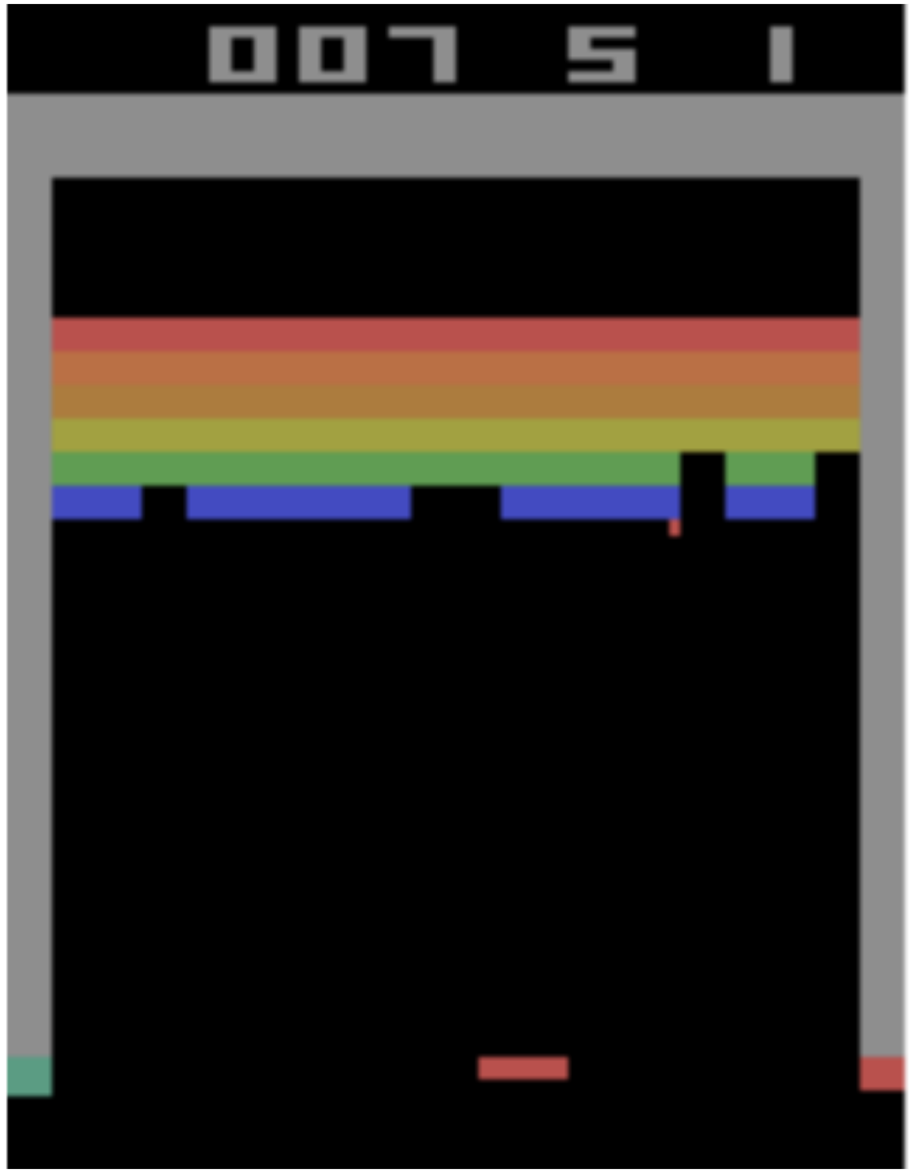
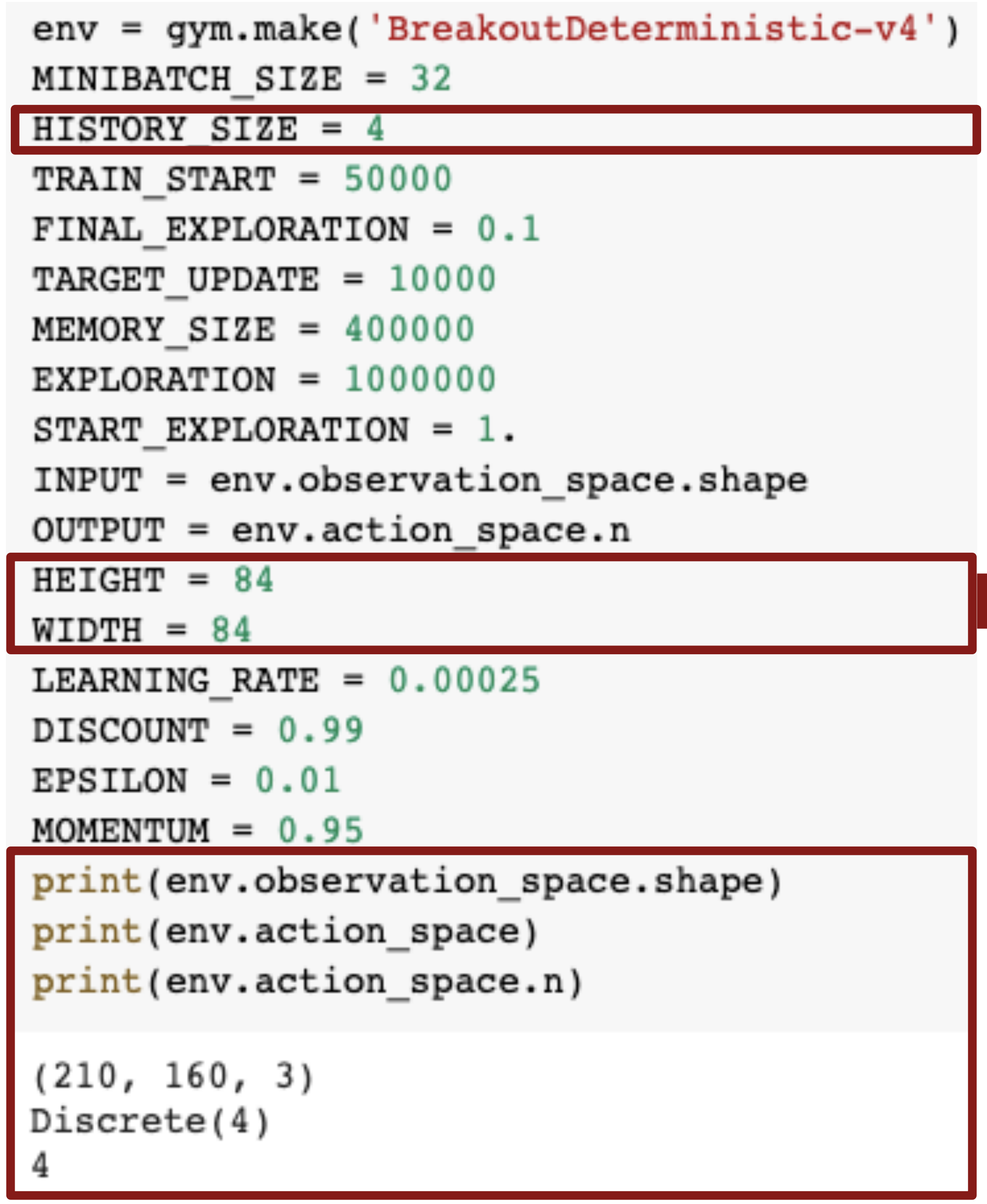
Reinforcement Learning

유승완, 박지윤, 김혜빈,
박정진, 양지현,이혜연

DQN을 이용한 벽돌깨기 (Breakout)

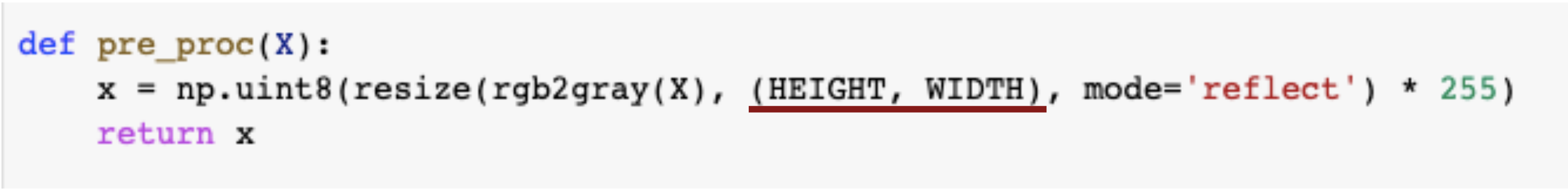


Hyperparameters



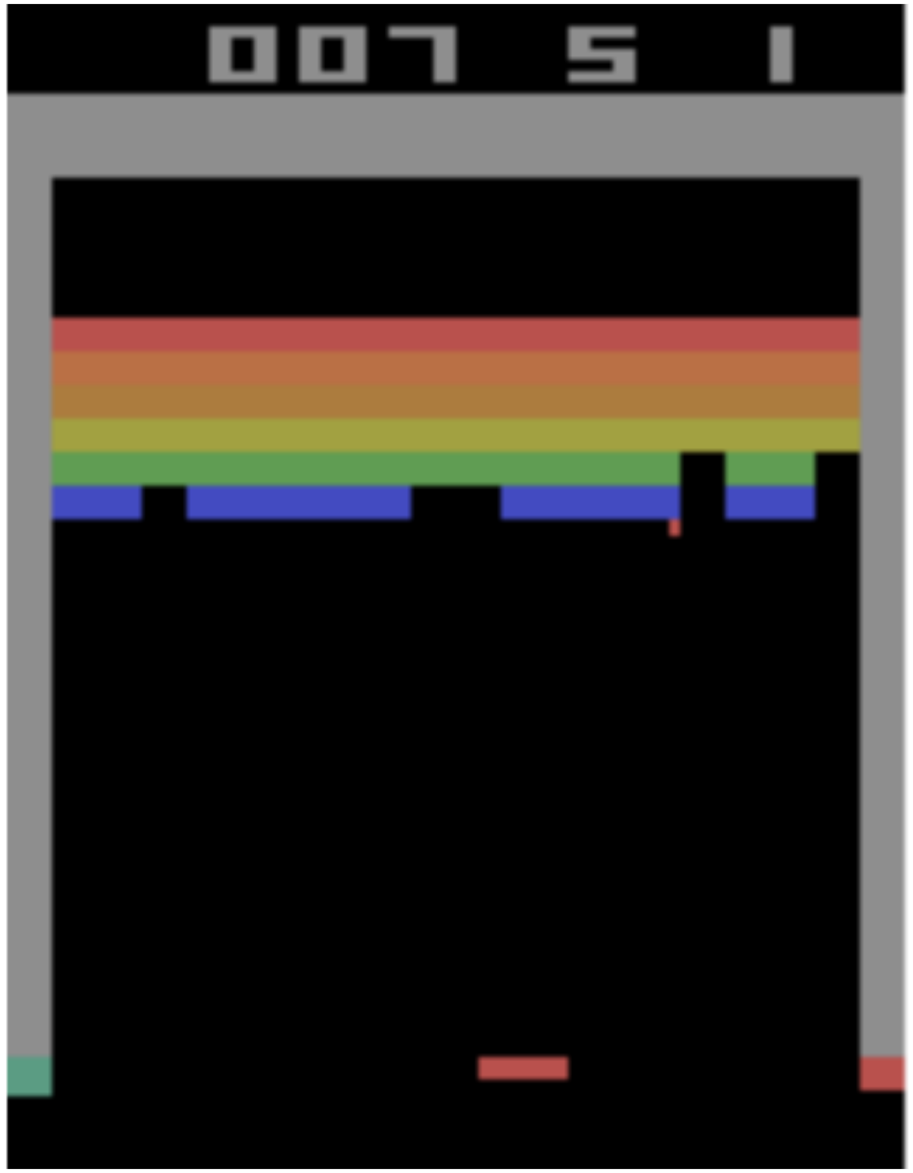
```
env = gym.make('BreakoutDeterministic-v4')
MINIBATCH_SIZE = 32
HISTORY_SIZE = 4
TRAIN_START = 50000
FINAL_EXPLORATION = 0.1
TARGET_UPDATE = 10000
MEMORY_SIZE = 400000
EXPLORATION = 1000000
START_EXPLORATION = 1.
INPUT = env.observation_space.shape
OUTPUT = env.action_space.n
HEIGHT = 84
WIDTH = 84
LEARNING_RATE = 0.00025
DISCOUNT = 0.99
EPSILON = 0.01
MOMENTUM = 0.95
print(env.observation_space.shape)
print(env.action_space)
print(env.action_space.n)
```

```
(210, 160, 3)
Discrete(4)
4
```

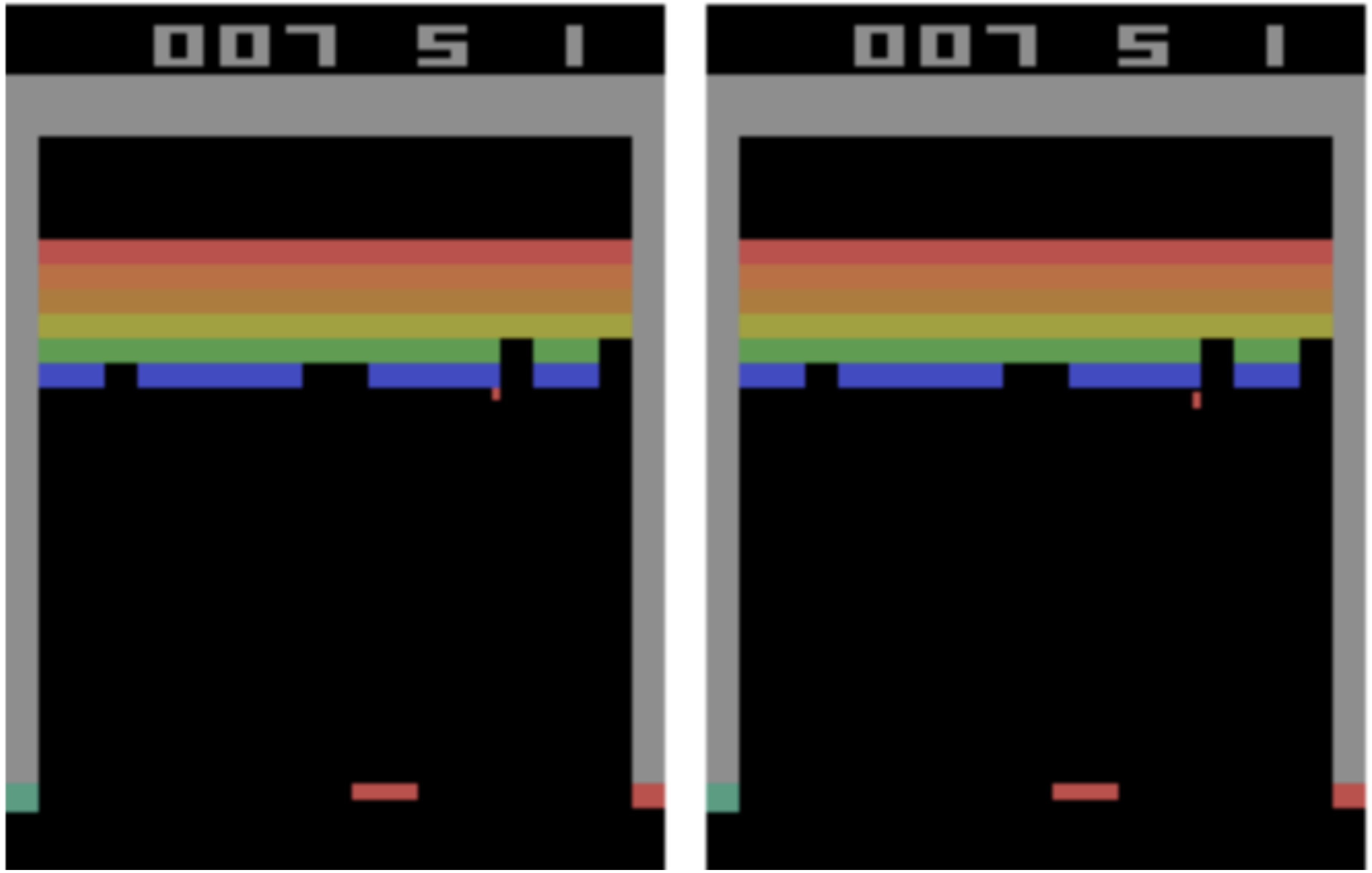


```
def pre_proc(X):
    x = np.uint8(resize(rgb2gray(X), (HEIGHT, WIDTH), mode='reflect') * 255)
    return x
```

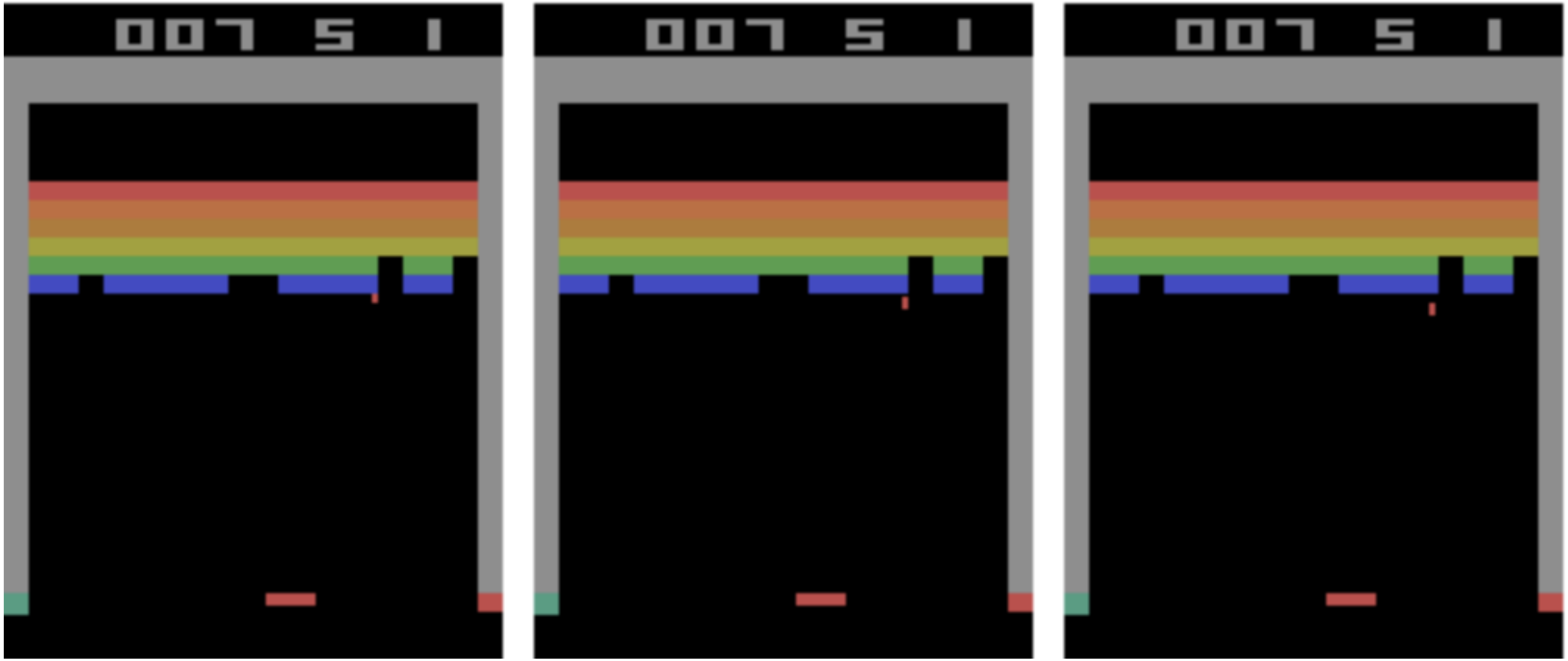
DQN을 이용한 벽돌깨기 (Breakout)



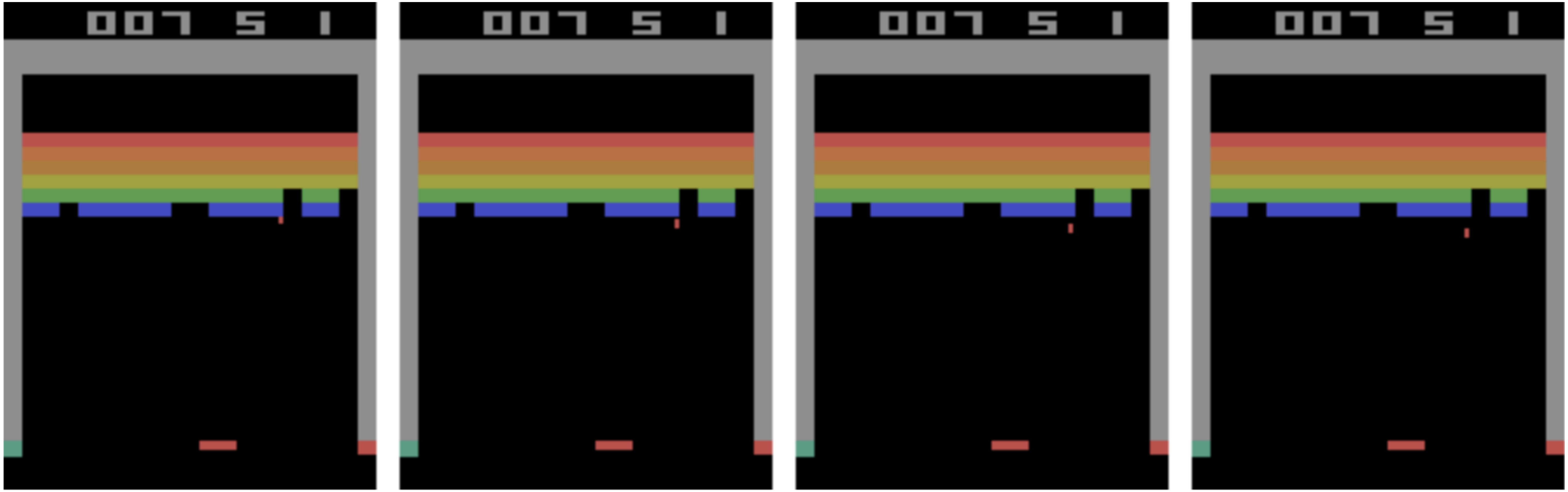
DQN을 이용한 벽돌깨기 (Breakout)



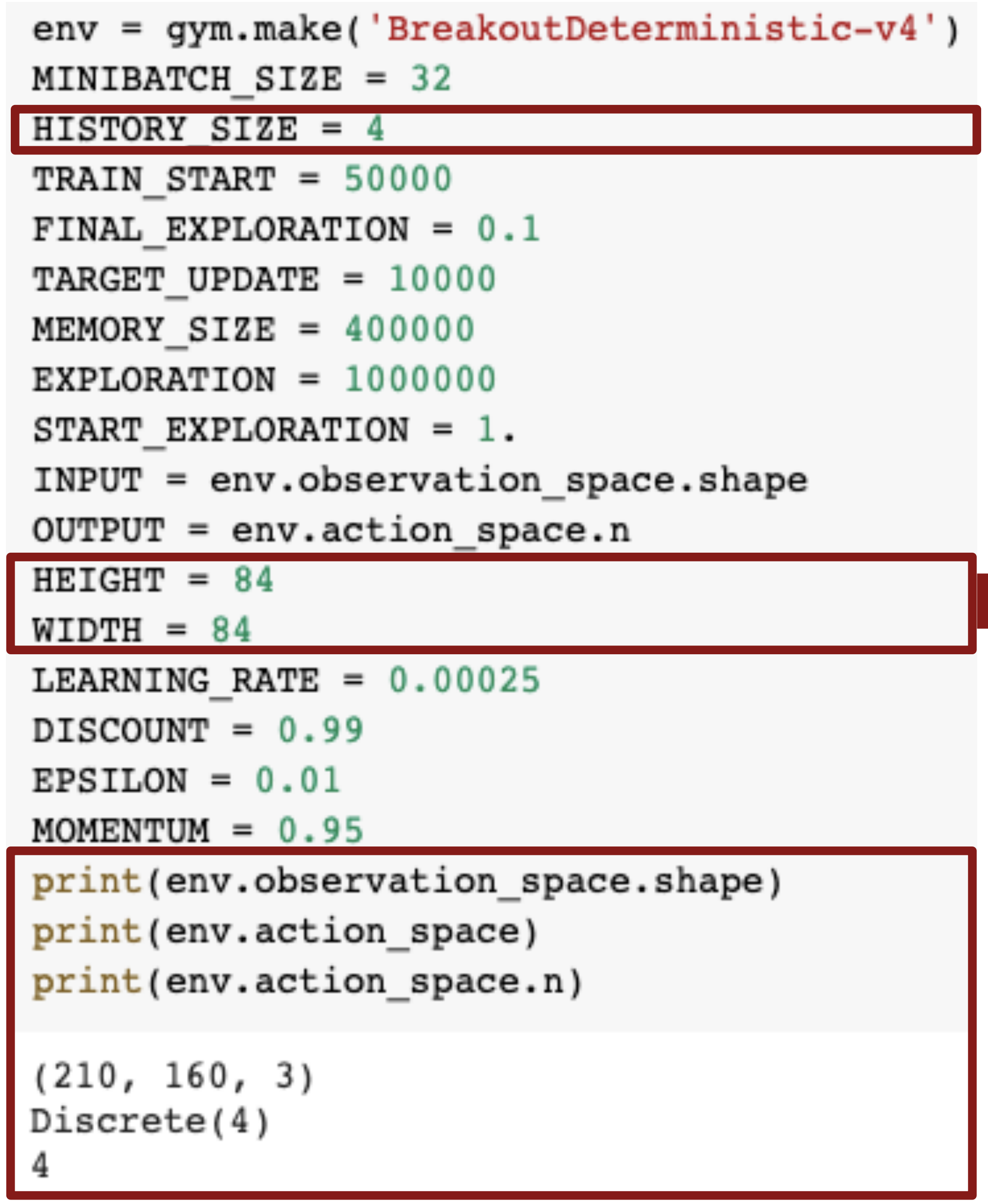
DQN을 이용한 벽돌깨기 (Breakout)



DQN을 이용한 벽돌깨기 (Breakout)

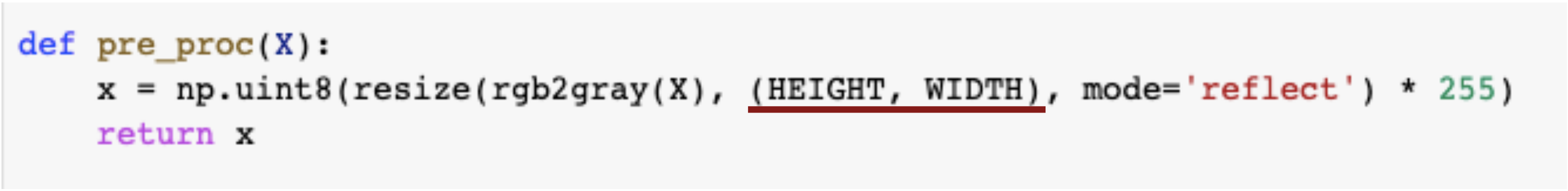


Hyperparameters



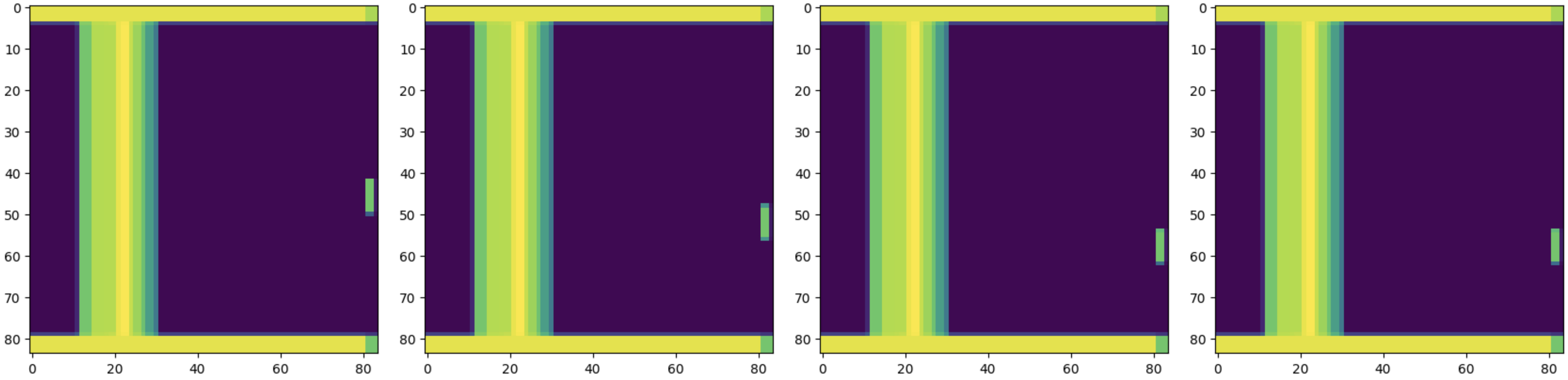
```
env = gym.make('BreakoutDeterministic-v4')
MINIBATCH_SIZE = 32
HISTORY_SIZE = 4
TRAIN_START = 50000
FINAL_EXPLORATION = 0.1
TARGET_UPDATE = 10000
MEMORY_SIZE = 400000
EXPLORATION = 1000000
START_EXPLORATION = 1.
INPUT = env.observation_space.shape
OUTPUT = env.action_space.n
HEIGHT = 84
WIDTH = 84
LEARNING_RATE = 0.00025
DISCOUNT = 0.99
EPSILON = 0.01
MOMENTUM = 0.95
print(env.observation_space.shape)
print(env.action_space)
print(env.action_space.n)

(210, 160, 3)
Discrete(4)
4
```

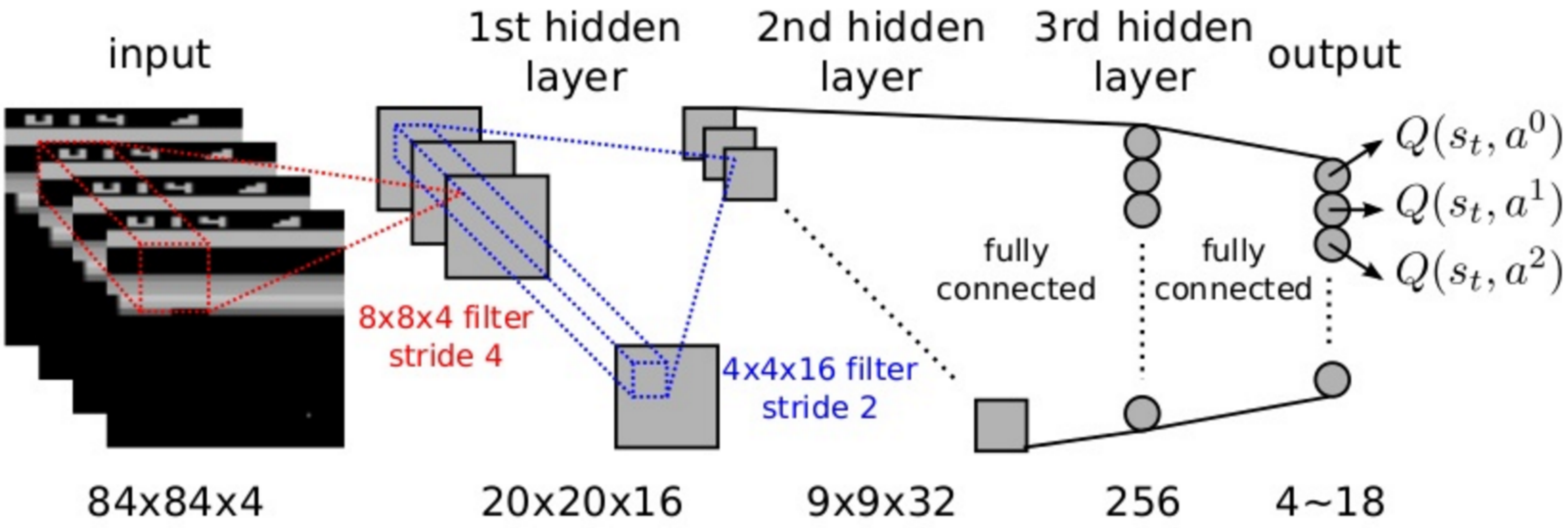


```
def pre_proc(X):
    x = np.uint8(resize(rgb2gray(X), (HEIGHT, WIDTH), mode='reflect') * 255)
    return x
```

DQN을 이용한 벽돌깨기 (Breakout)



CNN

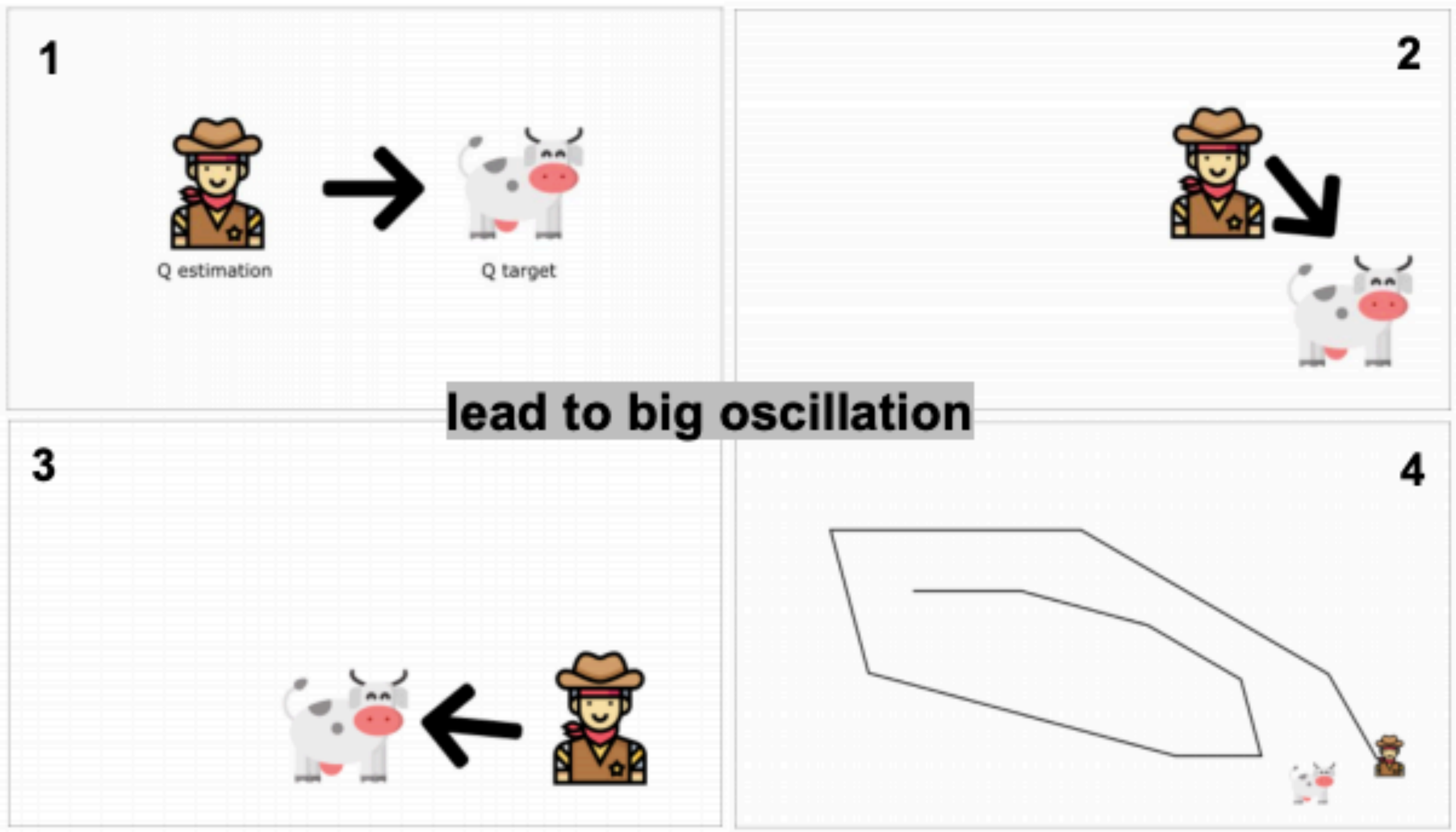

input
1st hidden layer
2nd hidden layer
3rd hidden layer
output
$Q(s_t, a^0)$
$Q(s_t, a^1)$
$Q(s_t, a^2)$
fully connected
fully connected
8x8x4 filter
stride 4
4x4x16 filter
stride 2
84x84x4
20x20x16
9x9x32
256
4~18

Q*-value: Target value

$$Q^*(s,a) = \max_{\pi} \mathbb{E}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots \middle| s_t = s,\ a_t = a,\ \pi\right]$$

루트	첫번째 보상	두번째 보상	총 보상
A	1	10	11
B	5	-1	4

Main & Target Network

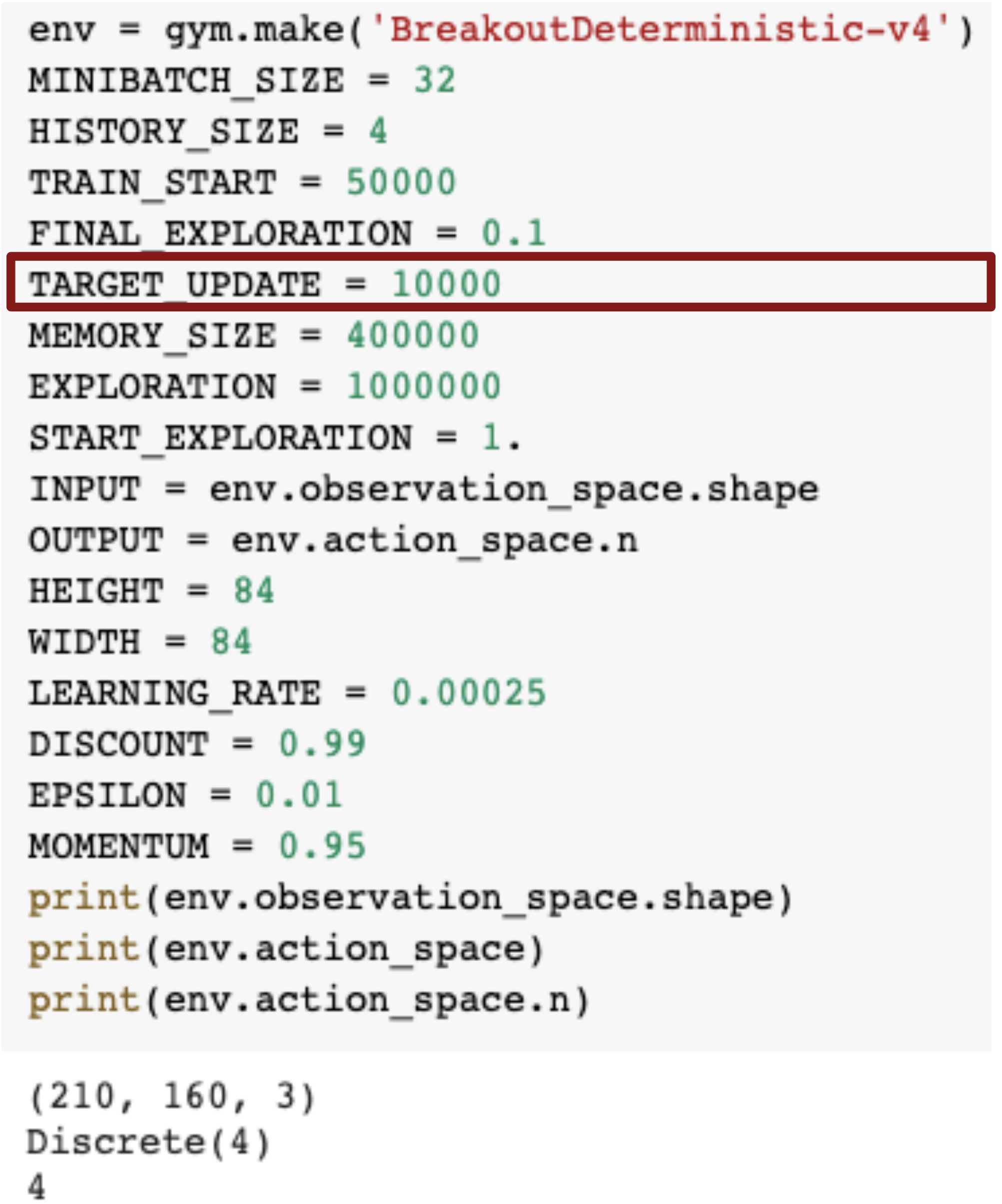


lead to big oscillation

<Fixed Q-targets>

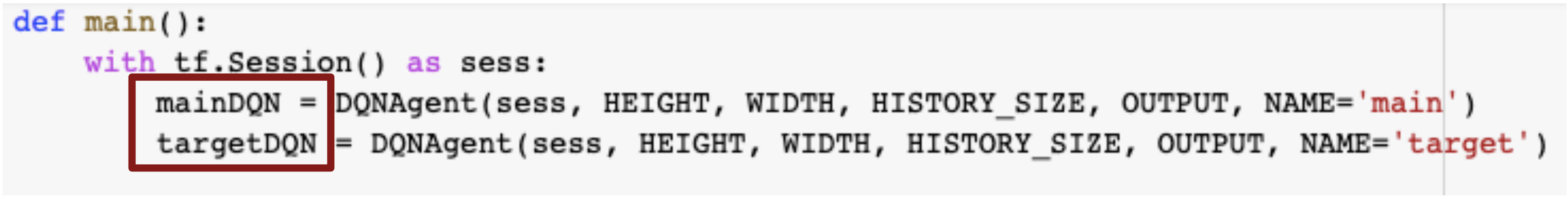
- 학습의 불안정함을 줄이기 위해 같은 구조이지만 다른 parameter를 가진 target network를 만든다.
- Target network parameters는 매 C step마다 Q network parameters로 업데이트 된다.

Hyperparameters



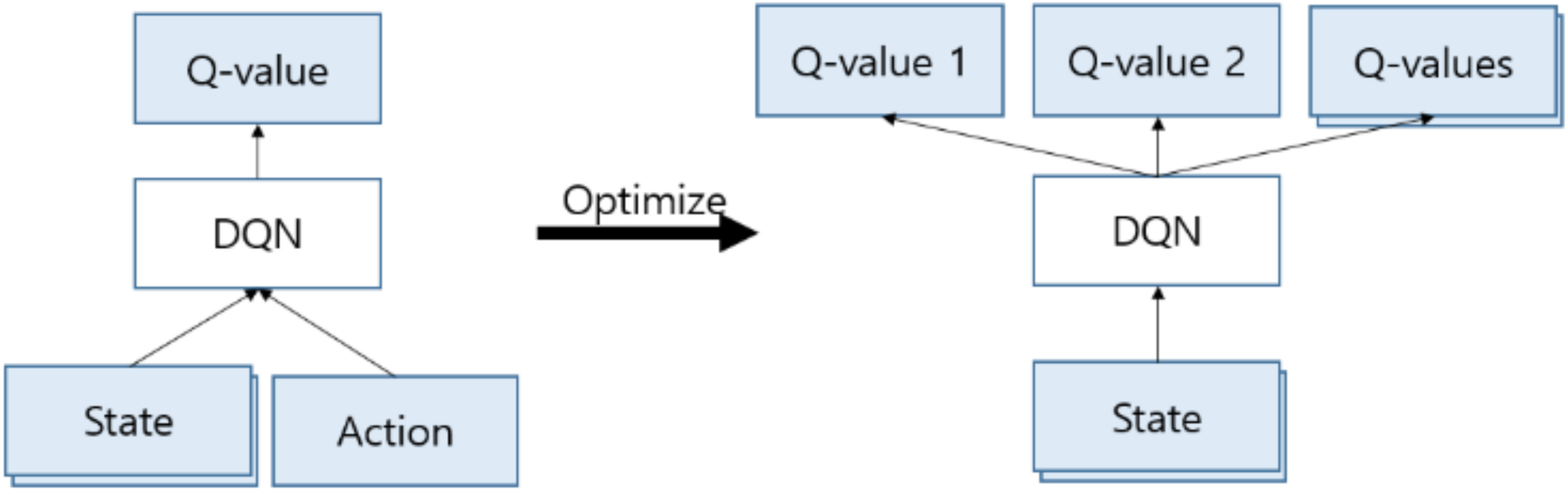
```
env = gym.make('BreakoutDeterministic-v4')
MINIBATCH_SIZE = 32
HISTORY_SIZE = 4
TRAIN_START = 50000
FINAL_EXPLORATION = 0.1
TARGET_UPDATE = 10000
MEMORY_SIZE = 400000
EXPLORATION = 1000000
START_EXPLORATION = 1.
INPUT = env.observation_space.shape
OUTPUT = env.action_space.n
HEIGHT = 84
WIDTH = 84
LEARNING_RATE = 0.00025
DISCOUNT = 0.99
EPSILON = 0.01
MOMENTUM = 0.95
print(env.observation_space.shape)
print(env.action_space)
print(env.action_space.n)
```

```
(210, 160, 3)
Discrete(4)
4
```



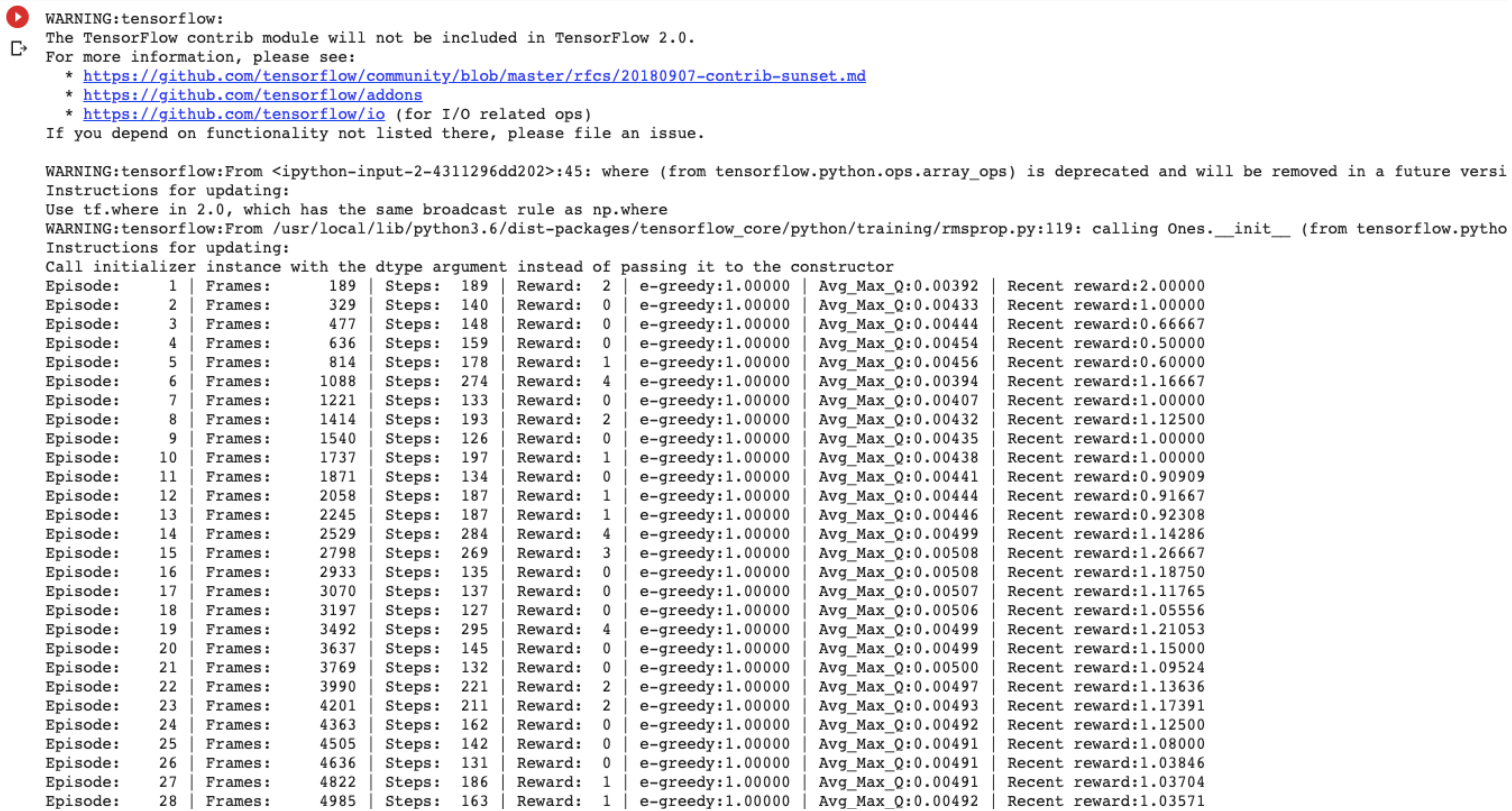
```
def main():
    with tf.Session() as sess:
        mainDQN = DQNAgent(sess, HEIGHT, WIDTH, HISTORY_SIZE, OUTPUT, NAME='main')
        targetDQN = DQNAgent(sess, HEIGHT, WIDTH, HISTORY_SIZE, OUTPUT, NAME='target')
```

Q-value



$$Q(s_t, a_t) \leftarrow R_{t+1} + \gamma Q(s_{t+1}, a') \qquad Q(s_t, a_t) \leftarrow R_{t+1} + \max_a \gamma Q(s_{t+1}, a)$$

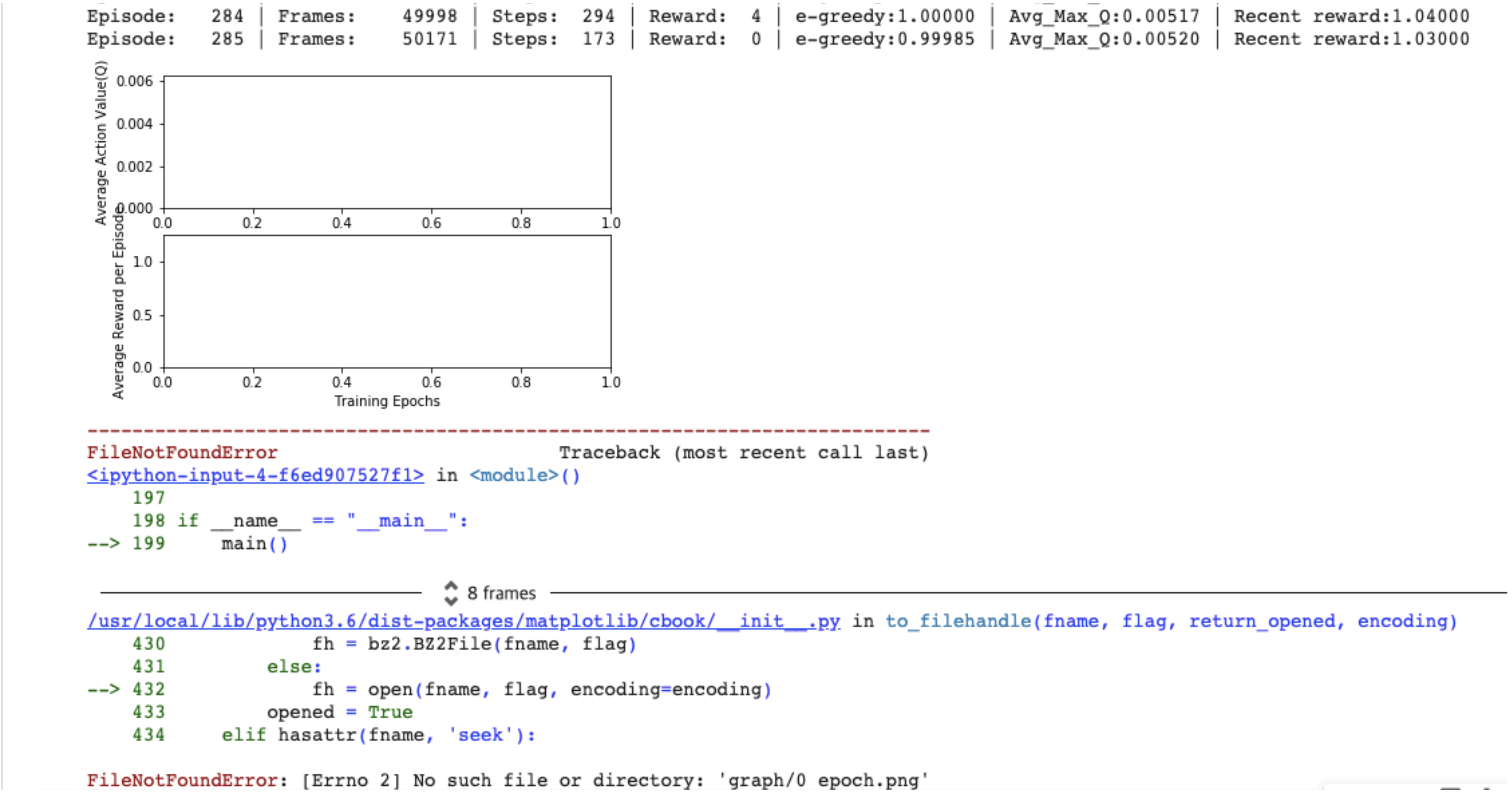
DQN을 이용한 벽돌깨기



```
WARNING:tensorflow:
The TensorFlow contrib module will not be included in TensorFlow 2.0.
For more information, please see:
  * https://github.com/tensorflow/community/blob/master/rfcs/20180907-contrib-sunset.md
  * https://github.com/tensorflow/addons
  * https://github.com/tensorflow/io (for I/O related ops)
If you depend on functionality not listed there, please file an issue.

WARNING:tensorflow:From <ipython-input-2-4311296dd202>:45: where (from tensorflow.python.ops.array_ops) is deprecated and will be removed in a future versi
Instructions for updating:
Use tf.where in 2.0, which has the same broadcast rule as np.where
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow_core/python/training/rmsprop.py:119: calling Ones.__init__ (from tensorflow.pytho
Instructions for updating:
Call initializer instance with the dtype argument instead of passing it to the constructor
Episode:     1 | Frames:      189 | Steps:  189 | Reward:  2 | e-greedy:1.00000 | Avg_Max_Q:0.00392 | Recent reward:2.00000
Episode:     2 | Frames:      329 | Steps:  140 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00433 | Recent reward:1.00000
Episode:     3 | Frames:      477 | Steps:  148 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00444 | Recent reward:0.66667
Episode:     4 | Frames:      636 | Steps:  159 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00454 | Recent reward:0.50000
Episode:     5 | Frames:      814 | Steps:  178 | Reward:  1 | e-greedy:1.00000 | Avg_Max_Q:0.00456 | Recent reward:0.60000
Episode:     6 | Frames:     1088 | Steps:  274 | Reward:  4 | e-greedy:1.00000 | Avg_Max_Q:0.00394 | Recent reward:1.16667
Episode:     7 | Frames:     1221 | Steps:  133 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00407 | Recent reward:1.00000
Episode:     8 | Frames:     1414 | Steps:  193 | Reward:  2 | e-greedy:1.00000 | Avg_Max_Q:0.00432 | Recent reward:1.12500
Episode:     9 | Frames:     1540 | Steps:  126 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00435 | Recent reward:1.00000
Episode:    10 | Frames:     1737 | Steps:  197 | Reward:  1 | e-greedy:1.00000 | Avg_Max_Q:0.00438 | Recent reward:1.00000
Episode:    11 | Frames:     1871 | Steps:  134 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00441 | Recent reward:0.90909
Episode:    12 | Frames:     2058 | Steps:  187 | Reward:  1 | e-greedy:1.00000 | Avg_Max_Q:0.00444 | Recent reward:0.91667
Episode:    13 | Frames:     2245 | Steps:  187 | Reward:  1 | e-greedy:1.00000 | Avg_Max_Q:0.00446 | Recent reward:0.92308
Episode:    14 | Frames:     2529 | Steps:  284 | Reward:  4 | e-greedy:1.00000 | Avg_Max_Q:0.00499 | Recent reward:1.14286
Episode:    15 | Frames:     2798 | Steps:  269 | Reward:  3 | e-greedy:1.00000 | Avg_Max_Q:0.00508 | Recent reward:1.26667
Episode:    16 | Frames:     2933 | Steps:  135 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00508 | Recent reward:1.18750
Episode:    17 | Frames:     3070 | Steps:  137 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00507 | Recent reward:1.11765
Episode:    18 | Frames:     3197 | Steps:  127 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00506 | Recent reward:1.05556
Episode:    19 | Frames:     3492 | Steps:  295 | Reward:  4 | e-greedy:1.00000 | Avg_Max_Q:0.00499 | Recent reward:1.21053
Episode:    20 | Frames:     3637 | Steps:  145 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00499 | Recent reward:1.15000
Episode:    21 | Frames:     3769 | Steps:  132 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00500 | Recent reward:1.09524
Episode:    22 | Frames:     3990 | Steps:  221 | Reward:  2 | e-greedy:1.00000 | Avg_Max_Q:0.00497 | Recent reward:1.13636
Episode:    23 | Frames:     4201 | Steps:  211 | Reward:  2 | e-greedy:1.00000 | Avg_Max_Q:0.00493 | Recent reward:1.17391
Episode:    24 | Frames:     4363 | Steps:  162 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00492 | Recent reward:1.12500
Episode:    25 | Frames:     4505 | Steps:  142 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00491 | Recent reward:1.08000
Episode:    26 | Frames:     4636 | Steps:  131 | Reward:  0 | e-greedy:1.00000 | Avg_Max_Q:0.00491 | Recent reward:1.03846
Episode:    27 | Frames:     4822 | Steps:  186 | Reward:  1 | e-greedy:1.00000 | Avg_Max_Q:0.00491 | Recent reward:1.03704
Episode:    28 | Frames:     4985 | Steps:  163 | Reward:  1 | e-greedy:1.00000 | Avg_Max_Q:0.00492 | Recent reward:1.03571
```

DQN을 이용한 뚝배기깨기



```
Episode:   284 | Frames:    49998 | Steps:  294 | Reward:  4 | e-greedy:1.00000 | Avg_Max_Q:0.00517 | Recent reward:1.04000
Episode:   285 | Frames:    50171 | Steps:  173 | Reward:  0 | e-greedy:0.99985 | Avg_Max_Q:0.00520 | Recent reward:1.03000
```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
<ipython-input-4-f6ed907527f1> in <module>()
    197
    198 if __name__ == "__main__":
--> 199     main()

/usr/local/lib/python3.6/dist-packages/matplotlib/cbook/__init__.py in to_filehandle(fname, flag, return_opened, encoding)
    430             fh = bz2.BZ2File(fname, flag)
    431         else:
--> 432             fh = open(fname, flag, encoding=encoding)
    433         opened = True
    434     elif hasattr(fname, 'seek'):

FileNotFoundError: [Errno 2] No such file or directory: 'graph/0 epoch.png'
```

?_?

나도 취뽀했다~!~!

KOREAN
DESSERT
CAFE