

PROJECT 3: COLLABORATION AND COMPETITION
UDACITY DEEP REINFORCEMENT LEARNING
CLEMENTE FORTUNA

INTRODUCTION

The solution implements a multi-agent deep deterministic policy gradient (MADDPG) algorithm. Focus areas include managing inputs/outputs for multiple agents, network expansion & regularization, computation graph trimming, priority experience replay, and hyperparameter tuning. The result was solution convergence in 2026 episodes.

SINGLE AGENT DDPG OVERVIEW

The figure below captures the single agent DDPG algorithm:

Algorithm 1 DDPG algorithm

Randomly initialize critic network $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ .
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer R
for episode = 1, M **do**
 Initialize a random process \mathcal{N} for action exploration
 Receive initial observation state s_1
 for t = 1, T **do**
 Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$ according to the current policy and exploration noise
 Execute action a_t and observe reward r_t and observe new state s_{t+1}
 Store transition (s_t, a_t, r_t, s_{t+1}) in R
 Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R
 Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
 Update critic by minimizing the loss: $L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
 Update the actor policy using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s_i}$$

Update the target networks:

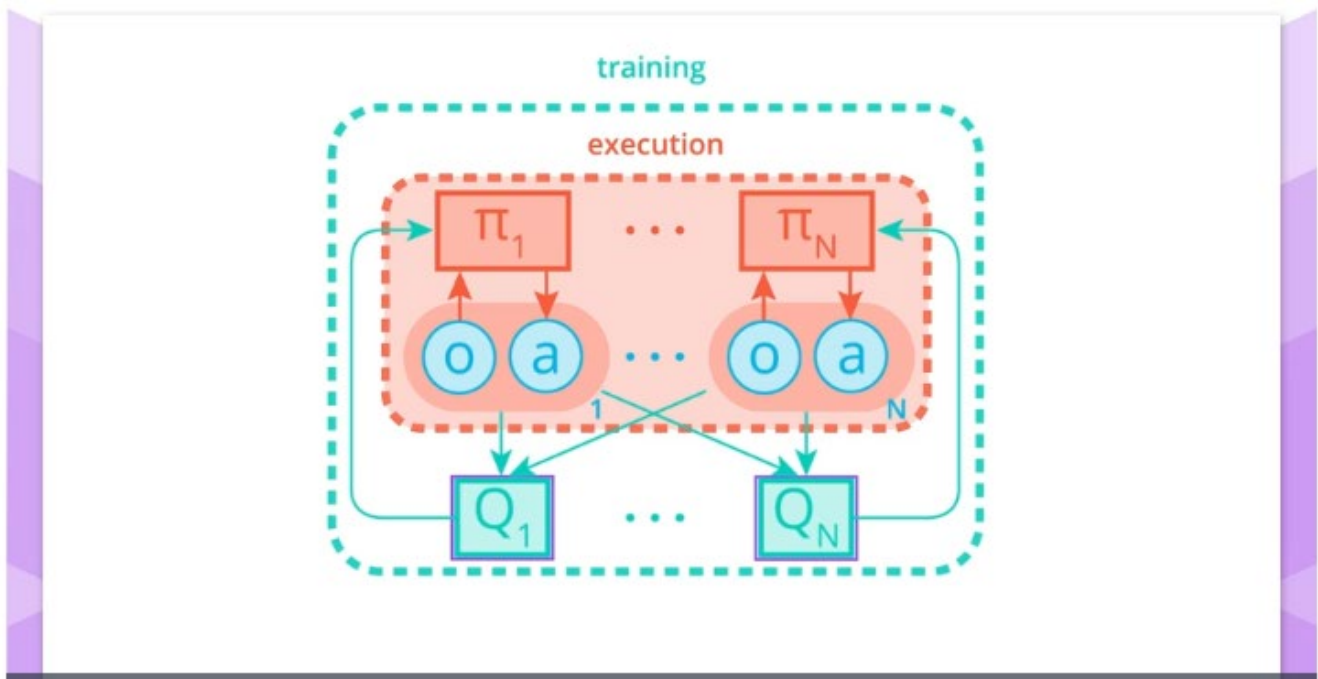
$$\begin{aligned}\theta^{Q'} &\leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'} \\ \theta^{\mu'} &\leftarrow \tau \theta^\mu + (1 - \tau) \theta^{\mu'}\end{aligned}$$

end for
end for

Action exploration includes noise generation in the selection method; this is in the **OUNoise** class and in **Agent.act()** line 40. The critic loss minimization was implemented for the multi-agent case as **pytorch.nn.functional.mse_loss()** in line 133 & 136 of **MADDPG.learn()**. The actor gradient maximization was implemented for multiple agents by computing the negative of the loss, in line 152 of **MADDPG.learn()**. Soft updates operate on the target networks, in **MADDPG.learn()** lines 160-161, and **MADDPG.soft_update()** for multiple agents.

EXTENSION TO MULTIPLE AGENTS

The figure below captures the relationship between multiple agents in training and execution:



Lines 9-19 of the **train()** function extract the states, actions, next states, rewards, and done flags for multiple agents. To act at each episode, **MADDPG.act_episode()** calls for each agent **Agent.act()**, which runs the respective state through the respective actor local network. After being batch added to the replay buffer, the experiences are sampled as lists of tensors. This is in **ReplayBuffer.sample()** lines 62-71 for random sampling and **ReplayBuffer.tree_sample()** lines 98-121 for prioritized experience replay.

Part 4.1 of the MADDPG paper describes a *centralized action-value function* that takes as input the actions, states, and observations of all agents. This amalgamation of actions/states/rewards is in **MADDPG.learn()** lines 107-109, 116, and 148. This is applied to the networks for both agents, in **MADDPG.learn()** lines 103, 117, 152, and 160-161.

Lastly, the figure above shows unique policies and value functions for each agent. This translates to unique actor local, actor target, critic local, and critic target networks for each agent. This is in **MADDPG.__init__()** line 27, **Agent.__init__()** lines 18-24, and in the **Actor()** and **Critic()** classes.

NETWORK DESIGN

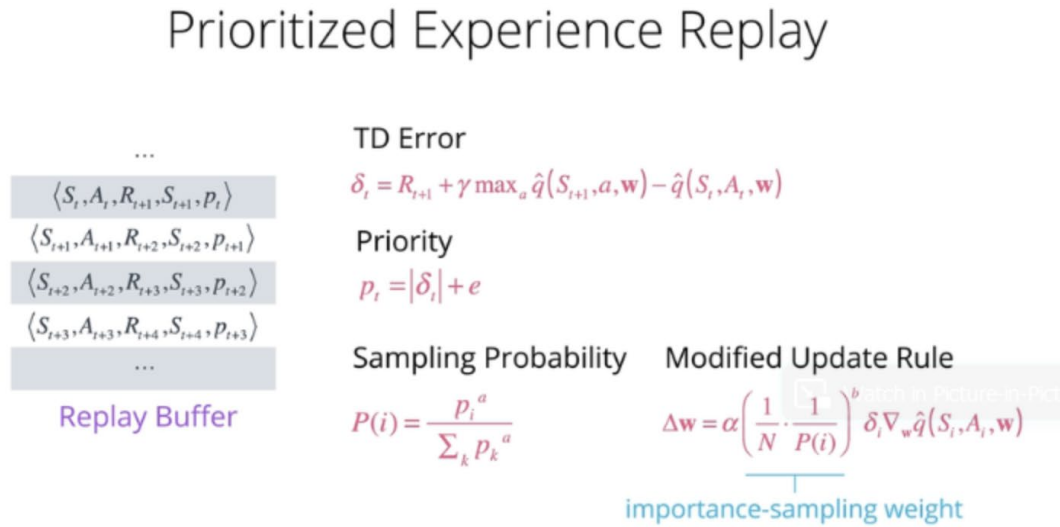
The design strategy was to start with the design from project 2, then expand the network, based on observed performance and peer recommendation. Descending fully connected layers sized to leverage GPU design at 512, 256, and 128 neurons were selected. All 3 layers were used for the Actor, while the first two were used for the Critic. Dropout at 0.25 was added between each layer for both Actor and Critic. This is in the **Actor()** and **Critic()** classes.

COMPUTATION GRAPH MANAGEMENT

Recalling the DDPG algorithm, we compute gradients for the local networks and perform hard or soft updating for the target networks. Hence we **detach()** tensors coming from the target networks. This is in **MADDPG.learn()** lines 116 and 117.

PRIORITIZED EXPERIENCE REPLAY

The figure below captures the essential elements of prioritized experienced replay:



The importance sampling weight was normalized by the maximum priority, for stability purposes. This is in **ReplayBuffer.tree_sample()** line 108.

It was critical to compute the TD errors during learning, rather than once an experience was added, as one may be tempted. It made sense then, to determine experience priority during learning. Since at this point the samples have already been drawn, this creates an experience flow in which during the early parts of learning experiences are more often sampled randomly, while later in learning experiences are more often sampled by priority (TD error). Experience addition/sampling is in **MADDPG.step()** lines 73 & 80, while TD error computation and priority updating is in **MADDPG.learn()** lines 128-132 and **ReplayBuffer.update_priorities()**, and lastly importance sampling weight computation is in **ReplayBuffer.tree_sample()** lines 107 & 108.

The b value was driven by the proximity of the episode reward to the solution criteria, 0.5. Hence, it was handled outside of the **Agent**, in the **train()** loop line 25.

Lastly, a sum tree data structure was implemented for the replay buffer memory. The sum tree made computation time tolerable. The code was sourced heavily from the following:

https://github.com/MorvanZhou/Reinforcement-learning-with-tensorflow/blob/master/contents/5.2_Prioritized_Replay_DQN/RL_brain.py

https://github.com/KailinTong/Udacity-Project-1-Navigation/blob/main/dqn_agent.py

HYPERPARAMETER TUNING AND SELECTION

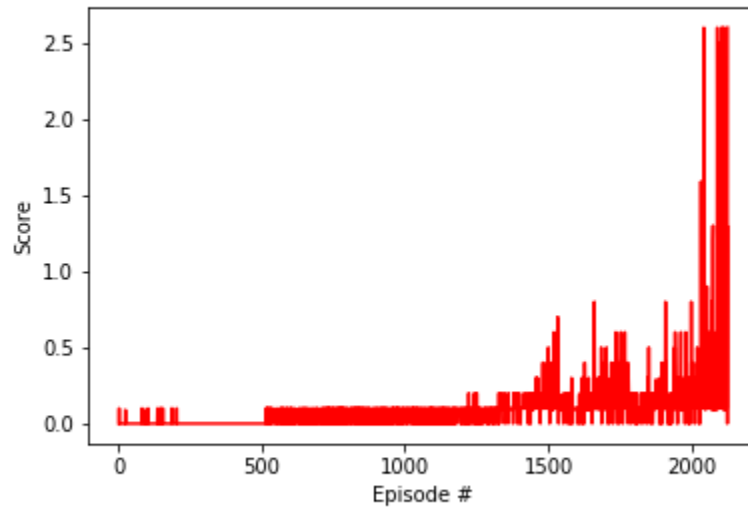
Hyperparameter selection meant the difference between solution convergence and divergence. The buffer size was $2 * 1e5$, multiplying the size from the DQN project by the number of agents. Although a large batch size of 1024 worked for the previous project, convergence was more easily achievable at size 256. A noise standard deviation of 0.1 simply worked better than 0.2. The discount factor γ , soft update factor τ , and learning rates were set according to peer recommendation, and worked nicely.

Modifying the stabilization e and prioritization extent a for prioritized experience replay (PER) was critical to ensuring PER convergence. Instability observations motivated increasing e from $1e-6$ to $1e-3$. Performance limitation around a score of 0.27 suggested the algorithm was stuck learning from the same experiences. This motivated decreasing a from 0.8 to 0.5 to flatten the priority distribution, preventing experiences with higher TD error from over-dominating experiences with lower TD error. These variables were set globally in the **MADDPG()** cell, passed through the **MADDPG()** class in the **__main__** cell, and handled in **ReplayBuffer.add()** line 39 and **ReplBuffer.update_priorities()** line 52.

RESULT AND CONCLUSION

This implementation took nearly an hour across ~2k episodes to achieve the score requirement:

```
Environment solved in 2026 episodes!  
Max score: 2.600      Average Score: 0.517
```



```
finish time: 3428.2594363689423
```

The weights are included as **checkpoint_actor_agent1_per.pth**, **checkpoint_actor_agent2_per.pth**, **checkpoint_critic_agent1_per.pth**, and **checkpoint_critic_agent2_per.pth**. The episode statistics and provided below as an amendment. Performance could be improved by further tuning hyperparameters (perhaps focusing more on PER hyperparameters), experimenting with other network structures, and experimenting with other learning intervals. Its possible that fixing the batch size closer to GPU capacity, say 1024, could provide an anchor for such adjustments.

AMENDMENT: EPISODE STATISTICS

Episode 0	Max10 score: 0.000	Average Score: 0.000	time: 0.10606
Episode 10	Max10 score: 0.100	Average Score: 0.009	time: 2.11616
Episode 20	Max10 score: 0.000	Average Score: 0.005	time: 4.64141
Episode 30	Max10 score: 0.090	Average Score: 0.006	time: 10.86767
Episode 40	Max10 score: 0.000	Average Score: 0.005	time: 16.31010
Episode 50	Max10 score: 0.000	Average Score: 0.004	time: 21.75554
Episode 60	Max10 score: 0.000	Average Score: 0.003	time: 27.30303
Episode 70	Max10 score: 0.000	Average Score: 0.003	time: 32.83838
Episode 80	Max10 score: 0.000	Average Score: 0.002	time: 38.37979
Episode 90	Max10 score: 0.100	Average Score: 0.003	time: 44.57978
Episode 100	Max10 score: 0.090	Average Score: 0.004	time: 50.84747
Episode 110	Max10 score: 0.100	Average Score: 0.004	time: 57.04645
Episode 120	Max10 score: 0.000	Average Score: 0.004	time: 62.73131
Episode 130	Max10 score: 0.000	Average Score: 0.003	time: 68.29999
Episode 140	Max10 score: 0.100	Average Score: 0.004	time: 74.42121
Episode 150	Max10 score: 0.100	Average Score: 0.006	time: 82.08382
Episode 160	Max10 score: 0.100	Average Score: 0.007	time: 88.19998
Episode 170	Max10 score: 0.000	Average Score: 0.007	time: 93.65757
Episode 180	Max10 score: 0.000	Average Score: 0.007	time: 99.37272
Episode 190	Max10 score: 0.100	Average Score: 0.008	time: 106.15756
Episode 200	Max10 score: 0.000	Average Score: 0.007	time: 111.92221
Episode 210	Max10 score: 0.100	Average Score: 0.007	time: 118.30808
Episode 220	Max10 score: 0.000	Average Score: 0.007	time: 124.03434
Episode 230	Max10 score: 0.000	Average Score: 0.007	time: 129.60504
Episode 240	Max10 score: 0.000	Average Score: 0.006	time: 135.16969
Episode 250	Max10 score: 0.000	Average Score: 0.004	time: 140.61212
Episode 260	Max10 score: 0.000	Average Score: 0.003	time: 146.18787
Episode 270	Max10 score: 0.000	Average Score: 0.003	time: 151.75454
Episode 280	Max10 score: 0.000	Average Score: 0.003	time: 157.33029
Episode 290	Max10 score: 0.000	Average Score: 0.001	time: 162.91010
Episode 300	Max10 score: 0.000	Average Score: 0.001	time: 168.49999
Episode 310	Max10 score: 0.000	Average Score: 0.000	time: 174.06363
Episode 320	Max10 score: 0.000	Average Score: 0.000	time: 179.63535
Episode 330	Max10 score: 0.000	Average Score: 0.000	time: 185.19494
Episode 340	Max10 score: 0.000	Average Score: 0.000	time: 190.74141
Episode 350	Max10 score: 0.000	Average Score: 0.000	time: 196.31212
Episode 360	Max10 score: 0.000	Average Score: 0.000	time: 201.85959
Episode 370	Max10 score: 0.000	Average Score: 0.000	time: 207.42121
Episode 380	Max10 score: 0.000	Average Score: 0.000	time: 212.99493
Episode 390	Max10 score: 0.000	Average Score: 0.000	time: 218.55959
Episode 400	Max10 score: 0.000	Average Score: 0.000	time: 224.17878
Episode 410	Max10 score: 0.000	Average Score: 0.000	time: 229.67777
Episode 420	Max10 score: 0.000	Average Score: 0.000	time: 235.19090
Episode 430	Max10 score: 0.000	Average Score: 0.000	time: 240.97272
Episode 440	Max10 score: 0.000	Average Score: 0.000	time: 246.46363
Episode 450	Max10 score: 0.000	Average Score: 0.000	time: 252.04646
Episode 460	Max10 score: 0.000	Average Score: 0.000	time: 257.59191
Episode 470	Max10 score: 0.000	Average Score: 0.000	time: 263.12525
Episode 480	Max10 score: 0.000	Average Score: 0.000	time: 268.63838
Episode 490	Max10 score: 0.000	Average Score: 0.000	time: 274.16767
Episode 500	Max10 score: 0.000	Average Score: 0.000	time: 279.69494
Episode 510	Max10 score: 0.000	Average Score: 0.000	time: 285.20606
Episode 520	Max10 score: 0.100	Average Score: 0.004	time: 293.16565
Episode 530	Max10 score: 0.100	Average Score: 0.008	time: 301.25353
Episode 540	Max10 score: 0.100	Average Score: 0.013	time: 310.02828
Episode 550	Max10 score: 0.100	Average Score: 0.014	time: 316.29898
Episode 560	Max10 score: 0.000	Average Score: 0.014	time: 321.91716

Episode 570	Max10 score: 0.100	Average Score: 0.018	time: 329.87272
Episode 580	Max10 score: 0.100	Average Score: 0.020	time: 336.78483
Episode 590	Max10 score: 0.100	Average Score: 0.025	time: 345.46565
Episode 600	Max10 score: 0.100	Average Score: 0.027	time: 352.32727
Episode 610	Max10 score: 0.100	Average Score: 0.030	time: 359.74141
Episode 620	Max10 score: 0.100	Average Score: 0.027	time: 365.99191
Episode 630	Max10 score: 0.100	Average Score: 0.024	time: 372.16261
Episode 640	Max10 score: 0.100	Average Score: 0.021	time: 379.02929
Episode 650	Max10 score: 0.100	Average Score: 0.023	time: 386.43838
Episode 660	Max10 score: 0.100	Average Score: 0.026	time: 394.65352
Episode 670	Max10 score: 0.100	Average Score: 0.023	time: 400.78282
Episode 680	Max10 score: 0.100	Average Score: 0.024	time: 408.25151
Episode 690	Max10 score: 0.100	Average Score: 0.020	time: 414.40302
Episode 700	Max10 score: 0.100	Average Score: 0.020	time: 421.27777
Episode 710	Max10 score: 0.100	Average Score: 0.018	time: 427.29493
Episode 720	Max10 score: 0.100	Average Score: 0.021	time: 435.32121
Episode 730	Max10 score: 0.100	Average Score: 0.025	time: 443.88282
Episode 740	Max10 score: 0.100	Average Score: 0.028	time: 452.53434
Episode 750	Max10 score: 0.100	Average Score: 0.031	time: 461.80706
Episode 760	Max10 score: 0.100	Average Score: 0.032	time: 469.79797
Episode 770	Max10 score: 0.100	Average Score: 0.037	time: 479.94949
Episode 780	Max10 score: 0.100	Average Score: 0.039	time: 488.59191
Episode 790	Max10 score: 0.100	Average Score: 0.044	time: 497.91414
Episode 800	Max10 score: 0.100	Average Score: 0.045	time: 505.68079
Episode 810	Max10 score: 0.100	Average Score: 0.046	time: 512.43635
Episode 820	Max10 score: 0.100	Average Score: 0.049	time: 522.37777
Episode 830	Max10 score: 0.100	Average Score: 0.050	time: 531.74343
Episode 840	Max10 score: 0.100	Average Score: 0.050	time: 540.44545
Episode 850	Max10 score: 0.100	Average Score: 0.051	time: 550.36161
Episode 860	Max10 score: 0.100	Average Score: 0.053	time: 559.64747
Episode 870	Max10 score: 0.100	Average Score: 0.049	time: 566.43737
Episode 880	Max10 score: 0.100	Average Score: 0.048	time: 574.63030
Episode 890	Max10 score: 0.100	Average Score: 0.043	time: 580.84040
Episode 900	Max10 score: 0.100	Average Score: 0.045	time: 589.94040
Episode 910	Max10 score: 0.100	Average Score: 0.047	time: 598.14040
Episode 920	Max10 score: 0.100	Average Score: 0.045	time: 606.97979
Episode 930	Max10 score: 0.100	Average Score: 0.044	time: 615.81413
Episode 940	Max10 score: 0.100	Average Score: 0.045	time: 625.41212
Episode 950	Max10 score: 0.100	Average Score: 0.039	time: 631.72020
Episode 960	Max10 score: 0.100	Average Score: 0.037	time: 640.03636
Episode 970	Max10 score: 0.100	Average Score: 0.043	time: 650.88888
Episode 980	Max10 score: 0.100	Average Score: 0.045	time: 660.46868
Episode 990	Max10 score: 0.100	Average Score: 0.047	time: 668.12323
Episode 1000	Max10 score: 0.100	Average Score: 0.049	time: 679.12020
Episode 1010	Max10 score: 0.100	Average Score: 0.052	time: 690.18988
Episode 1020	Max10 score: 0.100	Average Score: 0.056	time: 704.25353
Episode 1030	Max10 score: 0.100	Average Score: 0.056	time: 713.14444
Episode 1040	Max10 score: 0.100	Average Score: 0.056	time: 722.75151
Episode 1050	Max10 score: 0.100	Average Score: 0.060	time: 731.60000
Episode 1060	Max10 score: 0.100	Average Score: 0.058	time: 738.49292
Episode 1070	Max10 score: 0.100	Average Score: 0.054	time: 748.08181
Episode 1080	Max10 score: 0.100	Average Score: 0.051	time: 756.16868
Episode 1090	Max10 score: 0.100	Average Score: 0.052	time: 764.37675
Episode 1100	Max10 score: 0.100	Average Score: 0.052	time: 775.24040
Episode 1110	Max10 score: 0.100	Average Score: 0.049	time: 783.67777
Episode 1120	Max10 score: 0.100	Average Score: 0.044	time: 792.70000
Episode 1130	Max10 score: 0.100	Average Score: 0.045	time: 801.95959
Episode 1140	Max10 score: 0.100	Average Score: 0.044	time: 810.62525
Episode 1150	Max10 score: 0.100	Average Score: 0.045	time: 821.74848

Episode 1160	Max10 score: 0.100	Average Score: 0.047	time: 830.68989
Episode 1170	Max10 score: 0.100	Average Score: 0.047	time: 838.84544
Episode 1180	Max10 score: 0.100	Average Score: 0.051	time: 850.64545
Episode 1190	Max10 score: 0.100	Average Score: 0.057	time: 864.33029
Episode 1200	Max10 score: 0.100	Average Score: 0.055	time: 873.90707
Episode 1210	Max10 score: 0.100	Average Score: 0.059	time: 886.92626
Episode 1220	Max10 score: 0.100	Average Score: 0.061	time: 897.14342
Episode 1230	Max10 score: 0.200	Average Score: 0.062	time: 908.65959
Episode 1240	Max10 score: 0.100	Average Score: 0.065	time: 920.99393
Episode 1250	Max10 score: 0.200	Average Score: 0.068	time: 933.97171
Episode 1260	Max10 score: 0.100	Average Score: 0.070	time: 943.36060
Episode 1270	Max10 score: 0.100	Average Score: 0.072	time: 953.79292
Episode 1280	Max10 score: 0.100	Average Score: 0.073	time: 964.56262
Episode 1290	Max10 score: 0.100	Average Score: 0.070	time: 974.57373
Episode 1300	Max10 score: 0.100	Average Score: 0.073	time: 986.36060
Episode 1310	Max10 score: 0.100	Average Score: 0.072	time: 996.40909
Episode 1320	Max10 score: 0.100	Average Score: 0.074	time: 1007.76666
Episode 1330	Max10 score: 0.200	Average Score: 0.076	time: 1022.27979
Episode 1340	Max10 score: 0.200	Average Score: 0.079	time: 1037.48382
Episode 1350	Max10 score: 0.200	Average Score: 0.079	time: 1050.48282
Episode 1360	Max10 score: 0.200	Average Score: 0.083	time: 1063.28686
Episode 1370	Max10 score: 0.100	Average Score: 0.086	time: 1077.53636
Episode 1380	Max10 score: 0.200	Average Score: 0.088	time: 1093.62121
Episode 1390	Max10 score: 0.200	Average Score: 0.094	time: 1110.89494
Episode 1400	Max10 score: 0.200	Average Score: 0.097	time: 1127.49494
Episode 1410	Max10 score: 0.100	Average Score: 0.098	time: 1139.95151
Episode 1420	Max10 score: 0.200	Average Score: 0.100	time: 1157.51414
Episode 1430	Max10 score: 0.200	Average Score: 0.102	time: 1172.81110
Episode 1440	Max10 score: 0.200	Average Score: 0.102	time: 1191.15555
Episode 1450	Max10 score: 0.200	Average Score: 0.104	time: 1210.19797
Episode 1460	Max10 score: 0.300	Average Score: 0.109	time: 1234.00606
Episode 1470	Max10 score: 0.300	Average Score: 0.114	time: 1255.00000
Episode 1480	Max10 score: 0.200	Average Score: 0.116	time: 1273.42222
Episode 1490	Max10 score: 0.400	Average Score: 0.122	time: 1298.89594
Episode 1500	Max10 score: 0.500	Average Score: 0.136	time: 1337.27272
Episode 1510	Max10 score: 0.400	Average Score: 0.143	time: 1362.17676
Episode 1520	Max10 score: 0.600	Average Score: 0.147	time: 1387.08787
Episode 1530	Max10 score: 0.590	Average Score: 0.161	time: 1428.61919
Episode 1540	Max10 score: 0.700	Average Score: 0.169	time: 1456.20707
Episode 1550	Max10 score: 0.190	Average Score: 0.168	time: 1474.67373
Episode 1560	Max10 score: 0.200	Average Score: 0.164	time: 1490.55656
Episode 1570	Max10 score: 0.200	Average Score: 0.159	time: 1506.29696
Episode 1580	Max10 score: 0.200	Average Score: 0.155	time: 1523.15151
Episode 1590	Max10 score: 0.300	Average Score: 0.148	time: 1543.65555
Episode 1600	Max10 score: 0.100	Average Score: 0.134	time: 1559.05454
Episode 1610	Max10 score: 0.190	Average Score: 0.128	time: 1576.08484
Episode 1620	Max10 score: 0.300	Average Score: 0.125	time: 1598.80099
Episode 1630	Max10 score: 0.400	Average Score: 0.114	time: 1621.56767
Episode 1640	Max10 score: 0.100	Average Score: 0.105	time: 1638.17676
Episode 1650	Max10 score: 0.300	Average Score: 0.108	time: 1656.35555
Episode 1660	Max10 score: 0.400	Average Score: 0.114	time: 1686.28282
Episode 1670	Max10 score: 0.800	Average Score: 0.125	time: 1717.30606
Episode 1680	Max10 score: 0.300	Average Score: 0.132	time: 1742.94343
Episode 1690	Max10 score: 0.500	Average Score: 0.139	time: 1776.11111
Episode 1700	Max10 score: 0.400	Average Score: 0.146	time: 1803.89090
Episode 1710	Max10 score: 0.500	Average Score: 0.154	time: 1833.83636
Episode 1720	Max10 score: 0.400	Average Score: 0.156	time: 1858.06868
Episode 1730	Max10 score: 0.200	Average Score: 0.154	time: 1879.29090
Episode 1740	Max10 score: 0.600	Average Score: 0.169	time: 1921.33332

Episode 1750	Max10 score: 0.400	Average Score: 0.170	time: 1946.42323
Episode 1760	Max10 score: 0.600	Average Score: 0.174	time: 1978.93434
Episode 1770	Max10 score: 0.600	Average Score: 0.173	time: 2009.91818
Episode 1780	Max10 score: 0.400	Average Score: 0.173	time: 2036.18282
Episode 1790	Max10 score: 0.200	Average Score: 0.169	time: 2060.63737
Episode 1800	Max10 score: 0.200	Average Score: 0.165	time: 2082.84949
Episode 1810	Max10 score: 0.200	Average Score: 0.156	time: 2099.31010
Episode 1820	Max10 score: 0.100	Average Score: 0.150	time: 2114.85554
Episode 1830	Max10 score: 0.200	Average Score: 0.149	time: 2134.40201
Episode 1840	Max10 score: 0.200	Average Score: 0.136	time: 2156.27272
Episode 1850	Max10 score: 0.300	Average Score: 0.135	time: 2178.68585
Episode 1860	Max10 score: 0.500	Average Score: 0.129	time: 2204.67574
Episode 1870	Max10 score: 0.200	Average Score: 0.120	time: 2221.50707
Episode 1880	Max10 score: 0.290	Average Score: 0.120	time: 2248.86968
Episode 1890	Max10 score: 0.300	Average Score: 0.123	time: 2273.66666
Episode 1900	Max10 score: 0.400	Average Score: 0.122	time: 2295.76868
Episode 1910	Max10 score: 0.800	Average Score: 0.135	time: 2330.19898
Episode 1920	Max10 score: 0.400	Average Score: 0.140	time: 2353.20000
Episode 1930	Max10 score: 0.200	Average Score: 0.137	time: 2369.04747
Episode 1940	Max10 score: 0.200	Average Score: 0.138	time: 2386.71515
Episode 1950	Max10 score: 0.600	Average Score: 0.148	time: 2421.19796
Episode 1960	Max10 score: 0.300	Average Score: 0.149	time: 2445.99999
Episode 1970	Max10 score: 0.600	Average Score: 0.157	time: 2473.56867
Episode 1980	Max10 score: 0.600	Average Score: 0.161	time: 2502.50000
Episode 1990	Max10 score: 0.400	Average Score: 0.162	time: 2531.35554
Episode 2000	Max10 score: 0.800	Average Score: 0.170	time: 2563.41313
Episode 2010	Max10 score: 0.400	Average Score: 0.169	time: 2594.83433
Episode 2020	Max10 score: 0.500	Average Score: 0.174	time: 2621.87979
Episode 2030	Max10 score: 0.300	Average Score: 0.183	time: 2649.99191
Episode 2040	Max10 score: 1.590	Average Score: 0.222	time: 2728.76161
Episode 2050	Max10 score: 2.600	Average Score: 0.238	time: 2787.30909
Episode 2060	Max10 score: 0.900	Average Score: 0.250	time: 2829.35050
Episode 2070	Max10 score: 0.800	Average Score: 0.272	time: 2888.07272
Episode 2080	Max10 score: 1.300	Average Score: 0.290	time: 2946.61313
Episode 2090	Max10 score: 2.600	Average Score: 0.314	time: 3008.90606
Episode 2100	Max10 score: 2.500	Average Score: 0.361	time: 3112.15454
Episode 2110	Max10 score: 2.600	Average Score: 0.425	time: 3240.66767
Episode 2120	Max10 score: 2.600	Average Score: 0.482	time: 3357.39393
Episode 2126	Episode score: 2.600	Average Score: 0.517	time: 3427.980

Environment solved in 2026 episodes!
Max score: 2.600 Average Score: 0.517