Proyecto práctico

Nombres:

- Alejandro Mínguez Molina
- Juan Olivan Marquina

Consideraciones a tener en cuenta:

- El entorno sobre el que trabajaremos será *SpaceInvaders-v0* y el algoritmo que usaremos será *DQN*.
- Para nuestro ejercicio, una solución óptima será alcanzada cuando el agente consiga una **media de recompensa por encima de 20 puntos en modo test**. Por ello, esta media de la recompensa se calculará a partir del código de test en la última celda del notebook.

Este proyecto práctico consta de tres partes:

1) Implementar la red neuronal que se usará en la solución

2) Implementar las distintas piezas de la solución DQN

3) Justificar la respuesta en relación a los resultados obtenidos

IMPORTANTE:

- Si no se consigue una puntuación óptima, responder sobre la mejor puntuación obtenida.
- Para entrenamientos largos, recordad que podéis usar checkpoints de vuestros modelos para retomar los entrenamientos. En este caso, recordad cambiar los parámetros adecuadamente (sobre todo los relacionados con el proceso de exploración).
- Tened en cuenta que las versiones de librerías recomendadas son Tensorflow==1.13.1, Keras==2.2.4 y keras-rl==0.4.2

In [1]:
```
# Uncomment this line for installing keras-rl on Google collaboratory
!pip install keras-rl2 > /dev/null 2>&1
```

In [2]:
```
!pip install -U gym[atari,accept-rom-license]
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
public/simple/
Requirement already satisfied: gym[accept-rom-license,atari] in /usr/local/lib/pytho
n3.7/dist-packages (0.17.3)
WARNING: gym 0.17.3 does not provide the extra 'accept-rom-license'
Requirement already satisfied: cloudpickle<1.7.0,>=1.2.0 in /usr/local/lib/python3.
7/dist-packages (from gym[accept-rom-license,atari]) (1.3.0)
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages (from
gym[accept-rom-license,atari]) (1.7.3)
Requirement already satisfied: pyglet<=1.5.0,>=1.4.0 in /usr/local/lib/python3.7/dis
t-packages (from gym[accept-rom-license,atari]) (1.5.0)
Requirement already satisfied: numpy>=1.10.4 in /usr/local/lib/python3.7/dist-packag
es (from gym[accept-rom-license,atari]) (1.21.6)
Requirement already satisfied: atari-py~=0.2.0 in /usr/local/lib/python3.7/dist-pack
ages (from gym[accept-rom-license,atari]) (0.2.9)
Requirement already satisfied: opencv-python in /usr/local/lib/python3.7/dist-packag
es (from gym[accept-rom-license,atari]) (4.6.0.66)
```

```
Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (fro
m gym[accept-rom-license,atari]) (7.1.2)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from a
tari-py~=0.2.0->gym[accept-rom-license,atari]) (1.15.0)
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (fro
m pyglet<=1.5.0,>=1.4.0->gym[accept-rom-license,atari]) (0.16.0)
```

In [3]:
```
!pip install pyvirtualdisplay > /dev/null 2>&1
```

In [4]:
```
!apt-get install -y xvfb python-opengl ffmpeg > /dev/null 2>&1
```

In [5]:
```
!pip install gym==0.15.3
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
public/simple/
Collecting gym==0.15.3
  Downloading gym-0.15.3.tar.gz (1.6 MB)
     |████████████████████████████████| 1.6 MB 9.5 MB/s
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages (from
gym==0.15.3) (1.7.3)
Requirement already satisfied: numpy>=1.10.4 in /usr/local/lib/python3.7/dist-packag
es (from gym==0.15.3) (1.21.6)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from g
ym==0.15.3) (1.15.0)
Collecting pyglet<=1.3.2,>=1.2.0
  Downloading pyglet-1.3.2-py2.py3-none-any.whl (1.0 MB)
     |████████████████████████████████| 1.0 MB 58.0 MB/s
Collecting cloudpickle~=1.2.0
  Downloading cloudpickle-1.2.2-py2.py3-none-any.whl (25 kB)
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (fro
m pyglet<=1.3.2,>=1.2.0->gym==0.15.3) (0.16.0)
Building wheels for collected packages: gym
  Building wheel for gym (setup.py) ... done
  Created wheel for gym: filename=gym-0.15.3-py3-none-any.whl size=1644970 sha256=de
777bdb67356befc063218a419c6a510cfd4d8c8f155911c611274ddf9ae002
  Stored in directory: /root/.cache/pip/wheels/55/16/6b/2250ca4f9f050a4d27d8bed287e5
7bbb3c33fc4066f557cc75
Successfully built gym
Installing collected packages: pyglet, cloudpickle, gym
  Attempting uninstall: pyglet
    Found existing installation: pyglet 1.5.0
    Uninstalling pyglet-1.5.0:
      Successfully uninstalled pyglet-1.5.0
  Attempting uninstall: cloudpickle
    Found existing installation: cloudpickle 1.3.0
    Uninstalling cloudpickle-1.3.0:
      Successfully uninstalled cloudpickle-1.3.0
  Attempting uninstall: gym
    Found existing installation: gym 0.17.3
    Uninstalling gym-0.17.3:
      Successfully uninstalled gym-0.17.3
ERROR: pip's dependency resolver does not currently take into account all the packag
es that are installed. This behaviour is the source of the following dependency conf
licts.
tensorflow-probability 0.16.0 requires cloudpickle>=1.3, but you have cloudpickle 1.
2.2 which is incompatible.
Successfully installed cloudpickle-1.2.2 gym-0.15.3 pyglet-1.3.2
```

In [6]:
```
!wget http://www.atarimania.com/roms/Roms.rar &> /dev/null
```

In [7]:
```
!unrar x Roms.rar roms/ &> /dev/null
```

In [8]:
```
!python -m atari_py.import_roms roms/ &> /dev/null
```

Importar librerías

In [9]:
```
from __future__ import division

from PIL import Image
import numpy as np
import gym

from keras.models import Sequential
from keras.layers import Dense, Activation, Flatten, Convolution2D, Permute
from tensorflow.keras.optimizers import Adam
import keras.backend as K

from rl.agents.dqn import DQNAgent
from rl.policy import LinearAnnealedPolicy, BoltzmannQPolicy, EpsGreedyQPolicy
from rl.memory import SequentialMemory
from rl.core import Processor
from rl.callbacks import FileLogger, ModelIntervalCheckpoint
```

Configuración base

In [10]:
```
INPUT_SHAPE = (84, 84)
WINDOW_LENGTH = 4

env_name = 'SpaceInvaders-v0'
env = gym.make(env_name)

np.random.seed(123)
env.seed(123)
nb_actions = env.action_space.n
```

In [11]:
```
class AtariProcessor(Processor):
    def process_observation(self, observation):
        assert observation.ndim == 3  # (height, width, channel)
        img = Image.fromarray(observation)
        img = img.resize(INPUT_SHAPE).convert('L')
        processed_observation = np.array(img)
        assert processed_observation.shape == INPUT_SHAPE
        return processed_observation.astype('uint8')

    def process_state_batch(self, batch):
        processed_batch = batch.astype('float32') / 255.
        return processed_batch

    def process_reward(self, reward):
        return np.clip(reward, -1., 1.)
```

1) Implementación de la red neuronal

In [12]:
```
# Next, we build our model. We use the same model that was described by Mnih et al.
input_shape = (WINDOW_LENGTH,) + INPUT_SHAPE
model = Sequential()

if K.image_data_format() == 'channels_last':
    # (width, height, channels)
```

```
    model.add(Permute((2, 3, 1), input_shape=input_shape))
elif K.image_data_format() == 'channels_first':
    # (channels, width, height)
    model.add(Permute((1, 2, 3), input_shape=input_shape))
else:
    raise RuntimeError('Unknown image_dim_ordering.')

model.add(Convolution2D(32, (8, 8), strides=(4, 4)))
model.add(Activation('relu'))
model.add(Convolution2D(64, (4, 4), strides=(2, 2)))
model.add(Activation('relu'))
model.add(Convolution2D(64, (3, 3), strides=(1, 1)))
model.add(Activation('relu'))
model.add(Flatten())
model.add(Dense(512))
model.add(Activation('relu'))
model.add(Dense(nb_actions))
model.add(Activation('linear'))
print(model.summary())
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 permute (Permute)           (None, 84, 84, 4)         0

 conv2d (Conv2D)             (None, 20, 20, 32)        8224

 activation (Activation)     (None, 20, 20, 32)        0

 conv2d_1 (Conv2D)           (None, 9, 9, 64)          32832

 activation_1 (Activation)   (None, 9, 9, 64)          0

 conv2d_2 (Conv2D)           (None, 7, 7, 64)          36928

 activation_2 (Activation)   (None, 7, 7, 64)          0

 flatten (Flatten)           (None, 3136)              0

 dense (Dense)               (None, 512)               1606144

 activation_3 (Activation)   (None, 512)               0

 dense_1 (Dense)             (None, 6)                 3078

 activation_4 (Activation)   (None, 6)                 0

=================================================================
Total params: 1,687,206
Trainable params: 1,687,206
Non-trainable params: 0
_________________________________________________________________
None
```

2) Implementación de la solución DQN

In [13]:

```
memory = SequentialMemory(limit=1000000, window_length=WINDOW_LENGTH)
processor = AtariProcessor()
```

In [14]:

```
policy = LinearAnnealedPolicy(EpsGreedyQPolicy(), attr='eps',
                              value_max=1., value_min=.1, value_test=.05,
                              nb_steps=1000000)
```

In [15]:
```
dqn = DQNAgent(model=model, nb_actions=nb_actions, policy=policy,
               memory=memory, processor=processor,
               nb_steps_warmup=50000, gamma=.99,
               target_model_update=10000,
               train_interval=20)
dqn.compile(Adam(lr=.00025), metrics=['mae'])
```

```
/usr/local/lib/python3.7/dist-packages/keras/optimizer_v2/adam.py:105: UserWarning:
The `lr` argument is deprecated, use `learning_rate` instead.
  super(Adam, self).__init__(name, **kwargs)
```

In [16]:
```
# Training part
weights_filename = 'dqn_{}_weights.h5f'.format(env_name)
checkpoint_weights_filename = 'dqn_' + env_name + '_weights_{step}.h5f'
log_filename = 'dqn_{}_log.json'.format(env_name)
callbacks = [ModelIntervalCheckpoint(checkpoint_weights_filename, interval=250000)]
callbacks += [FileLogger(log_filename, interval=100)]

dqn.fit(env, callbacks=callbacks, nb_steps=1750000, log_interval=10000, visualize=Fa

dqn.save_weights(weights_filename, overwrite=True)
```

```
Training for 1750000 steps ...
Interval 1 (0 steps performed)
/usr/local/lib/python3.7/dist-packages/keras/engine/training_v1.py:2079: UserWarnin
g: `Model.state_updates` will be removed in a future version. This property should n
ot be used in TensorFlow 2.0, as `updates` are applied automatically.
  updates=self.state_updates,
10000/10000 [==============================] - 48s 4ms/step - reward: 0.0130
13 episodes - episode_reward: 9.538 [4.000, 25.000] - ale.lives: 2.135

Interval 2 (10000 steps performed)
10000/10000 [==============================] - 39s 4ms/step - reward: 0.0139
14 episodes - episode_reward: 10.143 [4.000, 28.000] - ale.lives: 2.169

Interval 3 (20000 steps performed)
10000/10000 [==============================] - 39s 4ms/step - reward: 0.0136
16 episodes - episode_reward: 8.312 [2.000, 17.000] - ale.lives: 2.124

Interval 4 (30000 steps performed)
10000/10000 [==============================] - 39s 4ms/step - reward: 0.0137
15 episodes - episode_reward: 9.533 [2.000, 23.000] - ale.lives: 2.099

Interval 5 (40000 steps performed)
10000/10000 [==============================] - 39s 4ms/step - reward: 0.0139
13 episodes - episode_reward: 10.308 [5.000, 17.000] - ale.lives: 2.084

Interval 6 (50000 steps performed)
10000/10000 [==============================] - 86s 9ms/step - reward: 0.0131
12 episodes - episode_reward: 11.333 [4.000, 20.000] - loss: 0.007 - mae: 0.033 - me
an_q: 0.049 - mean_eps: 0.950 - ale.lives: 1.953

Interval 7 (60000 steps performed)
10000/10000 [==============================] - 84s 8ms/step - reward: 0.0151
14 episodes - episode_reward: 10.286 [6.000, 18.000] - loss: 0.007 - mae: 0.054 - me
an_q: 0.073 - mean_eps: 0.942 - ale.lives: 2.171

Interval 8 (70000 steps performed)
10000/10000 [==============================] - 84s 8ms/step - reward: 0.0152
15 episodes - episode_reward: 10.067 [4.000, 21.000] - loss: 0.008 - mae: 0.085 - me
an_q: 0.110 - mean_eps: 0.933 - ale.lives: 2.158

Interval 9 (80000 steps performed)
10000/10000 [==============================] - 84s 8ms/step - reward: 0.0130
15 episodes - episode_reward: 8.800 [4.000, 18.000] - loss: 0.007 - mae: 0.098 - mea
```

```
n_q: 0.124 - mean_eps: 0.924 - ale.lives: 2.048

Interval 10 (90000 steps performed)
10000/10000 [==============================] - 84s 8ms/step - reward: 0.0137
14 episodes - episode_reward: 9.857 [2.000, 18.000] - loss: 0.007 - mae: 0.106 - mea
n_q: 0.134 - mean_eps: 0.915 - ale.lives: 2.151

Interval 11 (100000 steps performed)
10000/10000 [==============================] - 84s 8ms/step - reward: 0.0128
15 episodes - episode_reward: 8.467 [2.000, 13.000] - loss: 0.007 - mae: 0.118 - mea
n_q: 0.149 - mean_eps: 0.906 - ale.lives: 2.170

Interval 12 (110000 steps performed)
10000/10000 [==============================] - 84s 8ms/step - reward: 0.0144
13 episodes - episode_reward: 11.231 [3.000, 22.000] - loss: 0.007 - mae: 0.136 - me
an_q: 0.171 - mean_eps: 0.897 - ale.lives: 2.290

Interval 13 (120000 steps performed)
10000/10000 [==============================] - 85s 8ms/step - reward: 0.0139
14 episodes - episode_reward: 9.643 [3.000, 16.000] - loss: 0.007 - mae: 0.158 - mea
n_q: 0.197 - mean_eps: 0.888 - ale.lives: 2.159

Interval 14 (130000 steps performed)
10000/10000 [==============================] - 85s 8ms/step - reward: 0.0139
14 episodes - episode_reward: 10.500 [3.000, 22.000] - loss: 0.006 - mae: 0.172 - me
an_q: 0.214 - mean_eps: 0.879 - ale.lives: 2.072

Interval 15 (140000 steps performed)
10000/10000 [==============================] - 85s 8ms/step - reward: 0.0140
11 episodes - episode_reward: 12.364 [4.000, 20.000] - loss: 0.008 - mae: 0.193 - me
an_q: 0.239 - mean_eps: 0.870 - ale.lives: 2.037

Interval 16 (150000 steps performed)
10000/10000 [==============================] - 85s 8ms/step - reward: 0.0130
13 episodes - episode_reward: 10.000 [5.000, 14.000] - loss: 0.007 - mae: 0.226 - me
an_q: 0.280 - mean_eps: 0.861 - ale.lives: 2.119

Interval 17 (160000 steps performed)
10000/10000 [==============================] - 85s 8ms/step - reward: 0.0122
15 episodes - episode_reward: 8.000 [3.000, 15.000] - loss: 0.007 - mae: 0.246 - mea
n_q: 0.302 - mean_eps: 0.852 - ale.lives: 2.230

Interval 18 (170000 steps performed)
10000/10000 [==============================] - 85s 9ms/step - reward: 0.0143
13 episodes - episode_reward: 11.308 [5.000, 24.000] - loss: 0.007 - mae: 0.274 - me
an_q: 0.336 - mean_eps: 0.843 - ale.lives: 2.126

Interval 19 (180000 steps performed)
10000/10000 [==============================] - 85s 9ms/step - reward: 0.0141
14 episodes - episode_reward: 9.429 [3.000, 21.000] - loss: 0.007 - mae: 0.300 - mea
n_q: 0.368 - mean_eps: 0.834 - ale.lives: 2.124

Interval 20 (190000 steps performed)
10000/10000 [==============================] - 85s 9ms/step - reward: 0.0145
14 episodes - episode_reward: 11.143 [2.000, 17.000] - loss: 0.007 - mae: 0.331 - me
an_q: 0.404 - mean_eps: 0.825 - ale.lives: 2.021

Interval 21 (200000 steps performed)
10000/10000 [==============================] - 85s 9ms/step - reward: 0.0143
16 episodes - episode_reward: 8.750 [3.000, 17.000] - loss: 0.008 - mae: 0.326 - mea
n_q: 0.399 - mean_eps: 0.816 - ale.lives: 2.144

Interval 22 (210000 steps performed)
10000/10000 [==============================] - 86s 9ms/step - reward: 0.0142
13 episodes - episode_reward: 10.615 [5.000, 22.000] - loss: 0.008 - mae: 0.373 - me
an_q: 0.455 - mean_eps: 0.807 - ale.lives: 1.993

Interval 23 (220000 steps performed)
10000/10000 [==============================] - 86s 9ms/step - reward: 0.0136
```

```
16 episodes - episode_reward: 8.500 [4.000, 15.000] - loss: 0.008 - mae: 0.380 - mea
n_q: 0.462 - mean_eps: 0.798 - ale.lives: 2.076

Interval 24 (230000 steps performed)
10000/10000 [==============================] - 86s 9ms/step - reward: 0.0143
16 episodes - episode_reward: 9.000 [3.000, 17.000] - loss: 0.008 - mae: 0.412 - mea
n_q: 0.503 - mean_eps: 0.789 - ale.lives: 2.186

Interval 25 (240000 steps performed)
10000/10000 [==============================] - 86s 9ms/step - reward: 0.0132
16 episodes - episode_reward: 8.125 [4.000, 15.000] - loss: 0.008 - mae: 0.423 - mea
n_q: 0.516 - mean_eps: 0.780 - ale.lives: 2.066

Interval 26 (250000 steps performed)
10000/10000 [==============================] - 87s 9ms/step - reward: 0.0159
14 episodes - episode_reward: 11.929 [3.000, 22.000] - loss: 0.007 - mae: 0.427 - me
an_q: 0.519 - mean_eps: 0.771 - ale.lives: 2.083

Interval 27 (260000 steps performed)
10000/10000 [==============================] - 86s 9ms/step - reward: 0.0153
10 episodes - episode_reward: 15.300 [5.000, 26.000] - loss: 0.008 - mae: 0.430 - me
an_q: 0.521 - mean_eps: 0.762 - ale.lives: 2.225

Interval 28 (270000 steps performed)
10000/10000 [==============================] - 87s 9ms/step - reward: 0.0146
10 episodes - episode_reward: 13.900 [6.000, 24.000] - loss: 0.010 - mae: 0.446 - me
an_q: 0.541 - mean_eps: 0.753 - ale.lives: 2.097

Interval 29 (280000 steps performed)
10000/10000 [==============================] - 87s 9ms/step - reward: 0.0152
15 episodes - episode_reward: 10.200 [4.000, 19.000] - loss: 0.009 - mae: 0.500 - me
an_q: 0.608 - mean_eps: 0.744 - ale.lives: 2.140

Interval 30 (290000 steps performed)
10000/10000 [==============================] - 87s 9ms/step - reward: 0.0154
14 episodes - episode_reward: 11.286 [4.000, 20.000] - loss: 0.008 - mae: 0.518 - me
an_q: 0.627 - mean_eps: 0.735 - ale.lives: 2.053

Interval 31 (300000 steps performed)
10000/10000 [==============================] - 87s 9ms/step - reward: 0.0137
16 episodes - episode_reward: 8.250 [2.000, 18.000] - loss: 0.008 - mae: 0.530 - mea
n_q: 0.642 - mean_eps: 0.726 - ale.lives: 2.002

Interval 32 (310000 steps performed)
10000/10000 [==============================] - 88s 9ms/step - reward: 0.0142
16 episodes - episode_reward: 8.938 [2.000, 21.000] - loss: 0.008 - mae: 0.547 - mea
n_q: 0.666 - mean_eps: 0.717 - ale.lives: 1.958

Interval 33 (320000 steps performed)
10000/10000 [==============================] - 88s 9ms/step - reward: 0.0160
15 episodes - episode_reward: 10.000 [1.000, 22.000] - loss: 0.008 - mae: 0.565 - me
an_q: 0.683 - mean_eps: 0.708 - ale.lives: 2.014

Interval 34 (330000 steps performed)
10000/10000 [==============================] - 88s 9ms/step - reward: 0.0148
15 episodes - episode_reward: 10.533 [4.000, 18.000] - loss: 0.008 - mae: 0.564 - me
an_q: 0.682 - mean_eps: 0.699 - ale.lives: 2.128

Interval 35 (340000 steps performed)
10000/10000 [==============================] - 89s 9ms/step - reward: 0.0142
16 episodes - episode_reward: 9.250 [3.000, 23.000] - loss: 0.007 - mae: 0.605 - mea
n_q: 0.730 - mean_eps: 0.690 - ale.lives: 2.048

Interval 36 (350000 steps performed)
10000/10000 [==============================] - 89s 9ms/step - reward: 0.0155
17 episodes - episode_reward: 9.118 [3.000, 18.000] - loss: 0.008 - mae: 0.616 - mea
n_q: 0.744 - mean_eps: 0.681 - ale.lives: 2.013

Interval 37 (360000 steps performed)
```

```
10000/10000 [==============================] - 89s 9ms/step - reward: 0.0153
15 episodes - episode_reward: 9.933 [5.000, 20.000] - loss: 0.009 - mae: 0.651 - mea
n_q: 0.788 - mean_eps: 0.672 - ale.lives: 2.126

Interval 38 (370000 steps performed)
10000/10000 [==============================] - 91s 9ms/step - reward: 0.0159
14 episodes - episode_reward: 11.214 [5.000, 22.000] - loss: 0.008 - mae: 0.670 - me
an_q: 0.811 - mean_eps: 0.663 - ale.lives: 2.090

Interval 39 (380000 steps performed)
10000/10000 [==============================] - 91s 9ms/step - reward: 0.0176
13 episodes - episode_reward: 12.692 [6.000, 32.000] - loss: 0.008 - mae: 0.700 - me
an_q: 0.847 - mean_eps: 0.654 - ale.lives: 2.050

Interval 40 (390000 steps performed)
10000/10000 [==============================] - 89s 9ms/step - reward: 0.0166
14 episodes - episode_reward: 13.071 [5.000, 20.000] - loss: 0.009 - mae: 0.725 - me
an_q: 0.876 - mean_eps: 0.645 - ale.lives: 2.010

Interval 41 (400000 steps performed)
10000/10000 [==============================] - 90s 9ms/step - reward: 0.0151
15 episodes - episode_reward: 9.467 [6.000, 17.000] - loss: 0.010 - mae: 0.752 - mea
n_q: 0.908 - mean_eps: 0.636 - ale.lives: 1.955

Interval 42 (410000 steps performed)
10000/10000 [==============================] - 90s 9ms/step - reward: 0.0162
15 episodes - episode_reward: 11.333 [5.000, 21.000] - loss: 0.008 - mae: 0.764 - me
an_q: 0.924 - mean_eps: 0.627 - ale.lives: 1.871

Interval 43 (420000 steps performed)
10000/10000 [==============================] - 91s 9ms/step - reward: 0.0138
15 episodes - episode_reward: 9.200 [2.000, 22.000] - loss: 0.009 - mae: 0.803 - mea
n_q: 0.971 - mean_eps: 0.618 - ale.lives: 2.073

Interval 44 (430000 steps performed)
10000/10000 [==============================] - 90s 9ms/step - reward: 0.0166
15 episodes - episode_reward: 10.733 [5.000, 21.000] - loss: 0.010 - mae: 0.806 - me
an_q: 0.973 - mean_eps: 0.609 - ale.lives: 2.038

Interval 45 (440000 steps performed)
10000/10000 [==============================] - 91s 9ms/step - reward: 0.0172
14 episodes - episode_reward: 12.214 [4.000, 25.000] - loss: 0.010 - mae: 0.822 - me
an_q: 0.993 - mean_eps: 0.600 - ale.lives: 1.922

Interval 46 (450000 steps performed)
10000/10000 [==============================] - 91s 9ms/step - reward: 0.0151
13 episodes - episode_reward: 11.385 [5.000, 17.000] - loss: 0.010 - mae: 0.822 - me
an_q: 0.992 - mean_eps: 0.591 - ale.lives: 2.126

Interval 47 (460000 steps performed)
10000/10000 [==============================] - 91s 9ms/step - reward: 0.0154
15 episodes - episode_reward: 10.467 [1.000, 19.000] - loss: 0.009 - mae: 0.819 - me
an_q: 0.989 - mean_eps: 0.582 - ale.lives: 2.080

Interval 48 (470000 steps performed)
10000/10000 [==============================] - 92s 9ms/step - reward: 0.0155
14 episodes - episode_reward: 11.000 [5.000, 18.000] - loss: 0.009 - mae: 0.866 - me
an_q: 1.046 - mean_eps: 0.573 - ale.lives: 1.883

Interval 49 (480000 steps performed)
10000/10000 [==============================] - 92s 9ms/step - reward: 0.0172
13 episodes - episode_reward: 13.385 [6.000, 25.000] - loss: 0.010 - mae: 0.897 - me
an_q: 1.085 - mean_eps: 0.564 - ale.lives: 2.014

Interval 50 (490000 steps performed)
10000/10000 [==============================] - 92s 9ms/step - reward: 0.0177
13 episodes - episode_reward: 13.538 [5.000, 26.000] - loss: 0.010 - mae: 0.902 - me
an_q: 1.089 - mean_eps: 0.555 - ale.lives: 1.997
```

```
Interval 51 (500000 steps performed)
10000/10000 [==============================] - 93s 9ms/step - reward: 0.0149
15 episodes - episode_reward: 10.133 [4.000, 21.000] - loss: 0.010 - mae: 0.909 - me
an_q: 1.097 - mean_eps: 0.546 - ale.lives: 2.029

Interval 52 (510000 steps performed)
10000/10000 [==============================] - 93s 9ms/step - reward: 0.0173
13 episodes - episode_reward: 12.077 [6.000, 19.000] - loss: 0.010 - mae: 0.926 - me
an_q: 1.118 - mean_eps: 0.537 - ale.lives: 2.132

Interval 53 (520000 steps performed)
10000/10000 [==============================] - 93s 9ms/step - reward: 0.0197
14 episodes - episode_reward: 14.857 [7.000, 23.000] - loss: 0.009 - mae: 0.946 - me
an_q: 1.142 - mean_eps: 0.528 - ale.lives: 2.040

Interval 54 (530000 steps performed)
10000/10000 [==============================] - 93s 9ms/step - reward: 0.0150
17 episodes - episode_reward: 9.118 [3.000, 21.000] - loss: 0.010 - mae: 0.962 - mea
n_q: 1.163 - mean_eps: 0.519 - ale.lives: 2.017

Interval 55 (540000 steps performed)
10000/10000 [==============================] - 94s 9ms/step - reward: 0.0158
14 episodes - episode_reward: 11.571 [5.000, 21.000] - loss: 0.009 - mae: 0.962 - me
an_q: 1.163 - mean_eps: 0.510 - ale.lives: 2.029

Interval 56 (550000 steps performed)
10000/10000 [==============================] - 95s 9ms/step - reward: 0.0167
16 episodes - episode_reward: 10.375 [6.000, 21.000] - loss: 0.010 - mae: 0.979 - me
an_q: 1.183 - mean_eps: 0.501 - ale.lives: 2.127

Interval 57 (560000 steps performed)
10000/10000 [==============================] - 96s 10ms/step - reward: 0.0148
15 episodes - episode_reward: 9.933 [4.000, 28.000] - loss: 0.011 - mae: 1.040 - mea
n_q: 1.259 - mean_eps: 0.492 - ale.lives: 2.028

Interval 58 (570000 steps performed)
10000/10000 [==============================] - 96s 10ms/step - reward: 0.0167
16 episodes - episode_reward: 9.625 [5.000, 17.000] - loss: 0.011 - mae: 1.072 - mea
n_q: 1.295 - mean_eps: 0.483 - ale.lives: 2.031

Interval 59 (580000 steps performed)
10000/10000 [==============================] - 95s 10ms/step - reward: 0.0153
15 episodes - episode_reward: 10.733 [3.000, 22.000] - loss: 0.010 - mae: 1.090 - me
an_q: 1.317 - mean_eps: 0.474 - ale.lives: 2.061

Interval 60 (590000 steps performed)
10000/10000 [==============================] - 96s 10ms/step - reward: 0.0161
13 episodes - episode_reward: 12.308 [6.000, 25.000] - loss: 0.011 - mae: 1.124 - me
an_q: 1.358 - mean_eps: 0.465 - ale.lives: 2.080

Interval 61 (600000 steps performed)
10000/10000 [==============================] - 96s 10ms/step - reward: 0.0163
14 episodes - episode_reward: 11.571 [3.000, 34.000] - loss: 0.011 - mae: 1.151 - me
an_q: 1.392 - mean_eps: 0.456 - ale.lives: 2.038

Interval 62 (610000 steps performed)
10000/10000 [==============================] - 96s 10ms/step - reward: 0.0172
13 episodes - episode_reward: 12.769 [3.000, 24.000] - loss: 0.011 - mae: 1.182 - me
an_q: 1.429 - mean_eps: 0.447 - ale.lives: 2.087

Interval 63 (620000 steps performed)
10000/10000 [==============================] - 97s 10ms/step - reward: 0.0167
14 episodes - episode_reward: 12.214 [5.000, 16.000] - loss: 0.010 - mae: 1.198 - me
an_q: 1.450 - mean_eps: 0.438 - ale.lives: 2.041

Interval 64 (630000 steps performed)
10000/10000 [==============================] - 97s 10ms/step - reward: 0.0172
16 episodes - episode_reward: 11.062 [5.000, 23.000] - loss: 0.011 - mae: 1.216 - me
an_q: 1.470 - mean_eps: 0.429 - ale.lives: 1.895
```

```
Interval 65 (640000 steps performed)
10000/10000 [==============================] - 97s 10ms/step - reward: 0.0175
14 episodes - episode_reward: 12.357 [5.000, 23.000] - loss: 0.011 - mae: 1.237 - me
an_q: 1.494 - mean_eps: 0.420 - ale.lives: 2.024

Interval 66 (650000 steps performed)
10000/10000 [==============================] - 98s 10ms/step - reward: 0.0172
17 episodes - episode_reward: 10.118 [4.000, 21.000] - loss: 0.011 - mae: 1.268 - me
an_q: 1.531 - mean_eps: 0.411 - ale.lives: 2.014

Interval 67 (660000 steps performed)
10000/10000 [==============================] - 98s 10ms/step - reward: 0.0159
14 episodes - episode_reward: 11.286 [5.000, 28.000] - loss: 0.011 - mae: 1.293 - me
an_q: 1.562 - mean_eps: 0.402 - ale.lives: 2.022

Interval 68 (670000 steps performed)
10000/10000 [==============================] - 98s 10ms/step - reward: 0.0176
14 episodes - episode_reward: 12.000 [4.000, 29.000] - loss: 0.011 - mae: 1.323 - me
an_q: 1.597 - mean_eps: 0.393 - ale.lives: 1.994

Interval 69 (680000 steps performed)
10000/10000 [==============================] - 99s 10ms/step - reward: 0.0170
12 episodes - episode_reward: 14.500 [4.000, 34.000] - loss: 0.010 - mae: 1.319 - me
an_q: 1.594 - mean_eps: 0.384 - ale.lives: 2.053

Interval 70 (690000 steps performed)
10000/10000 [==============================] - 99s 10ms/step - reward: 0.0158
15 episodes - episode_reward: 10.333 [2.000, 21.000] - loss: 0.011 - mae: 1.371 - me
an_q: 1.657 - mean_eps: 0.375 - ale.lives: 2.002

Interval 71 (700000 steps performed)
10000/10000 [==============================] - 99s 10ms/step - reward: 0.0185
12 episodes - episode_reward: 15.750 [8.000, 28.000] - loss: 0.011 - mae: 1.417 - me
an_q: 1.711 - mean_eps: 0.366 - ale.lives: 2.045

Interval 72 (710000 steps performed)
10000/10000 [==============================] - 100s 10ms/step - reward: 0.0189
15 episodes - episode_reward: 12.733 [5.000, 29.000] - loss: 0.012 - mae: 1.438 - me
an_q: 1.734 - mean_eps: 0.357 - ale.lives: 1.976

Interval 73 (720000 steps performed)
10000/10000 [==============================] - 100s 10ms/step - reward: 0.0165
13 episodes - episode_reward: 12.231 [5.000, 26.000] - loss: 0.012 - mae: 1.459 - me
an_q: 1.762 - mean_eps: 0.348 - ale.lives: 2.018

Interval 74 (730000 steps performed)
10000/10000 [==============================] - 100s 10ms/step - reward: 0.0150
12 episodes - episode_reward: 13.250 [4.000, 24.000] - loss: 0.012 - mae: 1.517 - me
an_q: 1.834 - mean_eps: 0.339 - ale.lives: 2.018

Interval 75 (740000 steps performed)
10000/10000 [==============================] - 101s 10ms/step - reward: 0.0161
13 episodes - episode_reward: 12.538 [6.000, 22.000] - loss: 0.012 - mae: 1.505 - me
an_q: 1.817 - mean_eps: 0.330 - ale.lives: 2.098

Interval 76 (750000 steps performed)
10000/10000 [==============================] - 102s 10ms/step - reward: 0.0165
13 episodes - episode_reward: 12.308 [4.000, 25.000] - loss: 0.011 - mae: 1.520 - me
an_q: 1.836 - mean_eps: 0.321 - ale.lives: 2.138

Interval 77 (760000 steps performed)
10000/10000 [==============================] - 101s 10ms/step - reward: 0.0166
13 episodes - episode_reward: 12.846 [9.000, 19.000] - loss: 0.012 - mae: 1.553 - me
an_q: 1.876 - mean_eps: 0.312 - ale.lives: 2.229

Interval 78 (770000 steps performed)
10000/10000 [==============================] - 101s 10ms/step - reward: 0.0157
14 episodes - episode_reward: 11.357 [6.000, 20.000] - loss: 0.013 - mae: 1.579 - me
```

```
an_q: 1.906 - mean_eps: 0.303 - ale.lives: 2.052

Interval 79 (780000 steps performed)
10000/10000 [==============================] - 102s 10ms/step - reward: 0.0163
14 episodes - episode_reward: 11.214 [5.000, 31.000] - loss: 0.013 - mae: 1.577 - me
an_q: 1.903 - mean_eps: 0.294 - ale.lives: 1.983

Interval 80 (790000 steps performed)
10000/10000 [==============================] - 103s 10ms/step - reward: 0.0141
13 episodes - episode_reward: 11.077 [5.000, 20.000] - loss: 0.013 - mae: 1.594 - me
an_q: 1.927 - mean_eps: 0.285 - ale.lives: 2.101

Interval 81 (800000 steps performed)
10000/10000 [==============================] - 102s 10ms/step - reward: 0.0176
16 episodes - episode_reward: 11.250 [3.000, 20.000] - loss: 0.013 - mae: 1.626 - me
an_q: 1.961 - mean_eps: 0.276 - ale.lives: 2.118

Interval 82 (810000 steps performed)
10000/10000 [==============================] - 103s 10ms/step - reward: 0.0168
13 episodes - episode_reward: 13.077 [5.000, 20.000] - loss: 0.011 - mae: 1.648 - me
an_q: 1.989 - mean_eps: 0.267 - ale.lives: 2.304

Interval 83 (820000 steps performed)
10000/10000 [==============================] - 103s 10ms/step - reward: 0.0169
12 episodes - episode_reward: 13.917 [8.000, 21.000] - loss: 0.012 - mae: 1.658 - me
an_q: 2.000 - mean_eps: 0.258 - ale.lives: 2.003

Interval 84 (830000 steps performed)
10000/10000 [==============================] - 103s 10ms/step - reward: 0.0181
13 episodes - episode_reward: 13.615 [6.000, 20.000] - loss: 0.013 - mae: 1.666 - me
an_q: 2.009 - mean_eps: 0.249 - ale.lives: 2.087

Interval 85 (840000 steps performed)
10000/10000 [==============================] - 104s 10ms/step - reward: 0.0153
15 episodes - episode_reward: 10.733 [3.000, 16.000] - loss: 0.013 - mae: 1.677 - me
an_q: 2.024 - mean_eps: 0.240 - ale.lives: 2.120

Interval 86 (850000 steps performed)
10000/10000 [==============================] - 104s 10ms/step - reward: 0.0187
11 episodes - episode_reward: 16.455 [8.000, 23.000] - loss: 0.011 - mae: 1.634 - me
an_q: 1.974 - mean_eps: 0.231 - ale.lives: 2.012

Interval 87 (860000 steps performed)
10000/10000 [==============================] - 105s 10ms/step - reward: 0.0164
14 episodes - episode_reward: 11.429 [5.000, 24.000] - loss: 0.013 - mae: 1.659 - me
an_q: 2.001 - mean_eps: 0.222 - ale.lives: 1.962

Interval 88 (870000 steps performed)
10000/10000 [==============================] - 105s 10ms/step - reward: 0.0152
14 episodes - episode_reward: 11.000 [3.000, 19.000] - loss: 0.013 - mae: 1.687 - me
an_q: 2.037 - mean_eps: 0.213 - ale.lives: 2.150

Interval 89 (880000 steps performed)
10000/10000 [==============================] - 105s 11ms/step - reward: 0.0169
16 episodes - episode_reward: 10.938 [6.000, 20.000] - loss: 0.013 - mae: 1.737 - me
an_q: 2.096 - mean_eps: 0.204 - ale.lives: 2.111

Interval 90 (890000 steps performed)
10000/10000 [==============================] - 106s 11ms/step - reward: 0.0172
13 episodes - episode_reward: 12.769 [7.000, 25.000] - loss: 0.012 - mae: 1.733 - me
an_q: 2.088 - mean_eps: 0.195 - ale.lives: 2.035

Interval 91 (900000 steps performed)
10000/10000 [==============================] - 106s 11ms/step - reward: 0.0170
14 episodes - episode_reward: 12.429 [2.000, 21.000] - loss: 0.012 - mae: 1.734 - me
an_q: 2.091 - mean_eps: 0.186 - ale.lives: 1.994

Interval 92 (910000 steps performed)
10000/10000 [==============================] - 107s 11ms/step - reward: 0.0184
```

```
16 episodes - episode_reward: 11.375 [5.000, 19.000] - loss: 0.013 - mae: 1.759 - me
an_q: 2.119 - mean_eps: 0.177 - ale.lives: 2.114

Interval 93 (920000 steps performed)
10000/10000 [==============================] - 107s 11ms/step - reward: 0.0166
16 episodes - episode_reward: 10.438 [4.000, 16.000] - loss: 0.013 - mae: 1.792 - me
an_q: 2.161 - mean_eps: 0.168 - ale.lives: 2.127

Interval 94 (930000 steps performed)
10000/10000 [==============================] - 107s 11ms/step - reward: 0.0178
14 episodes - episode_reward: 13.071 [6.000, 32.000] - loss: 0.013 - mae: 1.810 - me
an_q: 2.181 - mean_eps: 0.159 - ale.lives: 2.069

Interval 95 (940000 steps performed)
10000/10000 [==============================] - 107s 11ms/step - reward: 0.0189
14 episodes - episode_reward: 12.429 [1.000, 28.000] - loss: 0.012 - mae: 1.823 - me
an_q: 2.198 - mean_eps: 0.150 - ale.lives: 2.181

Interval 96 (950000 steps performed)
10000/10000 [==============================] - 107s 11ms/step - reward: 0.0170
13 episodes - episode_reward: 14.231 [3.000, 26.000] - loss: 0.012 - mae: 1.818 - me
an_q: 2.190 - mean_eps: 0.141 - ale.lives: 2.024

Interval 97 (960000 steps performed)
10000/10000 [==============================] - 108s 11ms/step - reward: 0.0186
15 episodes - episode_reward: 11.133 [5.000, 17.000] - loss: 0.012 - mae: 1.839 - me
an_q: 2.218 - mean_eps: 0.132 - ale.lives: 2.096

Interval 98 (970000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0186
14 episodes - episode_reward: 13.857 [7.000, 22.000] - loss: 0.013 - mae: 1.847 - me
an_q: 2.228 - mean_eps: 0.123 - ale.lives: 2.121

Interval 99 (980000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0168
13 episodes - episode_reward: 13.154 [4.000, 24.000] - loss: 0.012 - mae: 1.875 - me
an_q: 2.262 - mean_eps: 0.114 - ale.lives: 1.999

Interval 100 (990000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0168
15 episodes - episode_reward: 10.267 [2.000, 27.000] - loss: 0.011 - mae: 1.854 - me
an_q: 2.236 - mean_eps: 0.105 - ale.lives: 1.972

Interval 101 (1000000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0197
16 episodes - episode_reward: 13.062 [4.000, 24.000] - loss: 0.012 - mae: 1.869 - me
an_q: 2.257 - mean_eps: 0.100 - ale.lives: 2.152

Interval 102 (1010000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0179
14 episodes - episode_reward: 13.286 [6.000, 23.000] - loss: 0.012 - mae: 1.915 - me
an_q: 2.309 - mean_eps: 0.100 - ale.lives: 2.032

Interval 103 (1020000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0158
14 episodes - episode_reward: 10.929 [5.000, 18.000] - loss: 0.013 - mae: 1.940 - me
an_q: 2.340 - mean_eps: 0.100 - ale.lives: 2.128

Interval 104 (1030000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0153
18 episodes - episode_reward: 8.778 [3.000, 23.000] - loss: 0.012 - mae: 1.916 - mea
n_q: 2.311 - mean_eps: 0.100 - ale.lives: 2.095

Interval 105 (1040000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0185
16 episodes - episode_reward: 11.750 [6.000, 19.000] - loss: 0.013 - mae: 1.952 - me
an_q: 2.354 - mean_eps: 0.100 - ale.lives: 2.105

Interval 106 (1050000 steps performed)
```

```
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0157
13 episodes - episode_reward: 11.308 [5.000, 21.000] - loss: 0.014 - mae: 1.964 - me
an_q: 2.368 - mean_eps: 0.100 - ale.lives: 2.090

Interval 107 (1060000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0191
15 episodes - episode_reward: 12.400 [7.000, 21.000] - loss: 0.013 - mae: 1.972 - me
an_q: 2.377 - mean_eps: 0.100 - ale.lives: 2.131

Interval 108 (1070000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0160
14 episodes - episode_reward: 12.500 [7.000, 21.000] - loss: 0.012 - mae: 2.009 - me
an_q: 2.426 - mean_eps: 0.100 - ale.lives: 1.939

Interval 109 (1080000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0186
16 episodes - episode_reward: 11.625 [5.000, 27.000] - loss: 0.012 - mae: 2.063 - me
an_q: 2.487 - mean_eps: 0.100 - ale.lives: 2.057

Interval 110 (1090000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0176
14 episodes - episode_reward: 12.071 [5.000, 22.000] - loss: 0.013 - mae: 2.100 - me
an_q: 2.534 - mean_eps: 0.100 - ale.lives: 2.011

Interval 111 (1100000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0177
15 episodes - episode_reward: 11.000 [5.000, 18.000] - loss: 0.014 - mae: 2.123 - me
an_q: 2.561 - mean_eps: 0.100 - ale.lives: 2.153

Interval 112 (1110000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0183
16 episodes - episode_reward: 12.625 [2.000, 22.000] - loss: 0.012 - mae: 2.115 - me
an_q: 2.550 - mean_eps: 0.100 - ale.lives: 2.157

Interval 113 (1120000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0198
14 episodes - episode_reward: 13.786 [7.000, 21.000] - loss: 0.012 - mae: 2.118 - me
an_q: 2.555 - mean_eps: 0.100 - ale.lives: 1.937

Interval 114 (1130000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0188
14 episodes - episode_reward: 12.643 [8.000, 19.000] - loss: 0.014 - mae: 2.130 - me
an_q: 2.569 - mean_eps: 0.100 - ale.lives: 2.036

Interval 115 (1140000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0187
15 episodes - episode_reward: 13.533 [3.000, 21.000] - loss: 0.013 - mae: 2.161 - me
an_q: 2.606 - mean_eps: 0.100 - ale.lives: 1.931

Interval 116 (1150000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0187
14 episodes - episode_reward: 13.286 [2.000, 27.000] - loss: 0.013 - mae: 2.152 - me
an_q: 2.595 - mean_eps: 0.100 - ale.lives: 2.184

Interval 117 (1160000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0191
14 episodes - episode_reward: 13.214 [5.000, 19.000] - loss: 0.013 - mae: 2.133 - me
an_q: 2.573 - mean_eps: 0.100 - ale.lives: 2.091

Interval 118 (1170000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0162
15 episodes - episode_reward: 11.067 [4.000, 20.000] - loss: 0.013 - mae: 2.135 - me
an_q: 2.575 - mean_eps: 0.100 - ale.lives: 1.958

Interval 119 (1180000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0199
14 episodes - episode_reward: 13.929 [7.000, 24.000] - loss: 0.013 - mae: 2.140 - me
an_q: 2.584 - mean_eps: 0.100 - ale.lives: 2.100
```

```
Interval 120 (1190000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0162
15 episodes - episode_reward: 10.933 [6.000, 19.000] - loss: 0.012 - mae: 2.147 - me
an_q: 2.589 - mean_eps: 0.100 - ale.lives: 1.924

Interval 121 (1200000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0184
14 episodes - episode_reward: 12.643 [8.000, 33.000] - loss: 0.014 - mae: 2.158 - me
an_q: 2.602 - mean_eps: 0.100 - ale.lives: 2.011

Interval 122 (1210000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0180
16 episodes - episode_reward: 12.000 [7.000, 23.000] - loss: 0.014 - mae: 2.203 - me
an_q: 2.656 - mean_eps: 0.100 - ale.lives: 2.117

Interval 123 (1220000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0182
15 episodes - episode_reward: 12.133 [6.000, 22.000] - loss: 0.013 - mae: 2.226 - me
an_q: 2.681 - mean_eps: 0.100 - ale.lives: 2.130

Interval 124 (1230000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0201
15 episodes - episode_reward: 13.000 [5.000, 22.000] - loss: 0.014 - mae: 2.248 - me
an_q: 2.710 - mean_eps: 0.100 - ale.lives: 2.144

Interval 125 (1240000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0189
15 episodes - episode_reward: 12.267 [5.000, 24.000] - loss: 0.013 - mae: 2.255 - me
an_q: 2.719 - mean_eps: 0.100 - ale.lives: 2.039

Interval 126 (1250000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0197
14 episodes - episode_reward: 14.643 [5.000, 27.000] - loss: 0.013 - mae: 2.248 - me
an_q: 2.711 - mean_eps: 0.100 - ale.lives: 1.834

Interval 127 (1260000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0179
16 episodes - episode_reward: 11.250 [3.000, 19.000] - loss: 0.015 - mae: 2.277 - me
an_q: 2.745 - mean_eps: 0.100 - ale.lives: 2.002

Interval 128 (1270000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0165
16 episodes - episode_reward: 10.312 [4.000, 17.000] - loss: 0.013 - mae: 2.302 - me
an_q: 2.774 - mean_eps: 0.100 - ale.lives: 2.014

Interval 129 (1280000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0178
14 episodes - episode_reward: 12.500 [5.000, 23.000] - loss: 0.014 - mae: 2.318 - me
an_q: 2.792 - mean_eps: 0.100 - ale.lives: 2.078

Interval 130 (1290000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0175
13 episodes - episode_reward: 13.692 [5.000, 22.000] - loss: 0.012 - mae: 2.306 - me
an_q: 2.780 - mean_eps: 0.100 - ale.lives: 2.083

Interval 131 (1300000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0178
13 episodes - episode_reward: 13.462 [3.000, 28.000] - loss: 0.014 - mae: 2.295 - me
an_q: 2.767 - mean_eps: 0.100 - ale.lives: 2.071

Interval 132 (1310000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0177
16 episodes - episode_reward: 11.125 [6.000, 15.000] - loss: 0.014 - mae: 2.305 - me
an_q: 2.779 - mean_eps: 0.100 - ale.lives: 2.063

Interval 133 (1320000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0188
14 episodes - episode_reward: 13.071 [4.000, 23.000] - loss: 0.013 - mae: 2.333 - me
an_q: 2.810 - mean_eps: 0.100 - ale.lives: 2.140
```

```
Interval 134 (1330000 steps performed)
10000/10000 [==============================] - 109s 11ms/step - reward: 0.0188
12 episodes - episode_reward: 15.917 [8.000, 32.000] - loss: 0.014 - mae: 2.384 - me
an_q: 2.872 - mean_eps: 0.100 - ale.lives: 2.090

Interval 135 (1340000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0195
17 episodes - episode_reward: 11.706 [6.000, 19.000] - loss: 0.014 - mae: 2.390 - me
an_q: 2.877 - mean_eps: 0.100 - ale.lives: 2.146

Interval 136 (1350000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0157
14 episodes - episode_reward: 10.929 [5.000, 29.000] - loss: 0.014 - mae: 2.391 - me
an_q: 2.881 - mean_eps: 0.100 - ale.lives: 1.971

Interval 137 (1360000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0180
14 episodes - episode_reward: 12.714 [7.000, 21.000] - loss: 0.014 - mae: 2.390 - me
an_q: 2.878 - mean_eps: 0.100 - ale.lives: 2.078

Interval 138 (1370000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0151
12 episodes - episode_reward: 11.667 [6.000, 17.000] - loss: 0.015 - mae: 2.451 - me
an_q: 2.951 - mean_eps: 0.100 - ale.lives: 2.093

Interval 139 (1380000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0180
13 episodes - episode_reward: 14.615 [4.000, 27.000] - loss: 0.014 - mae: 2.476 - me
an_q: 2.979 - mean_eps: 0.100 - ale.lives: 1.872

Interval 140 (1390000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0196
14 episodes - episode_reward: 14.286 [5.000, 23.000] - loss: 0.014 - mae: 2.490 - me
an_q: 2.997 - mean_eps: 0.100 - ale.lives: 2.065

Interval 141 (1400000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0198
14 episodes - episode_reward: 13.786 [5.000, 26.000] - loss: 0.015 - mae: 2.522 - me
an_q: 3.037 - mean_eps: 0.100 - ale.lives: 2.076

Interval 142 (1410000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0207
14 episodes - episode_reward: 14.571 [6.000, 26.000] - loss: 0.015 - mae: 2.514 - me
an_q: 3.027 - mean_eps: 0.100 - ale.lives: 2.063

Interval 143 (1420000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0178
10 episodes - episode_reward: 16.600 [2.000, 28.000] - loss: 0.015 - mae: 2.532 - me
an_q: 3.049 - mean_eps: 0.100 - ale.lives: 1.857

Interval 144 (1430000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0191
14 episodes - episode_reward: 14.714 [6.000, 25.000] - loss: 0.014 - mae: 2.548 - me
an_q: 3.066 - mean_eps: 0.100 - ale.lives: 2.027

Interval 145 (1440000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0181
14 episodes - episode_reward: 13.286 [7.000, 29.000] - loss: 0.016 - mae: 2.578 - me
an_q: 3.106 - mean_eps: 0.100 - ale.lives: 2.122

Interval 146 (1450000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0207
14 episodes - episode_reward: 14.500 [4.000, 24.000] - loss: 0.015 - mae: 2.582 - me
an_q: 3.111 - mean_eps: 0.100 - ale.lives: 2.103

Interval 147 (1460000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0210
13 episodes - episode_reward: 15.923 [7.000, 26.000] - loss: 0.015 - mae: 2.554 - me
```

```
an_q: 3.076 - mean_eps: 0.100 - ale.lives: 2.062

Interval 148 (1470000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0209
12 episodes - episode_reward: 17.500 [10.000, 41.000] - loss: 0.014 - mae: 2.571 - m
ean_q: 3.098 - mean_eps: 0.100 - ale.lives: 2.059

Interval 149 (1480000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0189
14 episodes - episode_reward: 12.786 [6.000, 24.000] - loss: 0.015 - mae: 2.594 - me
an_q: 3.127 - mean_eps: 0.100 - ale.lives: 2.108

Interval 150 (1490000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0198
9 episodes - episode_reward: 22.667 [12.000, 41.000] - loss: 0.015 - mae: 2.598 - me
an_q: 3.131 - mean_eps: 0.100 - ale.lives: 2.189

Interval 151 (1500000 steps performed)
10000/10000 [==============================] - 110s 11ms/step - reward: 0.0202
13 episodes - episode_reward: 16.462 [6.000, 26.000] - loss: 0.015 - mae: 2.586 - me
an_q: 3.115 - mean_eps: 0.100 - ale.lives: 2.018

Interval 152 (1510000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0192
13 episodes - episode_reward: 14.308 [9.000, 25.000] - loss: 0.015 - mae: 2.594 - me
an_q: 3.126 - mean_eps: 0.100 - ale.lives: 2.013

Interval 153 (1520000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0175
13 episodes - episode_reward: 13.231 [6.000, 21.000] - loss: 0.015 - mae: 2.624 - me
an_q: 3.163 - mean_eps: 0.100 - ale.lives: 2.086

Interval 154 (1530000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0187
11 episodes - episode_reward: 18.091 [7.000, 31.000] - loss: 0.015 - mae: 2.621 - me
an_q: 3.158 - mean_eps: 0.100 - ale.lives: 1.990

Interval 155 (1540000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0194
11 episodes - episode_reward: 17.091 [8.000, 23.000] - loss: 0.014 - mae: 2.625 - me
an_q: 3.161 - mean_eps: 0.100 - ale.lives: 2.153

Interval 156 (1550000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0205
14 episodes - episode_reward: 13.571 [6.000, 22.000] - loss: 0.015 - mae: 2.588 - me
an_q: 3.115 - mean_eps: 0.100 - ale.lives: 2.051

Interval 157 (1560000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0177
13 episodes - episode_reward: 15.000 [5.000, 21.000] - loss: 0.015 - mae: 2.611 - me
an_q: 3.145 - mean_eps: 0.100 - ale.lives: 2.025

Interval 158 (1570000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0193
11 episodes - episode_reward: 16.727 [6.000, 29.000] - loss: 0.017 - mae: 2.607 - me
an_q: 3.138 - mean_eps: 0.100 - ale.lives: 1.993

Interval 159 (1580000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0166
13 episodes - episode_reward: 13.462 [3.000, 27.000] - loss: 0.016 - mae: 2.614 - me
an_q: 3.150 - mean_eps: 0.100 - ale.lives: 1.965

Interval 160 (1590000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0208
14 episodes - episode_reward: 14.857 [5.000, 24.000] - loss: 0.017 - mae: 2.634 - me
an_q: 3.174 - mean_eps: 0.100 - ale.lives: 2.118

Interval 161 (1600000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0192
```

```
12 episodes - episode_reward: 15.500 [6.000, 32.000] - loss: 0.017 - mae: 2.694 - me
an_q: 3.248 - mean_eps: 0.100 - ale.lives: 1.989

Interval 162 (1610000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0226
13 episodes - episode_reward: 17.615 [8.000, 29.000] - loss: 0.018 - mae: 2.717 - me
an_q: 3.272 - mean_eps: 0.100 - ale.lives: 2.000

Interval 163 (1620000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0197
11 episodes - episode_reward: 17.273 [11.000, 23.000] - loss: 0.017 - mae: 2.735 - m
ean_q: 3.295 - mean_eps: 0.100 - ale.lives: 2.062

Interval 164 (1630000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0200
13 episodes - episode_reward: 15.538 [4.000, 24.000] - loss: 0.016 - mae: 2.743 - me
an_q: 3.305 - mean_eps: 0.100 - ale.lives: 2.058

Interval 165 (1640000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0192
11 episodes - episode_reward: 17.364 [12.000, 25.000] - loss: 0.016 - mae: 2.717 - m
ean_q: 3.275 - mean_eps: 0.100 - ale.lives: 2.155

Interval 166 (1650000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0179
12 episodes - episode_reward: 15.000 [4.000, 31.000] - loss: 0.018 - mae: 2.780 - me
an_q: 3.352 - mean_eps: 0.100 - ale.lives: 1.961

Interval 167 (1660000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0195
13 episodes - episode_reward: 14.692 [10.000, 27.000] - loss: 0.016 - mae: 2.765 - m
ean_q: 3.328 - mean_eps: 0.100 - ale.lives: 1.925

Interval 168 (1670000 steps performed)
10000/10000 [==============================] - 111s 11ms/step - reward: 0.0205
12 episodes - episode_reward: 17.500 [5.000, 31.000] - loss: 0.017 - mae: 2.787 - me
an_q: 3.356 - mean_eps: 0.100 - ale.lives: 2.072

Interval 169 (1680000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0213
12 episodes - episode_reward: 18.583 [9.000, 32.000] - loss: 0.016 - mae: 2.798 - me
an_q: 3.371 - mean_eps: 0.100 - ale.lives: 1.943

Interval 170 (1690000 steps performed)
10000/10000 [==============================] - 113s 11ms/step - reward: 0.0219
12 episodes - episode_reward: 17.000 [6.000, 32.000] - loss: 0.018 - mae: 2.805 - me
an_q: 3.378 - mean_eps: 0.100 - ale.lives: 2.097

Interval 171 (1700000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0210
13 episodes - episode_reward: 17.077 [6.000, 32.000] - loss: 0.019 - mae: 2.816 - me
an_q: 3.392 - mean_eps: 0.100 - ale.lives: 2.050

Interval 172 (1710000 steps performed)
10000/10000 [==============================] - 113s 11ms/step - reward: 0.0208
14 episodes - episode_reward: 14.214 [8.000, 28.000] - loss: 0.017 - mae: 2.784 - me
an_q: 3.354 - mean_eps: 0.100 - ale.lives: 2.168

Interval 173 (1720000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0209
13 episodes - episode_reward: 16.462 [6.000, 24.000] - loss: 0.017 - mae: 2.809 - me
an_q: 3.385 - mean_eps: 0.100 - ale.lives: 2.017

Interval 174 (1730000 steps performed)
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0212
13 episodes - episode_reward: 16.231 [10.000, 22.000] - loss: 0.017 - mae: 2.827 - m
ean_q: 3.401 - mean_eps: 0.100 - ale.lives: 2.008

Interval 175 (1740000 steps performed)
```

```
10000/10000 [==============================] - 112s 11ms/step - reward: 0.0209
done, took 17450.412 seconds
```

In [17]:
```
# Testing part to calculate the mean reward
weights_filename = 'dqn_{}_weights.h5f'.format(env_name)
dqn.load_weights(weights_filename)
dqn.test(env, nb_episodes=10, visualize=False)
```

```
Testing for 10 episodes ...
Episode 1: reward: 14.000, steps: 612
Episode 2: reward: 22.000, steps: 1027
Episode 3: reward: 16.000, steps: 805
Episode 4: reward: 14.000, steps: 601
Episode 5: reward: 24.000, steps: 1006
Episode 6: reward: 10.000, steps: 489
Episode 7: reward: 15.000, steps: 574
Episode 8: reward: 13.000, steps: 629
Episode 9: reward: 13.000, steps: 618
Episode 10: reward: 16.000, steps: 841
```

Out[17]:
```
<keras.callbacks.History at 0x7fe727f8ed10>
```

3) Justificación de los parámetros seleccionados y de los resultados obtenidos

Explicación de los parámetros seleccionados:

Memory:

- limit=1000000 como el problema es más complejo, el límite de la memoria se aumentado a mil transiciones, se va ir acumulando transiciones hasta llenar la memoria.
- Window_lenght=4, cada vez que vayamos hacer una batch de esa memora para actrtualizar el modelo,, tendremos en cuenta esos 4 franmes para intentar capturar esa tendencia.

Processor:

- Será necesaria para el agente a nivel de observación y de normailizacion de la recompensa.

Policy, exploración y explotación basada en epsilon:

- como llevaremos a cabo el proceso de exploración y explotación, tenemos en cuenta los siguientes hiperparámetros:
  - value_max=1., valor con el va comenzar epsilon.
  - value_min=.1, hasta donde llegara el valor epsilon.
  - value_test=.05, cuando el modelo este entrenado tomarà acciones aleatorias en un conjunto muy reducido para salir de mínimos locales, y le va ir permitiendo al agente de ser más robusto.
  - nb_steps=1000000, proceso de exploración.

DQN, deifnición del agente:

- model
- nb_actions
- policy
- memory

- processor
- nb_steps_warmup=50000, estos son los steps de calentamineto
- target_model_update=10000, cada 10k se hará el update del target model.
- train_interval=20, cada 20 steps se actualizará el modelo Q de nuestro agente.
- gamma=0.99, se da este valor para la importancia de la recompensa futura y se utiliza en la ecuación de Bellman para el discount factor.

DQN compile, de "Adam", con un learning rate de .00025 (lr=.00025), y la metrica de meaning absolute error (mae) para el valor de regresion y hacer la compartitva entre la recompensa estimada por cada acción con la esperada.

La parte del entrenamiento:

Parte delos callback sirven para almecenar los checkpoint y logs de lo que esta ocurriendo en el archivo json y que lo ponemos como parámetro en el fit, en el Callback-Model Interval checkpoint:

- interval=250000, cada estos steps se irá almacenando el checkpoint del modelo.

DQN FIT, ya tenemos el entorno creado, también el parámetro de callback y definimos el interval=10000 para mostrar el reumen del log, cada estos steps se mostrará en pantalla la información. Por otra parte, cada nb_steps=1750000 tendremos 1000000 de steps de exploración y 750000 de explotacion a continuacion.

Resultados obtenidos:

Como podemos ver el resultado máximo que hemos obtenido tras la parte de testeo en 10 episodios ha sido de 24 puntos, con esto cabe decir que hemos llegado al objetivo de 20 puntos.

Podriamos obtener una mejor puntación de las siguientes maneras:

- Monitoreando la recompensa durante el entrenamiento y está debe tener una tendencia al alza; este es el indicador más robusto.
- En el caso que la recompensa no sea acumulativa podriamos utilizar otra métrica.