2025-07-20 11:15:44.660633: E external/local\_xla/xla/stream\_executor/cuda/cuda\_fft.cc:477] Unable to register cuFFT factory: Attempting to register factory for plugin cuFFT when one has already been registered

WARNING: All log messages before absl::InitializeLog() is called are written to STDERR

E0000 00:00:1753010144.705632 23507 cuda\_dnn.cc:8310] Unable to register cuDNN factory: Attempting to register factory for plugin cuDNN when one has already been registered

E0000 00:00:1753010144.721100 23507 cuda\_blas.cc:1418] Unable to register cuBLAS factory: Attempting to register factory for plugin cuBLAS when one has already been registered

2025-07-20 11:15:44.775967: I tensorflow/core/platform/cpu\_feature\_guard.cc:210] This TensorFlow binary is optimized to use available CPU instructions in performance-critical operations.

To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.

[I 2025-07-20 11:15:51,826] A new study created in RDB with name: DQN-bc-tuning-modular-v1

Tuning DQN: 0% 0/50 [00:00<?, ?it/s]

[INFO] Starting Trial 0 with Seed: 300525

[DEBUG] Trial 0 | Timestep 20000 | Mean reward: 1727.5804340000002

[DEBUG] Trial 0 | Timestep 40000 | Mean reward: 1607.1340229999998

[DEBUG] Trial 0 | Timestep 60000 | Mean reward: 1841.165746

[DEBUG] Trial 0 | Timestep 80000 | Mean reward: 1745.706094

[DEBUG] Trial 0 | Timestep 100000 | Mean reward: -5.170405

[DEBUG] Trial 0 | Timestep 120000 | Mean reward: 1863.578783

[DEBUG] Trial 0 | Timestep 140000 | Mean reward: 1845.125174

[DEBUG] Trial 0 | Timestep 160000 | Mean reward: 1805.328579

[DEBUG] Trial 0 | Timestep 180000 | Mean reward: 1887.214304

[DEBUG] Trial 0 | Timestep 200000 | Mean reward: 1777.193293

[I 2025-07-20 11:21:07,071] Trial 0 finished with value: 1777.193293 and parameters: {'n\_layers': 2, 'layer\_size': 73, 'activation\_fn': 'tanh', 'learning\_rate': 0.00014154486471907698, 'buffer\_size': 53748, 'batch\_size': 32, 'gamma': 0.9155069008402772, 'tau': 0.008882048075020464, 'train\_freq': 4, 'exploration\_fraction': 0.19886522152830693, 'exploration\_final\_eps': 0.04036512226290992, 'target\_update\_interval': 500}. Best is trial 0 with value: 1777.193293.

Tuning DQN: 2% 1/50 [05:15<4:17:26, 315.24s/it, best\_val=1777.19]

[INFO] Starting Trial 1 with Seed: 101352

[DEBUG] Trial 1 | Timestep 20000 | Mean reward: -0.167644

[DEBUG] Trial 1 | Timestep 40000 | Mean reward: 35.485171

[DEBUG] Trial 1 | Timestep 60000 | Mean reward: 1878.0167840000001

[DEBUG] Trial 1 | Timestep 80000 | Mean reward: 16.875332

[DEBUG] Trial 1 | Timestep 100000 | Mean reward: 9.198864

[DEBUG] Trial 1 | Timestep 120000 | Mean reward: 2.084617

[DEBUG] Trial 1 | Timestep 140000 | Mean reward: -2.718061

[DEBUG] Trial 1 | Timestep 160000 | Mean reward: -2.613952

[DEBUG] Trial 1 | Timestep 180000 | Mean reward: 10.5964

[DEBUG] Trial 1 | Timestep 200000 | Mean reward: 1867.370987

[I 2025-07-20 11:26:05,061] Trial 1 finished with value: 1867.370987 and parameters: {'n\_layers': 2, 'layer\_size': 202, 'activation\_fn': 'relu', 'learning\_rate': 0.00015985037708422188, 'buffer\_size': 195160, 'batch\_size': 512, 'gamma': 0.9356083356418619, 'tau': 0.017599594231949168, 'train\_freq': 8, 'exploration\_fraction': 0.1644792410463663, 'exploration\_final\_eps': 0.06175988268766712, 'target\_update\_interval': 10000}. Best is trial 1 with value: 1867.370987.

Tuning DQN: 4% 2/50 [10:13<4:04:04, 305.09s/it, best\_val=1867.37]

[INFO] Starting Trial 2 with Seed: 493962

[DEBUG] Trial 2 | Timestep 20000 | Mean reward: 1.665616

[DEBUG] Trial 2 | Timestep 40000 | Mean reward: 1899.373688

[DEBUG] Trial 2 | Timestep 60000 | Mean reward: 1919.1498519999998

[DEBUG] Trial 2 | Timestep 80000 | Mean reward: 1829.8284649999998

[DEBUG] Trial 2 | Timestep 100000 | Mean reward: 12.406174

[DEBUG] Trial 2 | Timestep 120000 | Mean reward: 1798.7060009999998

[DEBUG] Trial 2 | Timestep 140000 | Mean reward: 1862.428385

[DEBUG] Trial 2 | Timestep 160000 | Mean reward: -50.13162

[DEBUG] Trial 2 | Timestep 180000 | Mean reward: -4.654406

[DEBUG] Trial 2 | Timestep 200000 | Mean reward: 1779.717109

[I 2025-07-20 11:30:27,561] Trial 2 finished with value: 1779.717109 and parameters: {'n\_layers': 4, 'layer\_size': 250, 'activation\_fn': 'tanh', 'learning\_rate': 0.0008014622189583533, 'buffer\_size': 93114, 'batch\_size': 64, 'gamma': 0.9063321303494731, 'tau': 0.0163239110606003, 'train\_freq': 16, 'exploration\_fraction': 0.3887643004825859, 'exploration\_final\_eps': 0.043385899911248325, 'target\_update\_interval': 2000}. Best is trial 1 with value: 1867.370987.

Tuning DQN: 6% 3/50 [14:35<3:43:45, 285.64s/it, best\_val=1867.37]

[INFO] Starting Trial 3 with Seed: 911974

[DEBUG] Trial 3 | Timestep 20000 | Mean reward: 1805.6012199999998

[DEBUG] Trial 3 | Timestep 40000 | Mean reward: 1947.115101

[DEBUG] Trial 3 | Timestep 60000 | Mean reward: 1831.1288240000001

[DEBUG] Trial 3 | Timestep 80000 | Mean reward: 1906.614303

[DEBUG] Trial 3 | Timestep 100000 | Mean reward: 1899.859177

[DEBUG] Trial 3 | Timestep 120000 | Mean reward: 1824.014641

[DEBUG] Trial 3 | Timestep 140000 | Mean reward: 1795.951206

[DEBUG] Trial 3 | Timestep 160000 | Mean reward: 1825.515912

[DEBUG] Trial 3 | Timestep 180000 | Mean reward: 1913.751294

[DEBUG] Trial 3 | Timestep 200000 | Mean reward: 1905.820014

[I 2025-07-20 11:36:41,588] Trial 3 finished with value: 1905.820014 and parameters: {'n\_layers': 4, 'layer\_size': 171, 'activation\_fn': 'relu', 'learning\_rate': 0.0009706174983140924, 'buffer\_size': 52764, 'batch\_size': 256, 'gamma': 0.9921260140130664, 'tau': 0.016569905149490673, 'train\_freq': 8, 'exploration\_fraction': 0.3630728045562289, 'exploration\_final\_eps': 0.08594960313425784, 'target\_update\_interval': 5000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 8% 4/50 [20:49<4:05:44, 320.54s/it, best\_val=1905.82]

[INFO] Starting Trial 4 with Seed: 102197

[DEBUG] Trial 4 | Timestep 20000 | Mean reward: -13.976121

[DEBUG] Trial 4 | Timestep 40000 | Mean reward: 134.01562

[DEBUG] Trial 4 | Timestep 60000 | Mean reward: -11.476069

[DEBUG] Trial 4 | Timestep 80000 | Mean reward: -13.07833

[DEBUG] Trial 4 | Timestep 100000 | Mean reward: -13.976121

[DEBUG] Trial 4 | Timestep 120000 | Mean reward: 1811.673383

[DEBUG] Trial 4 | Timestep 140000 | Mean reward: 1828.194469

[DEBUG] Trial 4 | Timestep 160000 | Mean reward: -5.254787

[DEBUG] Trial 4 | Timestep 180000 | Mean reward: 25.939498

[DEBUG] Trial 4 | Timestep 200000 | Mean reward: -4.182495

[I 2025-07-20 11:40:14,431] Trial 4 finished with value: -4.182495 and parameters: {'n\_layers': 2, 'layer\_size': 189, 'activation\_fn': 'tanh', 'learning\_rate': 0.000704722702682615, 'buffer\_size': 174219, 'batch\_size': 32, 'gamma': 0.9641638510280942, 'tau': 0.011553008130519946, 'train\_freq': 16, 'exploration\_fraction': 0.2558865221993636, 'exploration\_final\_eps': 0.03885106270869217, 'target\_update\_interval': 10000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 10% 5/50 [24:22<3:31:16, 281.70s/it, best\_val=1905.82]

[INFO] Starting Trial 5 with Seed: 728697

[DEBUG] Trial 5 | Timestep 20000 | Mean reward: 51.249144

[DEBUG] Trial 5 | Timestep 40000 | Mean reward: 19.114964

[PRUNE] Trial 5 pruned at 40000 steps.

[I 2025-07-20 11:40:48,512] Trial 5 pruned.

Tuning DQN: 12% 6/50 [24:56<2:24:50, 197.51s/it, best\_val=1905.82]

[INFO] Starting Trial 6 with Seed: 330277

[DEBUG] Trial 6 | Timestep 20000 | Mean reward: 1168.65218

[DEBUG] Trial 6 | Timestep 40000 | Mean reward: 1736.767484

[DEBUG] Trial 6 | Timestep 60000 | Mean reward: 1836.0610669999999

[PRUNE] Trial 6 pruned at 60000 steps.

[I 2025-07-20 11:42:18,758] Trial 6 pruned.

Tuning DQN: 14% 7/50 [26:26<1:56:25, 162.44s/it, best\_val=1905.82]

[INFO] Starting Trial 7 with Seed: 494401

[DEBUG] Trial 7 | Timestep 20000 | Mean reward: 1897.444006

[DEBUG] Trial 7 | Timestep 40000 | Mean reward: 24.235784

[DEBUG] Trial 7 | Timestep 60000 | Mean reward: -10.392759

[DEBUG] Trial 7 | Timestep 80000 | Mean reward: 1816.0042110000002

[DEBUG] Trial 7 | Timestep 100000 | Mean reward: 1929.40896

[DEBUG] Trial 7 | Timestep 120000 | Mean reward: 1774.9640339999999

[DEBUG] Trial 7 | Timestep 140000 | Mean reward: -0.717705

[DEBUG] Trial 7 | Timestep 160000 | Mean reward: 101.240547

[DEBUG] Trial 7 | Timestep 180000 | Mean reward: 11.019016

[DEBUG] Trial 7 | Timestep 200000 | Mean reward: 546.061839

[I 2025-07-20 11:48:14,994] Trial 7 finished with value: 546.061839 and parameters: {'n\_layers': 2, 'layer\_size': 235, 'activation\_fn': 'leaky\_relu', 'learning\_rate': 0.000114416142070478, 'buffer\_size': 99881, 'batch\_size': 64, 'gamma': 0.9664459143877335, 'tau': 0.007725971307820809, 'train\_freq': 4, 'exploration\_fraction': 0.1246306217904643, 'exploration\_final\_eps': 0.05564513871360352, 'target\_update\_interval': 10000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 16% 8/50 [32:23<2:36:53, 224.14s/it, best\_val=1905.82]

[INFO] Starting Trial 8 with Seed: 46980

[DEBUG] Trial 8 | Timestep 20000 | Mean reward: 814.2754449999999

[DEBUG] Trial 8 | Timestep 40000 | Mean reward: 1799.5406410000003

[DEBUG] Trial 8 | Timestep 60000 | Mean reward: 1794.399606

[PRUNE] Trial 8 pruned at 60000 steps.

[I 2025-07-20 11:49:48,550] Trial 8 pruned.

Tuning DQN: 18% 9/50 [33:56<2:05:15, 183.32s/it, best\_val=1905.82]

[INFO] Starting Trial 9 with Seed: 249147

[DEBUG] Trial 9 | Timestep 20000 | Mean reward: 267.054011

[DEBUG] Trial 9 | Timestep 40000 | Mean reward: -13.976121

[PRUNE] Trial 9 pruned at 40000 steps.

[I 2025-07-20 11:50:23,544] Trial 9 pruned.

Tuning DQN: 20% 10/50 [34:31<1:31:41, 137.53s/it, best\_val=1905.82]

[INFO] Starting Trial 10 with Seed: 141875

[DEBUG] Trial 10 | Timestep 20000 | Mean reward: 1910.8330209999997

[DEBUG] Trial 10 | Timestep 40000 | Mean reward: 11.768338

[DEBUG] Trial 10 | Timestep 60000 | Mean reward: 34.068247

[DEBUG] Trial 10 | Timestep 80000 | Mean reward: -50.023786

[DEBUG] Trial 10 | Timestep 100000 | Mean reward: 1867.811504

[DEBUG] Trial 10 | Timestep 120000 | Mean reward: 1853.8829589999998

[DEBUG] Trial 10 | Timestep 140000 | Mean reward: 11.671639

[DEBUG] Trial 10 | Timestep 160000 | Mean reward: 33.075938

[DEBUG] Trial 10 | Timestep 180000 | Mean reward: 12.718517

[DEBUG] Trial 10 | Timestep 200000 | Mean reward: 1588.230827

[I 2025-07-20 11:56:07,934] Trial 10 finished with value: 1588.230827 and parameters: {'n\_layers': 4, 'layer\_size': 134, 'activation\_fn': 'relu', 'learning\_rate': 0.0004609558780349344, 'buffer\_size': 50945, 'batch\_size': 256, 'gamma': 0.9989229763881984, 'tau': 0.014638847240524822, 'train\_freq': 8, 'exploration\_fraction': 0.4831237330130315, 'exploration\_final\_eps': 0.011153537349834361, 'target\_update\_interval': 5000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 22% 11/50 [40:16<2:10:32, 200.84s/it, best\_val=1905.82]

[INFO] Starting Trial 11 with Seed: 819030

[DEBUG] Trial 11 | Timestep 20000 | Mean reward: 1774.1907550000003

[DEBUG] Trial 11 | Timestep 40000 | Mean reward: 1849.5578159999998

[DEBUG] Trial 11 | Timestep 60000 | Mean reward: 1829.1932530000001

[DEBUG] Trial 11 | Timestep 80000 | Mean reward: 1922.515633

[DEBUG] Trial 11 | Timestep 100000 | Mean reward: 1859.365004

[DEBUG] Trial 11 | Timestep 120000 | Mean reward: 1833.731537

[DEBUG] Trial 11 | Timestep 140000 | Mean reward: 1904.635555

[DEBUG] Trial 11 | Timestep 160000 | Mean reward: 1841.8045950000003

[DEBUG] Trial 11 | Timestep 180000 | Mean reward: 1796.823356

[DEBUG] Trial 11 | Timestep 200000 | Mean reward: 10.347947

[I 2025-07-20 12:01:57,775] Trial 11 finished with value: 10.347947 and parameters: {'n\_layers': 3, 'layer\_size': 151, 'activation\_fn': 'relu', 'learning\_rate': 0.0004800298638772757, 'buffer\_size': 88025, 'batch\_size': 512, 'gamma': 0.9457378004702556, 'tau': 0.019924212030622085, 'train\_freq': 8, 'exploration\_fraction': 0.13927908764240743, 'exploration\_final\_eps': 0.07035831812348572, 'target\_update\_interval': 5000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 24% 12/50 [46:05<2:35:54, 246.17s/it, best\_val=1905.82]

[INFO] Starting Trial 12 with Seed: 963151

[DEBUG] Trial 12 | Timestep 20000 | Mean reward: 1945.9865390000002

[DEBUG] Trial 12 | Timestep 40000 | Mean reward: 1860.5905759999998

[DEBUG] Trial 12 | Timestep 60000 | Mean reward: 1735.4459749999999

[DEBUG] Trial 12 | Timestep 80000 | Mean reward: -1.003047

[DEBUG] Trial 12 | Timestep 100000 | Mean reward: 1796.068488

[DEBUG] Trial 12 | Timestep 120000 | Mean reward: 1846.327217

[DEBUG] Trial 12 | Timestep 140000 | Mean reward: 1877.835112

[DEBUG] Trial 12 | Timestep 160000 | Mean reward: 1854.350004

[DEBUG] Trial 12 | Timestep 180000 | Mean reward: 1837.599003

[DEBUG] Trial 12 | Timestep 200000 | Mean reward: 1800.00766

[I 2025-07-20 12:22:23,703] Trial 12 finished with value: 1800.00766 and parameters: {'n\_layers': 3, 'layer\_size': 102, 'activation\_fn': 'relu', 'learning\_rate': 0.000451720038604597, 'buffer\_size': 151277, 'batch\_size': 512, 'gamma': 0.9305064594710305, 'tau': 0.016309052437142132, 'train\_freq': 1, 'exploration\_fraction': 0.4653722252625217, 'exploration\_final\_eps': 0.0717375694781826, 'target\_update\_interval': 500}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 26% 13/50 [1:06:31<5:34:49, 542.97s/it, best\_val=1905.82]

[INFO] Starting Trial 13 with Seed: 603677

[DEBUG] Trial 13 | Timestep 20000 | Mean reward: -13.863282

[DEBUG] Trial 13 | Timestep 40000 | Mean reward: 33.122393

[PRUNE] Trial 13 pruned at 40000 steps.

[I 2025-07-20 12:23:03,948] Trial 13 pruned.

Tuning DQN: 28% 14/50 [1:07:12<3:54:40, 391.12s/it, best\_val=1905.82]

[INFO] Starting Trial 14 with Seed: 331081

[DEBUG] Trial 14 | Timestep 20000 | Mean reward: 1788.588062

[DEBUG] Trial 14 | Timestep 40000 | Mean reward: 1899.993525

[DEBUG] Trial 14 | Timestep 60000 | Mean reward: 1792.5514140000002

[DEBUG] Trial 14 | Timestep 80000 | Mean reward: -49.970551

[DEBUG] Trial 14 | Timestep 100000 | Mean reward: 1861.952387

[DEBUG] Trial 14 | Timestep 120000 | Mean reward: -4.033867

[DEBUG] Trial 14 | Timestep 140000 | Mean reward: 1826.594429

[DEBUG] Trial 14 | Timestep 160000 | Mean reward: -4.8319

[DEBUG] Trial 14 | Timestep 180000 | Mean reward: 1842.238533

[DEBUG] Trial 14 | Timestep 200000 | Mean reward: 139.06901

[I 2025-07-20 12:27:41,025] Trial 14 finished with value: 139.06901 and parameters: {'n\_layers': 3, 'layer\_size': 113, 'activation\_fn': 'leaky\_relu', 'learning\_rate': 0.0002814173470297955, 'buffer\_size': 106932, 'batch\_size': 128, 'gamma': 0.9841207873595, 'tau': 0.017530530335784836, 'train\_freq': 8, 'exploration\_fraction': 0.4256749457533978, 'exploration\_final\_eps': 0.09322070168791168, 'target\_update\_interval': 5000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 30% 15/50 [1:11:49<3:28:06, 356.75s/it, best\_val=1905.82]

[INFO] Starting Trial 15 with Seed: 328240

[DEBUG] Trial 15 | Timestep 20000 | Mean reward: 18.087656

[DEBUG] Trial 15 | Timestep 40000 | Mean reward: 1842.340042

[DEBUG] Trial 15 | Timestep 60000 | Mean reward: 1883.561812

[DEBUG] Trial 15 | Timestep 80000 | Mean reward: -10.392759

[DEBUG] Trial 15 | Timestep 100000 | Mean reward: 11.578232

[DEBUG] Trial 15 | Timestep 120000 | Mean reward: 1924.13907

[DEBUG] Trial 15 | Timestep 140000 | Mean reward: 1945.895394

[DEBUG] Trial 15 | Timestep 160000 | Mean reward: -49.72074899999999

[DEBUG] Trial 15 | Timestep 180000 | Mean reward: 1914.258417

[DEBUG] Trial 15 | Timestep 200000 | Mean reward: 1844.116406

[I 2025-07-20 12:34:11,678] Trial 15 finished with value: 1844.116406 and parameters: {'n\_layers': 4, 'layer\_size': 155, 'activation\_fn': 'elu', 'learning\_rate': 0.0009854339140397499, 'buffer\_size': 74218, 'batch\_size': 512, 'gamma': 0.9296590406357953, 'tau': 0.013867718324097028, 'train\_freq': 8, 'exploration\_fraction': 0.2851362882799871, 'exploration\_final\_eps': 0.09870654352328935, 'target\_update\_interval': 10000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 32% 16/50 [1:18:19<3:27:56, 366.95s/it, best\_val=1905.82]

[INFO] Starting Trial 16 with Seed: 357545

[DEBUG] Trial 16 | Timestep 20000 | Mean reward: -4.418181

[DEBUG] Trial 16 | Timestep 40000 | Mean reward: 1791.666911

[PRUNE] Trial 16 pruned at 40000 steps.

[I 2025-07-20 12:37:53,018] Trial 16 pruned.

Tuning DQN: 34% 17/50 [1:22:01<2:57:44, 323.17s/it, best\_val=1905.82]

[INFO] Starting Trial 17 with Seed: 53678

[DEBUG] Trial 17 | Timestep 20000 | Mean reward: 1173.82184

[DEBUG] Trial 17 | Timestep 40000 | Mean reward: -50.144828999999994

[PRUNE] Trial 17 pruned at 40000 steps.

[I 2025-07-20 12:38:30,401] Trial 17 pruned.

Tuning DQN: 36% 18/50 [1:22:38<2:06:33, 237.29s/it, best\_val=1905.82]

[INFO] Starting Trial 18 with Seed: 379414

[DEBUG] Trial 18 | Timestep 20000 | Mean reward: -4.612426

[DEBUG] Trial 18 | Timestep 40000 | Mean reward: 13.410505

[PRUNE] Trial 18 pruned at 40000 steps.

[I 2025-07-20 12:39:39,724] Trial 18 pruned.

Tuning DQN: 38% 19/50 [1:23:47<1:36:32, 186.84s/it, best\_val=1905.82]

[INFO] Starting Trial 19 with Seed: 761508

[DEBUG] Trial 19 | Timestep 20000 | Mean reward: 1568.007668

[DEBUG] Trial 19 | Timestep 40000 | Mean reward: 4.826413

[PRUNE] Trial 19 pruned at 40000 steps.

[I 2025-07-20 12:40:30,148] Trial 19 pruned.

Tuning DQN: 40% 20/50 [1:24:38<1:12:56, 145.88s/it, best\_val=1905.82]

[INFO] Starting Trial 20 with Seed: 619729

[DEBUG] Trial 20 | Timestep 20000 | Mean reward: 19.505214

[DEBUG] Trial 20 | Timestep 40000 | Mean reward: 31.989703000000002

[PRUNE] Trial 20 pruned at 40000 steps.

[I 2025-07-20 12:42:16,148] Trial 20 pruned.

Tuning DQN: 42% 21/50 [1:26:24<1:04:43, 133.91s/it, best\_val=1905.82]

[INFO] Starting Trial 21 with Seed: 93718

[DEBUG] Trial 21 | Timestep 20000 | Mean reward: 1794.511528

[DEBUG] Trial 21 | Timestep 40000 | Mean reward: 1810.742916

[PRUNE] Trial 21 pruned at 40000 steps.

[I 2025-07-20 12:43:20,738] Trial 21 pruned.

Tuning DQN: 44% 22/50 [1:27:28<52:47, 113.11s/it, best\_val=1905.82]

[INFO] Starting Trial 22 with Seed: 899753

[DEBUG] Trial 22 | Timestep 20000 | Mean reward: -4.532556

[DEBUG] Trial 22 | Timestep 40000 | Mean reward: 1937.013984

[DEBUG] Trial 22 | Timestep 60000 | Mean reward: 1913.590594

[DEBUG] Trial 22 | Timestep 80000 | Mean reward: -4.419557

[DEBUG] Trial 22 | Timestep 100000 | Mean reward: 1852.2437070000003

[DEBUG] Trial 22 | Timestep 120000 | Mean reward: 11.578232

[DEBUG] Trial 22 | Timestep 140000 | Mean reward: 2.155392

[DEBUG] Trial 22 | Timestep 160000 | Mean reward: 12.080201

[DEBUG] Trial 22 | Timestep 180000 | Mean reward: 1933.358329

[DEBUG] Trial 22 | Timestep 200000 | Mean reward: 133.083167

[I 2025-07-20 12:50:44,977] Trial 22 finished with value: 133.083167 and parameters: {'n\_layers': 4, 'layer\_size': 214, 'activation\_fn': 'elu', 'learning\_rate': 0.0009743642831776056, 'buffer\_size': 68442, 'batch\_size': 512, 'gamma': 0.9392997919615342, 'tau': 0.015130569853782, 'train\_freq': 8, 'exploration\_fraction': 0.3878152962251167, 'exploration\_final\_eps': 0.08837556296835872, 'target\_update\_interval': 10000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 46% 23/50 [1:34:53<1:35:36, 212.47s/it, best\_val=1905.82]

[INFO] Starting Trial 23 with Seed: 563129

[DEBUG] Trial 23 | Timestep 20000 | Mean reward: 1785.724785

[DEBUG] Trial 23 | Timestep 40000 | Mean reward: 1787.9197969999998

[PRUNE] Trial 23 pruned at 40000 steps.

[I 2025-07-20 12:51:51,452] Trial 23 pruned.

Tuning DQN: 48% 24/50 [1:35:59<1:13:05, 168.67s/it, best\_val=1905.82]

[INFO] Starting Trial 24 with Seed: 237823

[DEBUG] Trial 24 | Timestep 20000 | Mean reward: 1815.8140969999997

[DEBUG] Trial 24 | Timestep 40000 | Mean reward: 111.357096

[PRUNE] Trial 24 pruned at 40000 steps.

[I 2025-07-20 12:52:47,444] Trial 24 pruned.

Tuning DQN: 50% 25/50 [1:36:55<56:11, 134.86s/it, best\_val=1905.82]

[INFO] Starting Trial 25 with Seed: 802773

[DEBUG] Trial 25 | Timestep 20000 | Mean reward: 17.051001

[DEBUG] Trial 25 | Timestep 40000 | Mean reward: -4.48139

[PRUNE] Trial 25 pruned at 40000 steps.

[I 2025-07-20 12:53:47,233] Trial 25 pruned.

Tuning DQN: 52% 26/50 [1:37:55<44:56, 112.34s/it, best\_val=1905.82]

[INFO] Starting Trial 26 with Seed: 551487

[DEBUG] Trial 26 | Timestep 20000 | Mean reward: 1960.118044

[DEBUG] Trial 26 | Timestep 40000 | Mean reward: 33.518163

[DEBUG] Trial 26 | Timestep 60000 | Mean reward: 55.312553

[DEBUG] Trial 26 | Timestep 80000 | Mean reward: 1831.0198950000001

[DEBUG] Trial 26 | Timestep 100000 | Mean reward: -13.976121

[DEBUG] Trial 26 | Timestep 120000 | Mean reward: -49.999655999999995

[DEBUG] Trial 26 | Timestep 140000 | Mean reward: 1823.604592

[DEBUG] Trial 26 | Timestep 160000 | Mean reward: 1816.342469

[DEBUG] Trial 26 | Timestep 180000 | Mean reward: 36.91871

[DEBUG] Trial 26 | Timestep 200000 | Mean reward: 1828.439525

[I 2025-07-20 13:00:10,016] Trial 26 finished with value: 1828.439525 and parameters: {'n\_layers': 3, 'layer\_size': 211, 'activation\_fn': 'relu', 'learning\_rate': 0.00037184802006546874, 'buffer\_size': 59182, 'batch\_size': 512, 'gamma': 0.9553935545263759, 'tau': 0.013111585646216638, 'train\_freq': 8, 'exploration\_fraction': 0.29328697587592006, 'exploration\_final\_eps': 0.02540832708033338, 'target\_update\_interval': 10000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 54% 27/50 [1:44:18<1:14:09, 193.48s/it, best\_val=1905.82]

[INFO] Starting Trial 27 with Seed: 767847

[DEBUG] Trial 27 | Timestep 20000 | Mean reward: 1800.0390639999998

[DEBUG] Trial 27 | Timestep 40000 | Mean reward: 1805.8160070000001

[PRUNE] Trial 27 pruned at 40000 steps.

[I 2025-07-20 13:01:11,092] Trial 27 pruned.

Tuning DQN: 56% 28/50 [1:45:19<56:22, 153.75s/it, best\_val=1905.82]

[INFO] Starting Trial 28 with Seed: 914128

[DEBUG] Trial 28 | Timestep 20000 | Mean reward: 20.796363

[DEBUG] Trial 28 | Timestep 40000 | Mean reward: -2.840157

[PRUNE] Trial 28 pruned at 40000 steps.

[I 2025-07-20 13:04:19,131] Trial 28 pruned.

Tuning DQN: 58% 29/50 [1:48:27<57:24, 164.04s/it, best\_val=1905.82]

[INFO] Starting Trial 29 with Seed: 627558

[DEBUG] Trial 29 | Timestep 20000 | Mean reward: 15.616725

[DEBUG] Trial 29 | Timestep 40000 | Mean reward: -4.236732

[PRUNE] Trial 29 pruned at 40000 steps.

[I 2025-07-20 13:05:12,414] Trial 29 pruned.

Tuning DQN: 60% 30/50 [1:49:20<43:36, 130.81s/it, best\_val=1905.82]

[INFO] Starting Trial 30 with Seed: 732840

[DEBUG] Trial 30 | Timestep 20000 | Mean reward: -2.8109179999999996

[DEBUG] Trial 30 | Timestep 40000 | Mean reward: 32.875737

[PRUNE] Trial 30 pruned at 40000 steps.

[I 2025-07-20 13:06:14,521] Trial 30 pruned.

Tuning DQN: 62% 31/50 [1:50:22<34:53, 110.20s/it, best\_val=1905.82]

[INFO] Starting Trial 31 with Seed: 391320

[DEBUG] Trial 31 | Timestep 20000 | Mean reward: 1801.515137

[DEBUG] Trial 31 | Timestep 40000 | Mean reward: 1836.207933

[PRUNE] Trial 31 pruned at 40000 steps.

[I 2025-07-20 13:07:17,021] Trial 31 pruned.

Tuning DQN: 64% 32/50 [1:51:25<28:46, 95.89s/it, best\_val=1905.82]

[INFO] Starting Trial 32 with Seed: 515490

[DEBUG] Trial 32 | Timestep 20000 | Mean reward: 1874.6028450000001

[DEBUG] Trial 32 | Timestep 40000 | Mean reward: 1924.308671

[DEBUG] Trial 32 | Timestep 60000 | Mean reward: 1846.254462

[DEBUG] Trial 32 | Timestep 80000 | Mean reward: 1880.26169

[DEBUG] Trial 32 | Timestep 100000 | Mean reward: 35.437828

[DEBUG] Trial 32 | Timestep 120000 | Mean reward: 1855.044783

[DEBUG] Trial 32 | Timestep 140000 | Mean reward: 1828.06462

[DEBUG] Trial 32 | Timestep 160000 | Mean reward: 1888.831347

[DEBUG] Trial 32 | Timestep 180000 | Mean reward: 1867.946301

[DEBUG] Trial 32 | Timestep 200000 | Mean reward: 1814.91197

[I 2025-07-20 13:15:39,357] Trial 32 finished with value: 1814.91197 and parameters: {'n\_layers': 4, 'layer\_size': 256, 'activation\_fn': 'relu', 'learning\_rate': 0.00037831018061640023, 'buffer\_size': 60541, 'batch\_size': 512, 'gamma': 0.9004716322805386, 'tau': 0.010100107758950191, 'train\_freq': 8, 'exploration\_fraction': 0.2760075708883399, 'exploration\_final\_eps': 0.028417796220663517, 'target\_update\_interval': 10000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 66% 33/50 [1:59:47<1:01:43, 217.82s/it, best\_val=1905.82]

[INFO] Starting Trial 33 with Seed: 223282

[DEBUG] Trial 33 | Timestep 20000 | Mean reward: -2.7025850000000005

[DEBUG] Trial 33 | Timestep 40000 | Mean reward: 15.923778

[PRUNE] Trial 33 pruned at 40000 steps.

[I 2025-07-20 13:16:30,735] Trial 33 pruned.

Tuning DQN: 68% 34/50 [2:00:38<44:46, 167.89s/it, best\_val=1905.82]

[INFO] Starting Trial 34 with Seed: 226446

[DEBUG] Trial 34 | Timestep 20000 | Mean reward: 1765.303486

[DEBUG] Trial 34 | Timestep 40000 | Mean reward: 36.151188

[PRUNE] Trial 34 pruned at 40000 steps.

[I 2025-07-20 13:17:14,004] Trial 34 pruned.

Tuning DQN: 70% 35/50 [2:01:22<32:37, 130.50s/it, best\_val=1905.82]

[INFO] Starting Trial 35 with Seed: 891564

[DEBUG] Trial 35 | Timestep 20000 | Mean reward: 1742.555434

[DEBUG] Trial 35 | Timestep 40000 | Mean reward: 7.523934999999999

[PRUNE] Trial 35 pruned at 40000 steps.

[I 2025-07-20 13:18:14,293] Trial 35 pruned.

Tuning DQN: 72% 36/50 [2:02:22<25:32, 109.44s/it, best\_val=1905.82]

[INFO] Starting Trial 36 with Seed: 972777

[DEBUG] Trial 36 | Timestep 20000 | Mean reward: 17.665296

[DEBUG] Trial 36 | Timestep 40000 | Mean reward: 1836.733846

[PRUNE] Trial 36 pruned at 40000 steps.

[I 2025-07-20 13:18:58,976] Trial 36 pruned.

Tuning DQN: 74% 37/50 [2:03:07<19:30, 90.01s/it, best\_val=1905.82]

[INFO] Starting Trial 37 with Seed: 962605

[DEBUG] Trial 37 | Timestep 20000 | Mean reward: 1466.701723

[DEBUG] Trial 37 | Timestep 40000 | Mean reward: -49.699087

[PRUNE] Trial 37 pruned at 40000 steps.

[I 2025-07-20 13:19:36,662] Trial 37 pruned.

Tuning DQN: 76% 38/50 [2:03:44<14:51, 74.31s/it, best\_val=1905.82]

[INFO] Starting Trial 38 with Seed: 107063

[DEBUG] Trial 38 | Timestep 20000 | Mean reward: 1914.897121

[DEBUG] Trial 38 | Timestep 40000 | Mean reward: 1777.140676

[DEBUG] Trial 38 | Timestep 60000 | Mean reward: -10.392759

[DEBUG] Trial 38 | Timestep 80000 | Mean reward: 1804.682374

[DEBUG] Trial 38 | Timestep 100000 | Mean reward: 1832.139212

[DEBUG] Trial 38 | Timestep 120000 | Mean reward: 1831.107051

[DEBUG] Trial 38 | Timestep 140000 | Mean reward: 1951.439308

[DEBUG] Trial 38 | Timestep 160000 | Mean reward: 1825.661033

[DEBUG] Trial 38 | Timestep 180000 | Mean reward: 1853.852944

[DEBUG] Trial 38 | Timestep 200000 | Mean reward: 1891.700931

[I 2025-07-20 13:27:29,075] Trial 38 finished with value: 1891.700931 and parameters: {'n\_layers': 4, 'layer\_size': 183, 'activation\_fn': 'relu', 'learning\_rate': 0.00012345441510697755, 'buffer\_size': 68766, 'batch\_size': 64, 'gamma': 0.9342508639853551, 'tau': 0.017884532464032227, 'train\_freq': 4, 'exploration\_fraction': 0.2551124425463288, 'exploration\_final\_eps': 0.05583921957368507, 'target\_update\_interval': 1000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 78% 39/50 [2:11:37<35:31, 193.74s/it, best\_val=1905.82]

[INFO] Starting Trial 39 with Seed: 854546

[DEBUG] Trial 39 | Timestep 20000 | Mean reward: 34.211791

[DEBUG] Trial 39 | Timestep 40000 | Mean reward: 1818.771836

[PRUNE] Trial 39 pruned at 40000 steps.

[I 2025-07-20 13:28:40,363] Trial 39 pruned.

Tuning DQN: 80% 40/50 [2:12:48<26:10, 157.01s/it, best\_val=1905.82]

[INFO] Starting Trial 40 with Seed: 93144

[DEBUG] Trial 40 | Timestep 20000 | Mean reward: 1820.836015

[DEBUG] Trial 40 | Timestep 40000 | Mean reward: 1832.94221

[PRUNE] Trial 40 pruned at 40000 steps.

[I 2025-07-20 13:29:52,452] Trial 40 pruned.

Tuning DQN: 82% 41/50 [2:14:00<19:43, 131.53s/it, best\_val=1905.82]

[INFO] Starting Trial 41 with Seed: 803660

[DEBUG] Trial 41 | Timestep 20000 | Mean reward: -1.175183

[DEBUG] Trial 41 | Timestep 40000 | Mean reward: 1712.433285

[PRUNE] Trial 41 pruned at 40000 steps.

[I 2025-07-20 13:31:13,768] Trial 41 pruned.

Tuning DQN: 84% 42/50 [2:15:21<15:31, 116.47s/it, best\_val=1905.82]

[INFO] Starting Trial 42 with Seed: 642686

[DEBUG] Trial 42 | Timestep 20000 | Mean reward: 1868.008688

[DEBUG] Trial 42 | Timestep 40000 | Mean reward: 1808.07997

[DEBUG] Trial 42 | Timestep 60000 | Mean reward: 1904.849835

[DEBUG] Trial 42 | Timestep 80000 | Mean reward: 1809.5400710000001

[DEBUG] Trial 42 | Timestep 100000 | Mean reward: 12.259859999999998

[DEBUG] Trial 42 | Timestep 120000 | Mean reward: 1791.817021

[DEBUG] Trial 42 | Timestep 140000 | Mean reward: 1863.877606

[DEBUG] Trial 42 | Timestep 160000 | Mean reward: 1829.7541040000003

[DEBUG] Trial 42 | Timestep 180000 | Mean reward: 1845.9223009999998

[DEBUG] Trial 42 | Timestep 200000 | Mean reward: 1896.239122

[I 2025-07-20 13:39:47,918] Trial 42 finished with value: 1896.239122 and parameters: {'n\_layers': 4, 'layer\_size': 228, 'activation\_fn': 'relu', 'learning\_rate': 0.00010691598322735156, 'buffer\_size': 53677, 'batch\_size': 64, 'gamma': 0.9913405241136456, 'tau': 0.015163767637068518, 'train\_freq': 4, 'exploration\_fraction': 0.32589993566195463, 'exploration\_final\_eps': 0.035845982119372136, 'target\_update\_interval': 1000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 86% 43/50 [2:23:56<27:30, 235.77s/it, best\_val=1905.82]

[INFO] Starting Trial 43 with Seed: 793710

[DEBUG] Trial 43 | Timestep 20000 | Mean reward: 1798.0512069999997

[DEBUG] Trial 43 | Timestep 40000 | Mean reward: 1853.96774

[DEBUG] Trial 43 | Timestep 60000 | Mean reward: 1819.116489

[DEBUG] Trial 43 | Timestep 80000 | Mean reward: 1834.925332

[DEBUG] Trial 43 | Timestep 100000 | Mean reward: 1746.590562

[DEBUG] Trial 43 | Timestep 120000 | Mean reward: 1915.434348

[DEBUG] Trial 43 | Timestep 140000 | Mean reward: 1863.098736

[DEBUG] Trial 43 | Timestep 160000 | Mean reward: 1863.1628119999998

[DEBUG] Trial 43 | Timestep 180000 | Mean reward: 1862.38566

[DEBUG] Trial 43 | Timestep 200000 | Mean reward: 1816.165318

[I 2025-07-20 13:48:38,836] Trial 43 finished with value: 1816.165318 and parameters: {'n\_layers': 4, 'layer\_size': 238, 'activation\_fn': 'relu', 'learning\_rate': 0.00010827837830981889, 'buffer\_size': 75054, 'batch\_size': 64, 'gamma': 0.9922243367789981, 'tau': 0.015059768417929964, 'train\_freq': 4, 'exploration\_fraction': 0.3161261865849684, 'exploration\_final\_eps': 0.04182802898018915, 'target\_update\_interval': 1000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 88% 44/50 [2:32:47<32:25, 324.32s/it, best\_val=1905.82]

[INFO] Starting Trial 44 with Seed: 289955

[DEBUG] Trial 44 | Timestep 20000 | Mean reward: 1899.424717

[DEBUG] Trial 44 | Timestep 40000 | Mean reward: 1819.3201180000003

[DEBUG] Trial 44 | Timestep 60000 | Mean reward: 1881.532672

[DEBUG] Trial 44 | Timestep 80000 | Mean reward: 1808.425502

[DEBUG] Trial 44 | Timestep 100000 | Mean reward: 1829.359924

[DEBUG] Trial 44 | Timestep 120000 | Mean reward: 1813.403029

[DEBUG] Trial 44 | Timestep 140000 | Mean reward: 1812.236651

[DEBUG] Trial 44 | Timestep 160000 | Mean reward: 1783.6583280000002

[DEBUG] Trial 44 | Timestep 180000 | Mean reward: 1828.0089070000001

[DEBUG] Trial 44 | Timestep 200000 | Mean reward: -1.890392

[I 2025-07-20 13:56:36,205] Trial 44 finished with value: -1.890392 and parameters: {'n\_layers': 4, 'layer\_size': 186, 'activation\_fn': 'relu', 'learning\_rate': 0.00010322752906620033, 'buffer\_size': 50217, 'batch\_size': 64, 'gamma': 0.9941137160774899, 'tau': 0.016621708681077766, 'train\_freq': 4, 'exploration\_fraction': 0.3498704676530086, 'exploration\_final\_eps': 0.036690215145456354, 'target\_update\_interval': 1000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 90% 45/50 [2:40:44<30:51, 370.23s/it, best\_val=1905.82]

[INFO] Starting Trial 45 with Seed: 816283

[DEBUG] Trial 45 | Timestep 20000 | Mean reward: 46.553396

[DEBUG] Trial 45 | Timestep 40000 | Mean reward: 1816.2595950000002

[PRUNE] Trial 45 pruned at 40000 steps.

[I 2025-07-20 13:58:00,476] Trial 45 pruned.

Tuning DQN: 92% 46/50 [2:42:08<18:57, 284.44s/it, best\_val=1905.82]

[INFO] Starting Trial 46 with Seed: 799565

[DEBUG] Trial 46 | Timestep 20000 | Mean reward: 1905.829461

[DEBUG] Trial 46 | Timestep 40000 | Mean reward: 1918.906215

[DEBUG] Trial 46 | Timestep 60000 | Mean reward: 1752.596372

[DEBUG] Trial 46 | Timestep 80000 | Mean reward: 1806.698274

[DEBUG] Trial 46 | Timestep 100000 | Mean reward: -11.569949

[DEBUG] Trial 46 | Timestep 120000 | Mean reward: 1792.233211

[DEBUG] Trial 46 | Timestep 140000 | Mean reward: 1859.7422450000001

[DEBUG] Trial 46 | Timestep 160000 | Mean reward: 1562.20378

[DEBUG] Trial 46 | Timestep 180000 | Mean reward: 1933.3133289999998

[DEBUG] Trial 46 | Timestep 200000 | Mean reward: 1789.828805

[I 2025-07-20 14:05:59,384] Trial 46 finished with value: 1789.828805 and parameters: {'n\_layers': 4, 'layer\_size': 167, 'activation\_fn': 'relu', 'learning\_rate': 0.00016242687427784215, 'buffer\_size': 198958, 'batch\_size': 64, 'gamma': 0.998448480184605, 'tau': 0.014719017073797296, 'train\_freq': 4, 'exploration\_fraction': 0.3715326057171486, 'exploration\_final\_eps': 0.06747452155794706, 'target\_update\_interval': 1000}. Best is trial 3 with value: 1905.820014.

Tuning DQN: 94% 47/50 [2:50:07<17:08, 342.78s/it, best\_val=1905.82]

[INFO] Starting Trial 47 with Seed: 914204

[DEBUG] Trial 47 | Timestep 20000 | Mean reward: 1831.3783570000003

[DEBUG] Trial 47 | Timestep 40000 | Mean reward: 1748.2737879999997

[PRUNE] Trial 47 pruned at 40000 steps.

[I 2025-07-20 14:07:20,183] Trial 47 pruned.

Tuning DQN: 96% 48/50 [2:51:28<08:48, 264.19s/it, best\_val=1905.82]

[INFO] Starting Trial 48 with Seed: 56935

[DEBUG] Trial 48 | Timestep 20000 | Mean reward: -4.666863

[DEBUG] Trial 48 | Timestep 40000 | Mean reward: 11.386694999999998

[PRUNE] Trial 48 pruned at 40000 steps.

[I 2025-07-20 14:07:57,479] Trial 48 pruned.

Tuning DQN: 98% 49/50 [2:52:05<03:16, 196.12s/it, best\_val=1905.82]

[INFO] Starting Trial 49 with Seed: 768783

[DEBUG] Trial 49 | Timestep 20000 | Mean reward: 314.81689

[DEBUG] Trial 49 | Timestep 40000 | Mean reward: 402.46838

[PRUNE] Trial 49 pruned at 40000 steps.

[I 2025-07-20 14:08:48,488] Trial 49 pruned.

Tuning DQN: 100% 50/50 [2:52:56<00:00, 207.53s/it, best\_val=1905.82]

================================================================================

All trial data saved to: hyperparameter\_results\_final/DQN\_best\_params\_seeded.json

--- INDIVIDUAL TRIAL RESULTS (BEST TO WORST) ---

================================================================================

--- Rank 1: Trial #3 (Seed: 911974) ---

Value (Mean Reward): 1905.8200

Params:

n\_layers : 4

layer\_size : 171

activation\_fn : relu

learning\_rate : 0.0009706174983140924

buffer\_size : 52764

batch\_size : 256

gamma : 0.9921260140130664

tau : 0.016569905149490673

train\_freq : 8

exploration\_fraction: 0.3630728045562289

exploration\_final\_eps: 0.08594960313425784

target\_update\_interval: 5000

--- Rank 2: Trial #42 (Seed: 642686) ---

Value (Mean Reward): 1896.2391

Params:

n\_layers : 4

layer\_size : 228

activation\_fn : relu

learning\_rate : 0.00010691598322735156

buffer\_size : 53677

batch\_size : 64

gamma : 0.9913405241136456

tau : 0.015163767637068518

train\_freq : 4

exploration\_fraction: 0.32589993566195463

exploration\_final\_eps: 0.035845982119372136

target\_update\_interval: 1000

--- Rank 3: Trial #38 (Seed: 107063) ---

Value (Mean Reward): 1891.7009

Params:

n\_layers : 4

layer\_size : 183

activation\_fn : relu

learning\_rate : 0.00012345441510697755

buffer\_size : 68766

batch\_size : 64

gamma : 0.9342508639853551

tau : 0.017884532464032227

train\_freq : 4

exploration\_fraction: 0.2551124425463288

exploration\_final\_eps: 0.05583921957368507

target\_update\_interval: 1000

--- Rank 4: Trial #1 (Seed: 101352) ---

Value (Mean Reward): 1867.3710

Params:

n\_layers : 2

layer\_size : 202

activation\_fn : relu

learning\_rate : 0.00015985037708422188

buffer\_size : 195160

batch\_size : 512

gamma : 0.9356083356418619

tau : 0.017599594231949168

train\_freq : 8

exploration\_fraction: 0.1644792410463663

exploration\_final\_eps: 0.06175988268766712

target\_update\_interval: 10000

--- Rank 5: Trial #15 (Seed: 328240) ---

Value (Mean Reward): 1844.1164

Params:

n\_layers : 4

layer\_size : 155

activation\_fn : elu

learning\_rate : 0.0009854339140397499

buffer\_size : 74218

batch\_size : 512

gamma : 0.9296590406357953

tau : 0.013867718324097028

train\_freq : 8

exploration\_fraction: 0.2851362882799871

exploration\_final\_eps: 0.09870654352328935

target\_update\_interval: 10000

--- Rank 6: Trial #26 (Seed: 551487) ---

Value (Mean Reward): 1828.4395

Params:

n\_layers : 3

layer\_size : 211

activation\_fn : relu

learning\_rate : 0.00037184802006546874

buffer\_size : 59182

batch\_size : 512

gamma : 0.9553935545263759

tau : 0.013111585646216638

train\_freq : 8

exploration\_fraction: 0.29328697587592006

exploration\_final\_eps: 0.02540832708033338

target\_update\_interval: 10000

--- Rank 7: Trial #43 (Seed: 793710) ---

Value (Mean Reward): 1816.1653

Params:

n\_layers : 4

layer\_size : 238

activation\_fn : relu

learning\_rate : 0.00010827837830981889

buffer\_size : 75054

batch\_size : 64

gamma : 0.9922243367789981

tau : 0.015059768417929964

train\_freq : 4

exploration\_fraction: 0.3161261865849684

exploration\_final\_eps: 0.04182802898018915

target\_update\_interval: 1000

--- Rank 8: Trial #32 (Seed: 515490) ---

Value (Mean Reward): 1814.9120

Params:

n\_layers : 4

layer\_size : 256

activation\_fn : relu

learning\_rate : 0.00037831018061640023

buffer\_size : 60541

batch\_size : 512

gamma : 0.9004716322805386

tau : 0.010100107758950191

train\_freq : 8

exploration\_fraction: 0.2760075708883399

exploration\_final\_eps: 0.028417796220663517

target\_update\_interval: 10000

--- Rank 9: Trial #12 (Seed: 963151) ---

Value (Mean Reward): 1800.0077

Params:

n\_layers : 3

layer\_size : 102

activation\_fn : relu

learning\_rate : 0.000451720038604597

buffer\_size : 151277

batch\_size : 512

gamma : 0.9305064594710305

tau : 0.016309052437142132

train\_freq : 1

exploration\_fraction: 0.4653722252625217

exploration\_final\_eps: 0.0717375694781826

target\_update\_interval: 500

--- Rank 10: Trial #46 (Seed: 799565) ---

Value (Mean Reward): 1789.8288

Params:

n\_layers : 4

layer\_size : 167

activation\_fn : relu

learning\_rate : 0.00016242687427784215

buffer\_size : 198958

batch\_size : 64

gamma : 0.998448480184605

tau : 0.014719017073797296

train\_freq : 4

exploration\_fraction: 0.3715326057171486

exploration\_final\_eps: 0.06747452155794706

target\_update\_interval: 1000

--- Rank 11: Trial #2 (Seed: 493962) ---

Value (Mean Reward): 1779.7171

Params:

n\_layers : 4

layer\_size : 250

activation\_fn : tanh

learning\_rate : 0.0008014622189583533

buffer\_size : 93114

batch\_size : 64

gamma : 0.9063321303494731

tau : 0.0163239110606003

train\_freq : 16

exploration\_fraction: 0.3887643004825859

exploration\_final\_eps: 0.043385899911248325

target\_update\_interval: 2000

--- Rank 12: Trial #0 (Seed: 300525) ---

Value (Mean Reward): 1777.1933

Params:

n\_layers : 2

layer\_size : 73

activation\_fn : tanh

learning\_rate : 0.00014154486471907698

buffer\_size : 53748

batch\_size : 32

gamma : 0.9155069008402772

tau : 0.008882048075020464

train\_freq : 4

exploration\_fraction: 0.19886522152830693

exploration\_final\_eps: 0.04036512226290992

target\_update\_interval: 500

--- Rank 13: Trial #10 (Seed: 141875) ---

Value (Mean Reward): 1588.2308

Params:

n\_layers : 4

layer\_size : 134

activation\_fn : relu

learning\_rate : 0.0004609558780349344

buffer\_size : 50945

batch\_size : 256

gamma : 0.9989229763881984

tau : 0.014638847240524822

train\_freq : 8

exploration\_fraction: 0.4831237330130315

exploration\_final\_eps: 0.011153537349834361

target\_update\_interval: 5000

--- Rank 14: Trial #7 (Seed: 494401) ---

Value (Mean Reward): 546.0618

Params:

n\_layers : 2

layer\_size : 235

activation\_fn : leaky\_relu

learning\_rate : 0.000114416142070478

buffer\_size : 99881

batch\_size : 64

gamma : 0.9664459143877335

tau : 0.007725971307820809

train\_freq : 4

exploration\_fraction: 0.1246306217904643

exploration\_final\_eps: 0.05564513871360352

target\_update\_interval: 10000

--- Rank 15: Trial #14 (Seed: 331081) ---

Value (Mean Reward): 139.0690

Params:

n\_layers : 3

layer\_size : 113

activation\_fn : leaky\_relu

learning\_rate : 0.0002814173470297955

buffer\_size : 106932

batch\_size : 128

gamma : 0.9841207873595

tau : 0.017530530335784836

train\_freq : 8

exploration\_fraction: 0.4256749457533978

exploration\_final\_eps: 0.09322070168791168

target\_update\_interval: 5000

--- Rank 16: Trial #22 (Seed: 899753) ---

Value (Mean Reward): 133.0832

Params:

n\_layers : 4

layer\_size : 214

activation\_fn : elu

learning\_rate : 0.0009743642831776056

buffer\_size : 68442

batch\_size : 512

gamma : 0.9392997919615342

tau : 0.015130569853782

train\_freq : 8

exploration\_fraction: 0.3878152962251167

exploration\_final\_eps: 0.08837556296835872

target\_update\_interval: 10000

--- Rank 17: Trial #11 (Seed: 819030) ---

Value (Mean Reward): 10.3479

Params:

n\_layers : 3

layer\_size : 151

activation\_fn : relu

learning\_rate : 0.0004800298638772757

buffer\_size : 88025

batch\_size : 512

gamma : 0.9457378004702556

tau : 0.019924212030622085

train\_freq : 8

exploration\_fraction: 0.13927908764240743

exploration\_final\_eps: 0.07035831812348572

target\_update\_interval: 5000

--- Rank 18: Trial #44 (Seed: 289955) ---

Value (Mean Reward): -1.8904

Params:

n\_layers : 4

layer\_size : 186

activation\_fn : relu

learning\_rate : 0.00010322752906620033

buffer\_size : 50217

batch\_size : 64

gamma : 0.9941137160774899

tau : 0.016621708681077766

train\_freq : 4

exploration\_fraction: 0.3498704676530086

exploration\_final\_eps: 0.036690215145456354

target\_update\_interval: 1000

--- Rank 19: Trial #4 (Seed: 102197) ---

Value (Mean Reward): -4.1825

Params:

n\_layers : 2

layer\_size : 189

activation\_fn : tanh

learning\_rate : 0.000704722702682615

buffer\_size : 174219

batch\_size : 32

gamma : 0.9641638510280942

tau : 0.011553008130519946

train\_freq : 16

exploration\_fraction: 0.2558865221993636

exploration\_final\_eps: 0.03885106270869217

target\_update\_interval: 10000

================================================================================

--- Best parameters for DQN ---

Best value (mean reward): 1905.8200

Best trial was #3 with seed 911974

n\_layers: 4

layer\_size: 171

activation\_fn: relu

learning\_rate: 0.0009706174983140924

buffer\_size: 52764

batch\_size: 256

gamma: 0.9921260140130664

tau: 0.016569905149490673

train\_freq: 8

exploration\_fraction: 0.3630728045562289

exploration\_final\_eps: 0.08594960313425784

target\_update\_interval: 5000