6000000000000003e-05     evaluation reward: 5.22
```

```
1470
episode: 1470    score: 5.0    memory length: 327143    epsilon: 0.5502548800097635    step
s: 309    lr: 1.6000000000000003e-05    evaluation reward: 5.24
1471
episode: 1471    score: 4.0    memory length: 327384    epsilon: 0.5497777000097739    step
s: 241    lr: 1.6000000000000003e-05    evaluation reward: 5.21
1472
episode: 1472    score: 4.0    memory length: 327643    epsilon: 0.549264880009785    step
s: 259    lr: 1.6000000000000003e-05    evaluation reward: 5.22
1473
episode: 1473    score: 2.0    memory length: 327843    epsilon: 0.5488688800097936    step
s: 200    lr: 1.6000000000000003e-05    evaluation reward: 5.16
1474
episode: 1474    score: 6.0    memory length: 328215    epsilon: 0.5481323200098096    step
s: 372    lr: 1.6000000000000003e-05    evaluation reward: 5.2
1475
episode: 1475    score: 6.0    memory length: 328551    epsilon: 0.547467040009824    step
s: 336    lr: 1.6000000000000003e-05    evaluation reward: 5.15
1476
episode: 1476    score: 6.0    memory length: 328878    epsilon: 0.5468195800098381    step
s: 327    lr: 1.6000000000000003e-05    evaluation reward: 5.13
1477
episode: 1477    score: 3.0    memory length: 329087    epsilon: 0.5464057600098471    step
s: 209    lr: 1.6000000000000003e-05    evaluation reward: 5.11
1478
episode: 1478    score: 6.0    memory length: 329481    epsilon: 0.545625640009864    step
s: 394    lr: 1.6000000000000003e-05    evaluation reward: 5.08
1479
episode: 1479    score: 3.0    memory length: 329710    epsilon: 0.5451722200098739    step
s: 229    lr: 1.6000000000000003e-05    evaluation reward: 5.09
1480
episode: 1480    score: 7.0    memory length: 330114    epsilon: 0.5443723000098912    step
s: 404    lr: 1.6000000000000003e-05    evaluation reward: 5.07
1481
episode: 1481    score: 6.0    memory length: 330435    epsilon: 0.543736720009905    step
s: 321    lr: 1.6000000000000003e-05    evaluation reward: 5.07
1482
episode: 1482    score: 3.0    memory length: 330683    epsilon: 0.5432456800099157    step
s: 248    lr: 1.6000000000000003e-05    evaluation reward: 4.96
1483
episode: 1483    score: 6.0    memory length: 331041    epsilon: 0.5425368400099311    step
s: 358    lr: 1.6000000000000003e-05    evaluation reward: 4.99
1484
episode: 1484    score: 5.0    memory length: 331352    epsilon: 0.5419210600099444    step
s: 311    lr: 1.6000000000000003e-05    evaluation reward: 4.97
1485
episode: 1485    score: 3.0    memory length: 331565    epsilon: 0.5414993200099536    step
s: 213    lr: 1.6000000000000003e-05    evaluation reward: 4.96
1486
episode: 1486    score: 6.0    memory length: 331924    epsilon: 0.540788500009969    step
s: 359    lr: 1.6000000000000003e-05    evaluation reward: 4.96
1487
episode: 1487    score: 7.0    memory length: 332301    epsilon: 0.5400420400099852    step
s: 377    lr: 1.6000000000000003e-05    evaluation reward: 4.96
1488
episode: 1488    score: 4.0    memory length: 332577    epsilon: 0.5394955600099971    step
s: 276    lr: 1.6000000000000003e-05    evaluation reward: 4.93
1489
episode: 1489    score: 4.0    memory length: 332837    epsilon: 0.5389807600100083    step
s: 260    lr: 1.6000000000000003e-05    evaluation reward: 4.9
1490
episode: 1490    score: 3.0    memory length: 333085    epsilon: 0.5384897200100189    step
s: 248    lr: 1.6000000000000003e-05    evaluation reward: 4.86
1491
episode: 1491    score: 8.0    memory length: 333503    epsilon: 0.5376620800100369    step
s: 418    lr: 1.6000000000000003e-05    evaluation reward: 4.89
```

```
1492
episode: 1492   score: 7.0   memory length: 333877   epsilon: 0.536921560010053     step
s: 374    lr: 1.6000000000000003e-05     evaluation reward: 4.91
1493
episode: 1493   score: 8.0   memory length: 334329   epsilon: 0.5360266000100724    step
s: 452    lr: 1.6000000000000003e-05     evaluation reward: 4.96
1494
episode: 1494   score: 5.0   memory length: 334655   epsilon: 0.5353811200100864    step
s: 326    lr: 1.6000000000000003e-05     evaluation reward: 4.99
1495
episode: 1495   score: 6.0   memory length: 334975   epsilon: 0.5347475200101002    step
s: 320    lr: 1.6000000000000003e-05     evaluation reward: 4.99
1496
episode: 1496   score: 6.0   memory length: 335368   epsilon: 0.5339693800101171    step
s: 393    lr: 1.6000000000000003e-05     evaluation reward: 4.98
1497
episode: 1497   score: 7.0   memory length: 335796   epsilon: 0.5331219400101355    step
s: 428    lr: 1.6000000000000003e-05     evaluation reward: 5.02
1498
episode: 1498   score: 8.0   memory length: 336184   epsilon: 0.5323537000101521    step
s: 388    lr: 1.6000000000000003e-05     evaluation reward: 5.06
1499
episode: 1499   score: 4.0   memory length: 336441   epsilon: 0.5318448400101632    step
s: 257    lr: 1.6000000000000003e-05     evaluation reward: 5.04
1500
episode: 1500   score: 9.0   memory length: 336903   epsilon: 0.530930080010183     step
s: 462    lr: 1.6000000000000003e-05     evaluation reward: 5.08
1501
episode: 1501   score: 4.0   memory length: 337183   epsilon: 0.5303756800101951    step
s: 280    lr: 1.6000000000000003e-05     evaluation reward: 5.08
1502
episode: 1502   score: 4.0   memory length: 337430   epsilon: 0.5298866200102057    step
s: 247    lr: 1.6000000000000003e-05     evaluation reward: 5.11
1503
episode: 1503   score: 7.0   memory length: 337805   epsilon: 0.5291441200102218    step
s: 375    lr: 1.6000000000000003e-05     evaluation reward: 5.15
1504
episode: 1504   score: 8.0   memory length: 338268   epsilon: 0.5282273800102417    step
s: 463    lr: 1.6000000000000003e-05     evaluation reward: 5.18
1505
episode: 1505   score: 3.0   memory length: 338498   epsilon: 0.5277719800102516    step
s: 230    lr: 1.6000000000000003e-05     evaluation reward: 5.18
1506
episode: 1506   score: 4.0   memory length: 338767   epsilon: 0.5272393600102632    step
s: 269    lr: 1.6000000000000003e-05     evaluation reward: 5.21
1507
episode: 1507   score: 5.0   memory length: 339057   epsilon: 0.5266651600102756    step
s: 290    lr: 1.6000000000000003e-05     evaluation reward: 5.2
1508
episode: 1508   score: 5.0   memory length: 339364   epsilon: 0.5260573000102888    step
s: 307    lr: 1.6000000000000003e-05     evaluation reward: 5.19
1509
episode: 1509   score: 4.0   memory length: 339623   epsilon: 0.5255444800103    steps:
 259    lr: 1.6000000000000003e-05     evaluation reward: 5.18
1510
episode: 1510   score: 7.0   memory length: 340012   epsilon: 0.5247742600103167    step
s: 389    lr: 1.6000000000000003e-05     evaluation reward: 5.2
1511
episode: 1511   score: 6.0   memory length: 340390   epsilon: 0.5240258200103329    step
s: 378    lr: 1.6000000000000003e-05     evaluation reward: 5.22
1512
episode: 1512   score: 8.0   memory length: 340826   epsilon: 0.5231625400103517    step
s: 436    lr: 1.6000000000000003e-05     evaluation reward: 5.27
1513
episode: 1513   score: 7.0   memory length: 341232   epsilon: 0.5223586600103691    step
s: 406    lr: 1.6000000000000003e-05     evaluation reward: 5.27
```

```
1514
episode: 1514   score: 8.0   memory length: 341602   epsilon: 0.521626060010385    step
s: 370   lr: 1.6000000000000003e-05    evaluation reward: 5.29
1515
episode: 1515   score: 5.0   memory length: 341950   epsilon: 0.5209370200104    steps:
 348   lr: 1.6000000000000003e-05    evaluation reward: 5.32
1516
episode: 1516   score: 6.0   memory length: 342292   epsilon: 0.5202598600104147    step
s: 342   lr: 1.6000000000000003e-05    evaluation reward: 5.36
1517
episode: 1517   score: 5.0   memory length: 342600   epsilon: 0.5196500200104279    step
s: 308   lr: 1.6000000000000003e-05    evaluation reward: 5.37
1518
episode: 1518   score: 7.0   memory length: 342966   epsilon: 0.5189253400104437    step
s: 366   lr: 1.6000000000000003e-05    evaluation reward: 5.38
1519
episode: 1519   score: 4.0   memory length: 343225   epsilon: 0.5184125200104548    step
s: 259   lr: 1.6000000000000003e-05    evaluation reward: 5.35
1520
episode: 1520   score: 7.0   memory length: 343580   epsilon: 0.51770962001047    steps:
 355   lr: 1.6000000000000003e-05    evaluation reward: 5.37
1521
episode: 1521   score: 5.0   memory length: 343905   epsilon: 0.517066120010484    step
s: 325   lr: 1.6000000000000003e-05    evaluation reward: 5.36
1522
episode: 1522   score: 4.0   memory length: 344185   epsilon: 0.5165117200104961    step
s: 280   lr: 1.6000000000000003e-05    evaluation reward: 5.34
1523
episode: 1523   score: 7.0   memory length: 344606   epsilon: 0.5156781400105142    step
s: 421   lr: 1.6000000000000003e-05    evaluation reward: 5.38
1524
episode: 1524   score: 7.0   memory length: 345028   epsilon: 0.5148425800105323    step
s: 422   lr: 1.6000000000000003e-05    evaluation reward: 5.38
1525
episode: 1525   score: 7.0   memory length: 345456   epsilon: 0.5139951400105507    step
s: 428   lr: 1.6000000000000003e-05    evaluation reward: 5.41
1526
episode: 1526   score: 4.0   memory length: 345718   epsilon: 0.513476380010562    step
s: 262   lr: 1.6000000000000003e-05    evaluation reward: 5.38
1527
episode: 1527   score: 9.0   memory length: 346061   epsilon: 0.5127972400105767    step
s: 343   lr: 1.6000000000000003e-05    evaluation reward: 5.42
1528
episode: 1528   score: 3.0   memory length: 346311   epsilon: 0.5123022400105874    step
s: 250   lr: 1.6000000000000003e-05    evaluation reward: 5.36
1529
episode: 1529   score: 3.0   memory length: 346562   epsilon: 0.5118052600105982    step
s: 251   lr: 1.6000000000000003e-05    evaluation reward: 5.34
1530
episode: 1530   score: 3.0   memory length: 346775   epsilon: 0.5113835200106074    step
s: 213   lr: 1.6000000000000003e-05    evaluation reward: 5.33
1531
episode: 1531   score: 5.0   memory length: 347091   epsilon: 0.510757840010621    step
s: 316   lr: 1.6000000000000003e-05    evaluation reward: 5.33
1532
episode: 1532   score: 2.0   memory length: 347288   epsilon: 0.5103677800106294    step
s: 197   lr: 1.6000000000000003e-05    evaluation reward: 5.28
1533
episode: 1533   score: 2.0   memory length: 347468   epsilon: 0.5100113800106372    step
s: 180   lr: 1.6000000000000003e-05    evaluation reward: 5.25
1534
episode: 1534   score: 3.0   memory length: 347717   epsilon: 0.5095183600106479    step
s: 249   lr: 1.6000000000000003e-05    evaluation reward: 5.2
1535
episode: 1535   score: 6.0   memory length: 348076   epsilon: 0.5088075400106633    step
s: 359   lr: 1.6000000000000003e-05    evaluation reward: 5.22
```

```
1536
episode: 1536    score: 7.0    memory length: 348343    epsilon: 0.5082788800106748    step
s: 267    lr: 1.6000000000000003e-05    evaluation reward: 5.23
1537
episode: 1537    score: 11.0    memory length: 348823    epsilon: 0.5073284800106954    ste
ps: 480    lr: 1.6000000000000003e-05    evaluation reward: 5.31
1538
episode: 1538    score: 5.0    memory length: 349147    epsilon: 0.5066869600107093    step
s: 324    lr: 1.6000000000000003e-05    evaluation reward: 5.32
1539
episode: 1539    score: 6.0    memory length: 349489    epsilon: 0.506009800010724    step
s: 342    lr: 1.6000000000000003e-05    evaluation reward: 5.36
1540
episode: 1540    score: 4.0    memory length: 349768    epsilon: 0.505457380010736    step
s: 279    lr: 1.6000000000000003e-05    evaluation reward: 5.35
1541
episode: 1541    score: 4.0    memory length: 350029    epsilon: 0.5049406000107473    step
s: 261    lr: 1.6000000000000003e-05    evaluation reward: 5.39
1542
episode: 1542    score: 2.0    memory length: 350227    epsilon: 0.5045485600107558    step
s: 198    lr: 1.6000000000000003e-05    evaluation reward: 5.33
1543
episode: 1543    score: 4.0    memory length: 350486    epsilon: 0.5040357400107669    step
s: 259    lr: 1.6000000000000003e-05    evaluation reward: 5.31
1544
episode: 1544    score: 6.0    memory length: 350832    epsilon: 0.5033506600107818    step
s: 346    lr: 1.6000000000000003e-05    evaluation reward: 5.27
1545
episode: 1545    score: 7.0    memory length: 351229    epsilon: 0.5025646000107988    step
s: 397    lr: 1.6000000000000003e-05    evaluation reward: 5.32
1546
episode: 1546    score: 5.0    memory length: 351517    epsilon: 0.5019943600108112    step
s: 288    lr: 1.6000000000000003e-05    evaluation reward: 5.32
1547
episode: 1547    score: 3.0    memory length: 351730    epsilon: 0.5015726200108204    step
s: 213    lr: 1.6000000000000003e-05    evaluation reward: 5.29
1548
episode: 1548    score: 3.0    memory length: 351943    epsilon: 0.5011508800108295    step
s: 213    lr: 1.6000000000000003e-05    evaluation reward: 5.23
1549
episode: 1549    score: 4.0    memory length: 352216    epsilon: 0.5006103400108413    step
s: 273    lr: 1.6000000000000003e-05    evaluation reward: 5.26
1550
episode: 1550    score: 5.0    memory length: 352544    epsilon: 0.49996090001085425    ste
ps: 328    lr: 1.6000000000000003e-05    evaluation reward: 5.24
1551
episode: 1551    score: 10.0    memory length: 353049    epsilon: 0.4989610000108479    ste
ps: 505    lr: 1.6000000000000003e-05    evaluation reward: 5.26
1552
episode: 1552    score: 4.0    memory length: 353328    epsilon: 0.4984085800108444    step
s: 279    lr: 1.6000000000000003e-05    evaluation reward: 5.27
1553
episode: 1553    score: 8.0    memory length: 353733    epsilon: 0.49760668001083935    ste
ps: 405    lr: 1.6000000000000003e-05    evaluation reward: 5.27
1554
episode: 1554    score: 7.0    memory length: 354109    epsilon: 0.49686220001083464    ste
ps: 376    lr: 1.6000000000000003e-05    evaluation reward: 5.3
1555
episode: 1555    score: 3.0    memory length: 354318    epsilon: 0.496448380010832    step
s: 209    lr: 1.6000000000000003e-05    evaluation reward: 5.29
1556
episode: 1556    score: 6.0    memory length: 354693    epsilon: 0.4957058800108273    step
s: 375    lr: 1.6000000000000003e-05    evaluation reward: 5.29
1557
episode: 1557    score: 6.0    memory length: 355061    epsilon: 0.4949772400108227    step
s: 368    lr: 1.6000000000000003e-05    evaluation reward: 5.28
```

```
1558
episode: 1558    score: 5.0    memory length: 355392    epsilon: 0.49432186001081857    ste
ps: 331    lr: 1.6000000000000003e-05    evaluation reward: 5.28
1559
episode: 1559    score: 7.0    memory length: 355784    epsilon: 0.49354570001081366    ste
ps: 392    lr: 1.6000000000000003e-05    evaluation reward: 5.29
1560
episode: 1560    score: 4.0    memory length: 356045    epsilon: 0.4930289200108104    step
s: 261    lr: 1.6000000000000003e-05    evaluation reward: 5.29
1561
episode: 1561    score: 3.0    memory length: 356273    epsilon: 0.49257748001080753    ste
ps: 228    lr: 1.6000000000000003e-05    evaluation reward: 5.25
1562
episode: 1562    score: 3.0    memory length: 356500    epsilon: 0.4921280200108047    step
s: 227    lr: 1.6000000000000003e-05    evaluation reward: 5.24
1563
episode: 1563    score: 6.0    memory length: 356819    epsilon: 0.4914964000108007    step
s: 319    lr: 1.6000000000000003e-05    evaluation reward: 5.28
1564
episode: 1564    score: 5.0    memory length: 357129    epsilon: 0.4908826000107968    step
s: 310    lr: 1.6000000000000003e-05    evaluation reward: 5.28
1565
episode: 1565    score: 7.0    memory length: 357513    epsilon: 0.490122280010792    step
s: 384    lr: 1.6000000000000003e-05    evaluation reward: 5.32
1566
episode: 1566    score: 3.0    memory length: 357741    epsilon: 0.48967084001078914    ste
ps: 228    lr: 1.6000000000000003e-05    evaluation reward: 5.29
1567
episode: 1567    score: 7.0    memory length: 358146    epsilon: 0.48886894001078407    ste
ps: 405    lr: 1.6000000000000003e-05    evaluation reward: 5.31
1568
episode: 1568    score: 5.0    memory length: 358452    epsilon: 0.48826306001078024    ste
ps: 306    lr: 1.6000000000000003e-05    evaluation reward: 5.31
1569
episode: 1569    score: 3.0    memory length: 358663    epsilon: 0.4878452800107776    step
s: 211    lr: 1.6000000000000003e-05    evaluation reward: 5.29
1570
episode: 1570    score: 6.0    memory length: 359033    epsilon: 0.48711268001077296    ste
ps: 370    lr: 1.6000000000000003e-05    evaluation reward: 5.3
1571
episode: 1571    score: 4.0    memory length: 359312    epsilon: 0.48656026001076946    ste
ps: 279    lr: 1.6000000000000003e-05    evaluation reward: 5.3
1572
episode: 1572    score: 4.0    memory length: 359572    epsilon: 0.4860454600107662    step
s: 260    lr: 1.6000000000000003e-05    evaluation reward: 5.3
1573
episode: 1573    score: 4.0    memory length: 359852    epsilon: 0.4854910600107627    step
s: 280    lr: 1.6000000000000003e-05    evaluation reward: 5.32
1574
episode: 1574    score: 4.0    memory length: 360109    epsilon: 0.4849822000107595    step
s: 257    lr: 1.6000000000000003e-05    evaluation reward: 5.3
1575
episode: 1575    score: 7.0    memory length: 360516    epsilon: 0.4841763400107544    step
s: 407    lr: 1.6000000000000003e-05    evaluation reward: 5.31
1576
episode: 1576    score: 8.0    memory length: 360943    epsilon: 0.48333088001074903    ste
ps: 427    lr: 1.6000000000000003e-05    evaluation reward: 5.33
1577
episode: 1577    score: 4.0    memory length: 361240    epsilon: 0.4827428200107453    step
s: 297    lr: 1.6000000000000003e-05    evaluation reward: 5.34
1578
episode: 1578    score: 3.0    memory length: 361466    epsilon: 0.4822953400107425    step
s: 226    lr: 1.6000000000000003e-05    evaluation reward: 5.31
1579
episode: 1579    score: 5.0    memory length: 361776    epsilon: 0.4816815400107386    step
s: 310    lr: 1.6000000000000003e-05    evaluation reward: 5.33
```

```
1580
episode: 1580   score: 3.0   memory length: 361989   epsilon: 0.4812598000107359   step
s: 213   lr: 1.6000000000000003e-05   evaluation reward: 5.29
1581
episode: 1581   score: 5.0   memory length: 362314   epsilon: 0.48061630001073186   ste
ps: 325   lr: 1.6000000000000003e-05   evaluation reward: 5.28
1582
episode: 1582   score: 11.0   memory length: 362866   epsilon: 0.47952334001072494   st
eps: 552   lr: 1.6000000000000003e-05   evaluation reward: 5.36
1583
episode: 1583   score: 6.0   memory length: 363242   epsilon: 0.47877886001072023   ste
ps: 376   lr: 1.6000000000000003e-05   evaluation reward: 5.36
1584
episode: 1584   score: 7.0   memory length: 363648   epsilon: 0.47797498001071514   ste
ps: 406   lr: 1.6000000000000003e-05   evaluation reward: 5.38
1585
episode: 1585   score: 6.0   memory length: 364010   epsilon: 0.4772582200107106   step
s: 362   lr: 1.6000000000000003e-05   evaluation reward: 5.41
1586
episode: 1586   score: 5.0   memory length: 364300   epsilon: 0.476684020010707   step
s: 290   lr: 1.6000000000000003e-05   evaluation reward: 5.4
1587
episode: 1587   score: 4.0   memory length: 364580   epsilon: 0.47612962001070347   ste
ps: 280   lr: 1.6000000000000003e-05   evaluation reward: 5.37
1588
episode: 1588   score: 3.0   memory length: 364827   epsilon: 0.4756405600107004   step
s: 247   lr: 1.6000000000000003e-05   evaluation reward: 5.36
1589
episode: 1589   score: 5.0   memory length: 365154   epsilon: 0.4749931000106963   step
s: 327   lr: 1.6000000000000003e-05   evaluation reward: 5.37
1590
episode: 1590   score: 5.0   memory length: 365463   epsilon: 0.4743812800106924   step
s: 309   lr: 1.6000000000000003e-05   evaluation reward: 5.39
1591
episode: 1591   score: 2.0   memory length: 365666   epsilon: 0.47397934001068986   ste
ps: 203   lr: 1.6000000000000003e-05   evaluation reward: 5.33
1592
episode: 1592   score: 3.0   memory length: 365898   epsilon: 0.47351998001068696   ste
ps: 232   lr: 1.6000000000000003e-05   evaluation reward: 5.29
1593
episode: 1593   score: 4.0   memory length: 366197   epsilon: 0.4729279600106832   step
s: 299   lr: 1.6000000000000003e-05   evaluation reward: 5.25
1594
episode: 1594   score: 3.0   memory length: 366427   epsilon: 0.47247256001068033   ste
ps: 230   lr: 1.6000000000000003e-05   evaluation reward: 5.23
1595
episode: 1595   score: 5.0   memory length: 366717   epsilon: 0.4718983600106767   step
s: 290   lr: 1.6000000000000003e-05   evaluation reward: 5.22
1596
episode: 1596   score: 6.0   memory length: 367111   epsilon: 0.47111824001067176   ste
ps: 394   lr: 1.6000000000000003e-05   evaluation reward: 5.22
1597
episode: 1597   score: 6.0   memory length: 367487   epsilon: 0.47037376001066705   ste
ps: 376   lr: 1.6000000000000003e-05   evaluation reward: 5.21
1598
episode: 1598   score: 3.0   memory length: 367717   epsilon: 0.46991836001066417   ste
ps: 230   lr: 1.6000000000000003e-05   evaluation reward: 5.16
1599
episode: 1599   score: 3.0   memory length: 367948   epsilon: 0.4694609800106613   step
s: 231   lr: 1.6000000000000003e-05   evaluation reward: 5.15
1600
episode: 1600   score: 5.0   memory length: 368240   epsilon: 0.4688828200106576   step
s: 292   lr: 1.6000000000000003e-05   evaluation reward: 5.11
1601
episode: 1601   score: 6.0   memory length: 368600   epsilon: 0.4681700200106531   step
s: 360   lr: 1.6000000000000003e-05   evaluation reward: 5.13
```

```
1602
episode: 1602   score: 5.0   memory length: 368931   epsilon: 0.46751464001064896     ste
ps: 331    lr: 1.6000000000000003e-05     evaluation reward: 5.14
1603
episode: 1603   score: 6.0   memory length: 369273   epsilon: 0.4668374800106447     step
s: 342    lr: 1.6000000000000003e-05     evaluation reward: 5.13
1604
episode: 1604   score: 10.0   memory length: 369811   epsilon: 0.46577224001063794     st
eps: 538    lr: 1.6000000000000003e-05     evaluation reward: 5.15
1605
episode: 1605   score: 6.0   memory length: 370129   epsilon: 0.46514260001063396     ste
ps: 318    lr: 1.6000000000000003e-05     evaluation reward: 5.18
1606
episode: 1606   score: 6.0   memory length: 370461   epsilon: 0.4644852400106298     step
s: 332    lr: 1.6000000000000003e-05     evaluation reward: 5.2
1607
episode: 1607   score: 9.0   memory length: 370959   epsilon: 0.46349920001062356     ste
ps: 498    lr: 1.6000000000000003e-05     evaluation reward: 5.24
1608
episode: 1608   score: 6.0   memory length: 371317   epsilon: 0.4627903600106191     step
s: 358    lr: 1.6000000000000003e-05     evaluation reward: 5.25
1609
episode: 1609   score: 7.0   memory length: 371565   epsilon: 0.462299932001061597     ste
ps: 248    lr: 1.6000000000000003e-05     evaluation reward: 5.28
1610
episode: 1610   score: 7.0   memory length: 371924   epsilon: 0.46158850001061147     ste
ps: 359    lr: 1.6000000000000003e-05     evaluation reward: 5.28
1611
episode: 1611   score: 10.0   memory length: 372483   epsilon: 0.46048168001060447     st
eps: 559    lr: 1.6000000000000003e-05     evaluation reward: 5.32
1612
episode: 1612   score: 6.0   memory length: 372829   epsilon: 0.45979660001060013     ste
ps: 346    lr: 1.6000000000000003e-05     evaluation reward: 5.3
1613
episode: 1613   score: 3.0   memory length: 373075   epsilon: 0.45930952001059705     ste
ps: 246    lr: 1.6000000000000003e-05     evaluation reward: 5.26
1614
episode: 1614   score: 5.0   memory length: 373381   epsilon: 0.4587036400105932     step
s: 306    lr: 1.6000000000000003e-05     evaluation reward: 5.23
1615
episode: 1615   score: 10.0   memory length: 373876   epsilon: 0.457723540010587     step
s: 495    lr: 1.6000000000000003e-05     evaluation reward: 5.28
1616
episode: 1616   score: 6.0   memory length: 374223   epsilon: 0.45703648001058267     ste
ps: 347    lr: 1.6000000000000003e-05     evaluation reward: 5.28
1617
episode: 1617   score: 9.0   memory length: 374716   epsilon: 0.4560603400105765     step
s: 493    lr: 1.6000000000000003e-05     evaluation reward: 5.32
1618
episode: 1618   score: 14.0   memory length: 375305   epsilon: 0.4548941200105691     ste
ps: 589    lr: 1.6000000000000003e-05     evaluation reward: 5.39
1619
episode: 1619   score: 11.0   memory length: 375810   epsilon: 0.4538942200105628     ste
ps: 505    lr: 1.6000000000000003e-05     evaluation reward: 5.46
1620
episode: 1620   score: 7.0   memory length: 376203   epsilon: 0.45311608001055786     ste
ps: 393    lr: 1.6000000000000003e-05     evaluation reward: 5.46
1621
episode: 1621   score: 7.0   memory length: 376592   epsilon: 0.452345860010553     step
s: 389    lr: 1.6000000000000003e-05     evaluation reward: 5.48
1622
episode: 1622   score: 7.0   memory length: 376979   epsilon: 0.45157960001054814     ste
ps: 387    lr: 1.6000000000000003e-05     evaluation reward: 5.51
1623
episode: 1623   score: 4.0   memory length: 377238   epsilon: 0.4510667800105449     step
s: 259    lr: 1.6000000000000003e-05     evaluation reward: 5.48
```

```
1624
episode: 1624    score: 4.0    memory length: 377533    epsilon: 0.4504826800105412    step
s: 295    lr: 1.6000000000000003e-05    evaluation reward: 5.45
1625
episode: 1625    score: 3.0    memory length: 377763    epsilon: 0.4500272800105383    step
s: 230    lr: 1.6000000000000003e-05    evaluation reward: 5.41
1626
episode: 1626    score: 4.0    memory length: 378035    epsilon: 0.4494887200105349    step
s: 272    lr: 1.6000000000000003e-05    evaluation reward: 5.41
1627
episode: 1627    score: 5.0    memory length: 378325    epsilon: 0.4489145200105313    step
s: 290    lr: 1.6000000000000003e-05    evaluation reward: 5.37
1628
episode: 1628    score: 5.0    memory length: 378619    epsilon: 0.4483324000105276    step
s: 294    lr: 1.6000000000000003e-05    evaluation reward: 5.39
1629
episode: 1629    score: 9.0    memory length: 379068    epsilon: 0.447443380010522    step
s: 449    lr: 1.6000000000000003e-05    evaluation reward: 5.45
1630
episode: 1630    score: 13.0    memory length: 379550    epsilon: 0.44648902001051594    st
eps: 482    lr: 1.6000000000000003e-05    evaluation reward: 5.55
1631
episode: 1631    score: 3.0    memory length: 379761    epsilon: 0.4460712400105133    step
s: 211    lr: 1.6000000000000003e-05    evaluation reward: 5.53
1632
episode: 1632    score: 9.0    memory length: 380187    epsilon: 0.44522776001050796    ste
ps: 426    lr: 1.6000000000000003e-05    evaluation reward: 5.6
1633
episode: 1633    score: 4.0    memory length: 380447    epsilon: 0.4447129600105047    step
s: 260    lr: 1.6000000000000003e-05    evaluation reward: 5.62
1634
episode: 1634    score: 6.0    memory length: 380786    epsilon: 0.44404174001050045    ste
ps: 339    lr: 1.6000000000000003e-05    evaluation reward: 5.65
1635
episode: 1635    score: 6.0    memory length: 381136    epsilon: 0.44334874001049607    ste
ps: 350    lr: 1.6000000000000003e-05    evaluation reward: 5.65
1636
episode: 1636    score: 4.0    memory length: 381415    epsilon: 0.4427963200104926    step
s: 279    lr: 1.6000000000000003e-05    evaluation reward: 5.62
1637
episode: 1637    score: 5.0    memory length: 381722    epsilon: 0.4421884600104887    step
s: 307    lr: 1.6000000000000003e-05    evaluation reward: 5.56
1638
episode: 1638    score: 7.0    memory length: 382079    epsilon: 0.44148160001048425    ste
ps: 357    lr: 1.6000000000000003e-05    evaluation reward: 5.58
1639
episode: 1639    score: 9.0    memory length: 382536    epsilon: 0.44057674001047853    ste
ps: 457    lr: 1.6000000000000003e-05    evaluation reward: 5.61
1640
episode: 1640    score: 5.0    memory length: 382840    epsilon: 0.4399748200104747    step
s: 304    lr: 1.6000000000000003e-05    evaluation reward: 5.62
1641
episode: 1641    score: 9.0    memory length: 383364    epsilon: 0.43893730001046816    ste
ps: 524    lr: 1.6000000000000003e-05    evaluation reward: 5.67
1642
episode: 1642    score: 7.0    memory length: 383693    epsilon: 0.43828588001046404    ste
ps: 329    lr: 1.6000000000000003e-05    evaluation reward: 5.72
1643
episode: 1643    score: 7.0    memory length: 384065    epsilon: 0.4375493200104594    step
s: 372    lr: 1.6000000000000003e-05    evaluation reward: 5.75
1644
episode: 1644    score: 8.0    memory length: 384487    epsilon: 0.4367137600104541    step
s: 422    lr: 1.6000000000000003e-05    evaluation reward: 5.77
1645
episode: 1645    score: 4.0    memory length: 384729    epsilon: 0.43623460001045106    ste
ps: 242    lr: 1.6000000000000003e-05    evaluation reward: 5.74
```

```
1646
episode: 1646   score: 1.0   memory length: 384880   epsilon: 0.43593562001044917    ste
ps: 151    lr: 1.6000000000000003e-05    evaluation reward: 5.7
1647
episode: 1647   score: 5.0   memory length: 385189   epsilon: 0.4353238000104453    step
s: 309    lr: 1.6000000000000003e-05    evaluation reward: 5.72
1648
episode: 1648   score: 4.0   memory length: 385448   epsilon: 0.43481098001044205    ste
ps: 259    lr: 1.6000000000000003e-05    evaluation reward: 5.73
1649
episode: 1649   score: 8.0   memory length: 385860   epsilon: 0.4339952200104369    step
s: 412    lr: 1.6000000000000003e-05    evaluation reward: 5.77
1650
episode: 1650   score: 6.0   memory length: 386197   epsilon: 0.43332796001043267    ste
ps: 337    lr: 1.6000000000000003e-05    evaluation reward: 5.78
1651
episode: 1651   score: 4.0   memory length: 386452   epsilon: 0.4328230600104295    step
s: 255    lr: 1.6000000000000003e-05    evaluation reward: 5.72
1652
episode: 1652   score: 7.0   memory length: 386831   epsilon: 0.4320726400104247    step
s: 379    lr: 1.6000000000000003e-05    evaluation reward: 5.75
1653
episode: 1653   score: 6.0   memory length: 387189   epsilon: 0.43136380001042024    ste
ps: 358    lr: 1.6000000000000003e-05    evaluation reward: 5.73
1654
episode: 1654   score: 7.0   memory length: 387557   epsilon: 0.43063516001041563    ste
ps: 368    lr: 1.6000000000000003e-05    evaluation reward: 5.73
1655
episode: 1655   score: 4.0   memory length: 387834   epsilon: 0.43008670001041216    ste
ps: 277    lr: 1.6000000000000003e-05    evaluation reward: 5.74
1656
episode: 1656   score: 6.0   memory length: 388225   epsilon: 0.42931252001040726    ste
ps: 391    lr: 1.6000000000000003e-05    evaluation reward: 5.74
1657
episode: 1657   score: 6.0   memory length: 388578   epsilon: 0.42861358001040284    ste
ps: 353    lr: 1.6000000000000003e-05    evaluation reward: 5.74
1658
episode: 1658   score: 8.0   memory length: 389013   epsilon: 0.4277522800103974    step
s: 435    lr: 1.6000000000000003e-05    evaluation reward: 5.77
1659
episode: 1659   score: 3.0   memory length: 389222   epsilon: 0.42733846001039477    ste
ps: 209    lr: 1.6000000000000003e-05    evaluation reward: 5.73
1660
episode: 1660   score: 5.0   memory length: 389531   epsilon: 0.4267266400103909    step
s: 309    lr: 1.6000000000000003e-05    evaluation reward: 5.74
1661
episode: 1661   score: 7.0   memory length: 389958   epsilon: 0.42588118001038555    ste
ps: 427    lr: 1.6000000000000003e-05    evaluation reward: 5.78
1662
episode: 1662   score: 15.0   memory length: 390519   epsilon: 0.4247704000103785    ste
ps: 561    lr: 1.6000000000000003e-05    evaluation reward: 5.9
1663
episode: 1663   score: 5.0   memory length: 390828   epsilon: 0.42415858001037465    ste
ps: 309    lr: 1.6000000000000003e-05    evaluation reward: 5.89
1664
episode: 1664   score: 4.0   memory length: 391088   epsilon: 0.4236437800103714    step
s: 260    lr: 1.6000000000000003e-05    evaluation reward: 5.88
1665
episode: 1665   score: 6.0   memory length: 391422   epsilon: 0.4229824600103672    step
s: 334    lr: 1.6000000000000003e-05    evaluation reward: 5.87
1666
episode: 1666   score: 4.0   memory length: 391697   epsilon: 0.42243796001036377    ste
ps: 275    lr: 1.6000000000000003e-05    evaluation reward: 5.88
1667
episode: 1667   score: 6.0   memory length: 392040   epsilon: 0.42175882001035947    ste
ps: 343    lr: 1.6000000000000003e-05    evaluation reward: 5.87
```

```
1668
episode: 1668   score: 5.0   memory length: 392367   epsilon: 0.4211113600103554    step
s: 327   lr: 1.6000000000000003e-05     evaluation reward: 5.87
1669
episode: 1669   score: 9.0   memory length: 392838   epsilon: 0.4201787800103495    step
s: 471   lr: 1.6000000000000003e-05     evaluation reward: 5.93
1670
episode: 1670   score: 5.0   memory length: 393111   epsilon: 0.41963824001034605    ste
ps: 273   lr: 1.6000000000000003e-05     evaluation reward: 5.92
1671
episode: 1671   score: 4.0   memory length: 393369   epsilon: 0.4191274000103428    step
s: 258   lr: 1.6000000000000003e-05     evaluation reward: 5.92
1672
episode: 1672   score: 5.0   memory length: 393680   epsilon: 0.4185116200103389    step
s: 311   lr: 1.6000000000000003e-05     evaluation reward: 5.93
1673
episode: 1673   score: 5.0   memory length: 394006   epsilon: 0.41786614001033484    ste
ps: 326   lr: 1.6000000000000003e-05     evaluation reward: 5.94
1674
episode: 1674   score: 5.0   memory length: 394355   epsilon: 0.41717512001033047    ste
ps: 349   lr: 1.6000000000000003e-05     evaluation reward: 5.95
1675
episode: 1675   score: 4.0   memory length: 394634   epsilon: 0.416622700010327    step
s: 279   lr: 1.6000000000000003e-05     evaluation reward: 5.92
1676
episode: 1676   score: 11.0   memory length: 395236   epsilon: 0.41543074001031943    st
eps: 602   lr: 1.6000000000000003e-05     evaluation reward: 5.95
1677
episode: 1677   score: 5.0   memory length: 395531   epsilon: 0.41484664001031574    ste
ps: 295   lr: 1.6000000000000003e-05     evaluation reward: 5.96
1678
episode: 1678   score: 7.0   memory length: 395932   epsilon: 0.4140526600103107    step
s: 401   lr: 1.6000000000000003e-05     evaluation reward: 6.0
1679
episode: 1679   score: 6.0   memory length: 396259   epsilon: 0.4134052000103066    step
s: 327   lr: 1.6000000000000003e-05     evaluation reward: 6.01
1680
episode: 1680   score: 5.0   memory length: 396583   epsilon: 0.41276368001030256    ste
ps: 324   lr: 1.6000000000000003e-05     evaluation reward: 6.03
1681
episode: 1681   score: 8.0   memory length: 397021   epsilon: 0.41189644001029707    ste
ps: 438   lr: 1.6000000000000003e-05     evaluation reward: 6.06
1682
episode: 1682   score: 8.0   memory length: 397423   epsilon: 0.41110048001029204    ste
ps: 402   lr: 1.6000000000000003e-05     evaluation reward: 6.03
1683
episode: 1683   score: 6.0   memory length: 397781   epsilon: 0.41039164001028755    ste
ps: 358   lr: 1.6000000000000003e-05     evaluation reward: 6.03
1684
episode: 1684   score: 10.0   memory length: 398291   epsilon: 0.40938184001028116    st
eps: 510   lr: 1.6000000000000003e-05     evaluation reward: 6.06
1685
episode: 1685   score: 3.0   memory length: 398501   epsilon: 0.40896604001027853    ste
ps: 210   lr: 1.6000000000000003e-05     evaluation reward: 6.03
1686
episode: 1686   score: 6.0   memory length: 398830   epsilon: 0.4083146200102744    step
s: 329   lr: 1.6000000000000003e-05     evaluation reward: 6.04
1687
episode: 1687   score: 6.0   memory length: 399144   epsilon: 0.4076929000102705    step
s: 314   lr: 1.6000000000000003e-05     evaluation reward: 6.06
1688
episode: 1688   score: 7.0   memory length: 399516   epsilon: 0.4069563400102658    step
s: 372   lr: 1.6000000000000003e-05     evaluation reward: 6.1
1689
episode: 1689   score: 5.0   memory length: 399823   epsilon: 0.40634848001026197    ste
ps: 307   lr: 1.6000000000000003e-05     evaluation reward: 6.1
```

```
1690
episode: 1690   score: 5.0   memory length: 400130   epsilon: 0.4057406200102581   step
s: 307   lr: 6.400000000000001e-06   evaluation reward: 6.1
1691
episode: 1691   score: 5.0   memory length: 400454   epsilon: 0.405099100001025407   ste
ps: 324   lr: 6.400000000000001e-06   evaluation reward: 6.13
1692
episode: 1692   score: 8.0   memory length: 400860   epsilon: 0.404295220010249   step
s: 406   lr: 6.400000000000001e-06   evaluation reward: 6.18
1693
episode: 1693   score: 3.0   memory length: 401071   epsilon: 0.40387744001024634   ste
ps: 211   lr: 6.400000000000001e-06   evaluation reward: 6.17
1694
episode: 1694   score: 10.0   memory length: 401563   epsilon: 0.4029032800102402   ste
ps: 492   lr: 6.400000000000001e-06   evaluation reward: 6.24
1695
episode: 1695   score: 6.0   memory length: 401884   epsilon: 0.40226770001023615   ste
ps: 321   lr: 6.400000000000001e-06   evaluation reward: 6.25
1696
episode: 1696   score: 6.0   memory length: 402260   epsilon: 0.40152322001023144   ste
ps: 376   lr: 6.400000000000001e-06   evaluation reward: 6.25
1697
episode: 1697   score: 5.0   memory length: 402572   epsilon: 0.40090546001022753   ste
ps: 312   lr: 6.400000000000001e-06   evaluation reward: 6.24
1698
episode: 1698   score: 7.0   memory length: 402930   epsilon: 0.40019662001022305   ste
ps: 358   lr: 6.400000000000001e-06   evaluation reward: 6.28
1699
episode: 1699   score: 5.0   memory length: 403202   epsilon: 0.399658060001021964   ste
ps: 272   lr: 6.400000000000001e-06   evaluation reward: 6.3
1700
episode: 1700   score: 6.0   memory length: 403526   epsilon: 0.3990165400102156   step
s: 324   lr: 6.400000000000001e-06   evaluation reward: 6.31
1701
episode: 1701   score: 5.0   memory length: 403832   epsilon: 0.39841066001021175   ste
ps: 306   lr: 6.400000000000001e-06   evaluation reward: 6.3
1702
episode: 1702   score: 5.0   memory length: 404155   epsilon: 0.3977711200102077   step
s: 323   lr: 6.400000000000001e-06   evaluation reward: 6.3
1703
episode: 1703   score: 7.0   memory length: 404560   epsilon: 0.39696922001020263   ste
ps: 405   lr: 6.400000000000001e-06   evaluation reward: 6.31
1704
episode: 1704   score: 9.0   memory length: 405018   epsilon: 0.3960623800101969   step
s: 458   lr: 6.400000000000001e-06   evaluation reward: 6.3
1705
episode: 1705   score: 6.0   memory length: 405339   epsilon: 0.39542680001019287   ste
ps: 321   lr: 6.400000000000001e-06   evaluation reward: 6.3
1706
episode: 1706   score: 6.0   memory length: 405667   epsilon: 0.39477736001018876   ste
ps: 328   lr: 6.400000000000001e-06   evaluation reward: 6.3
1707
episode: 1707   score: 4.0   memory length: 405946   epsilon: 0.39422494001018527   ste
ps: 279   lr: 6.400000000000001e-06   evaluation reward: 6.25
1708
episode: 1708   score: 9.0   memory length: 406379   epsilon: 0.39336760001017984   ste
ps: 433   lr: 6.400000000000001e-06   evaluation reward: 6.28
1709
episode: 1709   score: 4.0   memory length: 406636   epsilon: 0.3928587400101766   step
s: 257   lr: 6.400000000000001e-06   evaluation reward: 6.25
1710
episode: 1710   score: 6.0   memory length: 406979   epsilon: 0.3921796000101723   step
s: 343   lr: 6.400000000000001e-06   evaluation reward: 6.24
1711
episode: 1711   score: 7.0   memory length: 407372   epsilon: 0.3914014600101674   step
s: 393   lr: 6.400000000000001e-06   evaluation reward: 6.21
```

```
1712
episode: 1712   score: 5.0   memory length: 407679   epsilon: 0.39079360001016356   ste
ps: 307   lr: 6.400000000000001e-06   evaluation reward: 6.2
1713
episode: 1713   score: 8.0   memory length: 408128   epsilon: 0.38990458001015793   ste
ps: 449   lr: 6.400000000000001e-06   evaluation reward: 6.25
1714
episode: 1714   score: 6.0   memory length: 408473   epsilon: 0.3892214800101536   step
s: 345   lr: 6.400000000000001e-06   evaluation reward: 6.26
1715
episode: 1715   score: 8.0   memory length: 408887   epsilon: 0.3884017600101484   step
s: 414   lr: 6.400000000000001e-06   evaluation reward: 6.24
1716
episode: 1716   score: 4.0   memory length: 409182   epsilon: 0.3878176600101447   step
s: 295   lr: 6.400000000000001e-06   evaluation reward: 6.22
1717
episode: 1717   score: 9.0   memory length: 409657   epsilon: 0.3868771600101388   step
s: 475   lr: 6.400000000000001e-06   evaluation reward: 6.22
1718
episode: 1718   score: 5.0   memory length: 409962   epsilon: 0.38627326001013496   ste
ps: 305   lr: 6.400000000000001e-06   evaluation reward: 6.13
1719
episode: 1719   score: 4.0   memory length: 410219   epsilon: 0.38576440001013174   ste
ps: 257   lr: 6.400000000000001e-06   evaluation reward: 6.06
1720
episode: 1720   score: 6.0   memory length: 410596   epsilon: 0.385017940010127   step
s: 377   lr: 6.400000000000001e-06   evaluation reward: 6.05
1721
episode: 1721   score: 6.0   memory length: 410917   epsilon: 0.384382360010123   step
s: 321   lr: 6.400000000000001e-06   evaluation reward: 6.04
1722
episode: 1722   score: 7.0   memory length: 411316   epsilon: 0.383592340010118   step
s: 399   lr: 6.400000000000001e-06   evaluation reward: 6.04
1723
episode: 1723   score: 6.0   memory length: 411671   epsilon: 0.38288944001011355   ste
ps: 355   lr: 6.400000000000001e-06   evaluation reward: 6.06
1724
episode: 1724   score: 4.0   memory length: 411927   epsilon: 0.38238256001011034   ste
ps: 256   lr: 6.400000000000001e-06   evaluation reward: 6.06
1725
episode: 1725   score: 9.0   memory length: 412434   epsilon: 0.381378700010104   step
s: 507   lr: 6.400000000000001e-06   evaluation reward: 6.12
1726
episode: 1726   score: 8.0   memory length: 412863   epsilon: 0.3805292800100986   step
s: 429   lr: 6.400000000000001e-06   evaluation reward: 6.16
1727
episode: 1727   score: 5.0   memory length: 413158   epsilon: 0.3799451800100949   step
s: 295   lr: 6.400000000000001e-06   evaluation reward: 6.16
1728
episode: 1728   score: 5.0   memory length: 413448   epsilon: 0.3793709800100913   step
s: 290   lr: 6.400000000000001e-06   evaluation reward: 6.16
1729
episode: 1729   score: 8.0   memory length: 413830   epsilon: 0.3786146200100865   step
s: 382   lr: 6.400000000000001e-06   evaluation reward: 6.15
1730
episode: 1730   score: 6.0   memory length: 414208   epsilon: 0.37786618001008176   ste
ps: 378   lr: 6.400000000000001e-06   evaluation reward: 6.08
1731
episode: 1731   score: 5.0   memory length: 414535   epsilon: 0.37721872001007767   ste
ps: 327   lr: 6.400000000000001e-06   evaluation reward: 6.1
1732
episode: 1732   score: 9.0   memory length: 415021   epsilon: 0.3762564400100716   step
s: 486   lr: 6.400000000000001e-06   evaluation reward: 6.1
1733
episode: 1733   score: 9.0   memory length: 415468   epsilon: 0.375371380010066   step
s: 447   lr: 6.400000000000001e-06   evaluation reward: 6.15
```

```
1734
episode: 1734    score: 7.0    memory length: 415869    epsilon: 0.37457740001006096    ste
ps: 401    lr: 6.400000000000001e-06    evaluation reward: 6.16
1735
episode: 1735    score: 6.0    memory length: 416248    epsilon: 0.3738269800100562    step
s: 379    lr: 6.400000000000001e-06    evaluation reward: 6.16
1736
episode: 1736    score: 3.0    memory length: 416461    epsilon: 0.37340524001005354    ste
ps: 213    lr: 6.400000000000001e-06    evaluation reward: 6.15
1737
episode: 1737    score: 4.0    memory length: 416720    epsilon: 0.3728924200100503    step
s: 259    lr: 6.400000000000001e-06    evaluation reward: 6.14
1738
episode: 1738    score: 7.0    memory length: 417055    epsilon: 0.3722291200100461    step
s: 335    lr: 6.400000000000001e-06    evaluation reward: 6.14
1739
episode: 1739    score: 6.0    memory length: 417411    epsilon: 0.37152424001004164    ste
ps: 356    lr: 6.400000000000001e-06    evaluation reward: 6.11
1740
episode: 1740    score: 6.0    memory length: 417723    epsilon: 0.37090648001003773    ste
ps: 312    lr: 6.400000000000001e-06    evaluation reward: 6.12
1741
episode: 1741    score: 5.0    memory length: 417995    epsilon: 0.3703679200100343    step
s: 272    lr: 6.400000000000001e-06    evaluation reward: 6.08
1742
episode: 1742    score: 6.0    memory length: 418344    epsilon: 0.36967690001002995    ste
ps: 349    lr: 6.400000000000001e-06    evaluation reward: 6.07
1743
episode: 1743    score: 3.0    memory length: 418556    epsilon: 0.3692571400100273    step
s: 212    lr: 6.400000000000001e-06    evaluation reward: 6.03
1744
episode: 1744    score: 4.0    memory length: 418815    epsilon: 0.368744432001002405    ste
ps: 259    lr: 6.400000000000001e-06    evaluation reward: 5.99
1745
episode: 1745    score: 11.0    memory length: 419314    epsilon: 0.3677563000100178    ste
ps: 499    lr: 6.400000000000001e-06    evaluation reward: 6.06
1746
episode: 1746    score: 10.0    memory length: 419776    epsilon: 0.366841540010012    step
s: 462    lr: 6.400000000000001e-06    evaluation reward: 6.15
1747
episode: 1747    score: 4.0    memory length: 420035    epsilon: 0.36632872001000877    ste
ps: 259    lr: 6.400000000000001e-06    evaluation reward: 6.14
1748
episode: 1748    score: 6.0    memory length: 420394    epsilon: 0.36561790001000427    ste
ps: 359    lr: 6.400000000000001e-06    evaluation reward: 6.16
1749
episode: 1749    score: 2.0    memory length: 420573    epsilon: 0.36526348001000203    ste
ps: 179    lr: 6.400000000000001e-06    evaluation reward: 6.1
1750
episode: 1750    score: 15.0    memory length: 421083    epsilon: 0.36425368000999564    st
eps: 510    lr: 6.400000000000001e-06    evaluation reward: 6.19
1751
episode: 1751    score: 5.0    memory length: 421356    epsilon: 0.3637131400099922    step
s: 273    lr: 6.400000000000001e-06    evaluation reward: 6.2
1752
episode: 1752    score: 5.0    memory length: 421665    epsilon: 0.36310132000998835    ste
ps: 309    lr: 6.400000000000001e-06    evaluation reward: 6.18
1753
episode: 1753    score: 9.0    memory length: 422008    epsilon: 0.36242218000998405    ste
ps: 343    lr: 6.400000000000001e-06    evaluation reward: 6.21
1754
episode: 1754    score: 4.0    memory length: 422283    epsilon: 0.3618776800099806    step
s: 275    lr: 6.400000000000001e-06    evaluation reward: 6.18
1755
episode: 1755    score: 5.0    memory length: 422595    epsilon: 0.3612599200099767    step
s: 312    lr: 6.400000000000001e-06    evaluation reward: 6.19
```

```
1756
episode: 1756   score: 3.0   memory length: 422844   epsilon: 0.3607669000099736    step
s: 249   lr: 6.400000000000001e-06    evaluation reward: 6.16
1757
episode: 1757   score: 6.0   memory length: 423183   epsilon: 0.36009568000996933    ste
ps: 339   lr: 6.400000000000001e-06    evaluation reward: 6.16
1758
episode: 1758   score: 4.0   memory length: 423442   epsilon: 0.3595828600099661    step
s: 259   lr: 6.400000000000001e-06    evaluation reward: 6.12
1759
episode: 1759   score: 8.0   memory length: 423920   epsilon: 0.3586364200099601    step
s: 478   lr: 6.400000000000001e-06    evaluation reward: 6.17
1760
episode: 1760   score: 8.0   memory length: 424355   epsilon: 0.35777512000995465    ste
ps: 435   lr: 6.400000000000001e-06    evaluation reward: 6.2
1761
episode: 1761   score: 7.0   memory length: 424746   epsilon: 0.35700094000994975    ste
ps: 391   lr: 6.400000000000001e-06    evaluation reward: 6.2
1762
episode: 1762   score: 7.0   memory length: 425124   epsilon: 0.356252500009945    step
s: 378   lr: 6.400000000000001e-06    evaluation reward: 6.12
1763
episode: 1763   score: 8.0   memory length: 425485   epsilon: 0.3555377200099405    step
s: 361   lr: 6.400000000000001e-06    evaluation reward: 6.15
1764
episode: 1764   score: 4.0   memory length: 425779   epsilon: 0.3549556000099368    step
s: 294   lr: 6.400000000000001e-06    evaluation reward: 6.15
1765
episode: 1765   score: 8.0   memory length: 426184   epsilon: 0.35415370000993174    ste
ps: 405   lr: 6.400000000000001e-06    evaluation reward: 6.17
1766
episode: 1766   score: 9.0   memory length: 426691   epsilon: 0.3531498400099254    step
s: 507   lr: 6.400000000000001e-06    evaluation reward: 6.22
1767
episode: 1767   score: 7.0   memory length: 427096   epsilon: 0.3523479400099203    step
s: 405   lr: 6.400000000000001e-06    evaluation reward: 6.23
1768
episode: 1768   score: 8.0   memory length: 427531   epsilon: 0.35148664000991486    ste
ps: 435   lr: 6.400000000000001e-06    evaluation reward: 6.26
1769
episode: 1769   score: 5.0   memory length: 427827   epsilon: 0.35090056000991116    ste
ps: 296   lr: 6.400000000000001e-06    evaluation reward: 6.22
1770
episode: 1770   score: 5.0   memory length: 428149   epsilon: 0.3502630000099071    step
s: 322   lr: 6.400000000000001e-06    evaluation reward: 6.22
1771
episode: 1771   score: 8.0   memory length: 428606   epsilon: 0.3493581400099014    step
s: 457   lr: 6.400000000000001e-06    evaluation reward: 6.26
1772
episode: 1772   score: 5.0   memory length: 428897   epsilon: 0.34878196000989775    ste
ps: 291   lr: 6.400000000000001e-06    evaluation reward: 6.26
1773
episode: 1773   score: 9.0   memory length: 429369   epsilon: 0.34784740000989184    ste
ps: 472   lr: 6.400000000000001e-06    evaluation reward: 6.3
1774
episode: 1774   score: 9.0   memory length: 429787   epsilon: 0.3470197600098866    step
s: 418   lr: 6.400000000000001e-06    evaluation reward: 6.34
1775
episode: 1775   score: 7.0   memory length: 430236   epsilon: 0.346130740009881    step
s: 449   lr: 6.400000000000001e-06    evaluation reward: 6.37
1776
episode: 1776   score: 5.0   memory length: 430546   epsilon: 0.3455169400098771    step
s: 310   lr: 6.400000000000001e-06    evaluation reward: 6.31
1777
episode: 1777   score: 4.0   memory length: 430825   epsilon: 0.3449645200098736    step
s: 279   lr: 6.400000000000001e-06    evaluation reward: 6.3
```

```
1778
episode: 1778    score: 6.0    memory length: 431149    epsilon: 0.34432300000986954      ste
ps: 324    lr: 6.400000000000001e-06      evaluation reward: 6.29
1779
episode: 1779    score: 8.0    memory length: 431569    epsilon: 0.3434914000098643      step
s: 420    lr: 6.400000000000001e-06      evaluation reward: 6.31
1780
episode: 1780    score: 7.0    memory length: 431945    epsilon: 0.34274692000985957      ste
ps: 376    lr: 6.400000000000001e-06      evaluation reward: 6.33
1781
episode: 1781    score: 7.0    memory length: 432359    epsilon: 0.3419272000098544      step
s: 414    lr: 6.400000000000001e-06      evaluation reward: 6.32
1782
episode: 1782    score: 10.0    memory length: 432715    epsilon: 0.3412223200098499      ste
ps: 356    lr: 6.400000000000001e-06      evaluation reward: 6.34
1783
episode: 1783    score: 5.0    memory length: 433044    epsilon: 0.3405709000098458      step
s: 329    lr: 6.400000000000001e-06      evaluation reward: 6.33
1784
episode: 1784    score: 7.0    memory length: 433423    epsilon: 0.33982048000984105      ste
ps: 379    lr: 6.400000000000001e-06      evaluation reward: 6.3
1785
episode: 1785    score: 3.0    memory length: 433633    epsilon: 0.3394046800098384      step
s: 210    lr: 6.400000000000001e-06      evaluation reward: 6.3
1786
episode: 1786    score: 6.0    memory length: 434009    epsilon: 0.3386602000098337      step
s: 376    lr: 6.400000000000001e-06      evaluation reward: 6.3
1787
episode: 1787    score: 6.0    memory length: 434348    epsilon: 0.33798898000982946      ste
ps: 339    lr: 6.400000000000001e-06      evaluation reward: 6.3
1788
episode: 1788    score: 9.0    memory length: 434802    epsilon: 0.3370900600098238      step
s: 454    lr: 6.400000000000001e-06      evaluation reward: 6.32
1789
episode: 1789    score: 5.0    memory length: 435108    epsilon: 0.33648418000981994      ste
ps: 306    lr: 6.400000000000001e-06      evaluation reward: 6.32
1790
episode: 1790    score: 3.0    memory length: 435338    epsilon: 0.33602878000981706      ste
ps: 230    lr: 6.400000000000001e-06      evaluation reward: 6.3
1791
episode: 1791    score: 5.0    memory length: 435644    epsilon: 0.33542290000981323      ste
ps: 306    lr: 6.400000000000001e-06      evaluation reward: 6.3
1792
episode: 1792    score: 11.0    memory length: 436184    epsilon: 0.33435370000980646      st
eps: 540    lr: 6.400000000000001e-06      evaluation reward: 6.33
1793
episode: 1793    score: 9.0    memory length: 436645    epsilon: 0.3334409200098007      step
s: 461    lr: 6.400000000000001e-06      evaluation reward: 6.39
1794
episode: 1794    score: 5.0    memory length: 436949    epsilon: 0.3328390000097969      step
s: 304    lr: 6.400000000000001e-06      evaluation reward: 6.34
1795
episode: 1795    score: 9.0    memory length: 437452    epsilon: 0.3318430600097906      step
s: 503    lr: 6.400000000000001e-06      evaluation reward: 6.37
1796
episode: 1796    score: 6.0    memory length: 437824    epsilon: 0.3311065000097859      step
s: 372    lr: 6.400000000000001e-06      evaluation reward: 6.37
1797
episode: 1797    score: 7.0    memory length: 438202    epsilon: 0.3303580600097812      step
s: 378    lr: 6.400000000000001e-06      evaluation reward: 6.39
1798
episode: 1798    score: 6.0    memory length: 438521    epsilon: 0.3297264400097772      step
s: 319    lr: 6.400000000000001e-06      evaluation reward: 6.38
1799
episode: 1799    score: 4.0    memory length: 438765    epsilon: 0.32924332000977413      ste
ps: 244    lr: 6.400000000000001e-06      evaluation reward: 6.37
```

```
1800
episode: 1800    score: 7.0    memory length: 439170    epsilon: 0.32844142000976906     ste
ps: 405    lr: 6.400000000000001e-06     evaluation reward: 6.38
1801
episode: 1801    score: 4.0    memory length: 439449    epsilon: 0.32788900000976556     ste
ps: 279    lr: 6.400000000000001e-06     evaluation reward: 6.37
1802
episode: 1802    score: 7.0    memory length: 439854    epsilon: 0.3270871000097605     step
s: 405    lr: 6.400000000000001e-06     evaluation reward: 6.39
1803
episode: 1803    score: 10.0    memory length: 440391    epsilon: 0.32602384000975376     st
eps: 537    lr: 6.400000000000001e-06     evaluation reward: 6.42
1804
episode: 1804    score: 8.0    memory length: 440821    epsilon: 0.3251724400097484     step
s: 430    lr: 6.400000000000001e-06     evaluation reward: 6.41
1805
episode: 1805    score: 8.0    memory length: 441240    epsilon: 0.3243428200097431     step
s: 419    lr: 6.400000000000001e-06     evaluation reward: 6.43
1806
episode: 1806    score: 8.0    memory length: 441676    epsilon: 0.32347954000973766     ste
ps: 436    lr: 6.400000000000001e-06     evaluation reward: 6.45
1807
episode: 1807    score: 8.0    memory length: 442054    epsilon: 0.32273110000973293     ste
ps: 378    lr: 6.400000000000001e-06     evaluation reward: 6.49
1808
episode: 1808    score: 4.0    memory length: 442312    epsilon: 0.3222202600097297     step
s: 258    lr: 6.400000000000001e-06     evaluation reward: 6.44
1809
episode: 1809    score: 7.0    memory length: 442735    epsilon: 0.3213827200097244     step
s: 423    lr: 6.400000000000001e-06     evaluation reward: 6.47
1810
episode: 1810    score: 3.0    memory length: 442985    epsilon: 0.32088772000972127     ste
ps: 250    lr: 6.400000000000001e-06     evaluation reward: 6.44
1811
episode: 1811    score: 9.0    memory length: 443432    epsilon: 0.32000266000971567     ste
ps: 447    lr: 6.400000000000001e-06     evaluation reward: 6.46
1812
episode: 1812    score: 7.0    memory length: 443837    epsilon: 0.3192007600097106     step
s: 405    lr: 6.400000000000001e-06     evaluation reward: 6.48
1813
episode: 1813    score: 7.0    memory length: 444208    epsilon: 0.31846618000970595     ste
ps: 371    lr: 6.400000000000001e-06     evaluation reward: 6.47
1814
episode: 1814    score: 8.0    memory length: 444608    epsilon: 0.31767418000970093     ste
ps: 400    lr: 6.400000000000001e-06     evaluation reward: 6.49
1815
episode: 1815    score: 8.0    memory length: 445043    epsilon: 0.3168128800096955     step
s: 435    lr: 6.400000000000001e-06     evaluation reward: 6.49
1816
episode: 1816    score: 6.0    memory length: 445389    epsilon: 0.31612780000969115     ste
ps: 346    lr: 6.400000000000001e-06     evaluation reward: 6.51
1817
episode: 1817    score: 5.0    memory length: 445700    epsilon: 0.31551202000968726     ste
ps: 311    lr: 6.400000000000001e-06     evaluation reward: 6.47
1818
episode: 1818    score: 6.0    memory length: 446031    epsilon: 0.3148566400096831     step
s: 331    lr: 6.400000000000001e-06     evaluation reward: 6.48
1819
episode: 1819    score: 5.0    memory length: 446342    epsilon: 0.3142408600096792     step
s: 311    lr: 6.400000000000001e-06     evaluation reward: 6.49
1820
episode: 1820    score: 4.0    memory length: 446635    epsilon: 0.31366072000967554     ste
ps: 293    lr: 6.400000000000001e-06     evaluation reward: 6.47
1821
episode: 1821    score: 3.0    memory length: 446846    epsilon: 0.3132429400096729     step
s: 211    lr: 6.400000000000001e-06     evaluation reward: 6.44
```

```
1822
episode: 1822   score: 8.0   memory length: 447292   epsilon: 0.3123598600096673    step
s: 446    lr: 6.400000000000001e-06    evaluation reward: 6.45
1823
episode: 1823   score: 7.0   memory length: 447704   epsilon: 0.31154410000966215    ste
ps: 412    lr: 6.400000000000001e-06    evaluation reward: 6.46
1824
episode: 1824   score: 5.0   memory length: 448013   epsilon: 0.3109322800096583    step
s: 309    lr: 6.400000000000001e-06    evaluation reward: 6.47
1825
episode: 1825   score: 8.0   memory length: 448436   epsilon: 0.310094740009653    step
s: 423    lr: 6.400000000000001e-06    evaluation reward: 6.46
1826
episode: 1826   score: 11.0   memory length: 448950   epsilon: 0.30907702000964654    st
eps: 514    lr: 6.400000000000001e-06    evaluation reward: 6.49
1827
episode: 1827   score: 8.0   memory length: 449434   epsilon: 0.3081187000096405    step
s: 484    lr: 6.400000000000001e-06    evaluation reward: 6.52
1828
episode: 1828   score: 9.0   memory length: 449863   epsilon: 0.3072692800096351    step
s: 429    lr: 6.400000000000001e-06    evaluation reward: 6.56
1829
episode: 1829   score: 10.0   memory length: 450399   epsilon: 0.3062080000096284    ste
ps: 536    lr: 6.400000000000001e-06    evaluation reward: 6.58
1830
episode: 1830   score: 7.0   memory length: 450846   epsilon: 0.3053229400096228    step
s: 447    lr: 6.400000000000001e-06    evaluation reward: 6.59
1831
episode: 1831   score: 3.0   memory length: 451059   epsilon: 0.3049012000096201    step
s: 213    lr: 6.400000000000001e-06    evaluation reward: 6.57
1832
episode: 1832   score: 9.0   memory length: 451500   epsilon: 0.3040280200096146    step
s: 441    lr: 6.400000000000001e-06    evaluation reward: 6.57
1833
episode: 1833   score: 5.0   memory length: 451824   epsilon: 0.30338650000961054    ste
ps: 324    lr: 6.400000000000001e-06    evaluation reward: 6.53
1834
episode: 1834   score: 5.0   memory length: 452107   epsilon: 0.302826160009607    step
s: 283    lr: 6.400000000000001e-06    evaluation reward: 6.51
1835
episode: 1835   score: 5.0   memory length: 452403   epsilon: 0.3022400800096033    step
s: 296    lr: 6.400000000000001e-06    evaluation reward: 6.5
1836
episode: 1836   score: 8.0   memory length: 452826   epsilon: 0.301402540009598    step
s: 423    lr: 6.400000000000001e-06    evaluation reward: 6.55
1837
episode: 1837   score: 9.0   memory length: 453289   epsilon: 0.3004858000095922    step
s: 463    lr: 6.400000000000001e-06    evaluation reward: 6.6
1838
episode: 1838   score: 3.0   memory length: 453502   epsilon: 0.3000640600095895    step
s: 213    lr: 6.400000000000001e-06    evaluation reward: 6.56
1839
episode: 1839   score: 9.0   memory length: 453964   epsilon: 0.29914930000958373    ste
ps: 462    lr: 6.400000000000001e-06    evaluation reward: 6.59
1840
episode: 1840   score: 8.0   memory length: 454400   epsilon: 0.29828602000957827    ste
ps: 436    lr: 6.400000000000001e-06    evaluation reward: 6.61
1841
episode: 1841   score: 5.0   memory length: 454706   epsilon: 0.29768014000957443    ste
ps: 306    lr: 6.400000000000001e-06    evaluation reward: 6.61
1842
episode: 1842   score: 6.0   memory length: 455058   epsilon: 0.29698318000957    steps:
 352    lr: 6.400000000000001e-06    evaluation reward: 6.61
1843
episode: 1843   score: 5.0   memory length: 455365   epsilon: 0.2963753200095662    step
s: 307    lr: 6.400000000000001e-06    evaluation reward: 6.63
```

```
1844
episode: 1844   score: 8.0   memory length: 455716   epsilon: 0.2956803400095618    step
s: 351   lr: 6.400000000000001e-06    evaluation reward: 6.67
1845
episode: 1845   score: 4.0   memory length: 456015   epsilon: 0.29508832000955804   ste
ps: 299   lr: 6.400000000000001e-06    evaluation reward: 6.6
1846
episode: 1846   score: 10.0   memory length: 456423   epsilon: 0.2942804800095529   ste
ps: 408   lr: 6.400000000000001e-06    evaluation reward: 6.6
1847
episode: 1847   score: 7.0   memory length: 456785   epsilon: 0.29356372000095484   step
s: 362   lr: 6.400000000000001e-06    evaluation reward: 6.63
1848
episode: 1848   score: 6.0   memory length: 457130   epsilon: 0.29288062000954407   ste
ps: 345   lr: 6.400000000000001e-06    evaluation reward: 6.63
1849
episode: 1849   score: 10.0   memory length: 457649   epsilon: 0.29185300000953757   st
eps: 519   lr: 6.400000000000001e-06    evaluation reward: 6.71
1850
episode: 1850   score: 7.0   memory length: 458052   epsilon: 0.2910550600095325   step
s: 403   lr: 6.400000000000001e-06    evaluation reward: 6.63
1851
episode: 1851   score: 3.0   memory length: 458263   epsilon: 0.2906372800095299   step
s: 211   lr: 6.400000000000001e-06    evaluation reward: 6.61
1852
episode: 1852   score: 9.0   memory length: 458752   epsilon: 0.28966906000952375   ste
ps: 489   lr: 6.400000000000001e-06    evaluation reward: 6.65
1853
episode: 1853   score: 9.0   memory length: 459236   epsilon: 0.2887107400095177   step
s: 484   lr: 6.400000000000001e-06    evaluation reward: 6.65
1854
episode: 1854   score: 5.0   memory length: 459539   epsilon: 0.2881108000095139   step
s: 303   lr: 6.400000000000001e-06    evaluation reward: 6.66
1855
episode: 1855   score: 9.0   memory length: 460011   epsilon: 0.287176240009508   step
s: 472   lr: 6.400000000000001e-06    evaluation reward: 6.7
1856
episode: 1856   score: 4.0   memory length: 460252   epsilon: 0.28669906000950496   ste
ps: 241   lr: 6.400000000000001e-06    evaluation reward: 6.71
1857
episode: 1857   score: 6.0   memory length: 460629   epsilon: 0.28595260000950024   ste
ps: 377   lr: 6.400000000000001e-06    evaluation reward: 6.71
1858
episode: 1858   score: 7.0   memory length: 460999   epsilon: 0.2852200000094956   step
s: 370   lr: 6.400000000000001e-06    evaluation reward: 6.74
1859
episode: 1859   score: 9.0   memory length: 461458   epsilon: 0.28431118000948985   ste
ps: 459   lr: 6.400000000000001e-06    evaluation reward: 6.75
1860
episode: 1860   score: 5.0   memory length: 461749   epsilon: 0.2837350000094862   step
s: 291   lr: 6.400000000000001e-06    evaluation reward: 6.72
1861
episode: 1861   score: 7.0   memory length: 462154   epsilon: 0.28293310000948113   ste
ps: 405   lr: 6.400000000000001e-06    evaluation reward: 6.72
1862
episode: 1862   score: 7.0   memory length: 462513   epsilon: 0.28222228000947663   ste
ps: 359   lr: 6.400000000000001e-06    evaluation reward: 6.72
1863
episode: 1863   score: 9.0   memory length: 462975   epsilon: 0.28130752000947085   ste
ps: 462   lr: 6.400000000000001e-06    evaluation reward: 6.73
1864
episode: 1864   score: 11.0   memory length: 463531   epsilon: 0.2802066400094639   ste
ps: 556   lr: 6.400000000000001e-06    evaluation reward: 6.8
1865
episode: 1865   score: 7.0   memory length: 463899   epsilon: 0.27947800000945927   ste
ps: 368   lr: 6.400000000000001e-06    evaluation reward: 6.79
```

```
1866
episode: 1866   score: 4.0   memory length: 464192   epsilon: 0.2788978600094556    step
s: 293   lr: 6.400000000000001e-06     evaluation reward: 6.74
1867
episode: 1867   score: 10.0   memory length: 464670   epsilon: 0.2779514200094496    ste
ps: 478   lr: 6.400000000000001e-06     evaluation reward: 6.77
1868
episode: 1868   score: 7.0   memory length: 465045   epsilon: 0.2772089200094449    step
s: 375   lr: 6.400000000000001e-06     evaluation reward: 6.76
1869
episode: 1869   score: 6.0   memory length: 465436   epsilon: 0.27643474000944    steps:
 391   lr: 6.400000000000001e-06     evaluation reward: 6.77
1870
episode: 1870   score: 5.0   memory length: 465747   epsilon: 0.2758189600094361    step
s: 311   lr: 6.400000000000001e-06     evaluation reward: 6.77
1871
episode: 1871   score: 5.0   memory length: 466077   epsilon: 0.275165560009432    step
s: 330   lr: 6.400000000000001e-06     evaluation reward: 6.74
1872
episode: 1872   score: 8.0   memory length: 466495   epsilon: 0.27433792000942675    ste
ps: 418   lr: 6.400000000000001e-06     evaluation reward: 6.77
1873
episode: 1873   score: 5.0   memory length: 466838   epsilon: 0.27365878000942245    ste
ps: 343   lr: 6.400000000000001e-06     evaluation reward: 6.73
1874
episode: 1874   score: 2.0   memory length: 467019   epsilon: 0.2733004000094202    step
s: 181   lr: 6.400000000000001e-06     evaluation reward: 6.66
1875
episode: 1875   score: 10.0   memory length: 467503   epsilon: 0.2723420800094141    ste
ps: 484   lr: 6.400000000000001e-06     evaluation reward: 6.69
1876
episode: 1876   score: 10.0   memory length: 467979   epsilon: 0.27139960000940816    st
eps: 476   lr: 6.400000000000001e-06     evaluation reward: 6.74
1877
episode: 1877   score: 9.0   memory length: 468454   epsilon: 0.2704591000094022    step
s: 475   lr: 6.400000000000001e-06     evaluation reward: 6.79
1878
episode: 1878   score: 6.0   memory length: 468812   epsilon: 0.2697502600093977    step
s: 358   lr: 6.400000000000001e-06     evaluation reward: 6.79
1879
episode: 1879   score: 7.0   memory length: 469225   epsilon: 0.26893252000939255    ste
ps: 413   lr: 6.400000000000001e-06     evaluation reward: 6.78
1880
episode: 1880   score: 7.0   memory length: 469666   epsilon: 0.268059340009387    step
s: 441   lr: 6.400000000000001e-06     evaluation reward: 6.78
1881
episode: 1881   score: 5.0   memory length: 469990   epsilon: 0.26741782000938297    ste
ps: 324   lr: 6.400000000000001e-06     evaluation reward: 6.76
1882
episode: 1882   score: 10.0   memory length: 470425   epsilon: 0.2665565200093775    ste
ps: 435   lr: 6.400000000000001e-06     evaluation reward: 6.76
1883
episode: 1883   score: 10.0   memory length: 470919   epsilon: 0.26557840000937133    st
eps: 494   lr: 6.400000000000001e-06     evaluation reward: 6.81
1884
episode: 1884   score: 9.0   memory length: 471350   epsilon: 0.26472502000936593    ste
ps: 431   lr: 6.400000000000001e-06     evaluation reward: 6.83
1885
episode: 1885   score: 13.0   memory length: 471961   epsilon: 0.2635152400093583    ste
ps: 611   lr: 6.400000000000001e-06     evaluation reward: 6.93
1886
episode: 1886   score: 6.0   memory length: 472307   epsilon: 0.26283016000935394    ste
ps: 346   lr: 6.400000000000001e-06     evaluation reward: 6.93
1887
episode: 1887   score: 8.0   memory length: 472725   epsilon: 0.2620025200093487    step
s: 418   lr: 6.400000000000001e-06     evaluation reward: 6.95
```

```
1888
episode: 1888   score: 11.0   memory length: 473296   epsilon: 0.26087194000934155   st
eps: 571   lr: 6.400000000000001e-06   evaluation reward: 6.97
1889
episode: 1889   score: 9.0   memory length: 473790   epsilon: 0.25989382000933536   ste
ps: 494   lr: 6.400000000000001e-06   evaluation reward: 7.01
1890
episode: 1890   score: 6.0   memory length: 474167   epsilon: 0.25914736000933064   ste
ps: 377   lr: 6.400000000000001e-06   evaluation reward: 7.04
1891
episode: 1891   score: 6.0   memory length: 474539   epsilon: 0.258410800009326   step
s: 372   lr: 6.400000000000001e-06   evaluation reward: 7.05
1892
episode: 1892   score: 8.0   memory length: 474938   epsilon: 0.257620780009321   step
s: 399   lr: 6.400000000000001e-06   evaluation reward: 7.02
1893
episode: 1893   score: 8.0   memory length: 475369   epsilon: 0.2567674000093156   step
s: 431   lr: 6.400000000000001e-06   evaluation reward: 7.01
1894
episode: 1894   score: 10.0   memory length: 475960   epsilon: 0.2555972200093082   ste
ps: 591   lr: 6.400000000000001e-06   evaluation reward: 7.06
1895
episode: 1895   score: 6.0   memory length: 476286   epsilon: 0.2549517400093041   step
s: 326   lr: 6.400000000000001e-06   evaluation reward: 7.03
1896
episode: 1896   score: 7.0   memory length: 476661   epsilon: 0.2542092400092994   step
s: 375   lr: 6.400000000000001e-06   evaluation reward: 7.04
1897
episode: 1897   score: 6.0   memory length: 477060   epsilon: 0.2534192200092944   step
s: 399   lr: 6.400000000000001e-06   evaluation reward: 7.03
1898
episode: 1898   score: 7.0   memory length: 477464   epsilon: 0.25261930000928934   ste
ps: 404   lr: 6.400000000000001e-06   evaluation reward: 7.04
1899
episode: 1899   score: 9.0   memory length: 477921   epsilon: 0.2517144400092836   step
s: 457   lr: 6.400000000000001e-06   evaluation reward: 7.09
1900
episode: 1900   score: 12.0   memory length: 478372   epsilon: 0.25082146000927796   st
eps: 451   lr: 6.400000000000001e-06   evaluation reward: 7.14
1901
episode: 1901   score: 10.0   memory length: 478780   epsilon: 0.25001362000927285   st
eps: 408   lr: 6.400000000000001e-06   evaluation reward: 7.2
1902
episode: 1902   score: 6.0   memory length: 479135   epsilon: 0.2493107200092684   step
s: 355   lr: 6.400000000000001e-06   evaluation reward: 7.19
1903
episode: 1903   score: 3.0   memory length: 479365   epsilon: 0.24885532000926552   ste
ps: 230   lr: 6.400000000000001e-06   evaluation reward: 7.12
1904
episode: 1904   score: 7.0   memory length: 479743   epsilon: 0.2481068800092608   step
s: 378   lr: 6.400000000000001e-06   evaluation reward: 7.11
1905
episode: 1905   score: 5.0   memory length: 480067   epsilon: 0.24746536000925673   ste
ps: 324   lr: 6.400000000000001e-06   evaluation reward: 7.08
1906
episode: 1906   score: 9.0   memory length: 480496   epsilon: 0.24661594000925136   ste
ps: 429   lr: 6.400000000000001e-06   evaluation reward: 7.09
1907
episode: 1907   score: 5.0   memory length: 480825   epsilon: 0.24596452000924723   ste
ps: 329   lr: 6.400000000000001e-06   evaluation reward: 7.06
1908
episode: 1908   score: 8.0   memory length: 481242   epsilon: 0.245138860009242   step
s: 417   lr: 6.400000000000001e-06   evaluation reward: 7.1
1909
episode: 1909   score: 7.0   memory length: 481671   epsilon: 0.24428944000923664   ste
ps: 429   lr: 6.400000000000001e-06   evaluation reward: 7.1
```

```
1910
episode: 1910   score: 4.0   memory length: 481932   epsilon: 0.24377266000923337   ste
ps: 261   lr: 6.400000000000001e-06   evaluation reward: 7.11
1911
episode: 1911   score: 7.0   memory length: 482323   epsilon: 0.24299848000922847   ste
ps: 391   lr: 6.400000000000001e-06   evaluation reward: 7.09
1912
episode: 1912   score: 9.0   memory length: 482632   epsilon: 0.2423866600092246   step
s: 309   lr: 6.400000000000001e-06   evaluation reward: 7.11
1913
episode: 1913   score: 9.0   memory length: 483162   epsilon: 0.24133726000921796   ste
ps: 530   lr: 6.400000000000001e-06   evaluation reward: 7.13
1914
episode: 1914   score: 3.0   memory length: 483372   epsilon: 0.24092146000921533   ste
ps: 210   lr: 6.400000000000001e-06   evaluation reward: 7.08
1915
episode: 1915   score: 7.0   memory length: 483800   epsilon: 0.24007402000920997   ste
ps: 428   lr: 6.400000000000001e-06   evaluation reward: 7.07
1916
episode: 1916   score: 6.0   memory length: 484177   epsilon: 0.23932756000920524   ste
ps: 377   lr: 6.400000000000001e-06   evaluation reward: 7.07
1917
episode: 1917   score: 7.0   memory length: 484554   epsilon: 0.23858110000920052   ste
ps: 377   lr: 6.400000000000001e-06   evaluation reward: 7.09
1918
episode: 1918   score: 9.0   memory length: 484993   epsilon: 0.23771188000919502   ste
ps: 439   lr: 6.400000000000001e-06   evaluation reward: 7.12
1919
episode: 1919   score: 8.0   memory length: 485392   epsilon: 0.23692186000919002   ste
ps: 399   lr: 6.400000000000001e-06   evaluation reward: 7.15
1920
episode: 1920   score: 9.0   memory length: 485900   epsilon: 0.23591602000918366   ste
ps: 508   lr: 6.400000000000001e-06   evaluation reward: 7.2
1921
episode: 1921   score: 7.0   memory length: 486290   epsilon: 0.23514382000917877   ste
ps: 390   lr: 6.400000000000001e-06   evaluation reward: 7.24
1922
episode: 1922   score: 12.0   memory length: 486818   epsilon: 0.23409838000917216   st
eps: 528   lr: 6.400000000000001e-06   evaluation reward: 7.28
1923
episode: 1923   score: 9.0   memory length: 487230   epsilon: 0.233282620009167   step
s: 412   lr: 6.400000000000001e-06   evaluation reward: 7.3
1924
episode: 1924   score: 5.0   memory length: 487521   epsilon: 0.23270644000916335   ste
ps: 291   lr: 6.400000000000001e-06   evaluation reward: 7.3
1925
episode: 1925   score: 8.0   memory length: 487996   epsilon: 0.2317659400091574   step
s: 475   lr: 6.400000000000001e-06   evaluation reward: 7.3
1926
episode: 1926   score: 8.0   memory length: 488450   epsilon: 0.2308670200091517   step
s: 454   lr: 6.400000000000001e-06   evaluation reward: 7.27
1927
episode: 1927   score: 4.0   memory length: 488689   epsilon: 0.23039380000914872   ste
ps: 239   lr: 6.400000000000001e-06   evaluation reward: 7.23
1928
episode: 1928   score: 8.0   memory length: 489104   epsilon: 0.22957210000914352   ste
ps: 415   lr: 6.400000000000001e-06   evaluation reward: 7.22
1929
episode: 1929   score: 8.0   memory length: 489491   epsilon: 0.22880584000913867   ste
ps: 387   lr: 6.400000000000001e-06   evaluation reward: 7.2
1930
episode: 1930   score: 8.0   memory length: 489914   epsilon: 0.22796830000913337   ste
ps: 423   lr: 6.400000000000001e-06   evaluation reward: 7.21
1931
episode: 1931   score: 9.0   memory length: 490378   epsilon: 0.22704958000912756   ste
ps: 464   lr: 6.400000000000001e-06   evaluation reward: 7.27
```

```
1932
episode: 1932   score: 7.0   memory length: 490768   epsilon: 0.22627738000912268      ste
ps: 390    lr: 6.400000000000001e-06      evaluation reward: 7.25
1933
episode: 1933   score: 3.0   memory length: 490978   epsilon: 0.22586158000912004      ste
ps: 210    lr: 6.400000000000001e-06      evaluation reward: 7.23
1934
episode: 1934   score: 13.0   memory length: 491313   epsilon: 0.22519828000911585      st
eps: 335    lr: 6.400000000000001e-06      evaluation reward: 7.31
1935
episode: 1935   score: 13.0   memory length: 491972   epsilon: 0.2238934600091076      ste
ps: 659    lr: 6.400000000000001e-06      evaluation reward: 7.39
1936
episode: 1936   score: 12.0   memory length: 492577   epsilon: 0.2226955600091    steps:
 605    lr: 6.400000000000001e-06      evaluation reward: 7.43
1937
episode: 1937   score: 10.0   memory length: 493051   epsilon: 0.22175704000909408      st
eps: 474    lr: 6.400000000000001e-06      evaluation reward: 7.44
1938
episode: 1938   score: 6.0   memory length: 493411   epsilon: 0.22104424000908957      ste
ps: 360    lr: 6.400000000000001e-06      evaluation reward: 7.47
1939
episode: 1939   score: 9.0   memory length: 493928   epsilon: 0.2200205800090831    step
s: 517    lr: 6.400000000000001e-06      evaluation reward: 7.47
1940
episode: 1940   score: 7.0   memory length: 494319   epsilon: 0.2192464000090782    step
s: 391    lr: 6.400000000000001e-06      evaluation reward: 7.46
1941
episode: 1941   score: 7.0   memory length: 494710   epsilon: 0.2184722200090733    step
s: 391    lr: 6.400000000000001e-06      evaluation reward: 7.48
1942
episode: 1942   score: 12.0   memory length: 495278   epsilon: 0.21734758000906618      st
eps: 568    lr: 6.400000000000001e-06      evaluation reward: 7.54
1943
episode: 1943   score: 10.0   memory length: 495794   epsilon: 0.2163259000090597      ste
ps: 516    lr: 6.400000000000001e-06      evaluation reward: 7.59
1944
episode: 1944   score: 8.0   memory length: 496205   epsilon: 0.21551212000905456      ste
ps: 411    lr: 6.400000000000001e-06      evaluation reward: 7.59
1945
episode: 1945   score: 5.0   memory length: 496502   epsilon: 0.21492406000905084      ste
ps: 297    lr: 6.400000000000001e-06      evaluation reward: 7.6
1946
episode: 1946   score: 12.0   memory length: 496981   epsilon: 0.21397564000904484      st
eps: 479    lr: 6.400000000000001e-06      evaluation reward: 7.62
1947
episode: 1947   score: 4.0   memory length: 497262   epsilon: 0.21341926000904132      ste
ps: 281    lr: 6.400000000000001e-06      evaluation reward: 7.59
1948
episode: 1948   score: 7.0   memory length: 497637   epsilon: 0.21267676000903663      ste
ps: 375    lr: 6.400000000000001e-06      evaluation reward: 7.6
1949
episode: 1949   score: 4.0   memory length: 497899   epsilon: 0.21215800000903334      ste
ps: 262    lr: 6.400000000000001e-06      evaluation reward: 7.54
1950
episode: 1950   score: 6.0   memory length: 498234   epsilon: 0.21149470000902915      ste
ps: 335    lr: 6.400000000000001e-06      evaluation reward: 7.53
1951
episode: 1951   score: 8.0   memory length: 498660   epsilon: 0.2106512200090238    step
s: 426    lr: 6.400000000000001e-06      evaluation reward: 7.58
1952
episode: 1952   score: 11.0   memory length: 499071   epsilon: 0.20983744000901866      st
eps: 411    lr: 6.400000000000001e-06      evaluation reward: 7.6
1953
episode: 1953   score: 4.0   memory length: 499332   epsilon: 0.2093206600090154    step
s: 261    lr: 6.400000000000001e-06      evaluation reward: 7.55
```

```
1954
episode: 1954   score: 5.0   memory length: 499625   epsilon: 0.20874052000901172     ste
ps: 293    lr: 6.400000000000001e-06     evaluation reward: 7.55
1955
episode: 1955   score: 6.0   memory length: 499946   epsilon: 0.2081049400090077     step
s: 321    lr: 6.400000000000001e-06     evaluation reward: 7.52
1956
episode: 1956   score: 6.0   memory length: 500256   epsilon: 0.20749114000900382     ste
ps: 310    lr: 2.560000000000001e-06     evaluation reward: 7.54
1957
episode: 1957   score: 8.0   memory length: 500701   epsilon: 0.20661004000899824     ste
ps: 445    lr: 2.560000000000001e-06     evaluation reward: 7.56
1958
episode: 1958   score: 8.0   memory length: 501151   epsilon: 0.2057190400089926     step
s: 450    lr: 2.560000000000001e-06     evaluation reward: 7.57
1959
episode: 1959   score: 11.0   memory length: 501620   epsilon: 0.20479042000898673     st
eps: 469    lr: 2.560000000000001e-06     evaluation reward: 7.59
1960
episode: 1960   score: 7.0   memory length: 502002   epsilon: 0.20403406000898194     ste
ps: 382    lr: 2.560000000000001e-06     evaluation reward: 7.61
1961
episode: 1961   score: 9.0   memory length: 502469   epsilon: 0.2031094000089761     step
s: 467    lr: 2.560000000000001e-06     evaluation reward: 7.63
1962
episode: 1962   score: 9.0   memory length: 502903   epsilon: 0.20225008000897066     ste
ps: 434    lr: 2.560000000000001e-06     evaluation reward: 7.65
1963
episode: 1963   score: 10.0   memory length: 503370   epsilon: 0.2013254200089648     ste
ps: 467    lr: 2.560000000000001e-06     evaluation reward: 7.66
1964
episode: 1964   score: 7.0   memory length: 503776   epsilon: 0.20052154000895972     ste
ps: 406    lr: 2.560000000000001e-06     evaluation reward: 7.62
1965
episode: 1965   score: 6.0   memory length: 504156   epsilon: 0.19976914000895496     ste
ps: 380    lr: 2.560000000000001e-06     evaluation reward: 7.61
1966
episode: 1966   score: 8.0   memory length: 504593   epsilon: 0.19890388000894949     ste
ps: 437    lr: 2.560000000000001e-06     evaluation reward: 7.65
1967
episode: 1967   score: 11.0   memory length: 505158   epsilon: 0.1977851800089424     ste
ps: 565    lr: 2.560000000000001e-06     evaluation reward: 7.66
1968
episode: 1968   score: 4.0   memory length: 505452   epsilon: 0.19720306000893872     ste
ps: 294    lr: 2.560000000000001e-06     evaluation reward: 7.63
1969
episode: 1969   score: 12.0   memory length: 505866   epsilon: 0.19638334000893354     st
eps: 414    lr: 2.560000000000001e-06     evaluation reward: 7.69
1970
episode: 1970   score: 3.0   memory length: 506116   epsilon: 0.1958883400089304     step
s: 250    lr: 2.560000000000001e-06     evaluation reward: 7.67
1971
episode: 1971   score: 5.0   memory length: 506408   epsilon: 0.19531018000892675     ste
ps: 292    lr: 2.560000000000001e-06     evaluation reward: 7.67
1972
episode: 1972   score: 6.0   memory length: 506749   epsilon: 0.19463500000892248     ste
ps: 341    lr: 2.560000000000001e-06     evaluation reward: 7.65
1973
episode: 1973   score: 5.0   memory length: 507056   epsilon: 0.19402714000891863     ste
ps: 307    lr: 2.560000000000001e-06     evaluation reward: 7.65
1974
episode: 1974   score: 7.0   memory length: 507445   epsilon: 0.19325692000891376     ste
ps: 389    lr: 2.560000000000001e-06     evaluation reward: 7.7
1975
episode: 1975   score: 6.0   memory length: 507768   epsilon: 0.1926173800089097     step
s: 323    lr: 2.560000000000001e-06     evaluation reward: 7.66
```

```
1976
episode: 1976   score: 14.0   memory length: 508257   epsilon: 0.19164916000890359    st
eps: 489    lr: 2.560000000000001e-06    evaluation reward: 7.7
1977
episode: 1977   score: 4.0   memory length: 508538   epsilon: 0.19109278000890006    ste
ps: 281    lr: 2.560000000000001e-06    evaluation reward: 7.65
1978
episode: 1978   score: 12.0   memory length: 509134   epsilon: 0.1899127000088926    ste
ps: 596    lr: 2.560000000000001e-06    evaluation reward: 7.71
1979
episode: 1979   score: 5.0   memory length: 509463   epsilon: 0.18926128000888848    ste
ps: 329    lr: 2.560000000000001e-06    evaluation reward: 7.69
1980
episode: 1980   score: 12.0   memory length: 510080   epsilon: 0.18803962000888075    st
eps: 617    lr: 2.560000000000001e-06    evaluation reward: 7.74
1981
episode: 1981   score: 6.0   memory length: 510438   epsilon: 0.18733078000887626    ste
ps: 358    lr: 2.560000000000001e-06    evaluation reward: 7.75
1982
episode: 1982   score: 6.0   memory length: 510784   epsilon: 0.18664570000887193    ste
ps: 346    lr: 2.560000000000001e-06    evaluation reward: 7.71
1983
episode: 1983   score: 10.0   memory length: 511333   epsilon: 0.18555868000886505    st
eps: 549    lr: 2.560000000000001e-06    evaluation reward: 7.71
1984
episode: 1984   score: 7.0   memory length: 511756   epsilon: 0.18472114000885975    ste
ps: 423    lr: 2.560000000000001e-06    evaluation reward: 7.69
1985
episode: 1985   score: 9.0   memory length: 512255   epsilon: 0.1837331200088535    step
s: 499    lr: 2.560000000000001e-06    evaluation reward: 7.65
1986
episode: 1986   score: 7.0   memory length: 512663   epsilon: 0.1829252800088484    step
s: 408    lr: 2.560000000000001e-06    evaluation reward: 7.66
1987
episode: 1987   score: 5.0   memory length: 512970   epsilon: 0.18231742000884454    ste
ps: 307    lr: 2.560000000000001e-06    evaluation reward: 7.63
1988
episode: 1988   score: 10.0   memory length: 513510   epsilon: 0.18124822000883778    st
eps: 540    lr: 2.560000000000001e-06    evaluation reward: 7.62
1989
episode: 1989   score: 5.0   memory length: 513801   epsilon: 0.18067204000883413    ste
ps: 291    lr: 2.560000000000001e-06    evaluation reward: 7.58
1990
episode: 1990   score: 4.0   memory length: 514082   epsilon: 0.1801156600088306    step
s: 281    lr: 2.560000000000001e-06    evaluation reward: 7.56
1991
episode: 1991   score: 6.0   memory length: 514427   epsilon: 0.1794325600088263    step
s: 345    lr: 2.560000000000001e-06    evaluation reward: 7.56
1992
episode: 1992   score: 14.0   memory length: 514883   epsilon: 0.17852968000882058    st
eps: 456    lr: 2.560000000000001e-06    evaluation reward: 7.62
1993
episode: 1993   score: 4.0   memory length: 515165   epsilon: 0.17797132000881705    ste
ps: 282    lr: 2.560000000000001e-06    evaluation reward: 7.58
1994
episode: 1994   score: 11.0   memory length: 515700   epsilon: 0.17691202000881034    st
eps: 535    lr: 2.560000000000001e-06    evaluation reward: 7.59
1995
episode: 1995   score: 6.0   memory length: 516040   epsilon: 0.17623882000880609    ste
ps: 340    lr: 2.560000000000001e-06    evaluation reward: 7.59
1996
episode: 1996   score: 8.0   memory length: 516438   epsilon: 0.1754507800088011    step
s: 398    lr: 2.560000000000001e-06    evaluation reward: 7.6
1997
episode: 1997   score: 13.0   memory length: 517051   epsilon: 0.17423704000879342    st
eps: 613    lr: 2.560000000000001e-06    evaluation reward: 7.67
```

```
1998
episode: 1998   score: 6.0   memory length: 517368   epsilon: 0.17360938000878945   ste
ps: 317   lr: 2.560000000000001e-06   evaluation reward: 7.66
1999
episode: 1999   score: 8.0   memory length: 517792   epsilon: 0.17276986000878414   ste
ps: 424   lr: 2.560000000000001e-06   evaluation reward: 7.65
2000
episode: 2000   score: 5.0   memory length: 518088   epsilon: 0.17218378000878043   ste
ps: 296   lr: 2.560000000000001e-06   evaluation reward: 7.58
2001
episode: 2001   score: 10.0   memory length: 518633   epsilon: 0.1711046800087736   ste
ps: 545   lr: 2.560000000000001e-06   evaluation reward: 7.58
2002
episode: 2002   score: 6.0   memory length: 518955   epsilon: 0.17046712000876957   ste
ps: 322   lr: 2.560000000000001e-06   evaluation reward: 7.58
2003
episode: 2003   score: 13.0   memory length: 519549   epsilon: 0.16929100000876213   st
eps: 594   lr: 2.560000000000001e-06   evaluation reward: 7.68
2004
episode: 2004   score: 8.0   memory length: 519994   epsilon: 0.16840990000875655   ste
ps: 445   lr: 2.560000000000001e-06   evaluation reward: 7.69
2005
episode: 2005   score: 7.0   memory length: 520399   epsilon: 0.16760800000875148   ste
ps: 405   lr: 2.560000000000001e-06   evaluation reward: 7.71
2006
episode: 2006   score: 11.0   memory length: 520942   epsilon: 0.16653286000874468   st
eps: 543   lr: 2.560000000000001e-06   evaluation reward: 7.73
2007
episode: 2007   score: 8.0   memory length: 521418   epsilon: 0.1655903800087387   step
s: 476   lr: 2.560000000000001e-06   evaluation reward: 7.76
2008
episode: 2008   score: 6.0   memory length: 521755   epsilon: 0.1649231200087345   step
s: 337   lr: 2.560000000000001e-06   evaluation reward: 7.74
2009
episode: 2009   score: 8.0   memory length: 522154   epsilon: 0.1641331000087295   step
s: 399   lr: 2.560000000000001e-06   evaluation reward: 7.75
2010
episode: 2010   score: 12.0   memory length: 522708   epsilon: 0.16303618000872255   st
eps: 554   lr: 2.560000000000001e-06   evaluation reward: 7.83
2011
episode: 2011   score: 18.0   memory length: 523252   epsilon: 0.16195906000871574   st
eps: 544   lr: 2.560000000000001e-06   evaluation reward: 7.94
2012
episode: 2012   score: 11.0   memory length: 523730   epsilon: 0.16101262000870975   st
eps: 478   lr: 2.560000000000001e-06   evaluation reward: 7.96
2013
episode: 2013   score: 6.0   memory length: 524052   epsilon: 0.16037506000870572   ste
ps: 322   lr: 2.560000000000001e-06   evaluation reward: 7.93
2014
episode: 2014   score: 11.0   memory length: 524557   epsilon: 0.1593751600086994   ste
ps: 505   lr: 2.560000000000001e-06   evaluation reward: 8.01
2015
episode: 2015   score: 9.0   memory length: 525033   epsilon: 0.15843268000869343   ste
ps: 476   lr: 2.560000000000001e-06   evaluation reward: 8.03
2016
episode: 2016   score: 13.0   memory length: 525543   epsilon: 0.15742288000868704   st
eps: 510   lr: 2.560000000000001e-06   evaluation reward: 8.1
2017
episode: 2017   score: 11.0   memory length: 526055   epsilon: 0.15640912000868062   st
eps: 512   lr: 2.560000000000001e-06   evaluation reward: 8.14
2018
episode: 2018   score: 6.0   memory length: 526418   epsilon: 0.15569038000867608   ste
ps: 363   lr: 2.560000000000001e-06   evaluation reward: 8.11
2019
episode: 2019   score: 12.0   memory length: 526855   epsilon: 0.1548251200086706   ste
ps: 437   lr: 2.560000000000001e-06   evaluation reward: 8.15
```

```
2020
episode: 2020   score: 8.0    memory length: 527308   epsilon: 0.15392818000866493    ste
ps: 453    lr: 2.560000000000001e-06    evaluation reward: 8.14
2021
episode: 2021   score: 13.0    memory length: 527920   epsilon: 0.15271642000865726    st
eps: 612    lr: 2.560000000000001e-06    evaluation reward: 8.2
2022
episode: 2022   score: 8.0    memory length: 528349   epsilon: 0.1518670000086519    step
s: 429    lr: 2.560000000000001e-06    evaluation reward: 8.16
2023
episode: 2023   score: 6.0    memory length: 528701   epsilon: 0.15117004000864748    ste
ps: 352    lr: 2.560000000000001e-06    evaluation reward: 8.13
2024
episode: 2024   score: 8.0    memory length: 529124   epsilon: 0.15033250000864218    ste
ps: 423    lr: 2.560000000000001e-06    evaluation reward: 8.16
2025
episode: 2025   score: 15.0    memory length: 529686   epsilon: 0.14921974000863514    st
eps: 562    lr: 2.560000000000001e-06    evaluation reward: 8.23
2026
episode: 2026   score: 8.0    memory length: 530090   epsilon: 0.14841982000863008    ste
ps: 404    lr: 2.560000000000001e-06    evaluation reward: 8.23
2027
episode: 2027   score: 7.0    memory length: 530511   epsilon: 0.1475862400086248    step
s: 421    lr: 2.560000000000001e-06    evaluation reward: 8.26
2028
episode: 2028   score: 9.0    memory length: 530964   epsilon: 0.14668930000861913    ste
ps: 453    lr: 2.560000000000001e-06    evaluation reward: 8.27
2029
episode: 2029   score: 10.0    memory length: 531457   epsilon: 0.14571316000861295    st
eps: 493    lr: 2.560000000000001e-06    evaluation reward: 8.29
2030
episode: 2030   score: 9.0    memory length: 531891   epsilon: 0.14485384000860752    ste
ps: 434    lr: 2.560000000000001e-06    evaluation reward: 8.3
2031
episode: 2031   score: 14.0    memory length: 532459   epsilon: 0.1437292000086004    ste
ps: 568    lr: 2.560000000000001e-06    evaluation reward: 8.35
2032
episode: 2032   score: 8.0    memory length: 532845   epsilon: 0.14296492000859556    ste
ps: 386    lr: 2.560000000000001e-06    evaluation reward: 8.36
2033
episode: 2033   score: 7.0    memory length: 533271   epsilon: 0.14212144000859023    ste
ps: 426    lr: 2.560000000000001e-06    evaluation reward: 8.4
2034
episode: 2034   score: 10.0    memory length: 533713   epsilon: 0.1412462800085847    ste
ps: 442    lr: 2.560000000000001e-06    evaluation reward: 8.37
2035
episode: 2035   score: 6.0    memory length: 534060   epsilon: 0.14055922000858034    ste
ps: 347    lr: 2.560000000000001e-06    evaluation reward: 8.3
2036
episode: 2036   score: 4.0    memory length: 534360   epsilon: 0.13996522000857659    ste
ps: 300    lr: 2.560000000000001e-06    evaluation reward: 8.22
2037
episode: 2037   score: 5.0    memory length: 534653   epsilon: 0.13938508000857291    ste
ps: 293    lr: 2.560000000000001e-06    evaluation reward: 8.17
2038
episode: 2038   score: 9.0    memory length: 535109   epsilon: 0.1384822000085672    step
s: 456    lr: 2.560000000000001e-06    evaluation reward: 8.2
2039
episode: 2039   score: 17.0    memory length: 535755   epsilon: 0.1372031200085591    ste
ps: 646    lr: 2.560000000000001e-06    evaluation reward: 8.28
2040
episode: 2040   score: 6.0    memory length: 536112   epsilon: 0.13649626000855464    ste
ps: 357    lr: 2.560000000000001e-06    evaluation reward: 8.27
2041
episode: 2041   score: 6.0    memory length: 536429   epsilon: 0.13586860000855067    ste
ps: 317    lr: 2.560000000000001e-06    evaluation reward: 8.26
```

```
2042
episode: 2042    score: 11.0    memory length: 536952    epsilon: 0.13483306000854411    st
eps: 523    lr: 2.560000000000001e-06    evaluation reward: 8.25
2043
episode: 2043    score: 6.0    memory length: 537290    epsilon: 0.13416382000853988    ste
ps: 338    lr: 2.560000000000001e-06    evaluation reward: 8.21
2044
episode: 2044    score: 7.0    memory length: 537716    epsilon: 0.13332034000853454    ste
ps: 426    lr: 2.560000000000001e-06    evaluation reward: 8.2
2045
episode: 2045    score: 8.0    memory length: 538171    epsilon: 0.132441944000852884    ste
ps: 455    lr: 2.560000000000001e-06    evaluation reward: 8.23
2046
episode: 2046    score: 13.0    memory length: 538663    epsilon: 0.13144528000852268    st
eps: 492    lr: 2.560000000000001e-06    evaluation reward: 8.24
2047
episode: 2047    score: 3.0    memory length: 538876    epsilon: 0.13102354000852    steps:
 213    lr: 2.560000000000001e-06    evaluation reward: 8.23
2048
episode: 2048    score: 9.0    memory length: 539328    epsilon: 0.13012858000851435    ste
ps: 452    lr: 2.560000000000001e-06    evaluation reward: 8.25
2049
episode: 2049    score: 4.0    memory length: 539608    epsilon: 0.12957418000851084    ste
ps: 280    lr: 2.560000000000001e-06    evaluation reward: 8.25
2050
episode: 2050    score: 10.0    memory length: 540102    epsilon: 0.12859606000850465    st
eps: 494    lr: 2.560000000000001e-06    evaluation reward: 8.29
2051
episode: 2051    score: 15.0    memory length: 540663    epsilon: 0.12748528000849763    st
eps: 561    lr: 2.560000000000001e-06    evaluation reward: 8.36
2052
episode: 2052    score: 4.0    memory length: 540910    epsilon: 0.12699622000849453    ste
ps: 247    lr: 2.560000000000001e-06    evaluation reward: 8.29
2053
episode: 2053    score: 11.0    memory length: 541440    epsilon: 0.1259468200084879    ste
ps: 530    lr: 2.560000000000001e-06    evaluation reward: 8.36
2054
episode: 2054    score: 19.0    memory length: 542008    epsilon: 0.12482218000848203    st
eps: 568    lr: 2.560000000000001e-06    evaluation reward: 8.5
2055
episode: 2055    score: 5.0    memory length: 542300    epsilon: 0.12424402000848242    ste
ps: 292    lr: 2.560000000000001e-06    evaluation reward: 8.49
2056
episode: 2056    score: 14.0    memory length: 542972    epsilon: 0.12291346000848333    st
eps: 672    lr: 2.560000000000001e-06    evaluation reward: 8.57
2057
episode: 2057    score: 12.0    memory length: 543491    epsilon: 0.12188584000848403    st
eps: 519    lr: 2.560000000000001e-06    evaluation reward: 8.61
2058
episode: 2058    score: 6.0    memory length: 543827    epsilon: 0.12122056000848448    ste
ps: 336    lr: 2.560000000000001e-06    evaluation reward: 8.59
2059
episode: 2059    score: 3.0    memory length: 544058    epsilon: 0.1207631800084848    step
s: 231    lr: 2.560000000000001e-06    evaluation reward: 8.51
2060
episode: 2060    score: 5.0    memory length: 544367    epsilon: 0.12015136000848521    ste
ps: 309    lr: 2.560000000000001e-06    evaluation reward: 8.49
2061
episode: 2061    score: 5.0    memory length: 544659    epsilon: 0.1195732000084856    step
s: 292    lr: 2.560000000000001e-06    evaluation reward: 8.45
2062
episode: 2062    score: 12.0    memory length: 545219    epsilon: 0.11846440000848636    st
eps: 560    lr: 2.560000000000001e-06    evaluation reward: 8.48
2063
episode: 2063    score: 6.0    memory length: 545565    epsilon: 0.11777932000848683    ste
ps: 346    lr: 2.560000000000001e-06    evaluation reward: 8.44
```

2064
episode: 2064   score: 10.0   memory length: 546047   epsilon: 0.11682496000848748   st
eps: 482   lr: 2.560000000000001e-06   evaluation reward: 8.47
2065
episode: 2065   score: 8.0   memory length: 546470   epsilon: 0.11598742000848805   ste
ps: 423   lr: 2.560000000000001e-06   evaluation reward: 8.49
2066
episode: 2066   score: 11.0   memory length: 547041   epsilon: 0.11485684000848882   st
eps: 571   lr: 2.560000000000001e-06   evaluation reward: 8.52
2067
episode: 2067   score: 9.0   memory length: 547497   epsilon: 0.11395396000848944   ste
ps: 456   lr: 2.560000000000001e-06   evaluation reward: 8.5
2068
episode: 2068   score: 9.0   memory length: 547906   epsilon: 0.11314414000848999   ste
ps: 409   lr: 2.560000000000001e-06   evaluation reward: 8.55
2069
episode: 2069   score: 11.0   memory length: 548463   epsilon: 0.11204128000849074   st
eps: 557   lr: 2.560000000000001e-06   evaluation reward: 8.54
2070
episode: 2070   score: 12.0   memory length: 548901   epsilon: 0.11117404000849133   st
eps: 438   lr: 2.560000000000001e-06   evaluation reward: 8.63
2071
episode: 2071   score: 10.0   memory length: 549393   epsilon: 0.110199880008492   step
s: 492   lr: 2.560000000000001e-06   evaluation reward: 8.68
2072
episode: 2072   score: 11.0   memory length: 549920   epsilon: 0.10915642000849271   st
eps: 527   lr: 2.560000000000001e-06   evaluation reward: 8.73
2073
episode: 2073   score: 4.0   memory length: 550177   epsilon: 0.10864756000849306   ste
ps: 257   lr: 2.560000000000001e-06   evaluation reward: 8.72
2074
episode: 2074   score: 8.0   memory length: 550604   epsilon: 0.10780210000849363   ste
ps: 427   lr: 2.560000000000001e-06   evaluation reward: 8.73
2075
episode: 2075   score: 9.0   memory length: 551029   epsilon: 0.10696060000849421   ste
ps: 425   lr: 2.560000000000001e-06   evaluation reward: 8.76
2076
episode: 2076   score: 9.0   memory length: 551480   epsilon: 0.10606762000849482   ste
ps: 451   lr: 2.560000000000001e-06   evaluation reward: 8.71
2077
episode: 2077   score: 10.0   memory length: 551933   epsilon: 0.10517068000849543   st
eps: 453   lr: 2.560000000000001e-06   evaluation reward: 8.77
2078
episode: 2078   score: 10.0   memory length: 552383   epsilon: 0.10427968000849604   st
eps: 450   lr: 2.560000000000001e-06   evaluation reward: 8.75
2079
episode: 2079   score: 6.0   memory length: 552733   epsilon: 0.10358668000849651   ste
ps: 350   lr: 2.560000000000001e-06   evaluation reward: 8.76
2080
episode: 2080   score: 18.0   memory length: 553293   epsilon: 0.10247788000849727   st
eps: 560   lr: 2.560000000000001e-06   evaluation reward: 8.82
2081
episode: 2081   score: 6.0   memory length: 553671   epsilon: 0.10172944000849778   ste
ps: 378   lr: 2.560000000000001e-06   evaluation reward: 8.82
2082
episode: 2082   score: 5.0   memory length: 553945   epsilon: 0.10118692000849815   ste
ps: 274   lr: 2.560000000000001e-06   evaluation reward: 8.81
2083
episode: 2083   score: 6.0   memory length: 554269   epsilon: 0.10054540000849858   ste
ps: 324   lr: 2.560000000000001e-06   evaluation reward: 8.77
2084
episode: 2084   score: 7.0   memory length: 554643   epsilon: 0.09980488000849909   ste
ps: 374   lr: 2.560000000000001e-06   evaluation reward: 8.77
2085
episode: 2085   score: 7.0   memory length: 555016   epsilon: 0.09906634000849959   ste
ps: 373   lr: 2.560000000000001e-06   evaluation reward: 8.75

```
2086
episode: 2086    score: 5.0    memory length: 555347    epsilon: 0.09841096000850004    ste
ps: 331    lr: 2.560000000000001e-06    evaluation reward: 8.73
2087
episode: 2087    score: 14.0    memory length: 555828    epsilon: 0.09745858000850069    st
eps: 481    lr: 2.560000000000001e-06    evaluation reward: 8.82
2088
episode: 2088    score: 17.0    memory length: 556390    epsilon: 0.09634582000850145    st
eps: 562    lr: 2.560000000000001e-06    evaluation reward: 8.89
2089
episode: 2089    score: 9.0    memory length: 556889    epsilon: 0.09535780000850212    ste
ps: 499    lr: 2.560000000000001e-06    evaluation reward: 8.93
2090
episode: 2090    score: 14.0    memory length: 557555    epsilon: 0.09403912000850302    st
eps: 666    lr: 2.560000000000001e-06    evaluation reward: 9.03
2091
episode: 2091    score: 11.0    memory length: 558087    epsilon: 0.09298576000850374    st
eps: 532    lr: 2.560000000000001e-06    evaluation reward: 9.08
2092
episode: 2092    score: 8.0    memory length: 558512    epsilon: 0.09214426000850431    ste
ps: 425    lr: 2.560000000000001e-06    evaluation reward: 9.02
2093
episode: 2093    score: 9.0    memory length: 558939    epsilon: 0.09129880000850489    ste
ps: 427    lr: 2.560000000000001e-06    evaluation reward: 9.07
2094
episode: 2094    score: 6.0    memory length: 559312    epsilon: 0.0905602600085054    step
s: 373    lr: 2.560000000000001e-06    evaluation reward: 9.02
2095
episode: 2095    score: 8.0    memory length: 559765    epsilon: 0.089663320008506    step
s: 453    lr: 2.560000000000001e-06    evaluation reward: 9.04
2096
episode: 2096    score: 12.0    memory length: 560356    epsilon: 0.0884931400085068    ste
ps: 591    lr: 2.560000000000001e-06    evaluation reward: 9.08
2097
episode: 2097    score: 6.0    memory length: 560667    epsilon: 0.08787736000850722    ste
ps: 311    lr: 2.560000000000001e-06    evaluation reward: 9.01
2098
episode: 2098    score: 16.0    memory length: 561225    epsilon: 0.08677252000850798    st
eps: 558    lr: 2.560000000000001e-06    evaluation reward: 9.11
2099
episode: 2099    score: 9.0    memory length: 561691    epsilon: 0.08584984000850861    ste
ps: 466    lr: 2.560000000000001e-06    evaluation reward: 9.12
2100
episode: 2100    score: 8.0    memory length: 562091    epsilon: 0.08505784000850915    ste
ps: 400    lr: 2.560000000000001e-06    evaluation reward: 9.15
2101
episode: 2101    score: 5.0    memory length: 562400    epsilon: 0.08444602000850956    ste
ps: 309    lr: 2.560000000000001e-06    evaluation reward: 9.1
2102
episode: 2102    score: 6.0    memory length: 562778    epsilon: 0.08369758000851008    ste
ps: 378    lr: 2.560000000000001e-06    evaluation reward: 9.1
2103
episode: 2103    score: 8.0    memory length: 563219    epsilon: 0.08282440000851067    ste
ps: 441    lr: 2.560000000000001e-06    evaluation reward: 9.05
2104
episode: 2104    score: 6.0    memory length: 563611    epsilon: 0.0820482400085112    step
s: 392    lr: 2.560000000000001e-06    evaluation reward: 9.03
2105
episode: 2105    score: 10.0    memory length: 564180    epsilon: 0.08092162000851197    st
eps: 569    lr: 2.560000000000001e-06    evaluation reward: 9.06
2106
episode: 2106    score: 6.0    memory length: 564499    epsilon: 0.0802900000085124    step
s: 319    lr: 2.560000000000001e-06    evaluation reward: 9.01
2107
episode: 2107    score: 7.0    memory length: 564921    epsilon: 0.07945444000851297    ste
ps: 422    lr: 2.560000000000001e-06    evaluation reward: 9.0
```

```
2108
episode: 2108   score: 11.0   memory length: 565473   epsilon: 0.07836148000851371    st
eps: 552    lr: 2.560000000000001e-06    evaluation reward: 9.05
2109
episode: 2109   score: 7.0   memory length: 565888   epsilon: 0.07753978000851428    ste
ps: 415    lr: 2.560000000000001e-06    evaluation reward: 9.04
2110
episode: 2110   score: 13.0   memory length: 566525   epsilon: 0.07627852000851514    st
eps: 637    lr: 2.560000000000001e-06    evaluation reward: 9.05
2111
episode: 2111   score: 10.0   memory length: 567070   epsilon: 0.07519942000851587    st
eps: 545    lr: 2.560000000000001e-06    evaluation reward: 8.97
2112
episode: 2112   score: 6.0   memory length: 567403   epsilon: 0.07454008000851632    ste
ps: 333    lr: 2.560000000000001e-06    evaluation reward: 8.92
2113
episode: 2113   score: 6.0   memory length: 567763   epsilon: 0.07382728000851681    ste
ps: 360    lr: 2.560000000000001e-06    evaluation reward: 8.92
2114
episode: 2114   score: 8.0   memory length: 568147   epsilon: 0.07306696000851733    ste
ps: 384    lr: 2.560000000000001e-06    evaluation reward: 8.89
2115
episode: 2115   score: 6.0   memory length: 568510   epsilon: 0.07234822000851782    ste
ps: 363    lr: 2.560000000000001e-06    evaluation reward: 8.86
2116
episode: 2116   score: 7.0   memory length: 568892   epsilon: 0.07159186000851833    ste
ps: 382    lr: 2.560000000000001e-06    evaluation reward: 8.8
2117
episode: 2117   score: 9.0   memory length: 569368   epsilon: 0.07064938000851897    ste
ps: 476    lr: 2.560000000000001e-06    evaluation reward: 8.78
2118
episode: 2118   score: 6.0   memory length: 569728   epsilon: 0.06993658000851946    ste
ps: 360    lr: 2.560000000000001e-06    evaluation reward: 8.78
2119
episode: 2119   score: 6.0   memory length: 570074   epsilon: 0.06925150000851993    ste
ps: 346    lr: 2.560000000000001e-06    evaluation reward: 8.72
2120
episode: 2120   score: 6.0   memory length: 570428   epsilon: 0.0685505800085204    step
s: 354    lr: 2.560000000000001e-06    evaluation reward: 8.7
2121
episode: 2121   score: 13.0   memory length: 570982   epsilon: 0.06745366000852115    st
eps: 554    lr: 2.560000000000001e-06    evaluation reward: 8.7
2122
episode: 2122   score: 8.0   memory length: 571410   epsilon: 0.06660622000852173    ste
ps: 428    lr: 2.560000000000001e-06    evaluation reward: 8.7
2123
episode: 2123   score: 10.0   memory length: 571779   epsilon: 0.06587560000852223    st
eps: 369    lr: 2.560000000000001e-06    evaluation reward: 8.74
2124
episode: 2124   score: 10.0   memory length: 572263   epsilon: 0.06491728000852288    st
eps: 484    lr: 2.560000000000001e-06    evaluation reward: 8.76
2125
episode: 2125   score: 12.0   memory length: 572863   epsilon: 0.0637292800085237    ste
ps: 600    lr: 2.560000000000001e-06    evaluation reward: 8.73
2126
episode: 2126   score: 5.0   memory length: 573175   epsilon: 0.06311152000852412    ste
ps: 312    lr: 2.560000000000001e-06    evaluation reward: 8.7
2127
episode: 2127   score: 9.0   memory length: 573614   epsilon: 0.06224230000852471    ste
ps: 439    lr: 2.560000000000001e-06    evaluation reward: 8.72
2128
episode: 2128   score: 9.0   memory length: 574103   epsilon: 0.06127408000852537    ste
ps: 489    lr: 2.560000000000001e-06    evaluation reward: 8.72
2129
episode: 2129   score: 10.0   memory length: 574645   epsilon: 0.0602009200085261    ste
ps: 542    lr: 2.560000000000001e-06    evaluation reward: 8.72
```

```
2130
episode: 2130   score: 12.0   memory length: 575153   epsilon: 0.05919508000852679     st
eps: 508    lr: 2.560000000000001e-06     evaluation reward: 8.75
2131
episode: 2131   score: 10.0   memory length: 575698   epsilon: 0.058115980008527524    s
teps: 545    lr: 2.560000000000001e-06     evaluation reward: 8.71
2132
episode: 2132   score: 7.0   memory length: 576077   epsilon: 0.057365560008528035    st
eps: 379    lr: 2.560000000000001e-06     evaluation reward: 8.7
2133
episode: 2133   score: 9.0   memory length: 576527   epsilon: 0.056474560008528064    ste
ps: 450    lr: 2.560000000000001e-06     evaluation reward: 8.72
2134
episode: 2134   score: 8.0   memory length: 576972   epsilon: 0.0555934600085292444    st
eps: 445    lr: 2.560000000000001e-06     evaluation reward: 8.7
2135
episode: 2135   score: 7.0   memory length: 577349   epsilon: 0.05484700000852975    ste
ps: 377    lr: 2.560000000000001e-06     evaluation reward: 8.71
2136
episode: 2136   score: 11.0   memory length: 577744   epsilon: 0.05406490000853029    st
eps: 395    lr: 2.560000000000001e-06     evaluation reward: 8.78
2137
episode: 2137   score: 6.0   memory length: 578100   epsilon: 0.05336002000853077    ste
ps: 356    lr: 2.560000000000001e-06     evaluation reward: 8.79
2138
episode: 2138   score: 7.0   memory length: 578508   epsilon: 0.05255218000853132    ste
ps: 408    lr: 2.560000000000001e-06     evaluation reward: 8.77
2139
episode: 2139   score: 6.0   memory length: 578902   epsilon: 0.05177206000853185    ste
ps: 394    lr: 2.560000000000001e-06     evaluation reward: 8.66
2140
episode: 2140   score: 7.0   memory length: 579284   epsilon: 0.051015700008532366    st
eps: 382    lr: 2.560000000000001e-06     evaluation reward: 8.67
2141
episode: 2141   score: 11.0   memory length: 579851   epsilon: 0.04989304000853313    st
eps: 567    lr: 2.560000000000001e-06     evaluation reward: 8.72
2142
episode: 2142   score: 14.0   memory length: 580379   epsilon: 0.048847600008533845    s
teps: 528    lr: 2.560000000000001e-06     evaluation reward: 8.75
2143
episode: 2143   score: 8.0   memory length: 580793   epsilon: 0.048027880008534404    st
eps: 414    lr: 2.560000000000001e-06     evaluation reward: 8.77
2144
episode: 2144   score: 9.0   memory length: 581222   epsilon: 0.047178460008534984    st
eps: 429    lr: 2.560000000000001e-06     evaluation reward: 8.79
2145
episode: 2145   score: 9.0   memory length: 581693   epsilon: 0.04624588000853562    ste
ps: 471    lr: 2.560000000000001e-06     evaluation reward: 8.8
2146
episode: 2146   score: 6.0   memory length: 582084   epsilon: 0.04547170000853615    ste
ps: 391    lr: 2.560000000000001e-06     evaluation reward: 8.73
2147
episode: 2147   score: 10.0   memory length: 582562   epsilon: 0.04452526000853679    st
eps: 478    lr: 2.560000000000001e-06     evaluation reward: 8.8
2148
episode: 2148   score: 19.0   memory length: 583005   epsilon: 0.04364812000853739    st
eps: 443    lr: 2.560000000000001e-06     evaluation reward: 8.9
2149
episode: 2149   score: 8.0   memory length: 583462   epsilon: 0.04274326000853801    ste
ps: 457    lr: 2.560000000000001e-06     evaluation reward: 8.94
2150
episode: 2150   score: 11.0   memory length: 583953   epsilon: 0.04177108000853867    st
eps: 491    lr: 2.560000000000001e-06     evaluation reward: 8.95
2151
episode: 2151   score: 8.0   memory length: 584368   epsilon: 0.04094938000853923    ste
ps: 415    lr: 2.560000000000001e-06     evaluation reward: 8.88
```

```
2152
episode: 2152   score: 11.0   memory length: 584850   epsilon: 0.03999502000853988      st
eps: 482    lr: 2.560000000000001e-06      evaluation reward: 8.95
2153
episode: 2153   score: 6.0   memory length: 585223   epsilon: 0.03925648000854039      ste
ps: 373    lr: 2.560000000000001e-06      evaluation reward: 8.9
2154
episode: 2154   score: 9.0   memory length: 585695   epsilon: 0.038321920008541024      st
eps: 472    lr: 2.560000000000001e-06      evaluation reward: 8.8
2155
episode: 2155   score: 9.0   memory length: 586168   epsilon: 0.03738538000854166      ste
ps: 473    lr: 2.560000000000001e-06      evaluation reward: 8.84
2156
episode: 2156   score: 10.0   memory length: 586678   epsilon: 0.03637558000854235      st
eps: 510    lr: 2.560000000000001e-06      evaluation reward: 8.8
2157
episode: 2157   score: 8.0   memory length: 587114   epsilon: 0.03551230000854294      ste
ps: 436    lr: 2.560000000000001e-06      evaluation reward: 8.76
2158
episode: 2158   score: 9.0   memory length: 587591   epsilon: 0.034567840008543585      st
eps: 477    lr: 2.560000000000001e-06      evaluation reward: 8.79
2159
episode: 2159   score: 10.0   memory length: 588131   epsilon: 0.033498640008544314      s
teps: 540    lr: 2.560000000000001e-06      evaluation reward: 8.86
2160
episode: 2160   score: 10.0   memory length: 588633   epsilon: 0.03250468000854499      st
eps: 502    lr: 2.560000000000001e-06      evaluation reward: 8.91
2161
episode: 2161   score: 9.0   memory length: 589087   epsilon: 0.031605760008545605      st
eps: 454    lr: 2.560000000000001e-06      evaluation reward: 8.95
2162
episode: 2162   score: 9.0   memory length: 589516   epsilon: 0.030756340008546185      st
eps: 429    lr: 2.560000000000001e-06      evaluation reward: 8.92
2163
episode: 2163   score: 9.0   memory length: 589943   epsilon: 0.02991088000854676      ste
ps: 427    lr: 2.560000000000001e-06      evaluation reward: 8.95
2164
episode: 2164   score: 4.0   memory length: 590187   epsilon: 0.02942776000854709      ste
ps: 244    lr: 2.560000000000001e-06      evaluation reward: 8.89
2165
episode: 2165   score: 5.0   memory length: 590535   epsilon: 0.02873872000854756      ste
ps: 348    lr: 2.560000000000001e-06      evaluation reward: 8.86
2166
episode: 2166   score: 11.0   memory length: 591070   epsilon: 0.027679420008548283      s
teps: 535    lr: 2.560000000000001e-06      evaluation reward: 8.86
2167
episode: 2167   score: 4.0   memory length: 591351   epsilon: 0.027123040008548663      st
eps: 281    lr: 2.560000000000001e-06      evaluation reward: 8.81
2168
episode: 2168   score: 11.0   memory length: 591920   epsilon: 0.02599642000854943      st
eps: 569    lr: 2.560000000000001e-06      evaluation reward: 8.83
2169
episode: 2169   score: 11.0   memory length: 592490   epsilon: 0.0248678200085502      ste
ps: 570    lr: 2.560000000000001e-06      evaluation reward: 8.83
2170
episode: 2170   score: 16.0   memory length: 593050   epsilon: 0.023759020008550957      s
teps: 560    lr: 2.560000000000001e-06      evaluation reward: 8.87
2171
episode: 2171   score: 19.0   memory length: 593736   epsilon: 0.022400740008551884      s
teps: 686    lr: 2.560000000000001e-06      evaluation reward: 8.96
2172
episode: 2172   score: 15.0   memory length: 594281   epsilon: 0.02132164000855262      st
eps: 545    lr: 2.560000000000001e-06      evaluation reward: 9.0
2173
episode: 2173   score: 10.0   memory length: 594808   epsilon: 0.02027818000855333      st
eps: 527    lr: 2.560000000000001e-06      evaluation reward: 9.06
```

```
2174
episode: 2174   score: 7.0   memory length: 595185   epsilon: 0.019531720008553384   ste
ps: 377    lr: 2.560000000000001e-06    evaluation reward: 9.05
2175
episode: 2175   score: 7.0   memory length: 595572   epsilon: 0.018765460008554363   st
eps: 387    lr: 2.560000000000001e-06    evaluation reward: 9.03
2176
episode: 2176   score: 6.0   memory length: 595906   epsilon: 0.018104140008554814   st
eps: 334    lr: 2.560000000000001e-06    evaluation reward: 9.0
2177
episode: 2177   score: 8.0   memory length: 596363   epsilon: 0.01719928000855543   ste
ps: 457    lr: 2.560000000000001e-06    evaluation reward: 8.98
2178
episode: 2178   score: 12.0   memory length: 596989   epsilon: 0.015959800008556277   s
teps: 626    lr: 2.560000000000001e-06    evaluation reward: 9.0
2179
episode: 2179   score: 11.0   memory length: 597504   epsilon: 0.014940100008556372   s
teps: 515    lr: 2.560000000000001e-06    evaluation reward: 9.05
2180
episode: 2180   score: 11.0   memory length: 597931   epsilon: 0.014094640008556208   s
teps: 427    lr: 2.560000000000001e-06    evaluation reward: 8.98
2181
episode: 2181   score: 5.0   memory length: 598225   epsilon: 0.013512520008556095   st
eps: 294    lr: 2.560000000000001e-06    evaluation reward: 8.97
2182
episode: 2182   score: 11.0   memory length: 598794   epsilon: 0.012385900008555877   s
teps: 569    lr: 2.560000000000001e-06    evaluation reward: 9.03
2183
episode: 2183   score: 12.0   memory length: 599304   epsilon: 0.01137610000855568   st
eps: 510    lr: 2.560000000000001e-06    evaluation reward: 9.09
2184
episode: 2184   score: 8.0   memory length: 599743   epsilon: 0.010506880008555512   st
eps: 439    lr: 2.560000000000001e-06    evaluation reward: 9.1
2185
episode: 2185   score: 3.0   memory length: 599958   epsilon: 0.01008118000855543   ste
ps: 215    lr: 2.560000000000001e-06    evaluation reward: 9.06
2186
episode: 2186   score: 14.0   memory length: 600610   epsilon: 0.0099998020008555413   s
teps: 652    lr: 1.0240000000000005e-06    evaluation reward: 9.15
2187
episode: 2187   score: 16.0   memory length: 601233   epsilon: 0.0099998020008555413   s
teps: 623    lr: 1.0240000000000005e-06    evaluation reward: 9.17
2188
episode: 2188   score: 7.0   memory length: 601628   epsilon: 0.0099998020008555413   st
eps: 395    lr: 1.0240000000000005e-06    evaluation reward: 9.07
2189
episode: 2189   score: 8.0   memory length: 602085   epsilon: 0.0099998020008555413   st
eps: 457    lr: 1.0240000000000005e-06    evaluation reward: 9.06
2190
episode: 2190   score: 5.0   memory length: 602377   epsilon: 0.0099998020008555413   st
eps: 292    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2191
episode: 2191   score: 7.0   memory length: 602803   epsilon: 0.0099998020008555413   st
eps: 426    lr: 1.0240000000000005e-06    evaluation reward: 8.93
2192
episode: 2192   score: 11.0   memory length: 603302   epsilon: 0.0099998020008555413   s
teps: 499    lr: 1.0240000000000005e-06    evaluation reward: 8.96
2193
episode: 2193   score: 16.0   memory length: 603920   epsilon: 0.0099998020008555413   s
teps: 618    lr: 1.0240000000000005e-06    evaluation reward: 9.03
2194
episode: 2194   score: 10.0   memory length: 604438   epsilon: 0.0099998020008555413   s
teps: 518    lr: 1.0240000000000005e-06    evaluation reward: 9.07
2195
episode: 2195   score: 11.0   memory length: 604979   epsilon: 0.0099998020008555413   s
teps: 541    lr: 1.0240000000000005e-06    evaluation reward: 9.1
```

```
2196
episode: 2196   score: 8.0   memory length: 605453   epsilon: 0.009998020008555413     st
eps: 474    lr: 1.0240000000000005e-06     evaluation reward: 9.06
2197
episode: 2197   score: 12.0   memory length: 606042   epsilon: 0.009998020008555413     s
teps: 589    lr: 1.0240000000000005e-06     evaluation reward: 9.12
2198
episode: 2198   score: 7.0   memory length: 606470   epsilon: 0.009998020008555413     st
eps: 428    lr: 1.0240000000000005e-06     evaluation reward: 9.03
2199
episode: 2199   score: 5.0   memory length: 606766   epsilon: 0.009998020008555413     st
eps: 296    lr: 1.0240000000000005e-06     evaluation reward: 8.99
2200
episode: 2200   score: 8.0   memory length: 607163   epsilon: 0.009998020008555413     st
eps: 397    lr: 1.0240000000000005e-06     evaluation reward: 8.99
2201
episode: 2201   score: 7.0   memory length: 607584   epsilon: 0.009998020008555413     st
eps: 421    lr: 1.0240000000000005e-06     evaluation reward: 9.01
2202
episode: 2202   score: 7.0   memory length: 607967   epsilon: 0.009998020008555413     st
eps: 383    lr: 1.0240000000000005e-06     evaluation reward: 9.02
2203
episode: 2203   score: 11.0   memory length: 608450   epsilon: 0.009998020008555413     s
teps: 483    lr: 1.0240000000000005e-06     evaluation reward: 9.05
2204
episode: 2204   score: 3.0   memory length: 608682   epsilon: 0.009998020008555413     st
eps: 232    lr: 1.0240000000000005e-06     evaluation reward: 9.02
2205
episode: 2205   score: 5.0   memory length: 609008   epsilon: 0.009998020008555413     st
eps: 326    lr: 1.0240000000000005e-06     evaluation reward: 8.97
2206
episode: 2206   score: 7.0   memory length: 609391   epsilon: 0.009998020008555413     st
eps: 383    lr: 1.0240000000000005e-06     evaluation reward: 8.98
2207
episode: 2207   score: 12.0   memory length: 609969   epsilon: 0.009998020008555413     s
teps: 578    lr: 1.0240000000000005e-06     evaluation reward: 9.03
2208
episode: 2208   score: 4.0   memory length: 610250   epsilon: 0.009998020008555413     st
eps: 281    lr: 1.0240000000000005e-06     evaluation reward: 8.96
2209
episode: 2209   score: 15.0   memory length: 610801   epsilon: 0.009998020008555413     s
teps: 551    lr: 1.0240000000000005e-06     evaluation reward: 9.04
2210
episode: 2210   score: 9.0   memory length: 611229   epsilon: 0.009998020008555413     st
eps: 428    lr: 1.0240000000000005e-06     evaluation reward: 9.0
2211
episode: 2211   score: 5.0   memory length: 611525   epsilon: 0.009998020008555413     st
eps: 296    lr: 1.0240000000000005e-06     evaluation reward: 8.95
2212
episode: 2212   score: 9.0   memory length: 612065   epsilon: 0.009998020008555413     st
eps: 540    lr: 1.0240000000000005e-06     evaluation reward: 8.98
2213
episode: 2213   score: 15.0   memory length: 612631   epsilon: 0.009998020008555413     s
teps: 566    lr: 1.0240000000000005e-06     evaluation reward: 9.07
2214
episode: 2214   score: 7.0   memory length: 612999   epsilon: 0.009998020008555413     st
eps: 368    lr: 1.0240000000000005e-06     evaluation reward: 9.06
2215
episode: 2215   score: 6.0   memory length: 613338   epsilon: 0.009998020008555413     st
eps: 339    lr: 1.0240000000000005e-06     evaluation reward: 9.06
2216
episode: 2216   score: 7.0   memory length: 613725   epsilon: 0.009998020008555413     st
eps: 387    lr: 1.0240000000000005e-06     evaluation reward: 9.06
2217
episode: 2217   score: 13.0   memory length: 614079   epsilon: 0.009998020008555413     s
teps: 354    lr: 1.0240000000000005e-06     evaluation reward: 9.1
```

```
2218
episode: 2218   score: 8.0    memory length: 614510   epsilon: 0.009998020008555413      st
eps: 431    lr: 1.0240000000000005e-06     evaluation reward: 9.12
2219
episode: 2219   score: 12.0    memory length: 615113   epsilon: 0.009998020008555413      s
teps: 603    lr: 1.0240000000000005e-06     evaluation reward: 9.18
2220
episode: 2220   score: 8.0    memory length: 615508   epsilon: 0.009998020008555413      st
eps: 395    lr: 1.0240000000000005e-06     evaluation reward: 9.2
2221
episode: 2221   score: 10.0    memory length: 615989   epsilon: 0.009998020008555413      s
teps: 481    lr: 1.0240000000000005e-06     evaluation reward: 9.17
2222
episode: 2222   score: 17.0    memory length: 616641   epsilon: 0.009998020008555413      s
teps: 652    lr: 1.0240000000000005e-06     evaluation reward: 9.26
2223
episode: 2223   score: 13.0    memory length: 617112   epsilon: 0.009998020008555413      s
teps: 471    lr: 1.0240000000000005e-06     evaluation reward: 9.29
2224
episode: 2224   score: 19.0    memory length: 617829   epsilon: 0.009998020008555413      s
teps: 717    lr: 1.0240000000000005e-06     evaluation reward: 9.38
2225
episode: 2225   score: 8.0    memory length: 618272   epsilon: 0.009998020008555413      st
eps: 443    lr: 1.0240000000000005e-06     evaluation reward: 9.34
2226
episode: 2226   score: 13.0    memory length: 618890   epsilon: 0.009998020008555413      s
teps: 618    lr: 1.0240000000000005e-06     evaluation reward: 9.42
2227
episode: 2227   score: 9.0    memory length: 619367   epsilon: 0.009998020008555413      st
eps: 477    lr: 1.0240000000000005e-06     evaluation reward: 9.42
2228
episode: 2228   score: 3.0    memory length: 619599   epsilon: 0.009998020008555413      st
eps: 232    lr: 1.0240000000000005e-06     evaluation reward: 9.36
2229
episode: 2229   score: 5.0    memory length: 619931   epsilon: 0.009998020008555413      st
eps: 332    lr: 1.0240000000000005e-06     evaluation reward: 9.31
2230
episode: 2230   score: 11.0    memory length: 620450   epsilon: 0.009998020008555413      s
teps: 519    lr: 1.0240000000000005e-06     evaluation reward: 9.3
2231
episode: 2231   score: 5.0    memory length: 620760   epsilon: 0.009998020008555413      st
eps: 310    lr: 1.0240000000000005e-06     evaluation reward: 9.25
2232
episode: 2232   score: 5.0    memory length: 621056   epsilon: 0.009998020008555413      st
eps: 296    lr: 1.0240000000000005e-06     evaluation reward: 9.23
2233
episode: 2233   score: 7.0    memory length: 621478   epsilon: 0.009998020008555413      st
eps: 422    lr: 1.0240000000000005e-06     evaluation reward: 9.21
2234
episode: 2234   score: 5.0    memory length: 621774   epsilon: 0.009998020008555413      st
eps: 296    lr: 1.0240000000000005e-06     evaluation reward: 9.18
2235
episode: 2235   score: 5.0    memory length: 622070   epsilon: 0.009998020008555413      st
eps: 296    lr: 1.0240000000000005e-06     evaluation reward: 9.16
2236
episode: 2236   score: 12.0    memory length: 622654   epsilon: 0.009998020008555413      s
teps: 584    lr: 1.0240000000000005e-06     evaluation reward: 9.17
2237
episode: 2237   score: 7.0    memory length: 623043   epsilon: 0.009998020008555413      st
eps: 389    lr: 1.0240000000000005e-06     evaluation reward: 9.18
2238
episode: 2238   score: 7.0    memory length: 623451   epsilon: 0.009998020008555413      st
eps: 408    lr: 1.0240000000000005e-06     evaluation reward: 9.18
2239
episode: 2239   score: 15.0    memory length: 624001   epsilon: 0.009998020008555413      s
teps: 550    lr: 1.0240000000000005e-06     evaluation reward: 9.27
```

```
2240
episode: 2240   score: 15.0   memory length: 624551   epsilon: 0.009998020008555413   s
teps: 550   lr: 1.0240000000000005e-06   evaluation reward: 9.35
2241
episode: 2241   score: 8.0   memory length: 624988   epsilon: 0.009998020008555413   st
eps: 437   lr: 1.0240000000000005e-06   evaluation reward: 9.32
2242
episode: 2242   score: 17.0   memory length: 625641   epsilon: 0.009998020008555413   s
teps: 653   lr: 1.0240000000000005e-06   evaluation reward: 9.35
2243
episode: 2243   score: 11.0   memory length: 626182   epsilon: 0.009998020008555413   s
teps: 541   lr: 1.0240000000000005e-06   evaluation reward: 9.38
2244
episode: 2244   score: 7.0   memory length: 626601   epsilon: 0.009998020008555413   st
eps: 419   lr: 1.0240000000000005e-06   evaluation reward: 9.36
2245
episode: 2245   score: 6.0   memory length: 626974   epsilon: 0.009998020008555413   st
eps: 373   lr: 1.0240000000000005e-06   evaluation reward: 9.33
2246
episode: 2246   score: 11.0   memory length: 627481   epsilon: 0.009998020008555413   s
teps: 507   lr: 1.0240000000000005e-06   evaluation reward: 9.38
2247
episode: 2247   score: 16.0   memory length: 628079   epsilon: 0.009998020008555413   s
teps: 598   lr: 1.0240000000000005e-06   evaluation reward: 9.44
2248
episode: 2248   score: 10.0   memory length: 628610   epsilon: 0.009998020008555413   s
teps: 531   lr: 1.0240000000000005e-06   evaluation reward: 9.35
2249
episode: 2249   score: 12.0   memory length: 629218   epsilon: 0.009998020008555413   s
teps: 608   lr: 1.0240000000000005e-06   evaluation reward: 9.39
2250
episode: 2250   score: 11.0   memory length: 629745   epsilon: 0.009998020008555413   s
teps: 527   lr: 1.0240000000000005e-06   evaluation reward: 9.39
2251
episode: 2251   score: 6.0   memory length: 630091   epsilon: 0.009998020008555413   st
eps: 346   lr: 1.0240000000000005e-06   evaluation reward: 9.37
2252
episode: 2252   score: 12.0   memory length: 630550   epsilon: 0.009998020008555413   s
teps: 459   lr: 1.0240000000000005e-06   evaluation reward: 9.38
2253
episode: 2253   score: 7.0   memory length: 630958   epsilon: 0.009998020008555413   st
eps: 408   lr: 1.0240000000000005e-06   evaluation reward: 9.39
2254
episode: 2254   score: 11.0   memory length: 631485   epsilon: 0.009998020008555413   s
teps: 527   lr: 1.0240000000000005e-06   evaluation reward: 9.41
2255
episode: 2255   score: 6.0   memory length: 631798   epsilon: 0.009998020008555413   st
eps: 313   lr: 1.0240000000000005e-06   evaluation reward: 9.38
2256
episode: 2256   score: 6.0   memory length: 632111   epsilon: 0.009998020008555413   st
eps: 313   lr: 1.0240000000000005e-06   evaluation reward: 9.34
2257
episode: 2257   score: 10.0   memory length: 632634   epsilon: 0.009998020008555413   s
teps: 523   lr: 1.0240000000000005e-06   evaluation reward: 9.36
2258
episode: 2258   score: 5.0   memory length: 632966   epsilon: 0.009998020008555413   st
eps: 332   lr: 1.0240000000000005e-06   evaluation reward: 9.32
2259
episode: 2259   score: 10.0   memory length: 633480   epsilon: 0.009998020008555413   s
teps: 514   lr: 1.0240000000000005e-06   evaluation reward: 9.32
2260
episode: 2260   score: 6.0   memory length: 633877   epsilon: 0.009998020008555413   st
eps: 397   lr: 1.0240000000000005e-06   evaluation reward: 9.28
2261
episode: 2261   score: 7.0   memory length: 634285   epsilon: 0.009998020008555413   st
eps: 408   lr: 1.0240000000000005e-06   evaluation reward: 9.26
```

```
2262
episode: 2262    score: 7.0    memory length: 634714    epsilon: 0.009998020008555413    st
eps: 429    lr: 1.0240000000000005e-06    evaluation reward: 9.24
2263
episode: 2263    score: 8.0    memory length: 635168    epsilon: 0.009998020008555413    st
eps: 454    lr: 1.0240000000000005e-06    evaluation reward: 9.23
2264
episode: 2264    score: 11.0    memory length: 635606    epsilon: 0.009998020008555413    s
teps: 438    lr: 1.0240000000000005e-06    evaluation reward: 9.3
2265
episode: 2265    score: 7.0    memory length: 636015    epsilon: 0.009998020008555413    st
eps: 409    lr: 1.0240000000000005e-06    evaluation reward: 9.32
2266
episode: 2266    score: 4.0    memory length: 636296    epsilon: 0.009998020008555413    st
eps: 281    lr: 1.0240000000000005e-06    evaluation reward: 9.25
2267
episode: 2267    score: 7.0    memory length: 636668    epsilon: 0.009998020008555413    st
eps: 372    lr: 1.0240000000000005e-06    evaluation reward: 9.28
2268
episode: 2268    score: 4.0    memory length: 636949    epsilon: 0.009998020008555413    st
eps: 281    lr: 1.0240000000000005e-06    evaluation reward: 9.21
2269
episode: 2269    score: 4.0    memory length: 637230    epsilon: 0.009998020008555413    st
eps: 281    lr: 1.0240000000000005e-06    evaluation reward: 9.14
2270
episode: 2270    score: 7.0    memory length: 637639    epsilon: 0.009998020008555413    st
eps: 409    lr: 1.0240000000000005e-06    evaluation reward: 9.05
2271
episode: 2271    score: 5.0    memory length: 637935    epsilon: 0.009998020008555413    st
eps: 296    lr: 1.0240000000000005e-06    evaluation reward: 8.91
2272
episode: 2272    score: 10.0    memory length: 638469    epsilon: 0.009998020008555413    s
teps: 534    lr: 1.0240000000000005e-06    evaluation reward: 8.86
2273
episode: 2273    score: 9.0    memory length: 638951    epsilon: 0.009998020008555413    st
eps: 482    lr: 1.0240000000000005e-06    evaluation reward: 8.85
2274
episode: 2274    score: 9.0    memory length: 639427    epsilon: 0.009998020008555413    st
eps: 476    lr: 1.0240000000000005e-06    evaluation reward: 8.87
2275
episode: 2275    score: 6.0    memory length: 639770    epsilon: 0.009998020008555413    st
eps: 343    lr: 1.0240000000000005e-06    evaluation reward: 8.86
2276
episode: 2276    score: 7.0    memory length: 640179    epsilon: 0.009998020008555413    st
eps: 409    lr: 1.0240000000000005e-06    evaluation reward: 8.87
2277
episode: 2277    score: 20.0    memory length: 640926    epsilon: 0.009998020008555413    s
teps: 747    lr: 1.0240000000000005e-06    evaluation reward: 8.99
2278
episode: 2278    score: 10.0    memory length: 641355    epsilon: 0.009998020008555413    s
teps: 429    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2279
episode: 2279    score: 7.0    memory length: 641746    epsilon: 0.009998020008555413    st
eps: 391    lr: 1.0240000000000005e-06    evaluation reward: 8.93
2280
episode: 2280    score: 7.0    memory length: 642155    epsilon: 0.009998020008555413    st
eps: 409    lr: 1.0240000000000005e-06    evaluation reward: 8.89
2281
episode: 2281    score: 11.0    memory length: 642571    epsilon: 0.009998020008555413    s
teps: 416    lr: 1.0240000000000005e-06    evaluation reward: 8.95
2282
episode: 2282    score: 8.0    memory length: 642990    epsilon: 0.009998020008555413    st
eps: 419    lr: 1.0240000000000005e-06    evaluation reward: 8.92
2283
episode: 2283    score: 6.0    memory length: 643303    epsilon: 0.009998020008555413    st
eps: 313    lr: 1.0240000000000005e-06    evaluation reward: 8.86
```

```
2284
episode: 2284   score: 4.0   memory length: 643582   epsilon: 0.009998020008555413   st
eps: 279    lr: 1.0240000000000005e-06    evaluation reward: 8.82
2285
episode: 2285   score: 7.0   memory length: 643987   epsilon: 0.009998020008555413   st
eps: 405    lr: 1.0240000000000005e-06    evaluation reward: 8.86
2286
episode: 2286   score: 8.0   memory length: 644417   epsilon: 0.009998020008555413   st
eps: 430    lr: 1.0240000000000005e-06    evaluation reward: 8.8
2287
episode: 2287   score: 7.0   memory length: 644822   epsilon: 0.009998020008555413   st
eps: 405    lr: 1.0240000000000005e-06    evaluation reward: 8.71
2288
episode: 2288   score: 6.0   memory length: 645158   epsilon: 0.009998020008555413   st
eps: 336    lr: 1.0240000000000005e-06    evaluation reward: 8.7
2289
episode: 2289   score: 5.0   memory length: 645452   epsilon: 0.009998020008555413   st
eps: 294    lr: 1.0240000000000005e-06    evaluation reward: 8.67
2290
episode: 2290   score: 10.0   memory length: 645981   epsilon: 0.009998020008555413   s
teps: 529    lr: 1.0240000000000005e-06    evaluation reward: 8.72
2291
episode: 2291   score: 11.0   memory length: 646555   epsilon: 0.009998020008555413   s
teps: 574    lr: 1.0240000000000005e-06    evaluation reward: 8.76
2292
episode: 2292   score: 11.0   memory length: 647073   epsilon: 0.009998020008555413   s
teps: 518    lr: 1.0240000000000005e-06    evaluation reward: 8.76
2293
episode: 2293   score: 8.0   memory length: 647501   epsilon: 0.009998020008555413   st
eps: 428    lr: 1.0240000000000005e-06    evaluation reward: 8.68
2294
episode: 2294   score: 6.0   memory length: 647814   epsilon: 0.009998020008555413   st
eps: 313    lr: 1.0240000000000005e-06    evaluation reward: 8.64
2295
episode: 2295   score: 8.0   memory length: 648220   epsilon: 0.009998020008555413   st
eps: 406    lr: 1.0240000000000005e-06    evaluation reward: 8.61
2296
episode: 2296   score: 9.0   memory length: 648659   epsilon: 0.009998020008555413   st
eps: 439    lr: 1.0240000000000005e-06    evaluation reward: 8.62
2297
episode: 2297   score: 11.0   memory length: 649059   epsilon: 0.009998020008555413   s
teps: 400    lr: 1.0240000000000005e-06    evaluation reward: 8.61
2298
episode: 2298   score: 16.0   memory length: 649654   epsilon: 0.009998020008555413   s
teps: 595    lr: 1.0240000000000005e-06    evaluation reward: 8.7
2299
episode: 2299   score: 8.0   memory length: 650111   epsilon: 0.009998020008555413   st
eps: 457    lr: 1.0240000000000005e-06    evaluation reward: 8.73
2300
episode: 2300   score: 7.0   memory length: 650482   epsilon: 0.009998020008555413   st
eps: 371    lr: 1.0240000000000005e-06    evaluation reward: 8.72
2301
episode: 2301   score: 16.0   memory length: 651055   epsilon: 0.009998020008555413   s
teps: 573    lr: 1.0240000000000005e-06    evaluation reward: 8.81
2302
episode: 2302   score: 5.0   memory length: 651368   epsilon: 0.009998020008555413   st
eps: 313    lr: 1.0240000000000005e-06    evaluation reward: 8.79
2303
episode: 2303   score: 10.0   memory length: 651877   epsilon: 0.009998020008555413   s
teps: 509    lr: 1.0240000000000005e-06    evaluation reward: 8.78
2304
episode: 2304   score: 16.0   memory length: 652439   epsilon: 0.009998020008555413   s
teps: 562    lr: 1.0240000000000005e-06    evaluation reward: 8.91
2305
episode: 2305   score: 8.0   memory length: 652876   epsilon: 0.009998020008555413   st
eps: 437    lr: 1.0240000000000005e-06    evaluation reward: 8.94
```

```
2306
episode: 2306   score: 10.0   memory length: 653428   epsilon: 0.009998020008555413   s
teps: 552    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2307
episode: 2307   score: 11.0   memory length: 653935   epsilon: 0.009998020008555413   s
teps: 507    lr: 1.0240000000000005e-06    evaluation reward: 8.96
2308
episode: 2308   score: 5.0   memory length: 654279   epsilon: 0.009998020008555413   st
eps: 344    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2309
episode: 2309   score: 15.0   memory length: 654960   epsilon: 0.009998020008555413   s
teps: 681    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2310
episode: 2310   score: 6.0   memory length: 655283   epsilon: 0.009998020008555413   st
eps: 323    lr: 1.0240000000000005e-06    evaluation reward: 8.94
2311
episode: 2311   score: 8.0   memory length: 655720   epsilon: 0.009998020008555413   st
eps: 437    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2312
episode: 2312   score: 11.0   memory length: 656282   epsilon: 0.009998020008555413   s
teps: 562    lr: 1.0240000000000005e-06    evaluation reward: 8.99
2313
episode: 2313   score: 9.0   memory length: 656725   epsilon: 0.009998020008555413   st
eps: 443    lr: 1.0240000000000005e-06    evaluation reward: 8.93
2314
episode: 2314   score: 10.0   memory length: 657195   epsilon: 0.009998020008555413   s
teps: 470    lr: 1.0240000000000005e-06    evaluation reward: 8.96
2315
episode: 2315   score: 7.0   memory length: 657562   epsilon: 0.009998020008555413   st
eps: 367    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2316
episode: 2316   score: 6.0   memory length: 657915   epsilon: 0.009998020008555413   st
eps: 353    lr: 1.0240000000000005e-06    evaluation reward: 8.96
2317
episode: 2317   score: 6.0   memory length: 658241   epsilon: 0.009998020008555413   st
eps: 326    lr: 1.0240000000000005e-06    evaluation reward: 8.89
2318
episode: 2318   score: 6.0   memory length: 658577   epsilon: 0.009998020008555413   st
eps: 336    lr: 1.0240000000000005e-06    evaluation reward: 8.87
2319
episode: 2319   score: 6.0   memory length: 658950   epsilon: 0.009998020008555413   st
eps: 373    lr: 1.0240000000000005e-06    evaluation reward: 8.81
2320
episode: 2320   score: 16.0   memory length: 659564   epsilon: 0.009998020008555413   s
teps: 614    lr: 1.0240000000000005e-06    evaluation reward: 8.89
2321
episode: 2321   score: 7.0   memory length: 659972   epsilon: 0.009998020008555413   st
eps: 408    lr: 1.0240000000000005e-06    evaluation reward: 8.86
2322
episode: 2322   score: 10.0   memory length: 660504   epsilon: 0.009998020008555413   s
teps: 532    lr: 1.0240000000000005e-06    evaluation reward: 8.79
2323
episode: 2323   score: 10.0   memory length: 661010   epsilon: 0.009998020008555413   s
teps: 506    lr: 1.0240000000000005e-06    evaluation reward: 8.76
2324
episode: 2324   score: 14.0   memory length: 661577   epsilon: 0.009998020008555413   s
teps: 567    lr: 1.0240000000000005e-06    evaluation reward: 8.71
2325
episode: 2325   score: 9.0   memory length: 662014   epsilon: 0.009998020008555413   st
eps: 437    lr: 1.0240000000000005e-06    evaluation reward: 8.72
2326
episode: 2326   score: 6.0   memory length: 662327   epsilon: 0.009998020008555413   st
eps: 313    lr: 1.0240000000000005e-06    evaluation reward: 8.65
2327
episode: 2327   score: 11.0   memory length: 662846   epsilon: 0.009998020008555413   s
teps: 519    lr: 1.0240000000000005e-06    evaluation reward: 8.67
```

```
2328
episode: 2328   score: 6.0    memory length: 663159   epsilon: 0.009998020008555413    st
eps: 313    lr: 1.0240000000000005e-06    evaluation reward: 8.7
2329
episode: 2329   score: 15.0    memory length: 663599   epsilon: 0.009998020008555413    s
teps: 440    lr: 1.0240000000000005e-06    evaluation reward: 8.8
2330
episode: 2330   score: 11.0    memory length: 663999   epsilon: 0.009998020008555413    s
teps: 400    lr: 1.0240000000000005e-06    evaluation reward: 8.8
2331
episode: 2331   score: 6.0    memory length: 664335   epsilon: 0.009998020008555413    st
eps: 336    lr: 1.0240000000000005e-06    evaluation reward: 8.81
2332
episode: 2332   score: 12.0    memory length: 664964   epsilon: 0.009998020008555413    s
teps: 629    lr: 1.0240000000000005e-06    evaluation reward: 8.88
2333
episode: 2333   score: 10.0    memory length: 665466   epsilon: 0.009998020008555413    s
teps: 502    lr: 1.0240000000000005e-06    evaluation reward: 8.91
2334
episode: 2334   score: 10.0    memory length: 666013   epsilon: 0.009998020008555413    s
teps: 547    lr: 1.0240000000000005e-06    evaluation reward: 8.96
2335
episode: 2335   score: 6.0    memory length: 666359   epsilon: 0.009998020008555413    st
eps: 346    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2336
episode: 2336   score: 11.0    memory length: 666761   epsilon: 0.009998020008555413    s
teps: 402    lr: 1.0240000000000005e-06    evaluation reward: 8.96
2337
episode: 2337   score: 8.0    memory length: 667193   epsilon: 0.009998020008555413    st
eps: 432    lr: 1.0240000000000005e-06    evaluation reward: 8.97
2338
episode: 2338   score: 8.0    memory length: 667636   epsilon: 0.009998020008555413    st
eps: 443    lr: 1.0240000000000005e-06    evaluation reward: 8.98
2339
episode: 2339   score: 4.0    memory length: 667915   epsilon: 0.009998020008555413    st
eps: 279    lr: 1.0240000000000005e-06    evaluation reward: 8.87
2340
episode: 2340   score: 10.0    memory length: 668446   epsilon: 0.009998020008555413    s
teps: 531    lr: 1.0240000000000005e-06    evaluation reward: 8.82
2341
episode: 2341   score: 9.0    memory length: 668860   epsilon: 0.009998020008555413    st
eps: 414    lr: 1.0240000000000005e-06    evaluation reward: 8.83
2342
episode: 2342   score: 5.0    memory length: 669154   epsilon: 0.009998020008555413    st
eps: 294    lr: 1.0240000000000005e-06    evaluation reward: 8.71
2343
episode: 2343   score: 14.0    memory length: 669827   epsilon: 0.009998020008555413    s
teps: 673    lr: 1.0240000000000005e-06    evaluation reward: 8.74
2344
episode: 2344   score: 11.0    memory length: 670229   epsilon: 0.009998020008555413    s
teps: 402    lr: 1.0240000000000005e-06    evaluation reward: 8.78
2345
episode: 2345   score: 9.0    memory length: 670634   epsilon: 0.009998020008555413    st
eps: 405    lr: 1.0240000000000005e-06    evaluation reward: 8.81
2346
episode: 2346   score: 11.0    memory length: 671034   epsilon: 0.009998020008555413    s
teps: 400    lr: 1.0240000000000005e-06    evaluation reward: 8.81
2347
episode: 2347   score: 12.0    memory length: 671571   epsilon: 0.009998020008555413    s
teps: 537    lr: 1.0240000000000005e-06    evaluation reward: 8.77
2348
episode: 2348   score: 11.0    memory length: 672078   epsilon: 0.009998020008555413    s
teps: 507    lr: 1.0240000000000005e-06    evaluation reward: 8.78
2349
episode: 2349   score: 11.0    memory length: 672495   epsilon: 0.009998020008555413    s
teps: 417    lr: 1.0240000000000005e-06    evaluation reward: 8.77
```

```
2350
episode: 2350   score: 7.0   memory length: 672904   epsilon: 0.009998020008555413     st
eps: 409    lr: 1.0240000000000005e-06     evaluation reward: 8.73
2351
episode: 2351   score: 7.0   memory length: 673313   epsilon: 0.009998020008555413     st
eps: 409    lr: 1.0240000000000005e-06     evaluation reward: 8.74
2352
episode: 2352   score: 7.0   memory length: 673732   epsilon: 0.009998020008555413     st
eps: 419    lr: 1.0240000000000005e-06     evaluation reward: 8.69
2353
episode: 2353   score: 6.0   memory length: 674105   epsilon: 0.009998020008555413     st
eps: 373    lr: 1.0240000000000005e-06     evaluation reward: 8.68
2354
episode: 2354   score: 7.0   memory length: 674514   epsilon: 0.009998020008555413     st
eps: 409    lr: 1.0240000000000005e-06     evaluation reward: 8.64
2355
episode: 2355   score: 9.0   memory length: 674993   epsilon: 0.009998020008555413     st
eps: 479    lr: 1.0240000000000005e-06     evaluation reward: 8.67
2356
episode: 2356   score: 12.0   memory length: 675455   epsilon: 0.009998020008555413     s
teps: 462    lr: 1.0240000000000005e-06     evaluation reward: 8.73
2357
episode: 2357   score: 7.0   memory length: 675823   epsilon: 0.009998020008555413     st
eps: 368    lr: 1.0240000000000005e-06     evaluation reward: 8.7
2358
episode: 2358   score: 10.0   memory length: 676370   epsilon: 0.009998020008555413     s
teps: 547    lr: 1.0240000000000005e-06     evaluation reward: 8.75
2359
episode: 2359   score: 7.0   memory length: 676779   epsilon: 0.009998020008555413     st
eps: 409    lr: 1.0240000000000005e-06     evaluation reward: 8.72
2360
episode: 2360   score: 16.0   memory length: 677463   epsilon: 0.009998020008555413     s
teps: 684    lr: 1.0240000000000005e-06     evaluation reward: 8.82
2361
episode: 2361   score: 8.0   memory length: 677900   epsilon: 0.009998020008555413     st
eps: 437    lr: 1.0240000000000005e-06     evaluation reward: 8.83
2362
episode: 2362   score: 6.0   memory length: 678273   epsilon: 0.009998020008555413     st
eps: 373    lr: 1.0240000000000005e-06     evaluation reward: 8.82
2363
episode: 2363   score: 7.0   memory length: 678682   epsilon: 0.009998020008555413     st
eps: 409    lr: 1.0240000000000005e-06     evaluation reward: 8.81
2364
episode: 2364   score: 10.0   memory length: 679152   epsilon: 0.009998020008555413     s
teps: 470    lr: 1.0240000000000005e-06     evaluation reward: 8.8
2365
episode: 2365   score: 7.0   memory length: 679561   epsilon: 0.009998020008555413     st
eps: 409    lr: 1.0240000000000005e-06     evaluation reward: 8.8
2366
episode: 2366   score: 6.0   memory length: 679934   epsilon: 0.009998020008555413     st
eps: 373    lr: 1.0240000000000005e-06     evaluation reward: 8.82
2367
episode: 2367   score: 4.0   memory length: 680200   epsilon: 0.009998020008555413     st
eps: 266    lr: 1.0240000000000005e-06     evaluation reward: 8.79
2368
episode: 2368   score: 11.0   memory length: 680602   epsilon: 0.009998020008555413     s
teps: 402    lr: 1.0240000000000005e-06     evaluation reward: 8.86
2369
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
Cell In[9], line 47
     45 # Start training after random sample generation
     46 if(frame >= train_frame): # You can set train_frame to a lower value while testi
ng your starts training earlier
---> 47     agent.train_policy_net(frame)
```

```
    48         # Update the target network only for Double DQN only
    49         if double_dqn and (frame % update_target_network_frequency)== 0:

File /kaggle/input/mp5dataset-2/assignment5_materials/agent_double.py:72, in Agent.train
_policy_net(self, frame)
    69 if self.epsilon > self.epsilon_min:
    70     self.epsilon -= self.epsilon_decay
---> 72 mini_batch = self.memory.sample_mini_batch(frame)
    73 mini_batch = np.array(mini_batch, dtype=object).transpose()
    75 history = np.stack(mini_batch[0], axis=0)

File /kaggle/input/mp5dataset-2/assignment5_materials/memory.py:30, in ReplayMemory.samp
le_mini_batch(self, frame)
    27     for j in range(HISTORY_SIZE + 1):
    28         sample.append(self.memory[i + j])
---> 30     sample = np.array(sample, dtype=object)
    31     mini_batch.append((np.stack(sample[:, 0], axis=0), sample[3, 1], sample[3, 2
], sample[3, 3]))
    33 return mini_batch

File /opt/conda/lib/python3.10/site-packages/torch/_tensor.py:1032, in Tensor.__array__
(self, dtype)
   1030     return self.numpy()
   1031 else:
-> 1032     return self.numpy().astype(dtype, copy=False)

KeyboardInterrupt:
Error in callback <function flush_figures at 0x78a150ce23b0> (for post_execute), with ar
guments args (),kwargs {}:
    ----------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
File /opt/conda/lib/python3.10/site-packages/matplotlib_inline/backend_inline.py:126, in
 flush_figures()
    123 if InlineBackend.instance().close_figures:
    124     # ignore the tracking, just draw and close all figures
    125     try:
--> 126         return show(True)
    127     except Exception as e:
    128         # safely show traceback if in IPython, else raise
    129         ip = get_ipython()

File /opt/conda/lib/python3.10/site-packages/matplotlib_inline/backend_inline.py:90, in
show(close, block)
    88 try:
    89     for figure_manager in Gcf.get_all_fig_managers():
---> 90         display(
    91             figure_manager.canvas.figure,
    92             metadata=_fetch_figure_metadata(figure_manager.canvas.figure)
    93         )
    94 finally:
    95     show._to_draw = []

File /opt/conda/lib/python3.10/site-packages/IPython/core/display_functions.py:298, in d
isplay(include, exclude, metadata, transient, display_id, raw, clear, *objs, **kwargs)
    296     publish_display_data(data=obj, metadata=metadata, **kwargs)
    297 else:
--> 298     format_dict, md_dict = format(obj, include=include, exclude=exclude)
    299     if not format_dict:
    300         # nothing to display (e.g. _ipython_display_ took over)
    301         continue

File /opt/conda/lib/python3.10/site-packages/IPython/core/formatters.py:179, in DisplayF
ormatter.format(self, obj, include, exclude)
    177 md = None
    178 try:
```

```
--> 179         data = formatter(obj)
    180 except:
    181     # FIXME: log the exception
    182     raise

File /opt/conda/lib/python3.10/site-packages/decorator.py:232, in decorate.<locals>.fun
(*args, **kw)
    230 if not kwsyntax:
    231     args, kw = fix(args, kw, sig)
--> 232 return caller(func, *(extras + args), **kw)

File /opt/conda/lib/python3.10/site-packages/IPython/core/formatters.py:223, in catch_fo
rmat_error(method, self, *args, **kwargs)
    221 """show traceback on failed format call"""
    222 try:
--> 223     r = method(self, *args, **kwargs)
    224 except NotImplementedError:
    225     # don't warn on NotImplementedErrors
    226     return self._check_return(None, args[0])

File /opt/conda/lib/python3.10/site-packages/IPython/core/formatters.py:340, in BaseForm
atter.__call__(self, obj)
    338         pass
    339     else:
--> 340         return printer(obj)
    341 # Finally look for special method names
    342 method = get_real_method(obj, self.print_method)

File /opt/conda/lib/python3.10/site-packages/IPython/core/pylabtools.py:152, in print_fi
gure(fig, fmt, bbox_inches, base64, **kwargs)
    149     from matplotlib.backend_bases import FigureCanvasBase
    150     FigureCanvasBase(fig)
--> 152 fig.canvas.print_figure(bytes_io, **kw)
    153 data = bytes_io.getvalue()
    154 if fmt == 'svg':

File /opt/conda/lib/python3.10/site-packages/matplotlib/backend_bases.py:2366, in Figure
CanvasBase.print_figure(self, filename, dpi, facecolor, edgecolor, orientation, format,
 bbox_inches, pad_inches, bbox_extra_artists, backend, **kwargs)
   2362 try:
   2363     # _get_renderer may change the figure dpi (as vector formats
   2364     # force the figure dpi to 72), so we need to set it again here.
   2365     with cbook._setattr_cm(self.figure, dpi=dpi):
-> 2366         result = print_method(
   2367             filename,
   2368             facecolor=facecolor,
   2369             edgecolor=edgecolor,
   2370             orientation=orientation,
   2371             bbox_inches_restore=_bbox_inches_restore,
   2372             **kwargs)
   2373 finally:
   2374     if bbox_inches and restore_bbox:

File /opt/conda/lib/python3.10/site-packages/matplotlib/backend_bases.py:2232, in Figure
CanvasBase._switch_canvas_and_return_print_method.<locals>.<lambda>(*args, **kwargs)
   2228     optional_kws = {  # Passed by print_figure for other renderers.
   2229         "dpi", "facecolor", "edgecolor", "orientation",
   2230         "bbox_inches_restore"}
   2231     skip = optional_kws - {*inspect.signature(meth).parameters}
-> 2232     print_method = functools.wraps(meth)(lambda *args, **kwargs: meth(
   2233         *args, **{k: v for k, v in kwargs.items() if k not in skip}))
   2234 else:  # Let third-parties do as they see fit.
   2235     print_method = meth

File /opt/conda/lib/python3.10/site-packages/matplotlib/backends/backend_agg.py:509, in
FigureCanvasAgg.print_png(self, filename_or_obj, metadata, pil_kwargs)
```

```
    462 def print_png(self, filename_or_obj, *, metadata=None, pil_kwargs=None):
    463     """
    464     Write the figure to a PNG file.
    465
(...)
    507         *metadata*, including the default 'Software' key.
    508     """
--> 509     self._print_pil(filename_or_obj, "png", pil_kwargs, metadata)

File /opt/conda/lib/python3.10/site-packages/matplotlib/backends/backend_agg.py:457, in
FigureCanvasAgg._print_pil(self, filename_or_obj, fmt, pil_kwargs, metadata)
    452 def _print_pil(self, filename_or_obj, fmt, pil_kwargs, metadata=None):
    453     """
    454     Draw the canvas, then save it using `.image.imsave` (to which
    455     *pil_kwargs* and *metadata* are forwarded).
    456     """
--> 457     FigureCanvasAgg.draw(self)
    458     mpl.image.imsave(
    459         filename_or_obj, self.buffer_rgba(), format=fmt, origin="upper",
    460         dpi=self.figure.dpi, metadata=metadata, pil_kwargs=pil_kwargs)

File /opt/conda/lib/python3.10/site-packages/matplotlib/backends/backend_agg.py:400, in
FigureCanvasAgg.draw(self)
    396 # Acquire a lock on the shared font cache.
    397 with RendererAgg.lock, \
    398      (self.toolbar._wait_cursor_for_draw_cm() if self.toolbar
    399       else nullcontext()):
--> 400     self.figure.draw(self.renderer)
    401     # A GUI class may be need to update a window using this draw, so
    402     # don't forget to call the superclass.
    403     super().draw()

File /opt/conda/lib/python3.10/site-packages/matplotlib/artist.py:95, in _finalize_raste
rization.<locals>.draw_wrapper(artist, renderer, *args, **kwargs)
     93 @wraps(draw)
     94 def draw_wrapper(artist, renderer, *args, **kwargs):
---> 95     result = draw(artist, renderer, *args, **kwargs)
     96     if renderer._rasterizing:
     97         renderer.stop_rasterizing()

File /opt/conda/lib/python3.10/site-packages/matplotlib/artist.py:72, in allow_rasteriza
tion.<locals>.draw_wrapper(artist, renderer)
     69     if artist.get_agg_filter() is not None:
     70         renderer.start_filter()
---> 72     return draw(artist, renderer)
     73 finally:
     74     if artist.get_agg_filter() is not None:

File /opt/conda/lib/python3.10/site-packages/matplotlib/figure.py:3175, in Figure.draw(s
elf, renderer)
   3172         # ValueError can occur when resizing a window.
   3174 self.patch.draw(renderer)
-> 3175 mimage._draw_list_compositing_images(
   3176     renderer, self, artists, self.suppressComposite)
   3178 for sfig in self.subfigs:
   3179     sfig.draw(renderer)

File /opt/conda/lib/python3.10/site-packages/matplotlib/image.py:131, in _draw_list_comp
ositing_images(renderer, parent, artists, suppress_composite)
    129 if not_composite or not has_images:
    130     for a in artists:
--> 131         a.draw(renderer)
    132 else:
    133     # Composite any adjacent images together
    134     image_group = []
```

```
File /opt/conda/lib/python3.10/site-packages/matplotlib/artist.py:72, in allow_rasteriza
tion.<locals>.draw_wrapper(artist, renderer)
     69         if artist.get_agg_filter() is not None:
     70             renderer.start_filter()
---> 72         return draw(artist, renderer)
     73     finally:
     74         if artist.get_agg_filter() is not None:

File /opt/conda/lib/python3.10/site-packages/matplotlib/axes/_base.py:3064, in _AxesBas
e.draw(self, renderer)
   3061     if artists_rasterized:
   3062         _draw_rasterized(self.figure, artists_rasterized, renderer)
-> 3064     mimage._draw_list_compositing_images(
   3065         renderer, self, artists, self.figure.suppressComposite)
   3067     renderer.close_group('axes')
   3068     self.stale = False

File /opt/conda/lib/python3.10/site-packages/matplotlib/image.py:131, in _draw_list_comp
ositing_images(renderer, parent, artists, suppress_composite)
    129     if not_composite or not has_images:
    130         for a in artists:
--> 131             a.draw(renderer)
    132     else:
    133         # Composite any adjacent images together
    134         image_group = []

File /opt/conda/lib/python3.10/site-packages/matplotlib/artist.py:72, in allow_rasteriza
tion.<locals>.draw_wrapper(artist, renderer)
     69         if artist.get_agg_filter() is not None:
     70             renderer.start_filter()
---> 72         return draw(artist, renderer)
     73     finally:
     74         if artist.get_agg_filter() is not None:

File /opt/conda/lib/python3.10/site-packages/matplotlib/lines.py:800, in Line2D.draw(sel
f, renderer)
    797             renderer.draw_path(gc, tpath, affine.frozen())
    798             gc.restore()
--> 800     if self._marker and self._markersize > 0:
    801         gc = renderer.new_gc()
    802         self._set_gc_clip(gc)

File /opt/conda/lib/python3.10/site-packages/matplotlib/markers.py:294, in MarkerStyle._
_bool__(self)
    293     def __bool__(self):
--> 294         return bool(len(self._path.vertices))

File /opt/conda/lib/python3.10/site-packages/matplotlib/path.py:211, in Path.vertices(se
lf)
    203         self._simplify_threshold = mpl.rcParams['path.simplify_threshold']
    204         self._should_simplify = (
    205             self._simplify_threshold > 0 and
    206             mpl.rcParams['path.simplify'] and
    207             len(self._vertices) >= 128 and
    208             (self._codes is None or np.all(self._codes <= Path.LINETO))
    209         )
--> 211     @property
    212     def vertices(self):
    213         """
    214         The list of vertices in the `Path` as an Nx2 numpy array.
    215         """
    216         return self._vertices

KeyboardInterrupt:
```

Visualize Agent Performance

BE AWARE THIS CODE BELOW MAY CRASH THE KERNEL IF YOU RUN THE SAME CELL TWICE.

Please save your model before running this portion of the code.

```python
In [11]: torch.save(agent.policy_net, "/kaggle/working/breakout_dqn_latest.pth")
```

```python
In [13]: # from gym.wrappers import Monitor # If importing monitor raises issues, try using `from
         from gym.wrappers import RecordVideo
         import glob
         import io
         import base64

         from IPython.display import HTML
         from IPython import display as ipythondisplay

         from pyvirtualdisplay import Display

         # Displaying the game live
         def show_state(env, step=0, info=""):
             plt.figure(3)
             plt.clf()
             plt.imshow(env.render(mode='rgb_array'))
             plt.title("%s | Step: %d %s" % ("Agent Playing",step, info))
             plt.axis('off')

             ipythondisplay.clear_output(wait=True)
             ipythondisplay.display(plt.gcf())

         # Recording the game and replaying the game afterwards
         def show_video():
             mp4list = glob.glob('/kaggle/working/video/*.mp4')
             if len(mp4list) > 0:
                 mp4 = mp4list[0]
                 video = io.open(mp4, 'r+b').read()
                 encoded = base64.b64encode(video)
                 ipythondisplay.display(HTML(data='''<video alt="test" autoplay
                         loop controls style="height: 400px;">
                         <source src="data:video/mp4;base64,{0}" type="video/mp4" />
                     </video>'''.format(encoded.decode('ascii'))))
             else:
                 print("Could not find video")


         def wrap_env(env):
             env = Monitor(env, '/kaggle/working/video', force=True)
             return env
```

```python
In [15]: !pip install moviepy


         import sys
         import gym
         import torch
         import pylab
         import random
         import numpy as np
         from collections import deque
         from datetime import datetime
         from copy import deepcopy
         import torch.nn as nn
```

```python
import torch.optim as optim
import torch.nn.functional as F
from torch.autograd import Variable
from torch.utils.tensorboard import SummaryWriter
from utils import find_max_lives, check_live, get_frame, get_init_state
from model import DQN, DQN_LSTM
from config import *

import matplotlib.pyplot as plt
# %load_ext autoreload
# %autoreload 2

from gym.wrappers import RecordVideo # If importing monitor raises issues, try using `fr
import glob
import io
import base64

from IPython.display import HTML
from IPython import display as ipythondisplay

from pyvirtualdisplay import Display

env = gym.make('BreakoutDeterministic-v4', render_mode='rgb_array')
state = env.reset()
number_lives = find_max_lives(env)
state_size = env.observation_space.shape
action_size = 3 #fire, left, and right

# Displaying the game live
def show_state(env, step=0, info=""):
    plt.figure(3)
    plt.clf()
    plt.imshow(env.render(mode='rgb_array'))
    plt.title("%s | Step: %d %s" % ("Agent Playing",step, info))
    plt.axis('off')

    ipythondisplay.clear_output(wait=True)
    ipythondisplay.display(plt.gcf())

# Recording the game and replaying the game afterwards
def show_video():
    mp4list = glob.glob('video/*.mp4')
    if len(mp4list) > 0:
        mp4 = mp4list[0]
        video = io.open(mp4, 'r+b').read()
        encoded = base64.b64encode(video)
        ipythondisplay.display(HTML(data='''<video alt="test" autoplay
                loop controls style="height: 400px;">
                <source src="data:video/mp4;base64,{0}" type="video/mp4" />
            </video>'''.format(encoded.decode('ascii'))))
    else:
        print("Could not find video")


def wrap_env(env):
    env = RecordVideo(env, '/kaggle/working/video')
    return env

from agent import Agent
action_size = 3

display = Display(visible=0, size=(300, 200))
display.start()

# Load agent
agent = Agent(action_size)
```

```python
agent.load_policy_net("/kaggle/working/breakout_dqn_latest.pth")
agent.epsilon = 0.0 # Set agent to only exploit the best action

env = wrap_env(env)

done = False
score = 0
step = 0
state, _ = env.reset()
next_state = state
life = number_lives
history = np.zeros([5, 84, 84], dtype=np.uint8)
get_init_state(history, state, HISTORY_SIZE)
frame = 0
while not done:
#     show_state(env,step) # uncommenting this provides another way to visualize the gam
    step += 1
    frame += 1

    # Perform a fire action if ball is no longer on screen
    if step > 1 and len(np.unique(next_state[:189] == state[:189])) < 2:
        action = 0
    else:
        action = agent.get_action(np.float32(history[:4, :, :]) / 255.)
    state = next_state

    next_state, reward, done, _, info = env.step(action + 1)

    frame_next_state = get_frame(next_state)
    history[4, :, :] = frame_next_state
    terminal_state = check_live(life, info['lives'])

    life = info['lives']
    r = np.clip(reward, -1, 1)
    r = reward

    # Store the transition in memory
    agent.memory.push(deepcopy(frame_next_state), action, r, terminal_state)
    # Start training after random sample generation
    score += reward

    history[:4, :, :] = history[1:, :, :]
env.close()
show_video()
display.stop()
```

```
Collecting moviepy
  Downloading moviepy-1.0.3.tar.gz (388 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 388.3/388.3 kB 10.2 MB/s eta 0:00:0000:01
  Preparing metadata (setup.py) ... done
Collecting decorator<5.0,>=4.0.2 (from moviepy)
  Downloading decorator-4.4.2-py2.py3-none-any.whl.metadata (4.2 kB)
Requirement already satisfied: tqdm<5.0,>=4.11.2 in /opt/conda/lib/python3.10/site-packa
ges (from moviepy) (4.66.1)
Requirement already satisfied: requests<3.0,>=2.8.1 in /opt/conda/lib/python3.10/site-pa
ckages (from moviepy) (2.31.0)
Collecting proglog<=1.0.0 (from moviepy)
  Downloading proglog-0.1.10-py3-none-any.whl.metadata (639 bytes)
Requirement already satisfied: numpy>=1.17.3 in /opt/conda/lib/python3.10/site-packages
 (from moviepy) (1.26.4)
Requirement already satisfied: imageio<3.0,>=2.5 in /opt/conda/lib/python3.10/site-packa
ges (from moviepy) (2.33.1)
Collecting imageio_ffmpeg>=0.2.0 (from moviepy)
  Downloading imageio_ffmpeg-0.4.9-py3-none-manylinux2010_x86_64.whl.metadata (1.7 kB)
Requirement already satisfied: pillow>=8.3.2 in /opt/conda/lib/python3.10/site-packages
 (from imageio<3.0,>=2.5->moviepy) (9.5.0)
```

```
Requirement already satisfied: setuptools in /root/.local/lib/python3.10/site-packages
 (from imageio_ffmpeg>=0.2.0->moviepy) (69.5.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.10/sit
e-packages (from requests<3.0,>=2.8.1->moviepy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages
 (from requests<3.0,>=2.8.1->moviepy) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.10/site-pack
ages (from requests<3.0,>=2.8.1->moviepy) (1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.10/site-pack
ages (from requests<3.0,>=2.8.1->moviepy) (2024.2.2)
Downloading decorator-4.4.2-py2.py3-none-any.whl (9.2 kB)
Downloading imageio_ffmpeg-0.4.9-py3-none-manylinux2010_x86_64.whl (26.9 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 26.9/26.9 MB 63.5 MB/s eta 0:00:00:00:0100:01
Downloading proglog-0.1.10-py3-none-any.whl (6.1 kB)
Building wheels for collected packages: moviepy
  Building wheel for moviepy (setup.py) ... done
  Created wheel for moviepy: filename=moviepy-1.0.3-py3-none-any.whl size=110720 sha256=
969c59913693b7587610bbcfbe47b2d0739b799a51960dc986f27791a4ebcae0
  Stored in directory: /root/.cache/pip/wheels/96/32/2d/e10123bd88fbfc02fed53cc18c80a171
d3c87479ed845fa7c1
Successfully built moviepy
Installing collected packages: proglog, imageio_ffmpeg, decorator, moviepy
  Attempting uninstall: decorator
    Found existing installation: decorator 5.1.1
    Uninstalling decorator-5.1.1:
      Successfully uninstalled decorator-5.1.1
Successfully installed decorator-4.4.2 imageio_ffmpeg-0.4.9 moviepy-1.0.3 proglog-0.1.10
/opt/conda/lib/python3.10/site-packages/gym/utils/passive_env_checker.py:233: Deprecatio
nWarning: `np.bool8` is a deprecated alias for `np.bool_`.  (Deprecated NumPy 1.24)
  if not isinstance(terminated, (bool, np.bool8)):
/opt/conda/lib/python3.10/site-packages/gym/wrappers/record_video.py:75: UserWarning: WA
RN: Overwriting existing videos at /kaggle/working/video folder (try specifying a differ
ent `video_folder` for the `RecordVideo` wrapper if this is not desired)
  logger.warn(
/opt/conda/lib/python3.10/site-packages/gym/utils/passive_env_checker.py:289: UserWarnin
g: WARN: No render fps was declared in the environment (env.metadata['render_fps'] is No
ne or not defined), rendering may occur at inconsistent fps.
  logger.warn(
Moviepy - Building video /kaggle/working/video/rl-video-episode-0.mp4.
Moviepy - Writing video /kaggle/working/video/rl-video-episode-0.mp4


---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[15], line 106
    103     action = agent.get_action(np.float32(history[:4, :, :]) / 255.)
    104 state = next_state
--> 106 next_state, reward, done, _, info = env.step(action + 1)
    108 frame_next_state = get_frame(next_state)
    109 history[4, :, :] = frame_next_state

File /opt/conda/lib/python3.10/site-packages/gym/wrappers/record_video.py:167, in Record
Video.step(self, action)
    165 if not self.is_vector_env:
    166     if terminateds or truncateds:
--> 167         self.close_video_recorder()
    168 elif terminateds[0] or truncateds[0]:
    169     self.close_video_recorder()

File /opt/conda/lib/python3.10/site-packages/gym/wrappers/record_video.py:180, in Record
Video.close_video_recorder(self)
    178 if self.recording:
    179     assert self.video_recorder is not None
--> 180     self.video_recorder.close()
    181 self.recording = False
    182 self.recorded_frames = 1
```

```
File /opt/conda/lib/python3.10/site-packages/gym/wrappers/monitoring/video_recorder.py:1
58, in VideoRecorder.close(self)
    156     logger.debug(f"Closing video encoder: path={self.path}")
    157     clip = ImageSequenceClip(self.recorded_frames, fps=self.frames_per_sec)
--> 158     clip.write_videofile(self.path)
    159 else:
    160     # No frames captured. Set metadata.
    161     if self.metadata is None:

File /opt/conda/lib/python3.10/site-packages/decorator.py:232, in decorate.<locals>.fun
(*args, **kw)
    230 if not kwsyntax:
    231     args, kw = fix(args, kw, sig)
--> 232 return caller(func, *(extras + args), **kw)

File /opt/conda/lib/python3.10/site-packages/moviepy/decorators.py:54, in requires_durat
ion(f, clip, *a, **k)
     52     raise ValueError("Attribute 'duration' not set")
     53 else:
---> 54     return f(clip, *a, **k)

File /opt/conda/lib/python3.10/site-packages/decorator.py:232, in decorate.<locals>.fun
(*args, **kw)
    230 if not kwsyntax:
    231     args, kw = fix(args, kw, sig)
--> 232 return caller(func, *(extras + args), **kw)

File /opt/conda/lib/python3.10/site-packages/moviepy/decorators.py:135, in use_clip_fps_
by_default(f, clip, *a, **k)
    130 new_a = [fun(arg) if (name=='fps') else arg
    131          for (arg, name) in zip(a, names)]
    132 new_kw = {k: fun(v) if k=='fps' else v
    133          for (k,v) in k.items()}
--> 135 return f(clip, *new_a, **new_kw)

File /opt/conda/lib/python3.10/site-packages/decorator.py:232, in decorate.<locals>.fun
(*args, **kw)
    230 if not kwsyntax:
    231     args, kw = fix(args, kw, sig)
--> 232 return caller(func, *(extras + args), **kw)

File /opt/conda/lib/python3.10/site-packages/moviepy/decorators.py:22, in convert_masks_
to_RGB(f, clip, *a, **k)
     20 if clip.ismask:
     21     clip = clip.to_RGB()
---> 22 return f(clip, *a, **k)

File /opt/conda/lib/python3.10/site-packages/moviepy/video/VideoClip.py:300, in VideoCli
p.write_videofile(self, filename, fps, codec, bitrate, audio, audio_fps, preset, audio_n
bytes, audio_codec, audio_bitrate, audio_bufsize, temp_audiofile, rewrite_audio, remove_
temp, write_logfile, verbose, threads, ffmpeg_params, logger)
    292 if make_audio:
    293     self.audio.write_audiofile(audiofile, audio_fps,
    294                                audio_nbytes, audio_bufsize,
    295                                audio_codec, bitrate=audio_bitrate,
    296                                write_logfile=write_logfile,
    297                                verbose=verbose,
    298                                logger=logger)
--> 300 ffmpeg_write_video(self, filename, fps, codec,
    301                    bitrate=bitrate,
    302                    preset=preset,
    303                    write_logfile=write_logfile,
    304                    audiofile=audiofile,
    305                    verbose=verbose, threads=threads,
    306                    ffmpeg_params=ffmpeg_params,
```

```
   307                          logger=logger)
   309 if remove_temp and make_audio:
   310     if os.path.exists(audiofile):

File /opt/conda/lib/python3.10/site-packages/moviepy/video/io/ffmpeg_writer.py:213, in f
fmpeg_write_video(clip, filename, fps, codec, bitrate, preset, withmask, write_logfile,
 audiofile, verbose, threads, ffmpeg_params, logger)
   211     logfile = None
   212 logger(message='Moviepy - Writing video %s\n' % filename)
--> 213 with FFMPEG_VideoWriter(filename, clip.size, fps, codec = codec,
   214                              preset=preset, bitrate=bitrate, logfile=logfile,
   215                              audiofile=audiofile, threads=threads,
   216                              ffmpeg_params=ffmpeg_params) as writer:
   218     nframes = int(clip.duration*fps)
   220     for t,frame in clip.iter_frames(logger=logger, with_times=True,
   221                                       fps=fps, dtype="uint8"):

File /opt/conda/lib/python3.10/site-packages/moviepy/video/io/ffmpeg_writer.py:88, in FF
MPEG_VideoWriter.__init__(self, filename, size, fps, codec, audiofile, preset, bitrate,
 withmask, logfile, threads, ffmpeg_params)
    77 self.ext = self.filename.split(".")[-1]
    79 # order is important
    80 cmd = [
    81     get_setting("FFMPEG_BINARY"),
    82     '-y',
    83     '-loglevel', 'error' if logfile == sp.PIPE else 'info',
    84     '-f', 'rawvideo',
    85     '-vcodec', 'rawvideo',
    86     '-s', '%dx%d' % (size[0], size[1]),
    87     '-pix_fmt', 'rgba' if withmask else 'rgb24',
---> 88     '-r', '%.02f' % fps,
    89     '-an', '-i', '-'
    90 ]
    91 if audiofile is not None:
    92     cmd.extend([
    93         '-i', audiofile,
    94         '-acodec', 'copy'
    95     ])

TypeError: must be real number, not NoneType
```

```python
display = Display(visible=0, size=(300, 200))
display.start()

# Load agent
# agent.load_policy_net("./save_model/breakout_dqn.pth")
agent.epsilon = 0.0 # Set agent to only exploit the best action

env = gym.make('BreakoutDeterministic-v4')
env = wrap_env(env)

done = False
score = 0
step = 0
state = env.reset()
next_state = state
life = number_lives
history = np.zeros([5, 84, 84], dtype=np.uint8)
get_init_state(history, state)

while not done:

    # Render breakout
    env.render()
#     show_state(env,step) # uncommenting this provides another way to visualize the gam
```

```python
        step += 1
        frame += 1

        # Perform a fire action if ball is no longer on screen
        if step > 1 and len(np.unique(next_state[:189] == state[:189])) < 2:
            action = 0
        else:
            action = agent.get_action(np.float32(history[:4, :, :]) / 255.)
        state = next_state

        next_state, reward, done, info = env.step(action + 1)

        frame_next_state = get_frame(next_state)
        history[4, :, :] = frame_next_state
        terminal_state = check_live(life, info['ale.lives'])

        life = info['ale.lives']
        r = np.clip(reward, -1, 1)
        r = reward

        # Store the transition in memory
        agent.memory.push(deepcopy(frame_next_state), action, r, terminal_state)
        # Start training after random sample generation
        score += reward

        history[:4, :, :] = history[1:, :, :]
env.close()
show_video()
display.stop()
```