```
Entrenando batch: 1
experimento: Experimento Arquitectura 9
arquitectura: 9
train steps: 450000
policy_nb_steps: 100000 agent_nb_steps_warmup: 3000
agent target model update: 10000
agent_batch_size: 32
agent train interval: 4
learning_rate: 0.0001
memory limit: 10000
optimizer type: Adam
history_reward: []
Ejecutando experimento ... Experimento_Arquitectura_9
input shape: (4, 84, 84)
4 (84, 84)
Usando arquitectura: 9
Layer (type)
                                                                                         Output Shape
permute 12 (Permute)
                                                                                     (None, 84, 84, 4)
dense 29 (Dense)
                                                                                         (None, 84, 84, 64)
dense_30 (Dense)
                                                                                                                                                                         2080
                                                                                        (None, 84, 84, 32)
dense_31 (Dense)
                                                                                        (None, 84, 84, 16)
                                                                                                                                                                        528
flatten_9 (Flatten)
                                                                                         (None, 112896)
dense_32 (Dense)
                                                                                                                                                                        677382
activation 42 (Activation)
Trainable params: 680,310
Non-trainable params: 0
Modelo inicial quardado en: 202001091630 qs Experimento Arquitectura 9 modelo inicial.h5
Training for 450000 steps ..
           812/450000: episode: 1, duration: 5.899s, episode steps: 812, steps per second: 138, episode reward: -21.000, mean reward: -0.026 (-1.000, 0.000), mean action: 2.
         012/300000: episode: 2, duration: 8.2958, episode steps: 129, steps per second: 130, episode reward: -21.000, mean reward: -0.105 [-1.000, 0.000], mean action: 2.2041/450000: episode: 2, duration: 8.295, episode steps: 1229, steps per second: 149, episode reward: -18.000, mean reward: -0.015 [-1.000, 0.000], mean action: 2.2065/450000: episode: 3, duration: 5.4698, episode steps: 824, steps per second: 151, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 2.
          3629/450000: episode: 4, duration: 5.309s, episode steps: 764, steps per second: 144, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2.
         4393/450000: episode: 5, duration: 5.335s, episode steps: 764, steps per second: 143, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2.
          5374/450000: episode: 6,
                                                                                    duration: 6.775s, episode steps: 981, steps per second: 145, episode reward: -20.000, mean reward: -0.020 [-1.000,
         6257/450000: episode: 7, duration: 6.147s, episode steps: 883, steps per second: 144, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 2,
          7118/450000: episode: 8, duration: 5.792s, episode steps: 861, steps per second: 149, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 2.
         7882/450000: episode: 9, duration: 5.365s, episode steps: 764, steps per second: 142, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2.8867/450000: episode: 10, duration: 6.617s, episode steps: 985, steps per second: 149, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action: 2
         9716/450000: episode: 11, duration: 5.663s, episode steps: 849, steps per second: 150, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 2
     10829/450000: episode: 12, duration: 26.559s, episode steps: 1113, steps per second: 42, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action: 11710/450000: episode: 13, duration: 23.627s, episode steps: 881, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 2
      12641/450000: episode: 14, duration: 24.995s, episode steps: 931, steps per second: 37, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
      13527/450000: episode: 15, duration: 23.826s, episode steps: 886, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 23.826s, episode steps: 886, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 23.826s, episode steps: 886, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 23.826s, episode steps: 886, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 23.826s, episode steps: 886, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 23.826s, episode steps: 886, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 23.826s, episode steps: 886, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean reward: -0.024 [-1.000, 0
      14320/450000: episode: 16, duration: 21.680s, episode steps: 793, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 2
      15191/450000: episode: 17, duration: 23.832s, episode steps: 871, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
      15955/450000: episode: 18, duration: 21.056s, episode steps: 764, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
      16719/450000: episode: 19, duration: 20.756s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2
     17592/450000: episode: 20, duration: 23.994s, episode steps: 873, steps per second: 36, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 218595/450000: episode: 21, duration: 27.296s, episode steps: 1003, steps per second: 37, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
      19724/450000: episode: 22, duration: 31.337s, episode steps: 1129, steps per second: 36, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
     20700/450000: episode: 23, duration: 26.755s, episode steps: 976, steps per second: 36, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action: 22017/450000: episode: 24, duration: 35.535s, episode steps: 1317, steps per second: 37, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean action:
     22888/450000: episode: 25, duration: 23.496s, episode steps: 871, steps per second: 37, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 23791/450000: episode: 26, duration: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, steps per second: 903, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, episode steps: 903, episode steps: 903, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 24.429s, episode steps: 903, episode steps: 903
     26462/450000: episode: 29, duration: 22.776s, episode steps: 841, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 27520/450000: episode: 30, duration: 28.619s, episode steps: 1058, steps per second: 37, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
     28505/450000: episode: 31, duration: 26.727s, episode steps: 985, steps per second: 37, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 22348/450000: episode: 32, duration: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 22.747s, episode steps: 843, steps per second: 27, episode reward: -20.000, mean reward: -0.000, episode steps: 843, steps per second: 27, episode steps: 843, steps per second: 843, episode steps: 843, episode step
     30345/450000: episode: 33, duration: 26.922s, episode steps: 997, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action: 231156/450000: episode: 34, duration: 21.803s, episode steps: 811, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 231156/450000: episode: 34, duration: 21.803s, episode steps: 811, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 231156/450000: episode: 34, duration: 21.803s, episode steps: 811, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 231156/450000: episode: 231156/450000: episode:
      32146/450000; episode: 35, duration: 26.727s, episode steps: 990, steps per second: 37, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
       33019/450000: episode: 36, duration: 23.469s, episode steps: 873, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
      33892/450000: episode: 37, duration: 23.662s, episode steps: 873, steps per second: 37, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
      34810/450000: episode: 38, duration: 24.803s, episode steps: 918, steps per second: 37, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 2
      35694/450000: episode: 39, duration: 23.831s, episode steps: 884, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 2
      36848/450000: episode: 40, duration: 31.171s, episode steps: 1154, steps per second: 37, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean action: 37671/450000: episode: 41, duration: 22.378s, episode steps: 823, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 2
      38693/450000: episode: 42, duration: 27.763s, episode steps: 1022, steps per second: 37, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
      39545/450000: episode: 43, duration: 23.080s, episode steps: 852, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 40475/450000: episode: 44, duration: 25.249s, episode steps: 930, steps per second: 37, episode reward: -20.000, mean reward: -0.025 [-1.000, 1.000], mean action: 2
      41299/450000: episode: 45, duration: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 22.233s, episode steps: 824, steps per second: 37, episode steps: 824, steps per second: 824, steps per s
      42229/450000: episode: 46, duration: 25.093s, episode steps: 930, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 42993/450000: episode: 47, duration: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 20.517s, episode steps: 764, steps per second: 27, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 27, episode steps: 27, episode step
      43937/450000: episode: 48, duration: 25.626s, episode steps: 944, steps per second: 37, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
      44929/450000: episode: 49. duration: 27.009s, episode steps: 992, steps per second: 37. episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2
      45979/450000: episode: 50, duration: 28.569s, episode steps: 1050, steps per second: 37, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 47032/450000: episode: 51, duration: 28.831s, episode steps: 1053, steps per second: 37, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
      48024/450000: episode: 52, duration: 26.984s, episode steps: 992, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action: 2
      48788/450000: episode: 53, duration: 21.029s, episode steps: 764, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 49640/450000: episode: 54, duration: 23.520s, episode steps: 852, steps per second: 36, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 250513/450000: episode: 55, duration: 24.083s, episode steps: 873, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 250513/450000: episode: 55, duration: 24.083s, episode steps: 873, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 250513/450000: episode: 55, duration: 24.083s, episode steps: 873, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 250513/450000: episode: 55, duration: 24.083s, episode steps: 873, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 250513/450000: episode: 55, duration: 24.083s, episode steps: 873, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 250513/450000: episode: 55, duration: 24.083s, episode steps: 873, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 250513/450000: episode: 55, duration: 250513/450000: episode: 250513/450000: episode: 2505
      51624/450000: episode: 56, duration: 30.382s, episode steps: 1111, steps per second: 37, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
      52532/450000: episode: 57, duration: 24.990s, episode steps: 908, steps per second: 36, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 2
      53572/450000: episode: 58, duration: 28.767s, episode steps: 1040, steps per second: 36, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action:
     54416/450000: episode: 59, duration: 23.002s, episode steps: 844, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 2 55276/450000: episode: 60, duration: 23.370s, episode steps: 860, steps per second: 37, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 25.603s, episode steps: 939, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 2 56215/450000: episode: 61, duration: 2 56215/
      57238/450000: episode: 62, duration: 28.190s, episode steps: 1023, steps per second: 36, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
```

```
58274/450000: episode: 63, duration: 29.206s, episode steps: 1036, steps per second: 35, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
59038/450000: episode: 64, duration: 21.133s, episode steps: 764, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 260109/450000: episode: 65, duration: 29.977s, episode steps: 1071, steps per second: 36, episode reward: -18.000, mean reward: -0.017 [-1.000, 1.000], mean action:
 61103/450000: episode: 66, duration: 27.054s, episode steps: 994, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action: 2
 61941/450000: episode: 67, duration: 22.731s, episode steps: 838, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 2
 62761/450000: episode: 68, duration: 22.348s, episode steps: 820, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
                  episode: 69, duration: 22.829s, episode steps: 842, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
 64446/450000: episode: 70, duration: 22.896s, episode steps: 843, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 265270/450000: episode: 71, duration: 22.520s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 265270/450000: episode: 71, duration: 22.520s, episode steps: 824, steps per second: 87, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 265270/450000: episode: 71, duration: 265270/450000: episode: 71, duratio
                                                                                                                                                                                     0.000], mean action: 2
 66053/450000: episode: 72, duration: 21.719s, episode steps: 783, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2
                  episode: 73, duration: 25.727s, episode steps: 921, steps per second: 36, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 2
 68009/450000: episode: 74, duration: 29.166s, episode steps: 1035, steps per second: 35, episode reward: -18.000, mean reward: -0.017
                                                                                                                                                                           [-1.000, 1.000], mean action:
                  episode: 75, duration: 23.681s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 2
 69781/450000: episode: 76, duration: 25.776s, episode steps: 930, steps per second: 36, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 2
                  episode:
                              77, duration: 25.998s, episode steps:
                                                                              940, steps per second: 36, episode reward:
                                                                                                                                     -21.000, mean reward: -0.022
                                                                                                                                                                                     0.000], mean action:
 71709/450000: episode: 78, duration: 27.558s, episode steps: 988, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000,
                                                                                                                                                                                     1.0001, mean action: 2
                  episode: 79, duration: 22.882s, episode steps:
                                                                              820, steps per second: 36, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                                          [-1.000, 0.000], mean action: 2
 73520/450000: episode: 80, duration: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action: 27.257s, episode steps: 991, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], 
                  episode: 81, duration: 24.284s, episode steps: 870, steps per second: 36, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                                                     1.000], mean action: 2
                                                                                                                                                                          [-1.000,
 75202/450000: episode: 82, duration: 22.550s, episode steps: 812, steps per second: 36, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 2
 76210/450000: episode: 83, duration: 28.120s, episode steps: 1008, steps per second: 36, episode reward: -19.000, mean reward: -0.019
                                                                                                                                                                           [-1.000, 1.000], mean action:
 77034/450000: episode: 84, duration: 22.786s, episode steps: 824, steps per second: 36, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 277877/450000: episode: 85, duration: 23.810s, episode steps: 843, steps per second: 35, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 2
 78719/450000: episode: 86, duration: 23.092s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 2
 79563/450000: episode: 87, duration: 23.163s, episode steps: 844, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 2
 80356/450000: episode: 88, duration: 22.250s, episode steps: 793, steps per second: 36, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 2
                  episode: 89, duration: 23.995s, episode steps:
                                                                              882, steps per second: 37, episode reward: -21.000, mean reward:
 82021/450000: episode: 90, duration: 21.705s, episode steps: 783, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2
                  episode: 91, duration: 23.886s, episode steps: 872, steps per second: 37, episode reward: -20.000, mean reward: -0.023 [-1.000,
                                                                                                                                                                                     1.000], mean action: 2
 83676/450000: episode: 92, duration: 22.102s, episode steps: 783, steps per second: 35, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2
                  episode: 93, duration: 22.777s, episode steps: 823, steps per second: 36, episode reward: -21.000, mean reward: -0.026 [-1.000,
                                                                                                                                                                                     0.000], mean action: 2
 85291/450000: episode: 94, duration: 21.673s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 2
                  episode: 95, duration: 29.395s, episode steps: 1077, steps per second: 37, episode reward: -19.000, mean reward: -0.018 [-1.000,
 87258/450000: episode: 96, duration: 24.114s, episode steps: 890, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 2
                  episode: 97, duration: 27.637s, episode steps: 1005, steps per second: 36, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
 89100/450000: episode: 98, duration: 23.040s, episode steps: 837, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 2
                  episode: 99, duration: 22.348s, episode steps: 820, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 2
 90789/450000: episode: 100, duration: 23.775s, episode steps: 869, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
                   episode: 101, duration: 21.996s, episode steps: 811, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000,
 91600/450000
                                                                                                                                                                                      0.000], mean action:
 92452/450000: episode: 102, duration: 23.362s, episode steps: 852, steps per second: 36, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
                  episode: 103, duration: 25.698s, episode steps: 932, steps per second: 36, episode reward: -21.000, mean reward: -0.023 [-1.000,
 94549/450000: episode: 104. duration: 31.974s. episode steps: 1165. steps per second: 36. episode reward: -19.000. mean reward: -0.016 [-1.000. 1.000]. mean action:
                                                                               920, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 95469/450000:
                  episode: 105, duration: 25.025s, episode steps:
 96579/450000: episode: 106, duration: 30.271s, episode steps: 1110, steps per second: 37, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
                  episode: 107, duration: 21.661s, episode steps: 793, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000,
 98164/450000: episode: 108, duration: 21.848s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
                  episode: 109, duration: 20.916s, episode steps:
                                                                               764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000,
 99949/450000: episode: 110, duration: 27.537s, episode steps: 1021, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
                  episode: 111, duration: 26.383s, episode steps: 976, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
101717/450000; episode: 112, duration: 21.684s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
102648/450000:
                  episode: 113, duration: 25.309s, episode steps: 931, steps per second: 37, episode reward: -20.000, mean reward: -0.021
                                                                                                                                                                                       1.000], mean action:
103461/450000: episode: 114, duration: 22.029s, episode steps: 813, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                                           [-1.000, 0.000], mean action:
                  episode: 115, duration: 24.602s, episode steps: 901, steps per second: 37, episode reward: -20.000, mean reward: -0.022
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                       1.000], mean action:
105232/450000: episode: 116, duration: 23.665s, episode steps: 870, steps per second: 37, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                      1.000], mean action:
                                                24.259s, episode steps:
                  episode: 117, duration:
                                                                               887, steps per second: 37, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                                                       1.000], mean action:
106963/450000: episode: 118. duration: 23.031s. episode steps: 844. steps per second: 37. episode reward: -21.000. mean reward: -0.025
                                                                                                                                                                           [-1.000.
                                                                                                                                                                                      0.0001, mean action:
                  episode: 119, duration: 21.430s, episode steps: 783, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                            [-1.000,
                                                                                                                                                                                       0.000], mean action:
108538/450000: episode: 120, duration: 21.353s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                           [-1.000, 0.000], mean action:
109321/450000:
                  episode: 121, duration: 21.552s, episode steps:
                                                                               783, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                                       0.000], mean action:
110115/450000: episode: 122, duration: 21.896s, episode steps: 794, steps per second: 36, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                                           [-1.000, 0.000], mean action:
                  episode: 123, duration: 21.726s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027
111873/450000: episode: 124, duration: 26.606s, episode steps: 966, steps per second: 36, episode reward: -19.000, mean reward: -0.020
                                                                                                                                                                           [-1.000, 1.000], mean action:
                  episode: 125, duration:
                                                26.401s, episode steps:
                                                                               962, steps per second:
                                                                                                             36, episode reward: -20.000, mean reward: -0.021
                                                                                                                                                                                       1.000], mean action:
113646/450000; episode: 126, duration: 22.323s, episode steps: 811, steps per second: 36, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                                           [-1.000.
                                                                                                                                                                                      0.0001, mean action:
                  episode: 127, duration: 23.832s, episode steps: 862, steps per second: 36, episode reward: -21.000, mean reward: -0.024
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                      0.000], mean action:
115375/450000: episode: 128, duration: 23.998s, episode steps: 867, steps per second: 36, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                                           [-1.000, 1.000], mean action:
                  episode: 129, duration: 24.917s, episode steps: 910, steps per second: 37, episode reward: -21.000, mean reward: -0.023
                                                                                                                                                                                      0.000], mean action:
117127/450000: episode: 130, duration: 23.192s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                      1.000], mean action:
                  episode: 131, duration: 21.763s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                                       0.000], mean action:
118768/450000: episode: 132. duration: 23.291s. episode steps: 849. steps per second: 36. episode reward: -21.000. mean reward: -0.025
                                                                                                                                                                           [-1.000, 0.000], mean action:
                  episode: 133, duration:
                                                26.928s, episode steps:
                                                                                     steps per second:
                                                                                                                  episode reward: -19.000, mean reward: -0.019
                                                                                                                                                                             -1.000,
                                                                                                                                                                                       1.000], mean action:
120529/450000; episode: 134, duration: 20.779s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                           [-1.000.
                                                                                                                                                                                      0.0001, mean action:
                  episode: 135, duration: 25.495s, episode steps: 930, steps per second: 36, episode reward: -20.000, mean reward: -0.022
122251/450000: episode: 136, duration: 21.503s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                      0.000], mean action:
                  episode: 137, duration: 22.974s, episode steps:
123101/450000:
                                                                               850, steps per second:
                                                                                                                  episode reward: -21.000, mean reward: -0.025
                                                                                                                                                                                       0.000], mean action:
124096/450000: episode: 138, duration: 26.905s, episode steps: 995, steps per second: 37, episode reward: -19.000, mean reward: -0.019
                                                                                                                                                                           [-1.000, 1.000], mean action:
                  episode: 139, duration: 22.858s, episode steps: 842, steps per second:
                                                                                                                                                                            [-1.000,
125702/450000; episode: 140, duration: 20.630s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                           [-1.000.
                                                                                                                                                                                      0.0001, mean action:
126544/450000:
                  episode: 141, duration: 22.760s, episode steps: 842, steps per second: 37,
                                                                                                                  episode reward: -20.000, mean reward: -0.024
                                                                                                                                                                            [-1.000,
                                                                                                                                                                                       1.000], mean action:
127336/450000: episode: 142, duration: 21.446s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                           [-1.000.
                                                                                                                                                                                      0.0001, mean action:
                  episode: 143, duration: 20.801s, episode steps:
                                                                                                                  episode reward: -21.000, mean reward: -0.027
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                       0.000], mean action
128968/450000: episode: 144, duration: 23.393s, episode steps: 868, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
                  episode: 145, duration:
                                                27.122s, episode steps:
                                                                               1010, steps per second:
                                                                                                              37, episode reward: -19.000, mean reward: -
130916/450000; episode: 146, duration: 25,274s, episode steps: 938, steps per second: 37, episode reward: -19,000, mean reward: -0.020 [-1,000, 1,000], mean action:
                  episode: 147, duration: 24.267s, episode steps: 898, steps per second: 37, episode reward: -20.000, mean reward: -0.022
131814/450000:
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                      1.000], mean action:
132713/450000: episode: 148, duration: 24.362s, episode steps: 899, steps per second: 37, episode reward: -20.000, mean reward: -0.022
                                                                                                                                                                           [-1.000, 1.000], mean action:
                                                                                                                                                                           r-1.000,
                                                                                                                  episode reward: -21.000, mean reward: -0.025
133567/450000.
                   episode: 149, duration: 23.055s, episode steps: 854,
                                                                                                                                                                                      0.000], mean action:
134466/450000: episode: 150, duration: 24.075s, episode steps: 899, steps per second: 37, episode reward: -20.000, mean reward: -0.022
                                                                                                                                                                           [-1.000, 1.000], mean action:
                  episode: 151, duration: 24.020s, episode steps: 880, steps per second: 37,
                                                                                                                  episode reward: -21.000, mean reward: -0.024
                                                                                                                                                                            [-1.000,
136314/450000: episode: 152, duration: 26.416s, episode steps: 968, steps per second: 37, episode reward: -19.000, mean reward: -0.020
                                                                                                                                                                           [-1.000, 1.000], mean action:
137194/450000:
                  episode: 153, duration: 24.113s, episode steps: 880,
                                                                                     steps per second: 36,
                                                                                                                  episode reward: -21.000, mean reward: -0.024
                                                                                                                                                                            -1.000,
138006/450000
                  episode: 154, duration: 21.903s, episode steps: 812, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                                           [-1.000, 0.000], mean action:
                  episode: 155, duration: 22.558s, episode steps: 824, steps per second: 37,
                                                                                                                                                                            [-1.000,
                                                                                                                                                                                       0.000], mean action:
139624/450000: episode: 156, duration: 21.372s, episode steps: 794, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                                           [-1.000, 0.000], mean action:
                                                                               792,
                   episode: 157, duration: 21.513s, episode steps:
140416/450000
                                                                                      steps per second: 37,
                                                                                                                  episode reward: -21.000,
                                                                                                                                                 mean reward: -0.027
                                                                                                                                                                            [-1.000.
                                                                                                                                                                                      0.000], mean action
141362/450000: episode: 158, duration: 25.839s, episode steps: 946, steps per second: 37, episode reward: -21.000, mean reward: -0.022
                                                                                                                                                                           [-1.000, 0.000], mean action:
                  episode: 159, duration: 26.412s, episode steps: 966, steps per second:
143279/450000; episode: 160, duration: 25.791s, episode steps: 951, steps per second: 37, episode reward: -19.000, mean reward: -0.020
                                                                                                                                                                           [-1.000, 1.000], mean action:
144123/450000:
                  episode: 161, duration: 22.776s, episode steps: 844, steps per second: 37,
                                                                                                                  episode reward: -20.000, mean reward: -0.024
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                       1.000], mean action:
145021/450000; episode: 162, duration: 24.313s, episode steps: 898, steps per second: 37, episode reward: -20.000, mean reward: -0.022
                                                                                                                                                                           [-1.000.
                                                                                                                                                                                      1.0001, mean action:
                  episode: 163, duration: 24.097s, episode steps: 889, steps per second: 37,
                                                                                                                  episode reward: -20.000, mean reward: -0.022
146773/450000: episode: 164, duration: 23.494s, episode steps: 863, steps per second: 37,
                                                                                                                  episode reward: -20.000, mean reward: -0.023
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                      1.000], mean action:
147787/450000
                   episode: 165, duration:
                                                27.528s, episode steps:
                                                                               1014.
                                                                                       steps per second:
                                                                                                              37, episode reward: -19.000,
                                                                                                                                                  mean reward: -0.019
                                                                                                                                                                             [-1.000
                                                                                                                                                                                        1.0001.
148598/450000: episode: 166, duration: 21.870s, episode steps: 811, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
                                                                                                                                                                            [-1.000,
                  episode: 167, duration: 24.013s, episode steps: 887, steps per second:
                                                                                                                  episode reward: -20.000, mean reward: -0.023
150356/450000; episode: 168, duration: 23.747s, episode steps: 871, steps per second: 37, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                                           [-1.000, 1.000], mean action:
151176/450000:
                  episode: 169, duration: 22.094s, episode steps: 820, steps per second: 37,
                                                                                                                  episode reward: -21.000, mean reward: -0.026
                                                                                                                                                                           [-1.000,
                                                                                                                                                                                      0.000], mean action:
152018/450000: episode: 170, duration: 22.827s, episode steps: 842, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 152782/450000: episode: 171, duration: 20.517s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
```

```
153861/450000: episode: 172, duration: 28.961s, episode steps: 1079, steps per second: 37, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action:
154653/450000: episode: 173, duration: 21.318s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 155615/450000: episode: 174, duration: 26.119s, episode steps: 962, steps per second: 37, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
156459/450000: episode: 175, duration: 23.203s, episode steps: 844, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
157299/450000: episode: 176, duration: 52774.022s, episode steps: 840, steps per second: 0, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action
158211/450000: episode: 177, duration: 25.231s, episode steps: 912, steps per second: 36, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
               episode: 178, duration: 24.746s, episode steps: 914, steps per second: 37, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
160106/450000: episode: 179. duration: 25.464s. episode steps: 981. steps per second: 39. episode reward: -20.000. mean reward: -0.020 [-1.000. 1.000], mean action:
               episode: 180, duration: 22.574s, episode steps: 861, steps per second: 38, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000, 1.000], mean action:
161897/450000: episode: 181, duration: 23.948s, episode steps: 930, steps per second: 39, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
162809/450000: episode: 182, duration: 23.694s, episode steps: 912, steps per second: 38, episode reward: -21.000, mean reward: -0.023
                                                                                                                                            [-1.000, 0.000], mean action:
163592/450000: episode: 183, duration: 20.140s, episode steps: 783, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 184, duration: 21.901s, episode steps: 848, steps per second: 39, episode reward: -21.000, mean reward: -0.025
                                                                                                                                                     0.000], mean action:
165294/450000: episode: 185, duration: 23.115s, episode steps: 854, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
               episode: 186, duration: 23.191s, episode steps: 861, steps per second:
                                                                                         37, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
166999/450000; episode: 187, duration: 23.067s, episode steps: 844, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
168114/450000: episode: 188, duration: 30.310s, episode steps: 1115, steps per second: 37, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
168934/450000: episode: 189, duration: 22.328s, episode steps: 820, steps per second: 37, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
               episode: 190, duration: 29.002s, episode steps: 1080, steps per second: 37, episode reward: -18.000, mean reward: -0.017 [-1.000, 1.000], mean action:
171035/450000: episode: 191, duration: 27.202s, episode steps: 1021, steps per second: 38, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
171922/450000: episode: 192, duration: 24.452s, episode steps: 887, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
172804/450000: episode: 193, duration: 23.928s, episode steps: 882, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 173688/450000: episode: 194, duration: 23.665s, episode steps: 884, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
                                        23.665s, episode steps:
174970/450000: episode: 195, duration: 34.266s, episode steps: 1282, steps per second: 37, episode reward: -17.000, mean reward: -0.013 [-1.000, 1.000], mean action:
175872/450000: episode: 196, duration: 24.185s, episode steps: 902, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
176714/450000: episode: 197, duration: 22.718s, episode steps: 842, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
               episode: 198, duration: 23.314s, episode steps: 872, steps per second: 37, episode reward: -20.000, mean reward: -0.023
178436/450000: episode: 199, duration: 23.118s, episode steps: 850, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
               episode: 200, duration: 23.044s, episode steps: 862, steps per second: 37, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                     1.000], mean action:
180140/450000: episode: 201, duration: 22.545s, episode steps: 842, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
               episode: 202, duration: 21.998s, episode steps: 822, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                            [-1.000,
                                                                                                                                                     0.000], mean action:
181946/450000: episode: 203, duration: 26.264s, episode steps: 984, steps per second: 37, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
               episode: 204, duration: 27.956s, episode steps: 1052, steps per second: 38, episode reward: -20.000, mean reward:
                                                                                                                                      -0.019
183898/450000: episode: 205, duration: 24.135s, episode steps: 900, steps per second: 37, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
               episode: 206, duration: 27.956s, episode steps: 1042, steps per second: 37, episode reward: -19.000, mean reward: -
                                                                                                                                            [-1.000, 1.000], mean action:
185831/450000; episode: 207, duration: 23.859s, episode steps: 891, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
186683/450000:
               episode: 208, duration: 22.806s, episode steps: 852, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
187647/450000: episode: 209, duration: 25.776s, episode steps: 964, steps per second: 37, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
               episode: 210, duration: 22.824s, episode steps: 842, steps per second: 37,
                                                                                             episode reward: -20.000, mean reward: -0.024
188489/450000.
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
189326/450000: episode: 211, duration: 22.335s, episode steps: 837, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
                                                                                                                                                      1.000], mean action:
               episode: 212, duration: 41.778s, episode steps: 1545, steps per second: 37, episode reward: -17.000, mean reward: -
191663/450000: episode: 213, duration: 21.095s, episode steps: 792, steps per second: 38, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 214, duration: 26.433s, episode steps: 982, steps per second: 37, episode reward: -20.000, mean reward: -0.020
192645/450000:
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
193409/450000; episode: 215, duration: 20.375s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 216, duration: 25.872s, episode steps: 968, steps per second: 37, episode reward: -19.000, mean reward: -0.020
                                                                                                                                                     1.000], mean action:
195169/450000: episode: 217, duration: 21.312s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 218, duration: 26.433s, episode steps:
                                                                949, steps per second: 36, episode reward: -19.000, mean reward: -0.020
                                                                                                                                                     1.000], mean action:
196910/450000: episode: 219, duration: 21.906s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 220, duration: 23.762s, episode steps: 867, steps per second: 36, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                     1.000], mean action:
198731/450000; episode: 221, duration: 25.958s, episode steps: 954, steps per second: 37, episode reward: -21.000, mean reward: -0.022
                                                                                                                                            [-1.000, 0.000], mean action:
199581/450000:
               episode: 222, duration: 22.879s, episode steps: 850, steps per second: 37, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000,
                                                                                                                                                     0.000], mean action:
200392/450000: episode: 223, duration: 21.690s, episode steps: 811, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 224, duration: 22.513s, episode steps: 839, steps per second: 37, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000,
                                                                                                                                                     0.000], mean action:
202041/450000: episode: 225, duration: 21.785s, episode steps: 810, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                            [-1.000,
                                                                                                                                                     0.000], mean action:
                                       26.148s, episode steps:
                                                                      steps per second: 37, episode reward: -21.000, mean reward: -0.021
               episode: 226, duration:
                                                                                                                                                     0.000], mean action:
203935/450000: episode: 227. duration: 24.476s. episode steps: 916. steps per second: 37. episode reward: -20.000. mean reward: -0.022
                                                                                                                                            [-1.000.
                                                                                                                                                     1.0001, mean action:
               episode: 228, duration: 24.113s, episode steps: 871, steps per second: 36, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
205658/450000: episode: 229, duration: 23.357s, episode steps: 852, steps per second: 36, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000, 0.000], mean action:
206422/450000:
               episode: 230, duration: 21.277s, episode steps: 764, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                     0.000], mean action:
207317/450000: episode: 231, duration: 24.718s, episode steps: 895, steps per second: 36, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 232, duration: 21.777s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027
208873/450000: episode: 233, duration: 21.023s, episode steps: 764, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 234, duration:
                                        27.126s, episode steps:
                                                                 993, steps per second: 37, episode reward: -19.000, mean reward: -0.019
                                                                                                                                                     1.000], mean action:
210888/450000; episode: 235, duration: 27,988s, episode steps: 1022, steps per second: 37, episode reward: -20,000, mean reward: -0.020 [-1,000, 1,000], mean action:
               episode: 236, duration: 23.133s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
212522/450000: episode: 237, duration: 21.409s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 238, duration: 28.813s, episode steps: 1053, steps per second: 37, episode reward: -19.000, mean reward: -
                                                                                                                                            [-1.000, 1.000], mean action:
214417/450000: episode: 239, duration: 23.393s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
               episode: 240, duration: 25.196s, episode steps: 920, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
216292/450000: episode: 241. duration: 26.421s. episode steps: 955. steps per second: 36. episode reward: -20.000. mean reward: -0.021 [-1.000. 1.000]. mean action:
               episode: 242, duration:
                                        26.624s, episode steps: 981, steps per second: 37, episode reward: -20.000, mean reward: -0.020
                                                                                                                                            [-1.000,
218237/450000; episode: 243, duration: 26.001s, episode steps: 964, steps per second: 37, episode reward: -20.000, mean reward: -0.021
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 244, duration: 27.386s, episode steps: 987, steps per second: 36, episode reward: -20.000, mean reward: -0.020
220063/450000: episode: 245, duration: 22.949s, episode steps: 839, steps per second: 37, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 246, duration: 21.070s, episode steps:
220827/450000:
                                                                764, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                     0.000], mean action:
221647/450000: episode: 247, duration: 22.381s, episode steps: 820, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 248, duration: 23.430s, episode steps: 844, steps per second: 36, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000,
223363/450000; episode: 249, duration: 23.765s, episode steps: 872, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
224245/450000:
               episode: 250, duration: 24.312s, episode steps: 882, steps per second: 36, episode reward: -21.000, mean reward: -0.024
                                                                                                                                            [-1.000,
                                                                                                                                                     0.000], mean action:
225197/450000: episode: 251, duration: 26.040s, episode steps: 952, steps per second: 37, episode reward: -19.000, mean reward: -0.020
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 252, duration: 21.897s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000,
                                                                                                                                                     0.000], mean action
227081/450000: episode: 253, duration: 30.262s, episode steps: 1092, steps per second: 36, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean action:
                                                                 821, steps per second: 36, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
               episode: 254, duration:
                                        22.716s, episode steps:
228886/450000: episode: 255, duration: 26.569s, episode steps: 984, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
               episode: 256, duration: 21.817s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
229678/450000:
230925/450000: episode: 257, duration: 33.783s, episode steps: 1247, steps per second: 37, episode reward: -17.000, mean reward: -0.014 [-1.000, 1.000], mean action:
231823/450000.
               episode: 258, duration: 24.235s, episode steps: 898, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
232920/450000: episode: 259, duration: 29.947s, episode steps: 1097, steps per second: 37, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action:
               episode: 260, duration: 22.847s, episode steps: 841, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
234768/450000: episode: 261, duration: 27.311s, episode steps: 1007, steps per second: 37, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
235666/450000:
               episode: 262, duration: 24.414s, episode steps: 898, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
236548/450000; episode: 263, duration: 23.853s, episode steps: 882, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
               episode: 264, duration: 20.936s, episode steps: 792, steps per second: 38, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
238190/450000: episode: 265, duration: 22.385s, episode steps: 850, steps per second: 38, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000, 0.000], mean action:
                                                                                                                                            r-1.000,
               episode: 266, duration: 25.218s, episode steps: 958,
                                                                      steps per second: 38, episode reward: -20.000, mean reward: -0.021
239148/450000
                                                                                                                                                     1.000], mean action:
240079/450000: episode: 267, duration: 24.000s, episode steps: 931, steps per second: 39, episode reward: -20.000, mean reward: -0.021
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 268, duration: 20.662s, episode steps:
                                                                 794, steps per second: 38, episode reward: -21.000, mean reward: -0.026
241814/450000; episode: 269, duration: 24.467s, episode steps: 941, steps per second: 38, episode reward: -21.000, mean reward: -0.022
                                                                                                                                            [-1.000, 0.000], mean action:
242805/450000:
               episode: 270, duration: 26.679s, episode steps: 991, steps per second: 37, episode reward: -20.000, mean reward: -0.020
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
243569/450000; episode: 271, duration: 19.949s, episode steps: 764, steps per second: 38, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 272, duration: 21.682s, episode steps: 838, steps per second: 39, episode reward: -20.000, mean reward: -0.024
245199/450000: episode: 273, duration: 20.318s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
                                                                                                                                             -1.000,
245991/450000
               episode: 274,
                                        20.539s, episode steps:
                                                                792.
                                                                      steps per second: 39,
                                                                                             episode reward: -21.000,
                                                                                                                       mean reward: -0.027
                                                                                                                                                     0.000], mean action:
246833/450000: episode: 275, duration: 21.838s, episode steps: 842, steps per second: 39, episode reward: -20.000, mean reward: -0.024
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
                                                                                                                                            [-1.000,
               episode: 276, duration: 21.879s, episode steps: 842, steps per second:
                                                                                         38, episode reward: -20.000, mean reward: -0.024
248596/450000; episode: 277, duration: 24.597s, episode steps: 921, steps per second: 37, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000, 1.000], mean action:
249438/450000:
               episode: 278, duration: 22.164s, episode steps: 842, steps per second: 38, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000, 0.000], mean action:
250280/450000: episode: 279, duration: 22.010s, episode steps: 842, steps per second: 38, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 251072/450000: episode: 280, duration: 20.471s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
```

```
251926/450000: episode: 281, duration: 21.936s, episode steps: 854, steps per second: 39, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
252921/450000: episode: 282, duration: 25.530s, episode steps: 995, steps per second: 39, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
               episode: 283, duration: 28.465s, episode steps: 1091, steps per second: 38, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action:
254926/450000: episode: 284, duration: 23.548s, episode steps: 914, steps per second: 39, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
256040/450000: episode: 285, duration: 28.718s, episode steps: 1114, steps per second: 39, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action:
256927/450000: episode: 286, duration: 22.829s, episode steps: 887, steps per second: 39, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
               episode: 287, duration: 21.072s, episode steps: 820, steps per second: 39, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
258767/450000: episode: 288, duration: 26.178s, episode steps: 1020, steps per second: 39, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
                                          .333s, episode steps: 1036, steps per second: 38, episode reward: -19.000, mean reward: -0.018 [-1.000,
259803/450000: episode: 289, duration: 27
                                                                                                                                                      1.000], mean action:
260567/450000: episode: 290, duration: 20.824s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
261561/450000: episode: 291, duration: 27.290s, episode steps: 994, steps per second: 36, episode reward: -20.000, mean reward: -0.020
                                                                                                                                           [-1.000, 1.000], mean action:
262440/450000: episode: 292, duration: 23.740s, episode steps: 879, steps per second: 37, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 263284/450000: episode: 293, duration: 22.849s, episode steps: 844, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
264048/450000: episode: 294, duration: 20.700s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 295, duration: 28.287s, episode steps: 1042, steps per second: 37, episode reward: -20.000, mean reward:
                                                                                                                                            [-1.000, 1.000], mean action:
265854/450000; episode: 296, duration: 20.809s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
266676/450000: episode: 297, duration: 22.281s, episode steps: 822, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                           [-1.000, 0.000], mean action:
267440/450000: episode: 298, duration: 20.557s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 299, duration: 21.635s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
269174/450000: episode: 300, duration: 25.841s, episode steps: 942, steps per second: 36, episode reward: -21.000, mean reward: -0.022
                                                                                                                                           [-1.000, 0.000], mean action:
270045/450000: episode: 301, duration: 23.621s, episode steps: 871, steps per second: 37, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
270915/450000: episode: 302, duration: 23.770s, episode steps: 870, steps per second: 37, episode reward: -20.000, mean reward: -0.023 [-1.000, 271815/450000: episode: 303, duration: 24.450s, episode steps: 900, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000,
                                                                                                                                                    1.000], mean action:
                                                                                                                                                     1.000], mean action:
272834/450000: episode: 304, duration: 28.119s, episode steps: 1019, steps per second: 36, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
273686/450000: episode: 305, duration: 23.561s, episode steps: 852, steps per second: 36, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
274594/450000: episode: 306, duration: 24.843s, episode steps: 908, steps per second: 37, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
               episode: 307, duration: 22.820s, episode steps:
                                                                824, steps per second: 36, episode reward: -21.000, mean reward: -0.025
276378/450000: episode: 308. duration: 26.416s. episode steps: 960. steps per second: 36. episode reward: -19.000. mean reward: -0.020
                                                                                                                                           [-1.000, 1.000], mean action:
               episode: 309, duration: 22.408s, episode steps: 820, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                     0.000], mean action:
277990/450000; episode: 310, duration: 21.925s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 311, duration: 22.499s, episode steps: 820, steps per second: 36, episode reward: -21.000, mean reward: -0.026
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
279733/450000: episode: 312, duration: 25.549s, episode steps: 923, steps per second: 36, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 313, duration: 21.772s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027
281308/450000: episode: 314, duration: 21.523s, episode steps: 783, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 315, duration: 23.371s, episode steps: 850, steps per second: 36, episode reward: -21.000, mean reward: -0.025
                                                                                                                                                     0.000], mean action:
282956/450000; episode: 316, duration: 21.971s, episode steps: 798, steps per second: 36, episode reward: -21.000, mean reward: -0.026
                                                                                                                                            [-1.000, 0.000], mean action:
283739/450000:
               episode: 317, duration: 21.665s, episode steps: 783, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
284610/450000: episode: 318, duration: 23.912s, episode steps: 871, steps per second: 36, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000.
                                                                                                                                                    1.0001, mean action:
               episode: 319, duration: 25.006s, episode steps: 902, steps per second: 36, episode reward: -20.000, mean reward: -0.022
285512/450000.
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
286470/450000: episode: 320, duration: 26.637s, episode steps: 958, steps per second: 36, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
               episode: 321, duration: 30.642s, episode steps: 1117, steps per second: 36, episode reward: -18.000, mean reward: -
                                                                                                                                      -0.016 [-1.000, 1.000], mean action:
288584/450000: episode: 322, duration: 27.390s, episode steps: 997, steps per second: 36, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
               episode: 323, duration: 29.347s, episode steps: 1060, steps per second: 36, episode reward: -19.000, mean reward: -0.018
                                                                                                                                            [-1.000,
290408/450000; episode: 324, duration: 21,241s, episode steps: 764, steps per second: 36, episode reward: -21,000, mean reward: -0.027 [-1,000, 0.000], mean action:
               episode: 325, duration: 25.172s, episode steps: 921, steps per second: 37, episode reward: -20.000, mean reward: -0.022
292230/450000: episode: 326, duration: 24.631s, episode steps: 901, steps per second: 37, episode reward: -21.000, mean reward: -0.023
                                                                                                                                           [-1.000, 0.000], mean action:
                                                                1100, steps per second: 37, episode reward: -19.000, mean reward: -0.017
               episode: 327, duration: 29.787s, episode steps:
                                                                                                                                            [-1.000,
                                                                                                                                                      1.000], mean action:
294230/450000: episode: 328, duration: 24.321s, episode steps: 900, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
               episode: 329, duration: 22.415s, episode steps: 826, steps per second: 37, episode reward: -21.000, mean reward: -0.025
                                                                                                                                                    0.000], mean action:
295906/450000; episode: 330, duration: 23,233s, episode steps: 850, steps per second: 37, episode reward: -21,000, mean reward: -0.025 [-1,000, 0.000], mean action:
297026/450000:
               episode: 331, duration: 30.230s, episode steps: 1120, steps per second: 37, episode reward: -19.000, mean reward: -0.017
                                                                                                                                            [-1.000, 1.000], mean action
298010/450000: episode: 332, duration: 26.686s, episode steps: 984, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
               episode: 333, duration: 24.262s, episode steps: 902, steps per second: 37, episode reward: -21.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
299704/450000: episode: 334, duration: 21.923s, episode steps: 792, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
                                       27.383s, episode steps:
               episode: 335, duration:
                                                                997, steps per second: 36, episode reward: -19.000, mean reward: -0.019
301603/450000: episode: 336. duration: 24.470s. episode steps: 902. steps per second: 37. episode reward: -20.000. mean reward: -0.022
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 337, duration: 23.371s, episode steps: 860, steps per second: 37, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
303412/450000: episode: 338, duration: 25.993s, episode steps: 949, steps per second: 37, episode reward: -20.000, mean reward: -0.021
                                                                                                                                           [-1.000, 1.000], mean action:
304283/450000:
               episode: 339, duration: 23.926s, episode steps: 871, steps per second: 36, episode reward: -20.000, mean reward: -0.023
305228/450000: episode: 340, duration: 26.294s, episode steps: 945, steps per second: 36, episode reward: -20.000, mean reward: -0.021
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 341, duration: 23.920s, episode steps: 871, steps per second: 36, episode reward: -20.000, mean reward: -0.023
306982/450000: episode: 342, duration: 24.045s, episode steps: 883, steps per second: 37, episode reward: -21.000, mean reward: -0.024
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 343, duration: 24.812s, episode steps:
                                                                902, steps per second: 36, episode reward: -20.000, mean reward: -0.022
                                                                                                                                                     1.000], mean action:
308710/450000; episode: 344, duration: 22.479s, episode steps: 826, steps per second: 37, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000.
                                                                                                                                                    0.0001, mean action:
               episode: 345, duration: 23.847s, episode steps: 880, steps per second: 37, episode reward: -21.000, mean reward: -0.024
                                                                                                                                            [-1.000,
310540/450000: episode: 346, duration: 25.658s, episode steps: 950, steps per second: 37, episode reward: -20.000, mean reward: -0.021
                                                                                                                                           [-1.000, 1.000], mean action:
               episode: 347, duration: 21.883s, episode steps: 811, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                     0.000], mean action:
312171/450000: episode: 348, duration: 22.402s, episode steps: 820, steps per second: 37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
               episode: 349, duration: 26.893s, episode steps: 979, steps per second: 36, episode reward: -19.000, mean reward: -0.019
                                                                                                                                                     1.000], mean action:
313914/450000: episode: 350. duration: 20.979s. episode steps: 764. steps per second: 36. episode reward: -21.000. mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
                                                                764, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                       20.986s, episode steps:
                                                                                                                                             -1.000,
                                                                                                                                                     0.000], mean action:
315442/450000; episode: 352, duration: 20.920s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 353, duration: 23.236s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024
317194/450000: episode: 354, duration: 24.972s, episode steps: 910, steps per second: 36, episode reward: -21.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
               episode: 355, duration: 22.154s, episode steps:
318005/450000:
                                                                811, steps per second: 37,
                                                                                            episode reward: -21.000, mean reward: -0.026
                                                                                                                                                     0.000], mean action:
318769/450000: episode: 356, duration: 20.961s, episode steps: 764, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 357, duration: 22.920s, episode steps: 837, steps per second:
                                                                                        37, episode reward: -20.000, mean reward: -0.024
                                                                                                                                            [-1.000,
320398/450000; episode: 358, duration: 21.645s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
321248/450000:
               episode: 359, duration: 23.442s, episode steps: 850, steps per second: 36, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
322208/450000: episode: 360, duration: 26.286s, episode steps: 960, steps per second: 37, episode reward: -20.000, mean reward: -0.021
                                                                                                                                            [-1.000.
                                                                                                                                                    1.0001, mean action:
               episode: 361, duration: 21.082s, episode steps: 764, steps per second: 36,
                                                                                            episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000,
                                                                                                                                                     0.000], mean action
323794/450000: episode: 362, duration: 22.526s, episode steps: 822, steps per second: 36, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
               episode: 363, duration:
                                       23.129s, episode steps: 843, steps per second: 36, episode reward: -21.000, mean reward: -0.025
                                                                                                                                                     0.000], mean action:
325401/450000; episode: 364, duration: 21.268s, episode steps: 764, steps per second: 36, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000.
                                                                                                                                                    0.0001, mean action:
326339/450000:
               episode: 365, duration: 25.551s, episode steps: 938, steps per second: 37, episode reward: -19.000, mean reward: -0.020
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
327163/450000: episode: 366, duration: 22.519s, episode steps: 824, steps per second: 37, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000.
                                                                                                                                                    0.0001, mean action:
                                                                                                                                            [-1.000,
                                                                                            episode reward: -21.000, mean reward: -0.027
               episode: 367, duration: 21.435s, episode steps: 792,
                                                                                                                                                    0.000], mean action:
327955/450000.
328892/450000: episode: 368, duration: 25.135s, episode steps: 937, steps per second: 37, episode reward: -19.000, mean reward: -0.020
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 369, duration: 20.713s, episode steps: 764, steps per second: 37,
                                                                                                                                            [-1.000,
330510/450000: episode: 370, duration: 22.851s, episode steps: 854, steps per second: 37, episode reward: -21.000, mean reward: -0.025
                                                                                                                                           [-1.000, 0.000], mean action:
331468/450000:
               episode: 371, duration:
                                       25.715s, episode steps: 958,
                                                                     steps per second: 37,
                                                                                            episode reward: -20.000, mean reward: -0.021
                                                                                                                                            -1.000,
                                                                                                                                                     1.000], mean action:
332260/450000; episode: 372, duration: 21.348s, episode steps: 792, steps per second: 37, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 373, duration: 23.026s, episode steps: 842, steps per second: 37,
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
333973/450000: episode: 374, duration: 24.051s, episode steps: 871, steps per second: 36, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 375, duration: 22.764s, episode steps: 842,
                                                                                            episode reward: -20.000,
334815/450000.
                                                                     steps per second: 37,
                                                                                                                      mean reward: -0.024
                                                                                                                                            [-1.000.
                                                                                                                                                     1.000], mean action
335640/450000: episode: 376, duration: 22.589s, episode steps: 825, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
               episode: 377, duration: 28.281s, episode steps: 1026, steps per second: 36, episode reward: -20.000, mean reward: -
337430/450000: episode: 378, duration: 20.814s, episode steps: 764, steps per second: 37, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
338334/450000:
               episode: 379, duration: 24.176s, episode steps: 904, steps per second: 37,
                                                                                            episode reward: -21.000, mean reward: -0.023 [-1.000,
                                                                                                                                                    0.000], mean action:
339243/450000; episode: 380, duration: 24.388s, episode steps: 909, steps per second: 37, episode reward: -21.000, mean reward: -0.023
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 381, duration: 22.600s, episode steps: 842, steps per second: 37,
                                                                                            episode reward: -20.000, mean reward: -0.024
340983/450000: episode: 382, duration: 23.940s, episode steps: 898, steps per second: 38, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000, 1.000], mean action:
                                                                                                                                            -1.000,
                                                                                                                                                     1.000],
341849/450000
                        383,
                             duration: 23.194s, episode steps:
                                                                      steps per second: 37,
                                                                                            episode reward: -20.000,
                                                                                                                      mean reward: -0.023
342747/450000: episode: 384, duration: 23.982s, episode steps: 898, steps per second: 37, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
                                                                                                                                            -1.000,
               episode: 385, duration: 24.168s, episode steps: 898, steps per second:
                                                                                            episode reward: -20.000, mean reward: -0.022
344623/450000; episode: 386, duration: 26.154s, episode steps: 978, steps per second: 37, episode reward: -19.000, mean reward: -0.019
                                                                                                                                           [-1.000, 1.000], mean action:
345541/450000:
               episode: 387, duration: 24.766s, episode steps: 918, steps per second: 37, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
346459/450000: episode: 388, duration: 25.121s, episode steps: 918, steps per second: 37, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 347379/450000: episode: 389, duration: 25.165s, episode steps: 920, steps per second: 37, episode reward: -19.000, mean reward: -0.021 [-1.000, 1.000], mean action:
```

```
348251/450000: episode: 390, duration: 23.949s, episode steps: 872, steps per second: 36, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
349088/450000: episode: 391, duration: 22.656s, episode steps: 837, steps per second: 37, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 350215/450000: episode: 392, duration: 30.610s, episode steps: 1127, steps per second: 37, episode reward: -17.000, mean reward: -0.015 [-1.000, 1.000], mean action:
351164/450000: episode: 393, duration: 25.608s, episode steps: 949, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
352048/450000: episode: 394, duration: 24.018s, episode steps: 884, steps per second: 37, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
352900/450000: episode: 395, duration: 22.932s, episode steps: 852, steps per second: 37, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
               episode: 396, duration: 26.287s, episode steps: 972, steps per second: 37, episode reward: -21.000, mean reward: -0.022
354839/450000: episode: 397, duration: 26.006s, episode steps: 967, steps per second: 37, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
                                          .315s, episode steps: 1024, steps per second: 37, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
               episode: 398, duration: 27
356967/450000: episode: 399, duration: 29.743s, episode steps: 1104, steps per second: 37, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
               episode: 400, duration: 23.075s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
358862/450000: episode: 401, duration: 28.958s, episode steps: 1053, steps per second: 36, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action:
               episode: 402, duration: 23.073s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
360546/450000: episode: 403, duration: 23.318s, episode steps: 842, steps per second: 36, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
               episode: 404, duration: 22.366s, episode steps: 820, steps per second:
                                                                                        37, episode reward: -21.000, mean reward: -0.026
                                                                                                                                                    0.000], mean action:
362334/450000; episode: 405, duration: 26.500s, episode steps: 968, steps per second: 37, episode reward: -21.000, mean reward: -0.022
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 406, duration: 23.402s, episode steps: 873, steps per second: 37, episode reward: -21.000, mean reward: -0.024
                                                                                                                                            [-1.000, 0.000], mean action:
364077/450000: episode: 407, duration: 22.831s, episode steps: 870, steps per second: 38, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
               episode: 408, duration: 20.895s, episode steps: 792, steps per second: 38, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
365758/450000: episode: 409, duration: 22.879s, episode steps: 889, steps per second: 39, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
366861/450000: episode: 410, duration: 28.381s, episode steps: 1103, steps per second: 39, episode reward: -18.000, mean reward:
                                                                                                                                      -0.016 [-1.000, 1.000], mean action:
367681/450000: episode: 411, duration: 21.091s, episode steps: 820, steps per second: 39, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 368523/450000: episode: 412, duration: 21.557s, episode steps: 842, steps per second: 39, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
                                       21.557s, episode steps:
369287/450000: episode: 413, duration: 19.590s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                           [-1.000, 0.000], mean action:
370051/450000: episode: 414, duration: 19.625s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
370891/450000: episode: 415, duration: 21.524s, episode steps: 840, steps per second: 39, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
               episode: 416, duration: 21.101s, episode steps: 820, steps per second: 39, episode reward: -21.000, mean reward: -0.026
372640/450000: episode: 417. duration: 23.927s. episode steps: 929. steps per second: 39. episode reward: -20.000. mean reward: -0.022
                                                                                                                                           [-1.000, 1.000], mean action:
               episode: 418, duration: 20.371s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                     0.000], mean action:
374196/450000; episode: 419, duration: 19.729s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 420, duration: 23.270s, episode steps:
                                                                876, steps per second: 38, episode reward: -21.000, mean reward: -0.024
                                                                                                                                                    0.000], mean action:
376101/450000: episode: 421, duration: 27.044s, episode steps: 1029, steps per second: 38, episode reward: -21.000, mean reward: -0.020
                                                                                                                                            [-1.000, 0.000], mean action:
                                                                764, steps per second: 38, episode reward: -21.000, mean reward: -0.027 [-1.000,
               episode: 422, duration: 19.961s, episode steps:
377657/450000: episode: 423, duration: 20.315s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 424, duration: 22.301s, episode steps:
                                                                 870, steps per second: 39, episode reward: -20.000, mean reward: -0.023
                                                                                                                                                     1.000], mean action:
379456/450000; episode: 425, duration: 23.908s, episode steps: 929, steps per second: 39, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000.
                                                                                                                                                    1.0001, mean action:
380358/450000:
               episode: 426, duration: 23.220s, episode steps: 902, steps per second: 39, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
381274/450000: episode: 427, duration: 23.493s, episode steps: 916, steps per second: 39, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000.
                                                                                                                                                    1.0001, mean action:
               episode: 428, duration: 20.340s, episode steps:
                                                                792, steps per second: 39, episode reward: -21.000, mean reward: -0.027
382066/450000.
                                                                                                                                            [-1.000,
                                                                                                                                                     0.000], mean action:
382974/450000: episode: 429, duration: 23.337s, episode steps: 908, steps per second: 39, episode reward: -21.000, mean reward: -0.023
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 430, duration: 24.771s, episode steps: 963, steps per second: 39, episode reward: -20.000, mean reward: -0.021
                                                                                                                                                     1.000], mean action:
384930/450000: episode: 431. duration: 25.495s. episode steps: 993. steps per second: 39. episode reward: -20.000. mean reward: -0.020
                                                                                                                                           [-1.000, 1.000], mean action:
               episode: 432, duration: 21.689s, episode steps: 843, steps per second: 39, episode reward: -20.000, mean reward: -0.024
                                                                                                                                            -1.000,
386627/450000; episode: 433, duration: 22.500s, episode steps: 854, steps per second: 38, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000.
                                                                                                                                                    0.0001, mean action:
               episode: 434, duration: 24.391s, episode steps: 926, steps per second: 38, episode reward: -20.000, mean reward: -0.022
                                                                                                                                                     1.000], mean action:
388317/450000: episode: 435, duration: 19.977s, episode steps: 764, steps per second: 38, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
               episode: 436, duration: 24.669s, episode steps:
                                                                962, steps per second: 39, episode reward: -21.000, mean reward: -0.022
                                                                                                                                                     0.000], mean action:
390149/450000: episode: 437, duration: 22.396s, episode steps: 870, steps per second: 39, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000, 1.000], mean action:
               episode: 438, duration: 19.702s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                     0.000], mean action:
391677/450000; episode: 439, duration: 19.605s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                           [-1.000, 0.000], mean action:
392725/450000:
               episode: 440, duration: 26.958s, episode steps: 1048, steps per second: 39, episode reward: -19.000, mean reward: -0.018
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action
393517/450000: episode: 441, duration: 20.355s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 442, duration: 23.745s, episode steps: 927, steps per second: 39, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
395346/450000: episode: 443, duration: 23.290s, episode steps: 902, steps per second: 39, episode reward: -20.000, mean reward: -0.022
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
                                       20.294s, episode steps:
               episode: 444, duration:
                                                                792, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                                     0.000], mean action:
397052/450000: episode: 445. duration: 23.476s. episode steps: 914. steps per second: 39. episode reward: -21.000. mean reward: -0.023
                                                                                                                                            [-1.000, 0.000], mean action:
               episode: 446, duration: 22.456s, episode steps: 870, steps per second: 39, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean action:
398804/450000; episode: 447, duration: 22.650s, episode steps: 882, steps per second: 39, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
               episode: 448, duration: 23.130s, episode steps: 898, steps per second: 39, episode reward: -20.000, mean reward: -0.022
400647/450000: episode: 449, duration: 24.207s, episode steps: 945, steps per second: 39, episode reward: -20.000, mean reward: -0.021
                                                                                                                                           [-1.000, 1.000], mean action:
               episode: 450, duration: 22.911s, episode steps: 889, steps per second: 39, episode reward: -20.000, mean reward: -0.022
402512/450000: episode: 451, duration: 25.203s, episode steps: 976, steps per second: 39, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
               episode: 452, duration:
                                       24.442s, episode steps:
                                                                949, steps per second: 39, episode reward: -20.000, mean reward: -0.021
                                                                                                                                                     1.000], mean action:
404574/450000; episode: 453, duration: 28.527s, episode steps: 1113, steps per second: 39, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action:
               episode: 454, duration: 21.086s, episode steps: 824, steps per second: 39, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
406190/450000: episode: 455, duration: 20.361s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 456, duration: 21.592s, episode steps:
                                                                 840, steps per second: 39, episode reward: -21.000, mean reward: -0.025
                                                                                                                                                    0.000], mean action:
408111/450000: episode: 457, duration: 27.735s, episode steps: 1081, steps per second: 39, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action:
                                                                 871, steps per second: 39, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
               episode: 458, duration: 22.384s, episode steps:
410015/450000: episode: 459. duration: 26.484s, episode steps: 1033. steps per second: 39. episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action:
                                                                 783, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 460, duration: 20.117s, episode steps:
411640/450000; episode: 461, duration: 21.613s, episode steps: 842, steps per second: 39, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
               episode: 462, duration: 21.667s, episode steps: 842, steps per second: 39, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
413419/450000: episode: 463, duration: 24.012s, episode steps: 937, steps per second: 39, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
               episode: 464, duration: 21.826s, episode steps:
414271/450000:
                                                                852, steps per second: 39, episode reward: -21.000, mean reward: -0.025
                                                                                                                                                     0.000], mean action:
415035/450000: episode: 465, duration: 19.634s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                           [-1.000, 0.000], mean action:
               episode: 466, duration: 21.927s, episode steps: 854, steps per second: 39, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000,
416709/450000; episode: 467, duration: 21.042s, episode steps: 820, steps per second: 39, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
417667/450000:
               episode: 468, duration: 24.528s, episode steps: 958, steps per second: 39, episode reward: -20.000, mean reward: -0.021
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
418509/450000: episode: 469, duration: 21.666s, episode steps: 842, steps per second: 39, episode reward: -20.000, mean reward: -0.024
                                                                                                                                           [-1.000, 1.000], mean action:
               episode: 470, duration: 22.114s, episode steps: 865, steps per second: 39, episode reward: -20.000, mean reward: -0.023
                                                                                                                                           [-1.000,
                                                                                                                                                    1.000], mean action:
420156/450000: episode: 471, duration: 20.132s, episode steps: 782, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 472, duration: 19.694s, episode steps:
                                                                764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
421945/450000; episode: 473, duration: 26.405s, episode steps: 1025, steps per second: 39, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
               episode: 474, duration: 20.129s, episode steps: 783, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
422728/450000:
423492/450000: episode: 475, duration: 19.643s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
               episode: 476, duration: 26.651s, episode steps: 1037, steps per second: 39, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
424529/450000:
425591/450000: episode: 477, duration: 27.196s, episode steps: 1062, steps per second: 39, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action:
               episode: 478, duration: 21.233s, episode steps: 824, steps per second: 39, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
427300/450000: episode: 479, duration: 22.829s, episode steps: 885, steps per second: 39, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
428202/450000:
               episode: 480, duration: 23.213s, episode steps: 902, steps per second: 39, episode reward: -20.000, mean reward: -0.022
429207/450000; episode: 481, duration: 25,698s, episode steps: 1005, steps per second: 39, episode reward: -21,000, mean reward: -0.021 [-1,000, 0.000], mean action:
               episode: 482, duration: 26.262s, episode steps: 1019, steps per second: 39, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
431294/450000: episode: 483, duration: 27.392s, episode steps: 1068, steps per second: 39, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action:
               episode: 484, duration: 26.172s, episode steps: 1019, steps per second: 39, episode reward: -20.000, mean reward: -0.020 [-1.000,
432313/450000.
                                                                                                                                                      1.000], mean action
433181/450000: episode: 485, duration: 22.291s, episode steps: 868, steps per second: 39, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
               episode: 486, duration: 22.347s, episode steps: 872, steps per second: 39, episode reward: -20.000, mean reward: -0.023
434905/450000; episode: 487, duration: 21.853s, episode steps: 852, steps per second: 39, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
435776/450000:
               episode: 488, duration: 22.400s, episode steps: 871, steps per second: 39, episode reward: -20.000, mean reward: -0.023
                                                                                                                                            [-1.000,
                                                                                                                                                    1.000], mean action:
436618/450000; episode: 489, duration: 21.694s, episode steps: 842, steps per second: 39, episode reward: -20.000, mean reward: -0.024
                                                                                                                                           [-1.000, 1.000], mean action:
               episode: 490, duration: 21.179s, episode steps: 824, steps per second: 39, episode reward: -21.000, mean reward: -0.025
438224/450000: episode: 491, duration: 20.076s, episode steps: 782, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
                                                                                                                                            -1.000,
                                                                                                                                                     1.000],
439066/450000
               episode: 492, duration: 21.581s, episode steps:
                                                                      steps per second: 39, episode reward: -20.000,
                                                                                                                      mean reward: -0.024
439858/450000: episode: 493, duration: 20.331s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000,
                                                                                                                                                    0.000], mean action:
                                                                                                                                            -1.000,
               episode: 494, duration: 23.446s, episode steps: 917,
                                                                     steps per second:
                                                                                        39, episode reward: -20.000, mean reward: -0.022
441600/450000; episode: 495, duration: 21.243s, episode steps: 825, steps per second: 39, episode reward: -21.000, mean reward: -0.025
                                                                                                                                            [-1.000, 0.000], mean action:
442364/450000:
               episode: 496, duration: 19.684s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027
                                                                                                                                            [-1.000, 0.000], mean action:
443295/450000: episode: 497, duration: 23.898s, episode steps: 931, steps per second: 39, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action: 444255/450000: episode: 498, duration: 24.757s, episode steps: 960, steps per second: 39, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
```

```
445181/450000: episode: 499, duration: 23.794s, episode steps: 926, steps per second: 39, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 445973/450000: episode: 500, duration: 20.343s, episode steps: 792, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 446893/450000: episode: 501, duration: 23.605s, episode steps: 920, steps per second: 39, episode reward: -19.000, mean reward: -0.021 [-1.000, 1.000], mean action:
447809/450000: episode: 501, duration: 23.505, episode steps: 920, steps per second: 39, episode reward: -19.000, mean reward: -0.021 [-1.000, 1.000], mean action: 447809/450000: episode: 502, duration: 23.519s, episode steps: 916, steps per second: 39, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 448573/450000: episode: 503, duration: 19.611s, episode steps: 764, steps per second: 39, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 449440/450000: episode: 504, duration: 22.303s, episode steps: 867, steps per second: 39, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: done, took 64662.943 seconds
                                      Reward Entrenamiento
      -17
                                                                                                                  1400
     -18
  -19
     -20
                                                        epoch
                                                                                                                                                                    epochs
Testing for 10 episodes ...
Episode 1: reward: -21.000, steps: 764
Episode 2: reward: -21.000, steps: 764
Episode 3: reward: -21.000, steps: 792
Episode 4: reward: -21.000, steps: 764
Episode 5: reward: -21.000, steps: 764
Episode 6: reward: -21.000, steps: 764
Episode 7: reward: -21.000, steps: 764
Episode 8: reward: -21.000, steps: 764
Episode 9: reward: -21.000, steps: 764
Episode 10: reward: -21.000, steps: 824 mean reward: -21.0 mean_steps: 772.8
mean_reward: -21.0
mean_steps 772.8

Pesos quardados en: 202001101028 gs_Experimento_Arquitectura_9--21.0_NOT_BEST_.hdf5W
mean_reward: -21.0 mean_steps: 772.8
```

La recompensa no mejoró, fin del entrenamiento  $\dots$  Tiempo de ejecución del experimento: 64714.0

1/10/2020, 10:50 AM 6 of 6