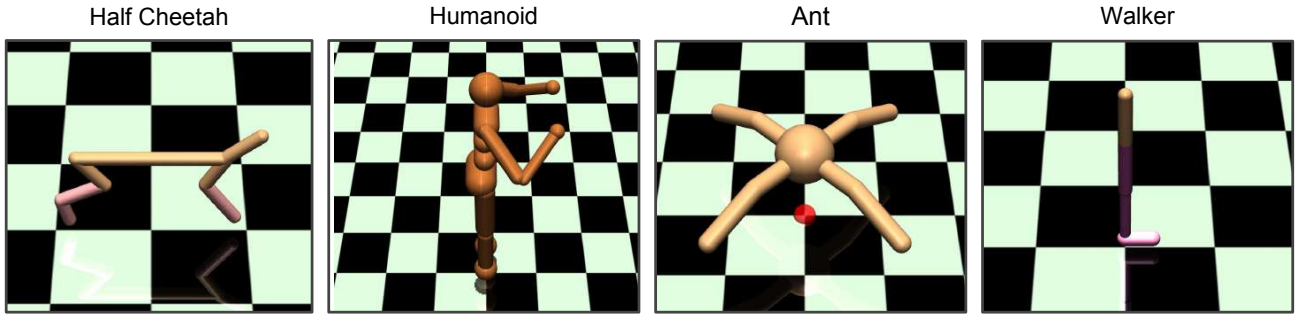

Figure 8. **Continuous control tasks**: left-to-right: the half-cheetah, humanoid, ant, and walker robots used in our evaluation.

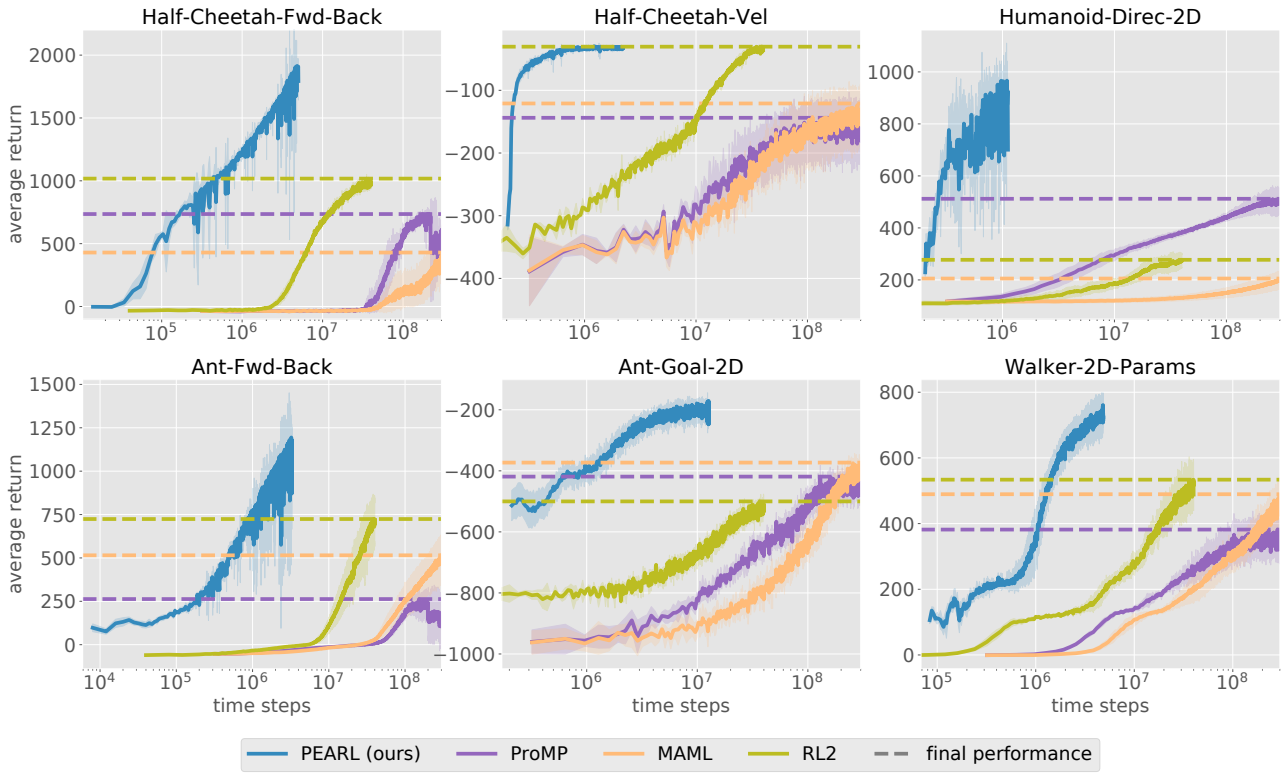

Figure 9. **Meta-learning continuous control**. Test task performance vs. samples collected during *meta-training*. While in the main paper we truncate the x-axis to better illustrate the performance of PEARL, here we plot PEARL against the on-policy methods run for the full number of time steps ($1e8$). Note that the x-axis is in **log scale**. PEARL is 20-100 times more sample efficient.

A. Experimental Details

Here we provide further details for the MuJoCo continuous control domains in Section 6.1. The agents used in these domains are visualized in Figure 8. The horizon in all tasks is 200 steps.

- Half-Cheetah-Dir: move forward and backward (2 tasks)
- Half-Cheetah-Vel: achieve a target velocity running forward (100 train tasks, 30 test tasks)
- Humanoid-Dir-2D: run in a target direction on 2D grid (100 train tasks, 30 test tasks)
- Ant-Fwd-Back: move forward and backward (2 tasks)
- Ant-Goal-2D: navigate to a target goal location on 2D grid (100 train tasks, 30 test tasks)
- Walker-2D-Params: agent is initialized with some system dynamics parameters randomized and must move forward (40 train tasks, 10 test tasks)

For these domains, the on-policy baseline approaches require many more samples to learn the benchmark tasks. Here in Figure 9 we plot the same data as in Figure 3 for the full number of time steps used by the baselines.