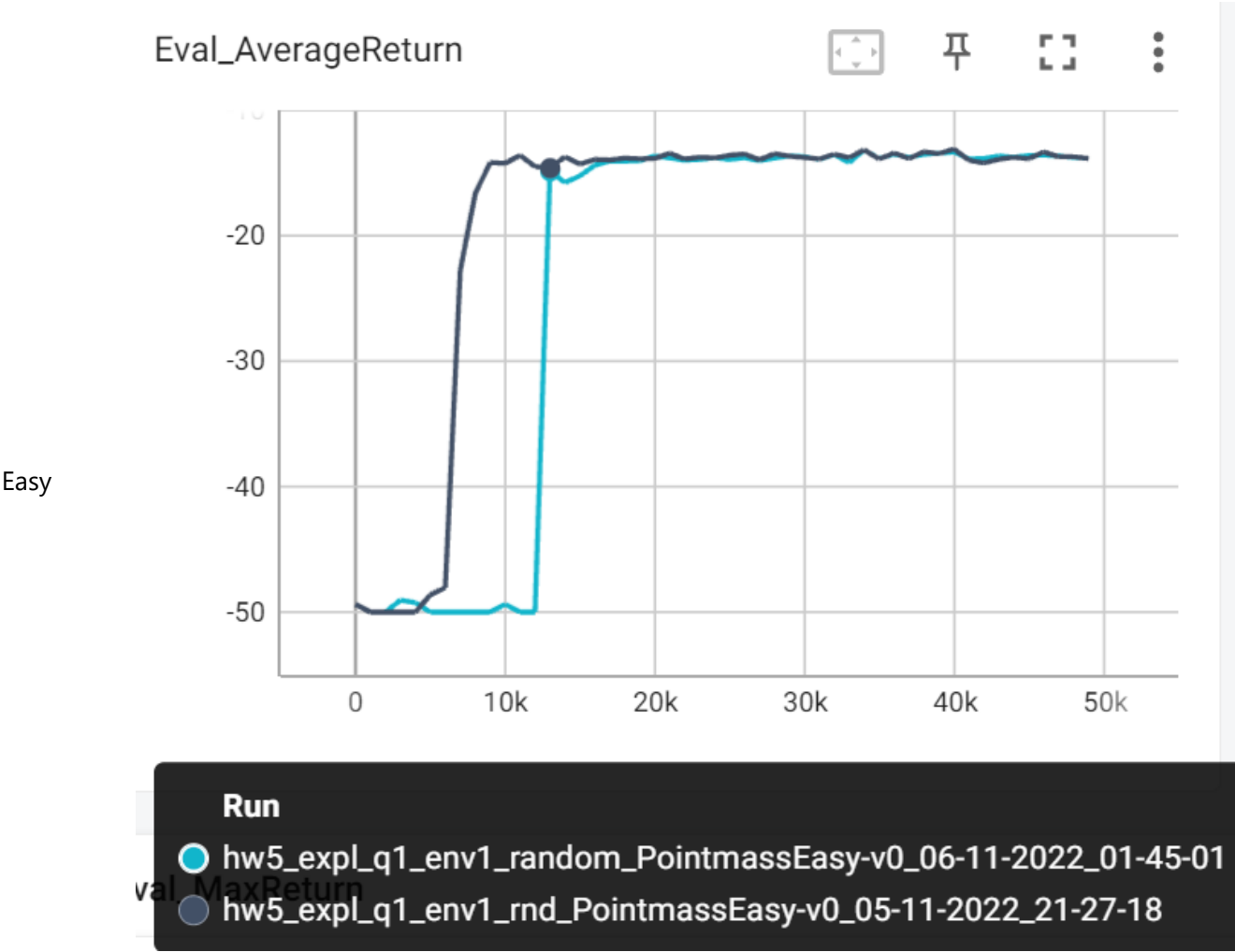


CS285 HW5 Report

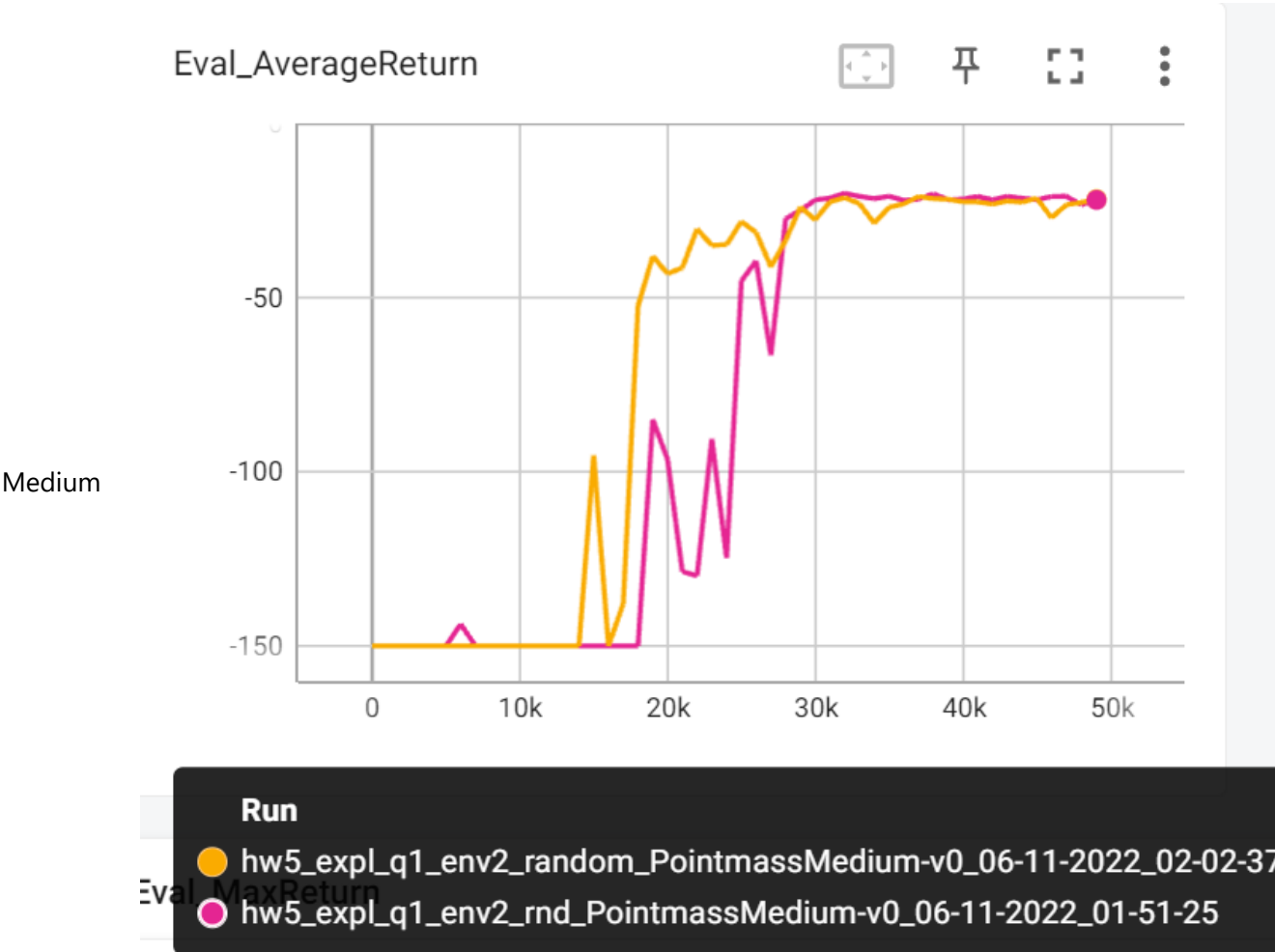
Part 1 "Unsupervised" RND and exploration performance

Performance Compare

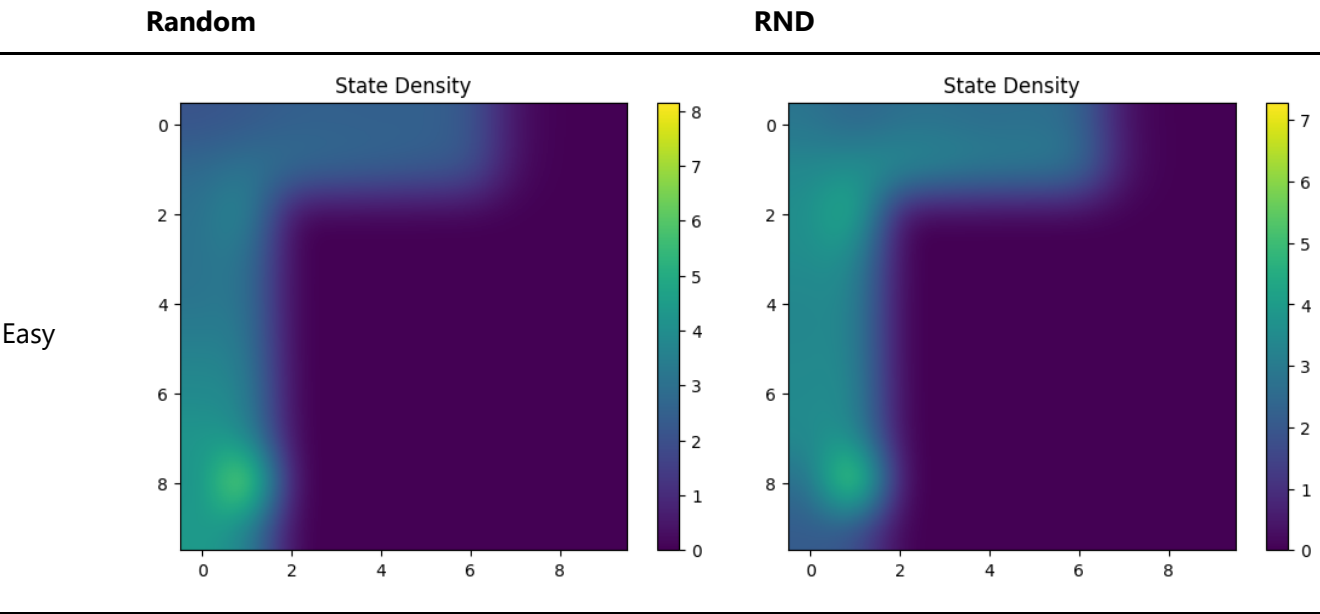
Eval Average

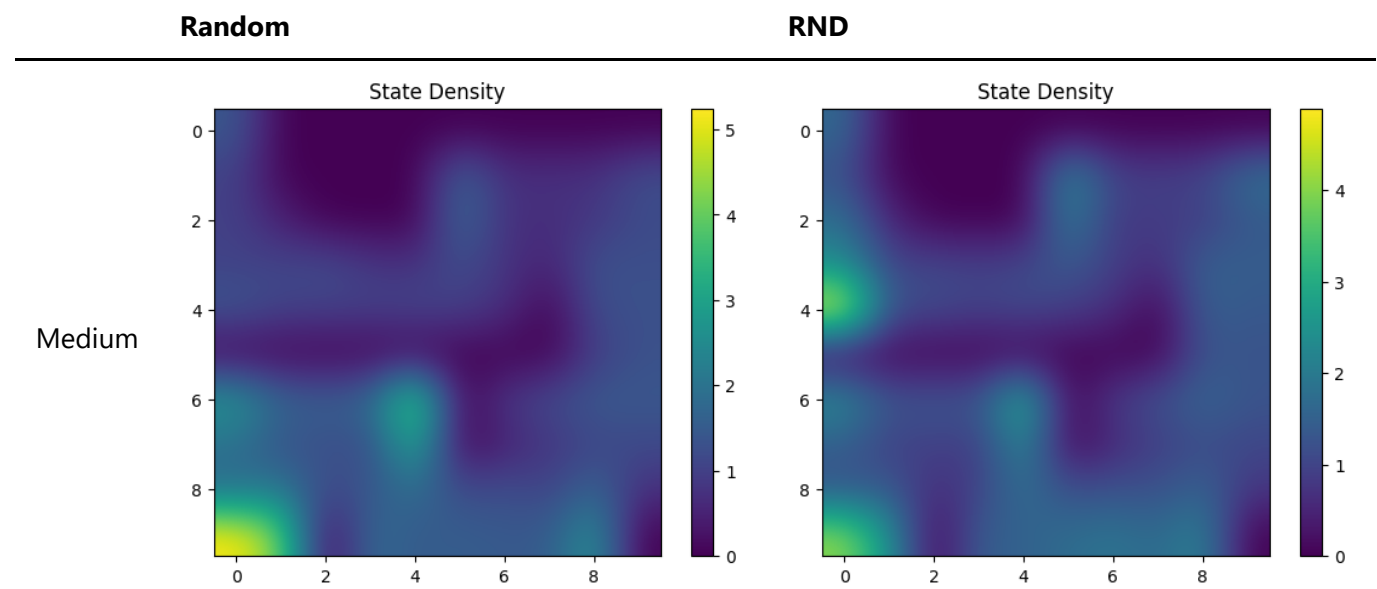


Eval Average

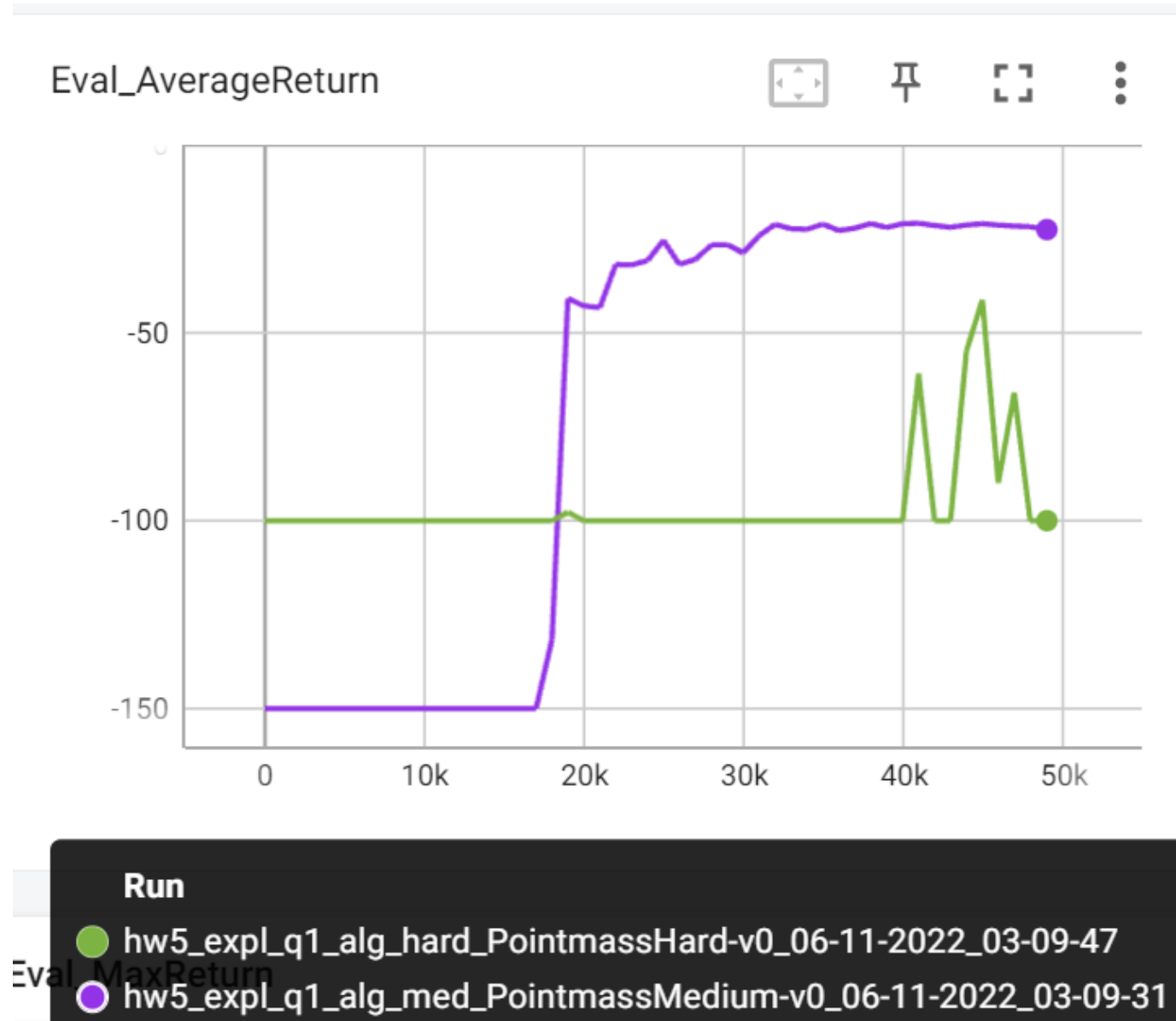


State Density Comparison



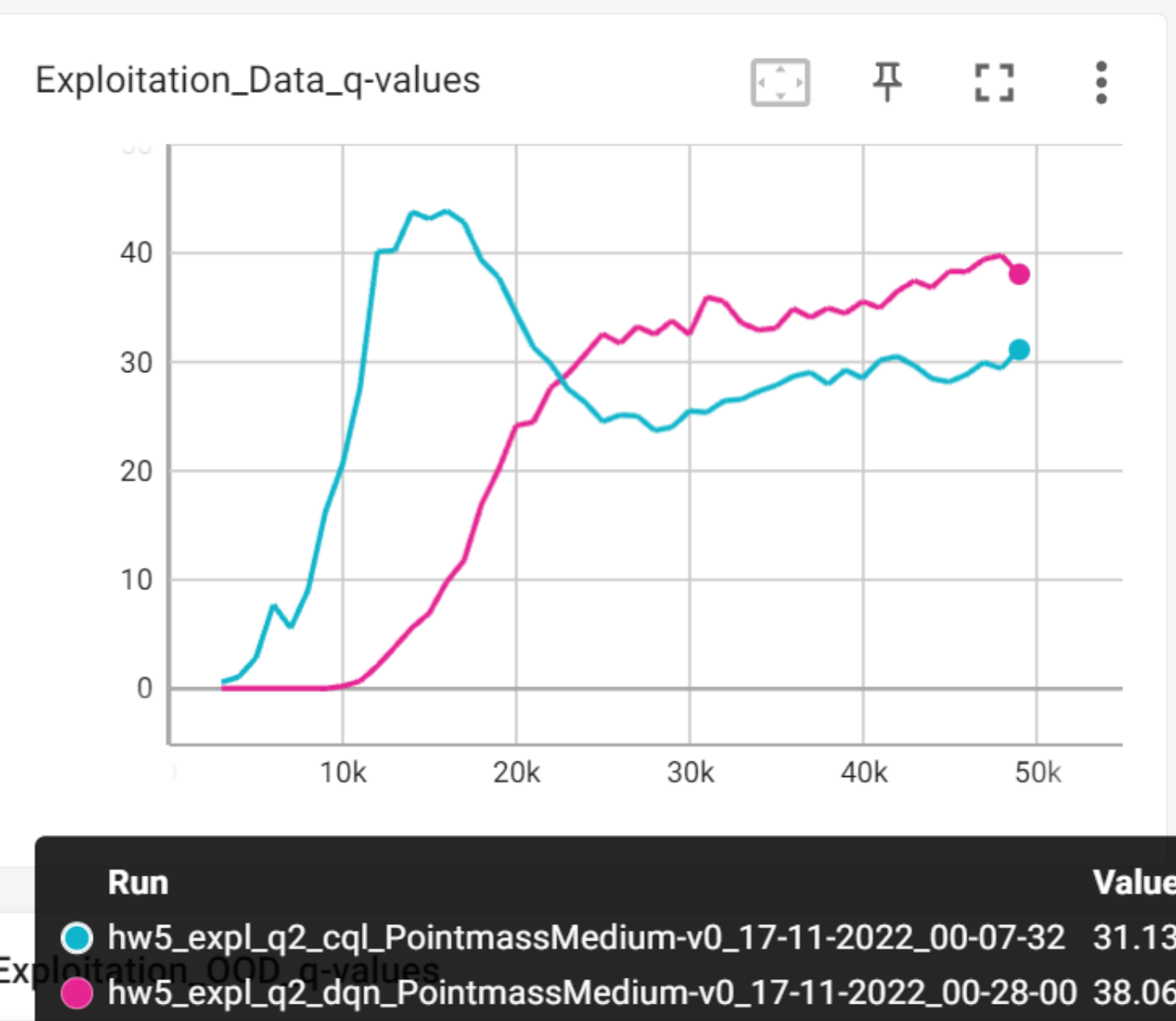


Subpart Custom Exploration: Boltzman

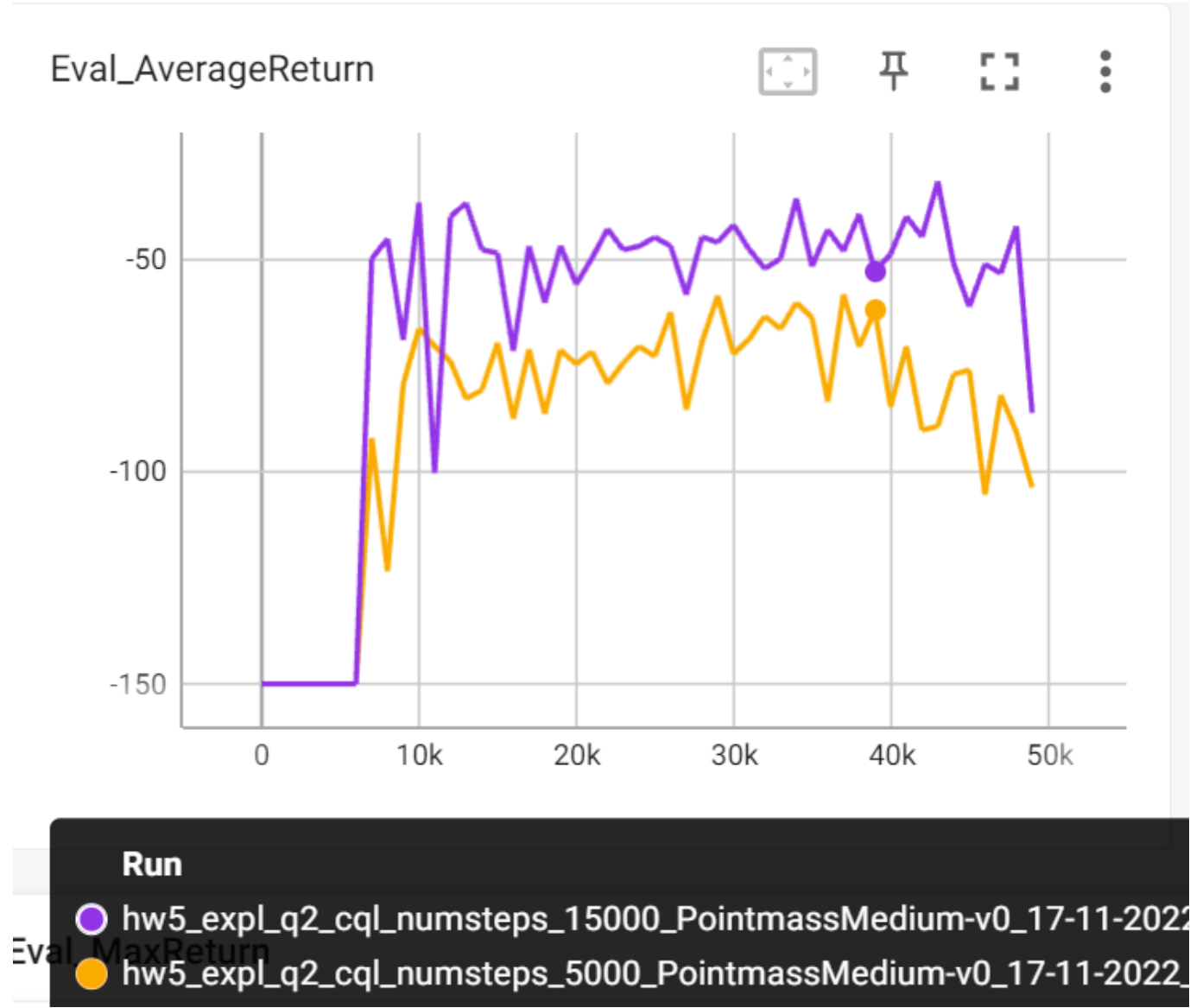


Part 2 Offline learning on exploration data

Subpart 1: Q value comparison

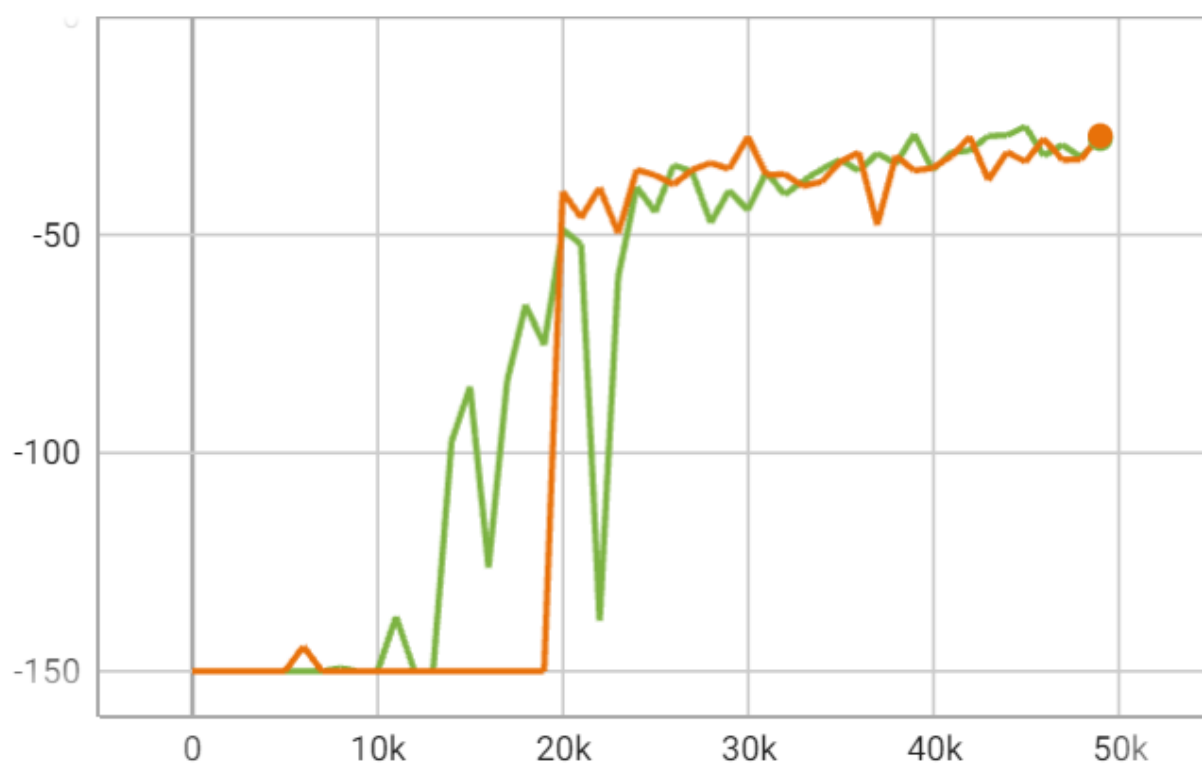


Subpart 2 Numsteps comparison:



Subpart 3: Alpha comparison:

Eval_AverageReturn

**Run**

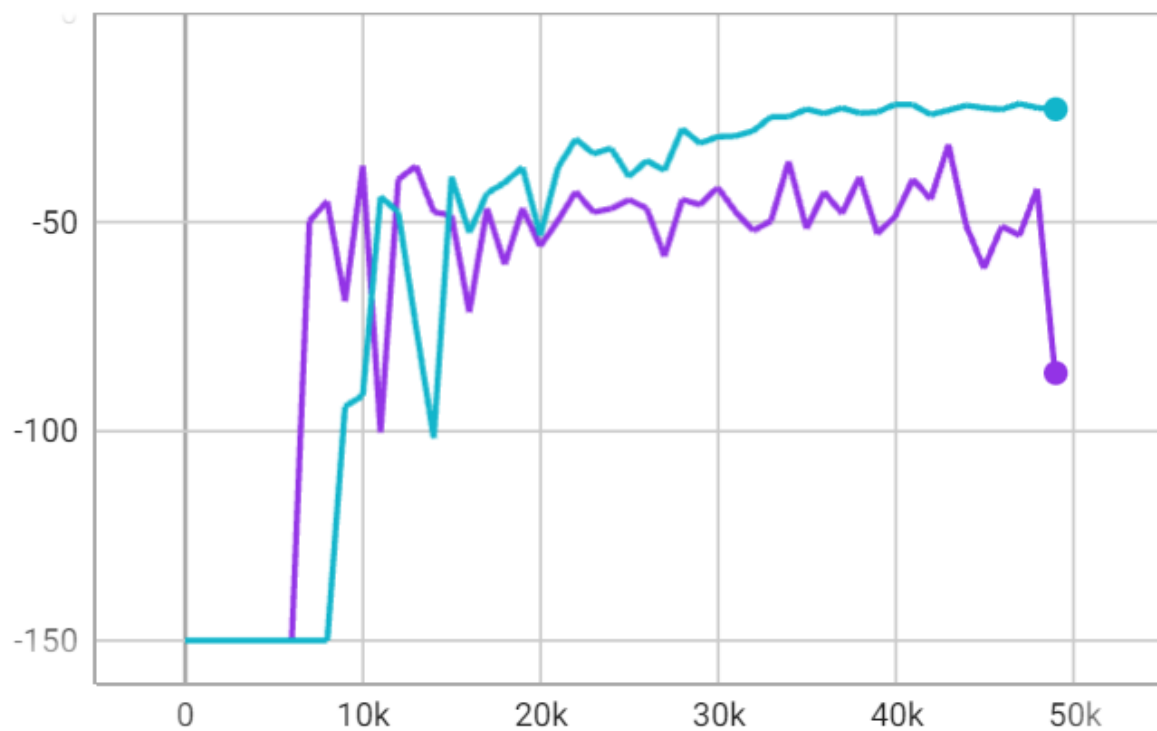
- hw5_expl_q2_alpha_0.2_PointmassMedium-v0_17-11-2022_03-46-58
- hw5_expl_q2_alpha_0.5_PointmassMedium-v0_17-11-2022_03-47-06

Alpha 0.2 performs the best while dqn performs the worst.

Part 3 "Supervised" exploration with mixed reward bonuses.

Compare to Q2(purely offline)

Eval_AverageReturn



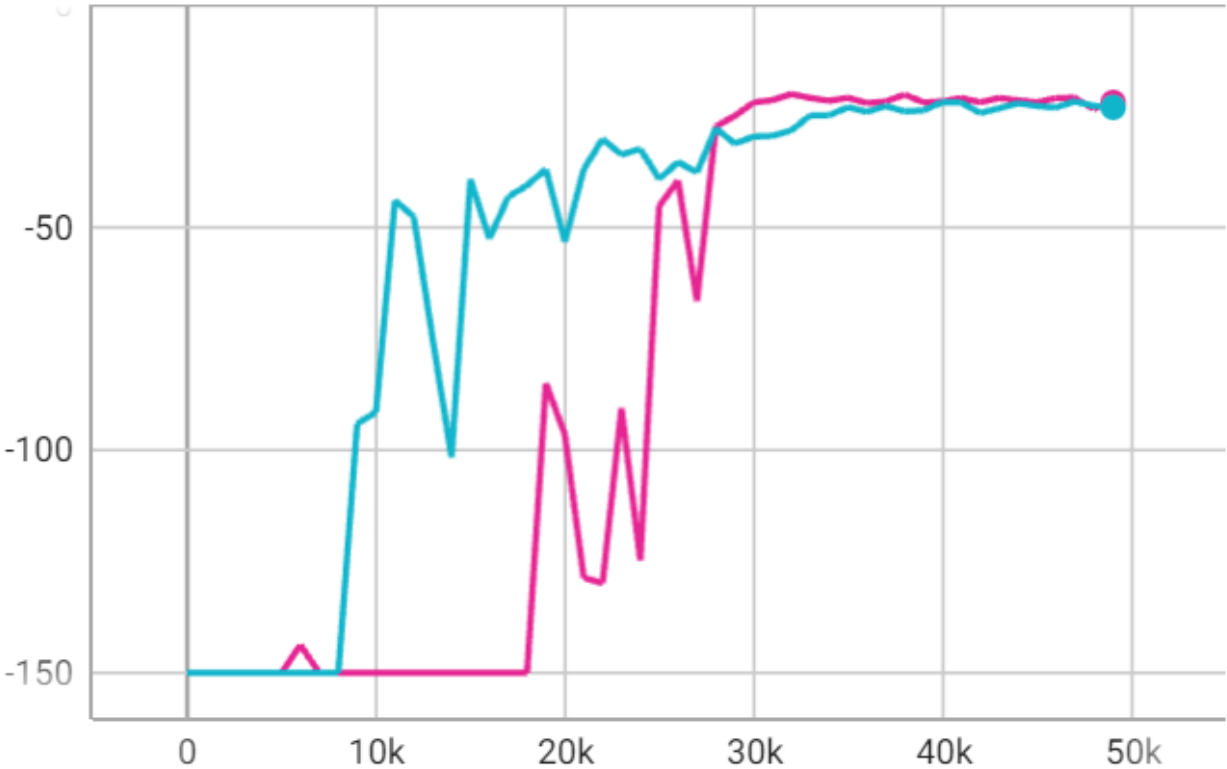
Run

- hw5_expl_q2_cql_numsteps_15000_PointmassMedium-v0_17-11-2022_0
- hw5_expl_q3_medium_cql_PointmassMedium-v0_17-11-2022_13-27-32

Clearly, mixed reward is the winner.

Compare to Q1(rnd with default exploration=10000steps)

Eval_AverageReturn



Run

- hw5_expl_q1_env2_rnd_PointmassMedium-v0_06-11-2022_01-51-25
- hw5_expl_q3_medium_cql_PointmassMedium-v0_17-11-2022_13-27-3

Even though the final result is close, but clearly CQL with mixed reward converges a lot faster than standard RND.

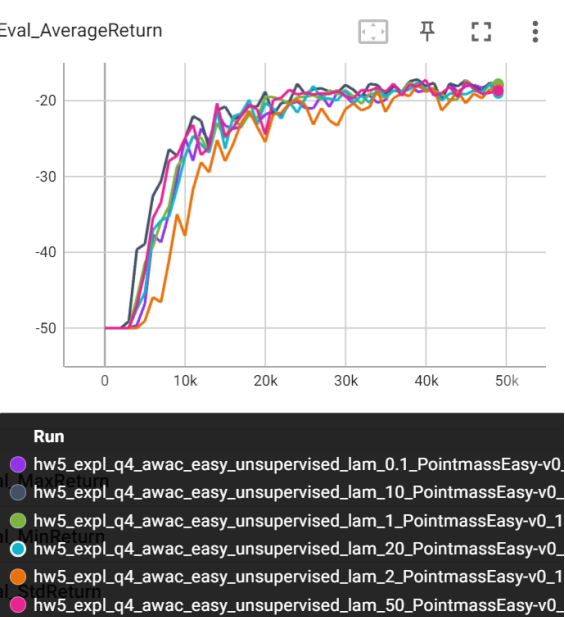
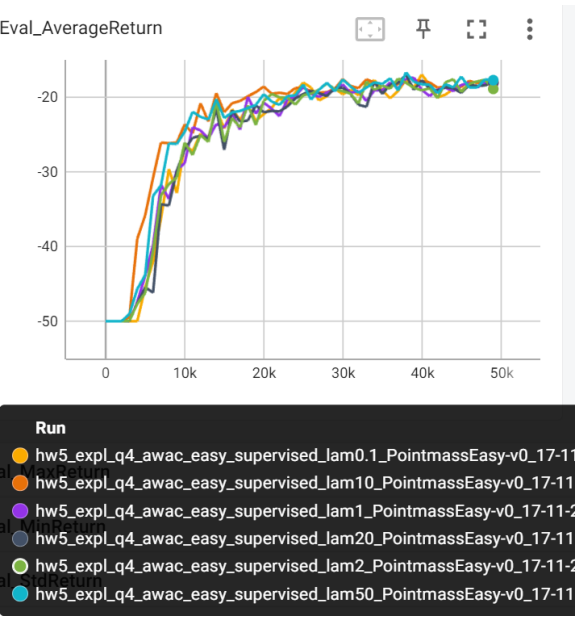
Part 4 Offline Learning with AWAC

Supervised	Unsupervised
------------	--------------

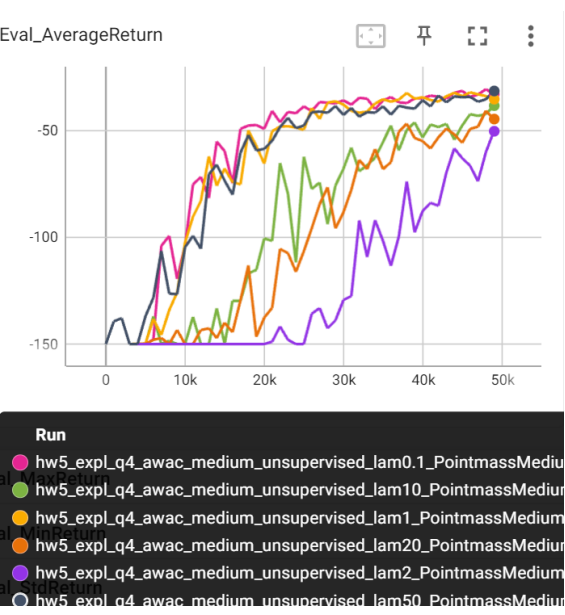
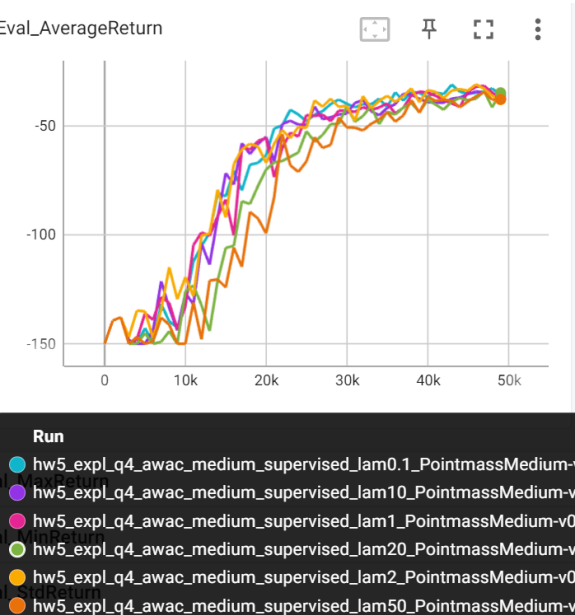
Supervised

Unsupervised

Easy



Medium



Best lambda: Easy-sup(10), Easy-unsup(10), Med-sup(2), Med_unsup(0.1)

Part 5 Offline Learning with IQL

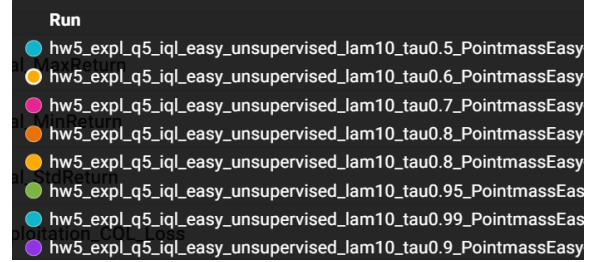
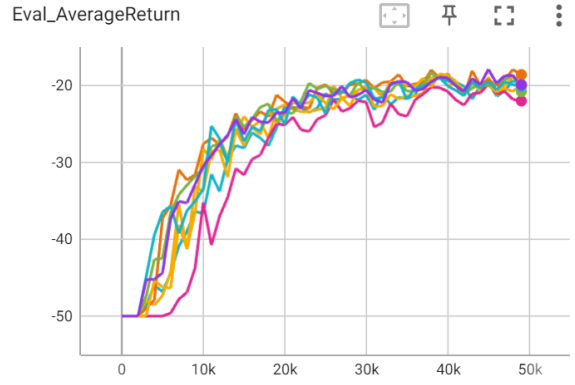
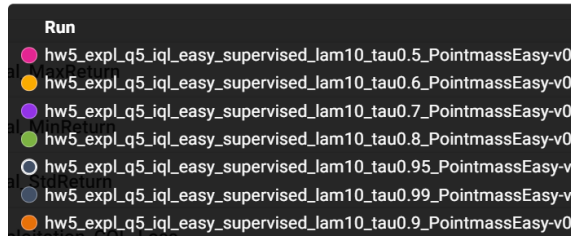
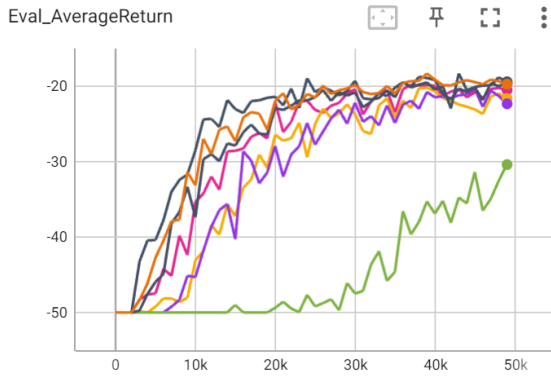
Supervised

Unsupervised

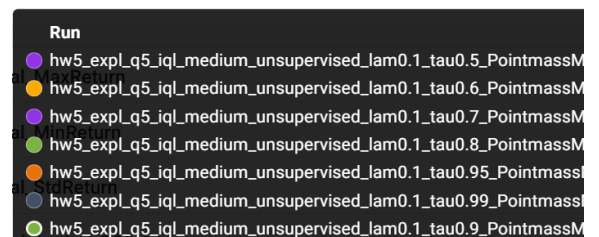
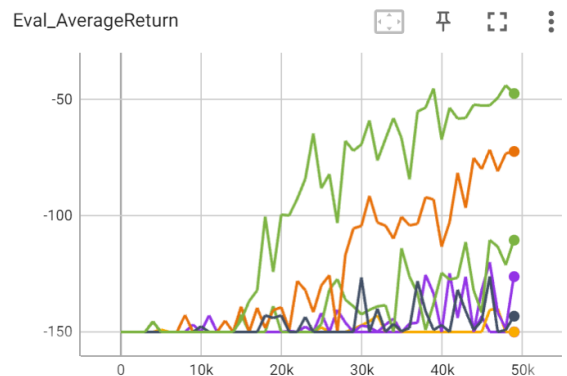
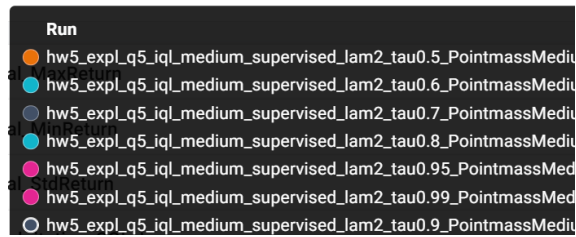
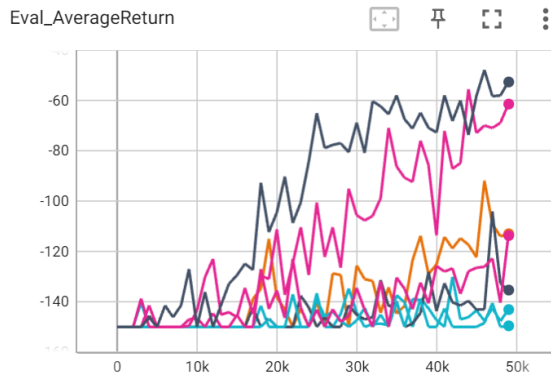
Supervised

Unsupervised

Easy



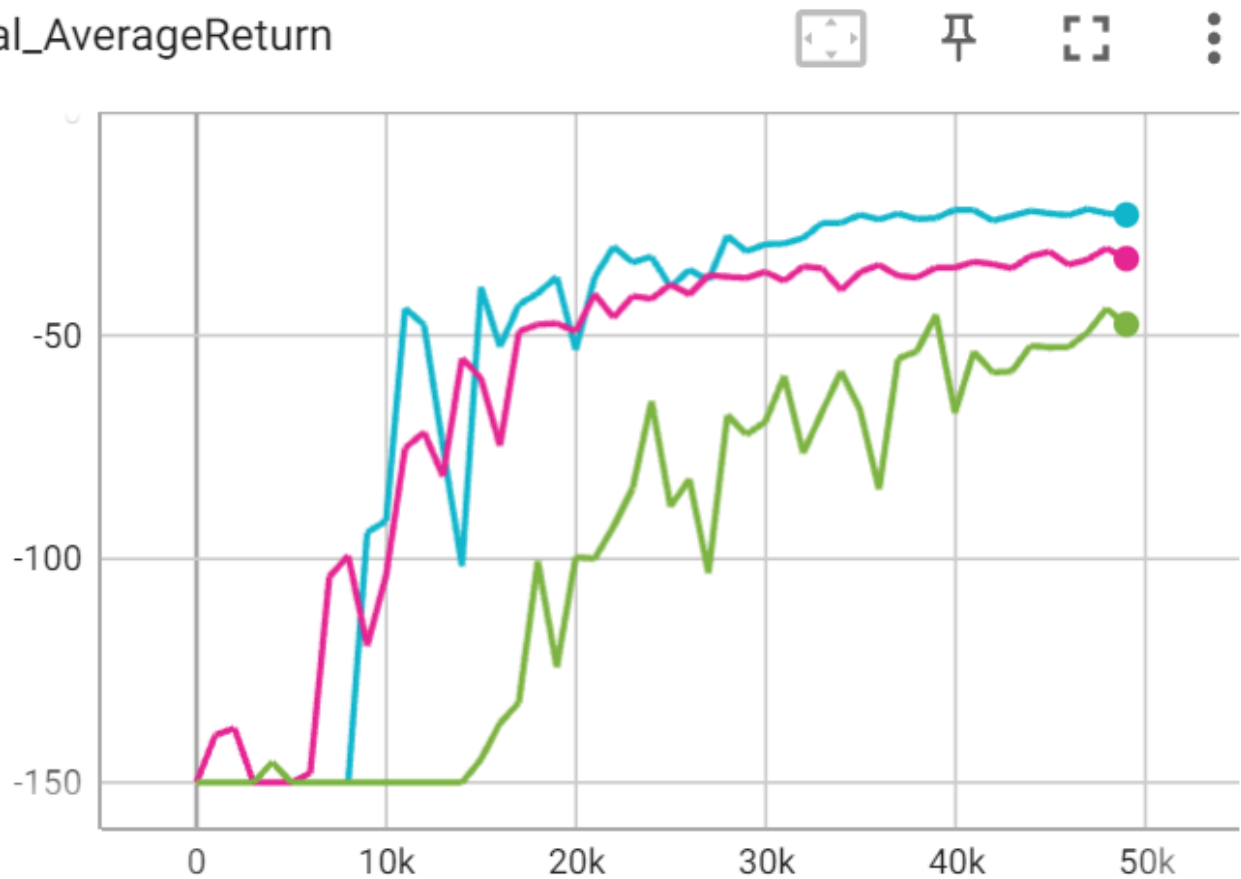
Medium



Best tau: Easy-sup(0.99), Easy-unsup(0.8), Med-sup(0.9), Med_unsup(0.9)

Final compare CQL, AWAC, IQL

Eval_AverageReturn



Run

- hw5_expl_q3_medium_cql_PointmassMedium-v0_17-11-2022_13-27-3
- hw5_expl_q4_awac_medium_unsupervised_lam0.1_PointmassMedium
- hw5_expl_q5_iql_medium_unsupervised_lam0.1_tau0.9_PointmassMe

From the plot, we can see that cql seems to perform the best in the end. AWAC is also really close and converges fast. IQL seems to perform the worst among all.