```
Training for 500000 steps ...
    764/500000: episode: 1, duration: 3.239s, episode steps: 764, steps per second: 236, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1.
   1677/500000: episode: 2, duration: 2.926s, episode steps: 913, steps per second: 312, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1.
   2591/500000: episode: 3, duration: 2.989s, episode steps: 914, steps per second: 306, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1.
   3415/500000: episode: 4, duration: 2.661s, episode steps: 824, steps per second: 310, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1.
   4260/500000: episode: 5, duration: 2.712s, episode steps: 845, steps per second: 312, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1.
   5293/500000: episode: 6, duration: 3.372s, episode steps: 1033, steps per second: 306, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action: 1
   6117/500000: episode: 7, duration: 2.625s, episode steps: 824, steps per second: 314, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1.
   6941/500000: episode: 8, duration: 2.611s, episode steps: 824, steps per second: 316, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1.
   8203/500000: episode: 9, duration: 4.118s, episode steps: 1262, steps per second: 306, episode reward: -17.000, mean reward: -0.013 [-1.000, 1.000], mean action: 1
   9134/500000: episode: 10, duration: 3.084s, episode steps: 931, steps per second: 302, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action: 1
   9977/500000: episode: 11, duration: 2.628s, episode steps: 843, steps per second: 321, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1.
  11106/500000: episode: 12, duration: 12.182s, episode steps: 1129, steps per second: 93, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action:
  12193/500000: episode: 13, duration: 9.944s, episode steps: 1087, steps per second: 109, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
  13045/500000: episode: 14, duration: 7.971s, episode steps: 852, steps per second: 107, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  14017/500000: episode: 15, duration: 9.843s, episode steps: 972, steps per second: 99, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action: 1.
  15117/500000: episode: 16, duration: 10.677s, episode steps: 1100, steps per second: 103, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
  15881/500000: episode: 17, duration: 7.053s, episode steps: 764, steps per second: 108, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1
  16923/500000: episode: 18, duration: 9.618s, episode steps: 1042, steps per second: 108, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action:
  17929/500000: episode: 19, duration: 9.282s, episode steps: 1006, steps per second: 108, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
  18860/500000: episode: 20, duration: 8.694s, episode steps: 931, steps per second: 107, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
  19624/500000: episode: 21, duration: 7.373s, episode steps: 764, steps per second: 104, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
  20416/500000: episode: 22, duration: 7.445s, episode steps: 792, steps per second: 106, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
  21268/500000: episode: 23, duration: 7.985s, episode steps: 852, steps per second: 107, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  22170/500000: episode: 24, duration: 8.577s, episode steps: 902, steps per second: 105, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
  23073/500000: episode: 25, duration: 8.850s, episode steps: 903, steps per second: 102, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
  23865/500000: episode: 26, duration: 7.885s, episode steps: 792, steps per second: 100, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
  24689/500000: episode: 27, duration: 8.110s, episode steps: 824, steps per second: 102, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  25633/500000: episode: 28, duration: 9.329s, episode steps: 944, steps per second: 101, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
  26486/500000: episode: 29, duration: 8.255s, episode steps: 853, steps per second: 103, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  27580/500000: episode: 30, duration: 10.189s, episode steps: 1094, steps per second: 107, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean action:
  28372/500000: episode: 31, duration: 7.301s, episode steps: 792, steps per second: 108, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
  29256/500000: episode: 32, duration: 9.364s, episode steps: 884, steps per second: 94, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1.
  30203/500000: episode: 33, duration: 10.531s, episode steps: 947, steps per second: 90, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action: 1.
  31087/500000: episode: 34, duration: 9.821s, episode steps: 884, steps per second: 90, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1.
  31985/500000: episode: 35, duration: 10.089s, episode steps: 898, steps per second: 89, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 1.
  32895/500000: episode: 36, duration: 9.949s, episode steps: 910, steps per second: 91, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1.
  33921/500000: episode: 37, duration: 11.252s, episode steps: 1026, steps per second: 91, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
  34805/500000: episode: 38, duration: 9.585s, episode steps: 884, steps per second: 92, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1.
  35719/500000: episode: 39, duration: 9.288s, episode steps: 914, steps per second: 98, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1.
  36877/500000: episode: 40, duration: 11.156s, episode steps: 1158, steps per second: 104, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean action:
  37779/500000: episode: 41, duration: 8.615s, episode steps: 902, steps per second: 105, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 1
  38832/500000: episode: 42, duration: 10.076s, episode steps: 1053, steps per second: 105, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
  39734/500000: episode: 43, duration: 8.713s, episode steps: 902, steps per second: 104, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 1
  40576/500000: episode: 44, duration: 8.284s, episode steps: 842, steps per second: 102, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action: 1
  41480/500000: episode: 45, duration: 8.649s, episode steps: 904, steps per second: 105, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1
  42426/500000: episode: 46, duration: 9.024s, episode steps: 946, steps per second: 105, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action: 1
  43360/500000: episode: 47, duration: 8.938s, episode steps: 934, steps per second: 104, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 1
  44124/500000: episode: 48, duration: 7.249s, episode steps: 764, steps per second: 105, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1
  45024/500000: episode: 49, duration: 8.503s, episode steps: 900, steps per second: 106, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 1
  45848/500000: episode: 50, duration: 7.705s, episode steps: 824, steps per second: 107, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
  46692/500000: episode: 51, duration: 7.937s, episode steps: 844, steps per second: 106, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
  47605/500000: episode: 52, duration: 8.525s, episode steps: 913, steps per second: 107, episode reward: -20.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1
  48580/500000: episode: 53, duration: 9.208s, episode steps: 975, steps per second: 106, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action: 1
  49344/500000: episode: 54, duration: 7.518s, episode steps: 764, steps per second: 102, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1
  50288/500000: episode: 55, duration: 9.506s, episode steps: 944, steps per second: 99, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action: 1.
  51493/500000: episode: 56, duration: 11.670s, episode steps: 1205, steps per second: 103, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean action:
  52425/500000: episode: 57, duration: 8.690s, episode steps: 932, steps per second: 107, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action: 1
  53433/500000: episode: 58, duration: 9.379s, episode steps: 1008, steps per second: 107, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
  54257/500000: episode: 59, duration: 7.944s, episode steps: 824, steps per second: 104, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
  55187/500000: episode: 60, duration: 8.972s, episode steps: 930, steps per second: 104, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 1
  56127/500000: episode: 61, duration: 8.898s, episode steps: 940, steps per second: 106, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action: 1
  57164/500000: episode: 62, duration: 9.827s, episode steps: 1037, steps per second: 106, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
  58048/500000: episode: 63, duration: 8.302s, episode steps: 884, steps per second: 106, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1
  59022/500000: episode: 64, duration: 9.367s, episode steps: 974, steps per second: 104, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action: 1
  59883/500000: episode: 65, duration: 8.395s, episode steps: 861, steps per second: 103, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 1
  60647/500000: episode: 66, duration: 7.536s, episode steps: 764, steps per second: 101, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1
  61546/500000: episode: 67, duration: 9.085s, episode steps: 899, steps per second: 99, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 1.
  62436/500000: episode: 68, duration: 8.939s, episode steps: 890, steps per second: 100, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action: 1
  63367/500000: episode: 69, duration: 9.203s, episode steps: 931, steps per second: 101, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action: 1
  64315/500000: episode: 70, duration: 9.066s, episode steps: 948, steps per second: 105, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action: 1
  65259/500000: episode: 71, duration: 9.144s, episode steps: 944, steps per second: 103, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action: 1
  66083/500000: episode: 72, duration: 7.837s, episode steps: 824, steps per second: 105, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
  66935/500000: episode: 73, duration: 9.274s, episode steps: 852, steps per second: 92, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1.
  67755/500000: episode: 74, duration: 9.227s, episode steps: 820, steps per second: 89, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action: 1.
  68920/500000: episode: 75, duration: 12.981s, episode steps: 1165, steps per second: 90, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean action: 1.
  69783/500000: episode: 76, duration: 9.659s, episode steps: 863, steps per second: 89, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action: 1.
  70632/500000: episode: 77, duration: 9.290s, episode steps: 849, steps per second: 91, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1.
  71516/500000: episode: 78, duration: 9.894s, episode steps: 884, steps per second: 89, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1.
  72550/500000: episode: 79, duration: 10.979s, episode steps: 1034, steps per second: 94, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action:
  73314/500000: episode: 80, duration: 7.895s, episode steps: 764, steps per second: 97, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1.
  74200/500000: episode: 81, duration: 8.959s, episode steps: 886, steps per second: 99, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1.
  75204/500000: episode: 82, duration: 10.218s, episode steps: 1004, steps per second: 98, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
  76315/500000: episode: 83, duration: 10.657s, episode steps: 1111, steps per second: 104, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean action:
  77139/500000: episode: 84, duration: 7.849s, episode steps: 824, steps per second: 105, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
  77993/500000: episode: 85, duration: 8.043s, episode steps: 854, steps per second: 106, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
  78905/500000: episode: 86, duration: 8.592s, episode steps: 912, steps per second: 106, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1
  79817/500000: episode: 87, duration: 8.530s, episode steps: 912, steps per second: 107, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1
  80795/500000: episode: 88, duration: 9.168s, episode steps: 978, steps per second: 107, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action: 1
  81679/500000: episode: 89, duration: 8.247s, episode steps: 884, steps per second: 107, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1
  82689/500000: episode: 90, duration: 9.310s, episode steps: 1010, steps per second: 108, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
  83733/500000: episode: 91, duration: 9.636s, episode steps: 1044, steps per second: 108, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
  84806/500000: episode: 92, duration: 9.824s, episode steps: 1073, steps per second: 109, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
  85693/500000: episode: 93, duration: 8.422s, episode steps: 887, steps per second: 105, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action: 1
  86457/500000: episode: 94, duration: 7.053s, episode steps: 764, steps per second: 108, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1
  87605/500000: episode: 95, duration: 10.472s, episode steps: 1148, steps per second: 110, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
  88397/500000: episode: 96, duration: 7.683s, episode steps: 792, steps per second: 103, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1
  89161/500000: episode: 97, duration: 7.577s, episode steps: 764, steps per second: 101, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action: 1
  90073/500000: episode: 98, duration: 8.802s, episode steps: 912, steps per second: 104, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action: 1
  90917/500000: episode: 99, duration: 7.826s, episode steps: 844, steps per second: 108, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
  91942/500000: episode: 100, duration: 9.343s, episode steps: 1025, steps per second: 110, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action:
  92888/500000: episode: 101, duration: 8.704s, episode steps: 946, steps per second: 109, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
  94178/500000: episode: 102, duration: 12.059s, episode steps: 1290, steps per second: 107, episode reward: -18.000, mean reward: -0.014 [-1.000, 1.000], mean action
  95062/500000: episode: 103, duration: 8.284s, episode steps: 884, steps per second: 107, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
  96096/500000: episode: 104, duration: 9.638s, episode steps: 1034, steps per second: 107, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action:
  96920/500000: episode: 105, duration: 7.732s, episode steps: 824, steps per second: 107, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  97822/500000: episode: 106, duration: 8.572s, episode steps: 902, steps per second: 105, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
  98706/500000: episode: 107, duration: 8.547s, episode steps: 884, steps per second: 103, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
  99653/500000: episode: 108, duration: 9.372s, episode steps: 947, steps per second: 101, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
```

```
100659/500000: episode: 109, duration: 10.072s, episode steps: 1006, steps per second: 100, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action
101543/500000: episode: 110, duration: 8.816s, episode steps: 884, steps per second: 100, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
102643/500000: episode: 111, duration: 10.567s, episode steps: 1100, steps per second: 104, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action
103527/500000: episode: 112, duration: 8.393s, episode steps: 884, steps per second: 105, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
104351/500000: episode: 113, duration: 8.862s, episode steps: 824, steps per second: 93, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
105175/500000: episode: 114, duration: 9.136s, episode steps: 824, steps per second: 90, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
106119/500000: episode: 115, duration: 10.396s, episode steps: 944, steps per second: 91, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
106943/500000: episode: 116, duration: 9.048s, episode steps: 824, steps per second: 91, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action: 1
107923/500000: episode: 117, duration: 10.566s, episode steps: 980, steps per second: 93, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
108930/500000: episode: 118, duration: 10.980s, episode steps: 1007, steps per second: 92, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
109970/500000: episode: 119, duration: 11.220s, episode steps: 1040, steps per second: 93, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
110974/500000: episode: 120, duration: 10.245s, episode steps: 1004, steps per second: 98, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
112298/500000: episode: 121, duration: 12.509s, episode steps: 1324, steps per second: 106, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean action
113393/500000: episode: 122, duration: 10.319s, episode steps: 1095, steps per second: 106, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action
114462/500000: episode: 123, duration: 10.276s, episode steps: 1069, steps per second: 104, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action
115436/500000: episode: 124, duration: 9.266s, episode steps: 974, steps per second: 105, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
116322/500000: episode: 125, duration: 8.480s, episode steps: 886, steps per second: 104, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
117146/500000: episode: 126, duration: 8.164s, episode steps: 824, steps per second: 101, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
118367/500000: episode: 127, duration: 11.245s, episode steps: 1221, steps per second: 109, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean action
119432/500000: episode: 128, duration: 9.759s, episode steps: 1065, steps per second: 109, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action:
120256/500000: episode: 129, duration: 7.533s, episode steps: 824, steps per second: 109, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
121080/500000: episode: 130, duration: 7.503s, episode steps: 824, steps per second: 110, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
121904/500000: episode: 131, duration: 7.426s, episode steps: 824, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
122946/500000: episode: 132, duration: 9.387s, episode steps: 1042, steps per second: 111, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
124251/500000: episode: 133, duration: 11.708s, episode steps: 1305, steps per second: 111, episode reward: -21.000, mean reward: -0.016 [-1.000, 0.000], mean action
125076/500000: episode: 134, duration: 7.453s, episode steps: 825, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
126080/500000: episode: 135, duration: 9.079s, episode steps: 1004, steps per second: 111, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
126904/500000: episode: 136, duration: 7.536s, episode steps: 824, steps per second: 109, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
128011/500000: episode: 137, duration: 9.979s, episode steps: 1107, steps per second: 111, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action:
129017/500000: episode: 138, duration: 9.025s, episode steps: 1006, steps per second: 111, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
130021/500000: episode: 139, duration: 9.013s, episode steps: 1004, steps per second: 111, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
130845/500000: episode: 140, duration: 7.343s, episode steps: 824, steps per second: 112, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
131669/500000: episode: 141, duration: 7.361s, episode steps: 824, steps per second: 112, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
132461/500000: episode: 142, duration: 7.132s, episode steps: 792, steps per second: 111, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
133285/500000: episode: 143, duration: 7.460s, episode steps: 824, steps per second: 110, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
134109/500000: episode: 144, duration: 7.472s, episode steps: 824, steps per second: 110, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
134947/500000: episode: 145, duration: 8.004s, episode steps: 838, steps per second: 105, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
135859/500000: episode: 146, duration: 8.471s, episode steps: 912, steps per second: 108, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
136772/500000: episode: 147, duration: 8.414s, episode steps: 913, steps per second: 109, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
137796/500000: episode: 148, duration: 9.487s, episode steps: 1024, steps per second: 108, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
138860/500000: episode: 149, duration: 9.604s, episode steps: 1064, steps per second: 111, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action:
139772/500000: episode: 150, duration: 8.236s, episode steps: 912, steps per second: 111, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
140855/500000: episode: 151, duration: 9.716s, episode steps: 1083, steps per second: 111, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action:
142128/500000: episode: 152, duration: 11.464s, episode steps: 1273, steps per second: 111, episode reward: -21.000, mean reward: -0.016 [-1.000, 0.000], mean action
143137/500000: episode: 153, duration: 9.127s, episode steps: 1009, steps per second: 110, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
144110/500000: episode: 154, duration: 8.677s, episode steps: 973, steps per second: 112, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
144934/500000: episode: 155, duration: 7.398s, episode steps: 824, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
146103/500000: episode: 156, duration: 10.500s, episode steps: 1169, steps per second: 111, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean action
147045/500000: episode: 157, duration: 8.566s, episode steps: 942, steps per second: 110, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
147989/500000: episode: 158, duration: 8.629s, episode steps: 944, steps per second: 109, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
148933/500000: episode: 159, duration: 8.742s, episode steps: 944, steps per second: 108, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
149757/500000: episode: 160, duration: 7.603s, episode steps: 824, steps per second: 108, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
150581/500000: episode: 161, duration: 7.744s, episode steps: 824, steps per second: 106, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
151648/500000: episode: 162, duration: 10.227s, episode steps: 1067, steps per second: 104, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action
152921/500000: episode: 163, duration: 11.582s, episode steps: 1273, steps per second: 110, episode reward: -21.000, mean reward: -0.016 [-1.000, 0.000], mean action
153782/500000: episode: 164, duration: 7.707s, episode steps: 861, steps per second: 112, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
154606/500000: episode: 165, duration: 7.446s, episode steps: 824, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
156028/500000: episode: 166, duration: 12.810s, episode steps: 1422, steps per second: 111, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean action
156967/500000: episode: 167, duration: 8.502s, episode steps: 939, steps per second: 110, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
158357/500000: episode: 168, duration: 12.445s, episode steps: 1390, steps per second: 112, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean action
159317/500000: episode: 169, duration: 8.653s, episode steps: 960, steps per second: 111, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
160424/500000: episode: 170, duration: 9.964s, episode steps: 1107, steps per second: 111, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action:
161246/500000: episode: 171, duration: 7.382s, episode steps: 822, steps per second: 111, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
162211/500000: episode: 172, duration: 8.703s, episode steps: 965, steps per second: 111, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
163175/500000: episode: 173, duration: 8.727s, episode steps: 964, steps per second: 110, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
164645/500000: episode: 174, duration: 13.456s, episode steps: 1470, steps per second: 109, episode reward: -18.000, mean reward: -0.012 [-1.000, 1.000], mean action
165818/500000: episode: 175, duration: 10.861s, episode steps: 1173, steps per second: 108, episode reward: -21.000, mean reward: -0.018 [-1.000, 0.000], mean action
167130/500000: episode: 176, duration: 11.993s, episode steps: 1312, steps per second: 109, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean action
168150/500000: episode: 177, duration: 9.479s, episode steps: 1020, steps per second: 108, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
169744/500000: episode: 178, duration: 14.302s, episode steps: 1594, steps per second: 111, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean action
170685/500000: episode: 179, duration: 8.441s, episode steps: 941, steps per second: 111, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
171589/500000: episode: 180, duration: 8.087s, episode steps: 904, steps per second: 112, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
172604/500000: episode: 181, duration: 8.953s, episode steps: 1015, steps per second: 113, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
173723/500000: episode: 182, duration: 9.912s, episode steps: 1119, steps per second: 113, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action:
174607/500000: episode: 183, duration: 7.780s, episode steps: 884, steps per second: 114, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
175610/500000: episode: 184, duration: 8.770s, episode steps: 1003, steps per second: 114, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
176582/500000: episode: 185, duration: 8.548s, episode steps: 972, steps per second: 114, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
178215/500000: episode: 186, duration: 14.589s, episode steps: 1633, steps per second: 112, episode reward: -20.000, mean reward: -0.012 [-1.000, 1.000], mean action
179339/500000: episode: 187, duration: 10.327s, episode steps: 1124, steps per second: 109, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action
180247/500000: episode: 188, duration: 8.227s, episode steps: 908, steps per second: 110, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
181472/500000: episode: 189, duration: 11.121s, episode steps: 1225, steps per second: 110, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean action
183125/500000: episode: 190, duration: 14.872s, episode steps: 1653, steps per second: 111, episode reward: -20.000, mean reward: -0.012 [-1.000, 1.000], mean action
184609/500000: episode: 191, duration: 12.952s, episode steps: 1484, steps per second: 115, episode reward: -20.000, mean reward: -0.013 [-1.000, 1.000], mean action
186067/500000: episode: 192, duration: 13.136s, episode steps: 1458, steps per second: 111, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean action
187295/500000: episode: 193, duration: 10.721s, episode steps: 1228, steps per second: 115, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean action
188582/500000: episode: 194, duration: 11.324s, episode steps: 1287, steps per second: 114, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean action
190237/500000: episode: 195, duration: 14.572s, episode steps: 1655, steps per second: 114, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean action
191707/500000: episode: 196, duration: 12.954s, episode steps: 1470, steps per second: 113, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean action
193024/500000: episode: 197, duration: 11.973s, episode steps: 1317, steps per second: 110, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean action
194120/500000: episode: 198, duration: 10.008s, episode steps: 1096, steps per second: 110, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action
195624/500000: episode: 199, duration: 13.793s, episode steps: 1504, steps per second: 109, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean action
197256/500000: episode: 200, duration: 14.789s, episode steps: 1632, steps per second: 110, episode reward: -19.000, mean reward: -0.012 [-1.000, 1.000], mean action
198822/500000: episode: 201, duration: 14.101s, episode steps: 1566, steps per second: 111, episode reward: -20.000, mean reward: -0.013 [-1.000, 1.000], mean action
200535/500000: episode: 202, duration: 15.403s, episode steps: 1713, steps per second: 111, episode reward: -20.000, mean reward: -0.012 [-1.000, 1.000], mean action
201956/500000: episode: 203, duration: 12.988s, episode steps: 1421, steps per second: 109, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean action
203879/500000: episode: 204, duration: 17.476s, episode steps: 1923, steps per second: 110, episode reward: -20.000, mean reward: -0.010 [-1.000, 1.000], mean action
205359/500000: episode: 205, duration: 13.542s, episode steps: 1480, steps per second: 109, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean action
206606/500000: episode: 206, duration: 11.663s, episode steps: 1247, steps per second: 107, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean action
207895/500000: episode: 207, duration: 12.068s, episode steps: 1289, steps per second: 107, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean action
209339/500000: episode: 208, duration: 13.522s, episode steps: 1444, steps per second: 107, episode reward: -21.000, mean reward: -0.015 [-1.000, 0.000], mean action
210660/500000: episode: 209, duration: 12.012s, episode steps: 1321, steps per second: 110, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean action
212079/500000: episode: 210, duration: 12.861s, episode steps: 1419, steps per second: 110, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean action
213057/500000: episode: 211, duration: 9.058s, episode steps: 978, steps per second: 108, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
214792/500000: episode: 212, duration: 15.885s, episode steps: 1735, steps per second: 109, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean action
216445/500000: episode: 213, duration: 15.266s, episode steps: 1653, steps per second: 108, episode reward: -20.000, mean reward: -0.012 [-1.000, 1.000], mean action
217888/500000: episode: 214, duration: 13.349s, episode steps: 1443, steps per second: 108, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean action
219450/500000: episode: 215, duration: 15.034s, episode steps: 1562, steps per second: 104, episode reward: -21.000, mean reward: -0.013 [-1.000, 0.000], mean action
221047/500000: episode: 216, duration: 15.042s, episode steps: 1597, steps per second: 106, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean action
222376/500000: episode: 217, duration: 12.263s, episode steps: 1329, steps per second: 108, episode reward: -21.000, mean reward: -0.016 [-1.000, 0.000], mean action
```

```
224260/500000: episode: 218, duration: 17.122s, episode steps: 1884, steps per second: 110, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean actior
225992/500000: episode: 219, duration: 15.725s, episode steps: 1732, steps per second: 110, episode reward: -19.000, mean reward: -0.011 [-1.000, 1.000], mean actior
227542/500000: episode: 220, duration: 13.987s, episode steps: 1550, steps per second: 111, episode reward: -19.000, mean reward: -0.012 [-1.000, 1.000], mean actior
228893/500000: episode: 221, duration: 12.293s, episode steps: 1351, steps per second: 110, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean actior
230213/500000: episode: 222, duration: 12.032s, episode steps: 1320, steps per second: 110, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean actior
231176/500000: episode: 223, duration: 8.841s, episode steps: 963, steps per second: 109, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
232326/500000: episode: 224, duration: 10.578s, episode steps: 1150, steps per second: 109, episode reward: -21.000, mean reward: -0.018 [-1.000, 0.000], mean actior
234293/500000: episode: 225, duration: 18.374s, episode steps: 1967, steps per second: 107, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean actior
235948/500000: episode: 226, duration: 15.663s, episode steps: 1655, steps per second: 106, episode reward: -20.000, mean reward: -0.012 [-1.000, 1.000], mean actior
237396/500000: episode: 227, duration: 13.136s, episode steps: 1448, steps per second: 110, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean actior
239055/500000: episode: 228, duration: 15.012s, episode steps: 1659, steps per second: 111, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean actior
240959/500000: episode: 229, duration: 17.284s, episode steps: 1904, steps per second: 110, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
242787/500000: episode: 230, duration: 16.706s, episode steps: 1828, steps per second: 109, episode reward: -19.000, mean reward: -0.010 [-1.000, 1.000], mean actior
244445/500000: episode: 231, duration: 15.283s, episode steps: 1658, steps per second: 108, episode reward: -19.000, mean reward: -0.011 [-1.000, 1.000], mean actior
246301/500000: episode: 232, duration: 17.312s, episode steps: 1856, steps per second: 107, episode reward: -21.000, mean reward: -0.011 [-1.000, 0.000], mean actior
247739/500000: episode: 233, duration: 13.321s, episode steps: 1438, steps per second: 108, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean actior
249596/500000: episode: 234, duration: 16.988s, episode steps: 1857, steps per second: 109, episode reward: -19.000, mean reward: -0.010 [-1.000, 1.000], mean actior
250904/500000: episode: 235, duration: 11.727s, episode steps: 1308, steps per second: 112, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean actior
252591/500000: episode: 236, duration: 15.119s, episode steps: 1687, steps per second: 112, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean actior
253997/500000: episode: 237, duration: 12.998s, episode steps: 1406, steps per second: 108, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean actior
255591/500000: episode: 238, duration: 14.235s, episode steps: 1594, steps per second: 112, episode reward: -19.000, mean reward: -0.012 [-1.000, 1.000], mean actior
257383/500000: episode: 239, duration: 16.141s, episode steps: 1792, steps per second: 111, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean actior
259274/500000: episode: 240, duration: 17.333s, episode steps: 1891, steps per second: 109, episode reward: -19.000, mean reward: -0.010 [-1.000, 1.000], mean actior
260990/500000: episode: 241, duration: 15.842s, episode steps: 1716, steps per second: 108, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean actior
262426/500000: episode: 242, duration: 13.183s, episode steps: 1436, steps per second: 109, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean actior
263832/500000: episode: 243, duration: 12.710s, episode steps: 1406, steps per second: 111, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean actior
265833/500000: episode: 244, duration: 17.919s, episode steps: 2001, steps per second: 112, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
267752/500000: episode: 245, duration: 17.163s, episode steps: 1919, steps per second: 112, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
269016/500000: episode: 246, duration: 11.600s, episode steps: 1264, steps per second: 109, episode reward: -18.000, mean reward: -0.014 [-1.000, 1.000], mean actior
270442/500000: episode: 247, duration: 12.746s, episode steps: 1426, steps per second: 112, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean actior
271640/500000: episode: 248, duration: 10.842s, episode steps: 1198, steps per second: 110, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean actior
273045/500000: episode: 249, duration: 12.885s, episode steps: 1405, steps per second: 109, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean actior
274662/500000: episode: 250, duration: 14.875s, episode steps: 1617, steps per second: 109, episode reward: -20.000, mean reward: -0.012 [-1.000, 1.000], mean actior
276366/500000: episode: 251, duration: 15.639s, episode steps: 1704, steps per second: 109, episode reward: -19.000, mean reward: -0.011 [-1.000, 1.000], mean actior
278264/500000: episode: 252, duration: 17.084s, episode steps: 1898, steps per second: 111, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean actior
279913/500000: episode: 253, duration: 15.012s, episode steps: 1649, steps per second: 110, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean actior
281901/500000: episode: 254, duration: 17.847s, episode steps: 1988, steps per second: 111, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
283874/500000: episode: 255, duration: 17.717s, episode steps: 1973, steps per second: 111, episode reward: -19.000, mean reward: -0.010 [-1.000, 1.000], mean actior
285864/500000: episode: 256, duration: 17.994s, episode steps: 1990, steps per second: 111, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
287935/500000: episode: 257, duration: 19.426s, episode steps: 2071, steps per second: 107, episode reward: -17.000, mean reward: -0.008 [-1.000, 1.000], mean actior
290463/500000: episode: 258, duration: 23.265s, episode steps: 2528, steps per second: 109, episode reward: -18.000, mean reward: -0.007 [-1.000, 1.000], mean actior
293047/500000: episode: 259, duration: 23.335s, episode steps: 2584, steps per second: 111, episode reward: -17.000, mean reward: -0.007 [-1.000, 1.000], mean actior
295071/500000: episode: 260, duration: 18.288s, episode steps: 2024, steps per second: 111, episode reward: -19.000, mean reward: -0.009 [-1.000, 1.000], mean actior
297208/500000: episode: 261, duration: 19.330s, episode steps: 2137, steps per second: 111, episode reward: -21.000, mean reward: -0.010 [-1.000, 0.000], mean actior
299524/500000: episode: 262, duration: 21.200s, episode steps: 2316, steps per second: 109, episode reward: -20.000, mean reward: -0.009 [-1.000, 1.000], mean actior
301391/500000: episode: 263, duration: 17.216s, episode steps: 1867, steps per second: 108, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean actior
303697/500000: episode: 264, duration: 21.415s, episode steps: 2306, steps per second: 108, episode reward: -17.000, mean reward: -0.007 [-1.000, 1.000], mean actior
305651/500000: episode: 265, duration: 17.531s, episode steps: 1954, steps per second: 111, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean actior
307626/500000: episode: 266, duration: 17.988s, episode steps: 1975, steps per second: 110, episode reward: -19.000, mean reward: -0.010 [-1.000, 1.000], mean actior
309556/500000: episode: 267, duration: 17.775s, episode steps: 1930, steps per second: 109, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
311991/500000: episode: 268, duration: 22.388s, episode steps: 2435, steps per second: 109, episode reward: -20.000, mean reward: -0.008 [-1.000, 1.000], mean actior
314113/500000: episode: 269, duration: 19.452s, episode steps: 2122, steps per second: 109, episode reward: -17.000, mean reward: -0.008 [-1.000, 1.000], mean actior
315969/500000: episode: 270, duration: 16.864s, episode steps: 1856, steps per second: 110, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean actior
318438/500000: episode: 271, duration: 22.762s, episode steps: 2469, steps per second: 108, episode reward: -18.000, mean reward: -0.007 [-1.000, 1.000], mean actior
320614/500000: episode: 272, duration: 20.273s, episode steps: 2176, steps per second: 107, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean actior
322585/500000: episode: 273, duration: 18.537s, episode steps: 1971, steps per second: 106, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
324907/500000: episode: 274, duration: 21.012s, episode steps: 2322, steps per second: 111, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean actior
327232/500000: episode: 275, duration: 21.113s, episode steps: 2325, steps per second: 110, episode reward: -19.000, mean reward: -0.008 [-1.000, 1.000], mean actior
329442/500000: episode: 276, duration: 20.371s, episode steps: 2210, steps per second: 108, episode reward: -18.000, mean reward: -0.008 [-1.000, 1.000], mean actior
330845/500000: episode: 277, duration: 12.874s, episode steps: 1403, steps per second: 109, episode reward: -21.000, mean reward: -0.015 [-1.000, 0.000], mean actior
332643/500000: episode: 278, duration: 16.509s, episode steps: 1798, steps per second: 109, episode reward: -21.000, mean reward: -0.012 [-1.000, 0.000], mean actior
335028/500000: episode: 279, duration: 21.722s, episode steps: 2385, steps per second: 110, episode reward: -18.000, mean reward: -0.008 [-1.000, 1.000], mean actior
337571/500000: episode: 280, duration: 23.440s, episode steps: 2543, steps per second: 108, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean actior
338875/500000: episode: 281, duration: 12.175s, episode steps: 1304, steps per second: 107, episode reward: -21.000, mean reward: -0.016 [-1.000, 0.000], mean actior
340899/500000: episode: 282, duration: 18.607s, episode steps: 2024, steps per second: 109, episode reward: -20.000, mean reward: -0.010 [-1.000, 1.000], mean actior
342420/500000: episode: 283, duration: 14.028s, episode steps: 1521, steps per second: 108, episode reward: -21.000, mean reward: -0.014 [-1.000, 0.000], mean actior
344920/500000: episode: 284, duration: 22.531s, episode steps: 2500, steps per second: 111, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean actior
346830/500000: episode: 285, duration: 17.181s, episode steps: 1910, steps per second: 111, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
349690/500000: episode: 286, duration: 26.290s, episode steps: 2860, steps per second: 109, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean actior
352159/500000: episode: 287, duration: 22.945s, episode steps: 2469, steps per second: 108, episode reward: -16.000, mean reward: -0.006 [-1.000, 1.000], mean actior
354074/500000: episode: 288, duration: 17.622s, episode steps: 1915, steps per second: 109, episode reward: -19.000, mean reward: -0.010 [-1.000, 1.000], mean actior
355979/500000: episode: 289, duration: 17.769s, episode steps: 1905, steps per second: 107, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
358164/500000: episode: 290, duration: 20.186s, episode steps: 2185, steps per second: 108, episode reward: -15.000, mean reward: -0.007 [-1.000, 1.000], mean actior
359887/500000: episode: 291, duration: 16.062s, episode steps: 1723, steps per second: 107, episode reward: -19.000, mean reward: -0.011 [-1.000, 1.000], mean actior
361479/500000: episode: 292, duration: 14.751s, episode steps: 1592, steps per second: 108, episode reward: -19.000, mean reward: -0.012 [-1.000, 1.000], mean actior
362971/500000: episode: 293, duration: 13.647s, episode steps: 1492, steps per second: 109, episode reward: -20.000, mean reward: -0.013 [-1.000, 1.000], mean actior
364979/500000: episode: 294, duration: 18.278s, episode steps: 2008, steps per second: 110, episode reward: -17.000, mean reward: -0.008 [-1.000, 1.000], mean actior
367195/500000: episode: 295, duration: 20.429s, episode steps: 2216, steps per second: 108, episode reward: -18.000, mean reward: -0.008 [-1.000, 1.000], mean actior
369234/500000: episode: 296, duration: 19.040s, episode steps: 2039, steps per second: 107, episode reward: -21.000, mean reward: -0.010 [-1.000, 0.000], mean actior
371404/500000: episode: 297, duration: 20.017s, episode steps: 2170, steps per second: 108, episode reward: -19.000, mean reward: -0.009 [-1.000, 1.000], mean actior
372866/500000: episode: 298, duration: 13.539s, episode steps: 1462, steps per second: 108, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean actior
375549/500000: episode: 299, duration: 24.451s, episode steps: 2683, steps per second: 110, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean actior
377430/500000: episode: 300, duration: 17.367s, episode steps: 1881, steps per second: 108, episode reward: -19.000, mean reward: -0.010 [-1.000, 1.000], mean actior
380496/500000: episode: 301, duration: 28.926s, episode steps: 3066, steps per second: 106, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean actior
382594/500000: episode: 302, duration: 19.826s, episode steps: 2098, steps per second: 106, episode reward: -15.000, mean reward: -0.007 [-1.000, 1.000], mean actior
385740/500000: episode: 303, duration: 29.077s, episode steps: 3146, steps per second: 108, episode reward: -14.000, mean reward: -0.004 [-1.000, 1.000], mean actior
387740/500000: episode: 304, duration: 18.844s, episode steps: 2000, steps per second: 106, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean actior
390022/500000: episode: 305, duration: 21.586s, episode steps: 2282, steps per second: 106, episode reward: -17.000, mean reward: -0.007 [-1.000, 1.000], mean actior
393102/500000: episode: 306, duration: 28.833s, episode steps: 3080, steps per second: 107, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean actior
396186/500000: episode: 307, duration: 28.384s, episode steps: 3084, steps per second: 109, episode reward: -17.000, mean reward: -0.006 [-1.000, 1.000], mean actior
399500/500000: episode: 308, duration: 30.990s, episode steps: 3314, steps per second: 107, episode reward: -17.000, mean reward: -0.005 [-1.000, 1.000], mean actior
401976/500000: episode: 309, duration: 23.471s, episode steps: 2476, steps per second: 105, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean actior
404607/500000: episode: 310, duration: 24.073s, episode steps: 2631, steps per second: 109, episode reward: -18.000, mean reward: -0.007 [-1.000, 1.000], mean actior
407871/500000: episode: 311, duration: 30.532s, episode steps: 3264, steps per second: 107, episode reward: -15.000, mean reward: -0.005 [-1.000, 1.000], mean actior
410321/500000: episode: 312, duration: 23.083s, episode steps: 2450, steps per second: 106, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean actior
412621/500000: episode: 313, duration: 21.438s, episode steps: 2300, steps per second: 107, episode reward: -17.000, mean reward: -0.007 [-1.000, 1.000], mean actior
414978/500000: episode: 314, duration: 21.534s, episode steps: 2357, steps per second: 109, episode reward: -17.000, mean reward: -0.007 [-1.000, 1.000], mean actior
418291/500000: episode: 315, duration: 30.564s, episode steps: 3313, steps per second: 108, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean actior
421752/500000: episode: 316, duration: 33.064s, episode steps: 3461, steps per second: 105, episode reward: -12.000, mean reward: -0.003 [-1.000, 1.000], mean actior
424874/500000: episode: 317, duration: 28.858s, episode steps: 3122, steps per second: 108, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean actior
428606/500000: episode: 318, duration: 34.436s, episode steps: 3732, steps per second: 108, episode reward: -14.000, mean reward: -0.004 [-1.000, 1.000], mean actior
431647/500000: episode: 319, duration: 28.269s, episode steps: 3041, steps per second: 108, episode reward: -16.000, mean reward: -0.005 [-1.000, 1.000], mean actior
434552/500000: episode: 320, duration: 26.395s, episode steps: 2905, steps per second: 110, episode reward: -17.000, mean reward: -0.006 [-1.000, 1.000], mean actior
437906/500000: episode: 321, duration: 30.621s, episode steps: 3354, steps per second: 110, episode reward: -15.000, mean reward: -0.004 [-1.000, 1.000], mean actior
441105/500000: episode: 322, duration: 29.208s, episode steps: 3199, steps per second: 110, episode reward: -13.000, mean reward: -0.004 [-1.000, 1.000], mean actior
444800/500000: episode: 323, duration: 33.124s, episode steps: 3695, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean action:
448603/500000: episode: 324, duration: 34.435s, episode steps: 3803, steps per second: 110, episode reward: -13.000, mean reward: -0.003 [-1.000, 1.000], mean actior
451172/500000: episode: 325, duration: 23.690s, episode steps: 2569, steps per second: 108, episode reward: -18.000, mean reward: -0.007 [-1.000, 1.000], mean actior
454363/500000: episode: 326, duration: 28.817s, episode steps: 3191, steps per second: 111, episode reward: -14.000, mean reward: -0.004 [-1.000, 1.000], mean actior
```
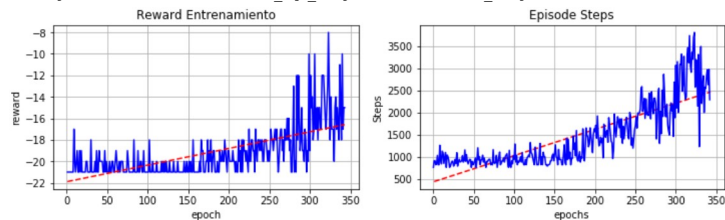
```
 457513/500000: episode: 327, duration: 29.420s, episode steps: 3150, steps per second: 107, episode reward: -17.000, mean reward: -0.005 [-1.000, 1.000], mean action
 459657/500000: episode: 328, duration: 19.947s, episode steps: 2144, steps per second: 107, episode reward: -17.000, mean reward: -0.008 [-1.000, 1.000], mean action
 462965/500000: episode: 329, duration: 30.256s, episode steps: 3308, steps per second: 109, episode reward: -17.000, mean reward: -0.005 [-1.000, 1.000], mean action
 464196/500000: episode: 330, duration: 11.201s, episode steps: 1231, steps per second: 110, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean action
 466884/500000: episode: 331, duration: 24.100s, episode steps: 2688, steps per second: 112, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean action
 470371/500000: episode: 332, duration: 32.285s, episode steps: 3487, steps per second: 108, episode reward: -14.000, mean reward: -0.004 [-1.000, 1.000], mean action
 472575/500000: episode: 333, duration: 20.299s, episode steps: 2204, steps per second: 109, episode reward: -17.000, mean reward: -0.008 [-1.000, 1.000], mean action
 475150/500000: episode: 334, duration: 23.209s, episode steps: 2575, steps per second: 111, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean action
 477978/500000: episode: 335, duration: 25.772s, episode steps: 2828, steps per second: 110, episode reward: -17.000, mean reward: -0.006 [-1.000, 1.000], mean action
 479978/500000: episode: 336, duration: 18.476s, episode steps: 2000, steps per second: 108, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean action
 482549/500000: episode: 337, duration: 23.616s, episode steps: 2571, steps per second: 109, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean action
 485088/500000: episode: 338, duration: 22.896s, episode steps: 2539, steps per second: 111, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean action
 487920/500000: episode: 339, duration: 25.656s, episode steps: 2832, steps per second: 110, episode reward: -18.000, mean reward: -0.006 [-1.000, 1.000], mean action
 490886/500000: episode: 340, duration: 27.607s, episode steps: 2966, steps per second: 107, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean action
 493512/500000: episode: 341, duration: 24.038s, episode steps: 2626, steps per second: 109, episode reward: -17.000, mean reward: -0.006 [-1.000, 1.000], mean action
 496482/500000: episode: 342, duration: 26.628s, episode steps: 2970, steps per second: 112, episode reward: -15.000, mean reward: -0.005 [-1.000, 1.000], mean action
 498770/500000: episode: 343, duration: 20.969s, episode steps: 2288, steps per second: 109, episode reward: -15.000, mean reward: -0.007 [-1.000, 1.000], mean action
done, took 4604.102 seconds
Pesos guardados en: 202001091122_dqn_PongDeterministic-v0_weights.h5f
```



```
Testing for 10 episodes ...
Episode 1: reward: -13.000, steps: 3497
Episode 2: reward: -12.000, steps: 3046
Episode 3: reward: -8.000, steps: 4034
Episode 4: reward: -19.000, steps: 2621
Episode 5: reward: -15.000, steps: 3005
Episode 6: reward: -15.000, steps: 3172
Episode 7: reward: -12.000, steps: 3208
Episode 8: reward: -16.000, steps: 2978
Episode 9: reward: -18.000, steps: 2729
Episode 10: reward: -16.000, steps: 3236
mean reward: -14.4
mean_steps: 3152.6
(-14.4, 3152.6)
```