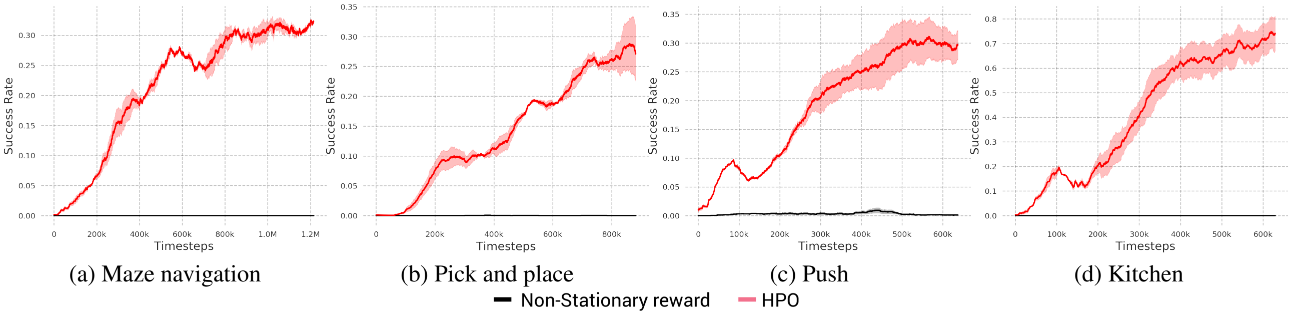
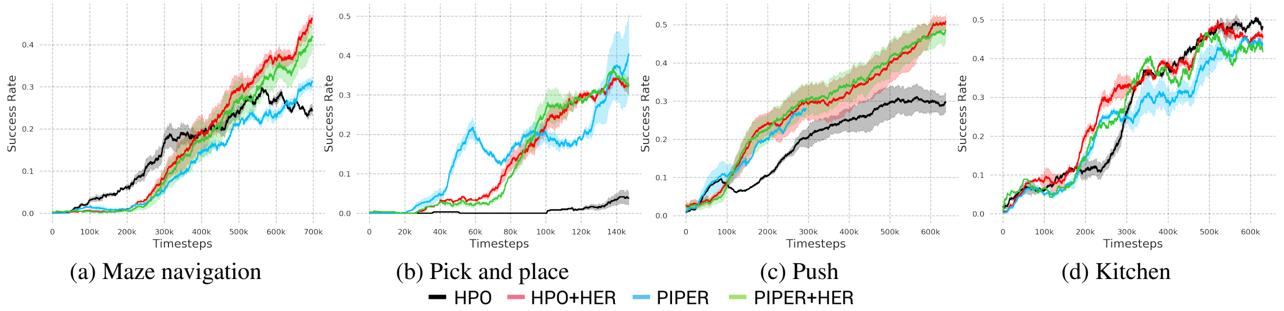
Additional Experiments



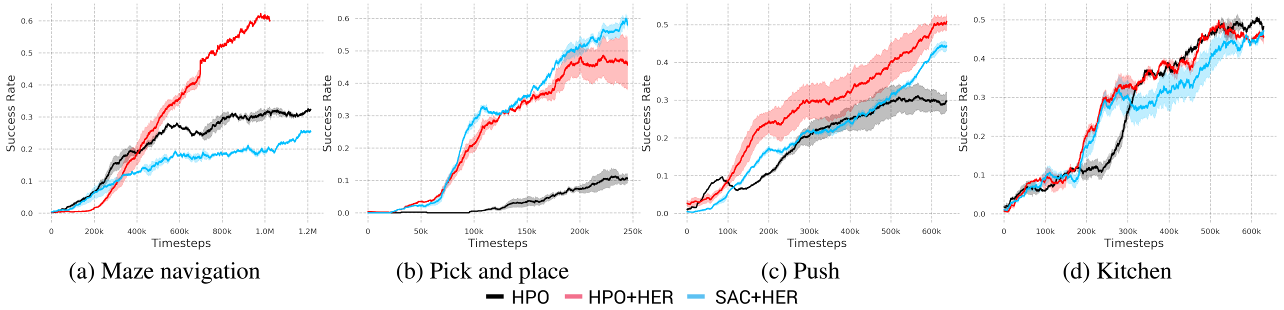
(a) Maze navigation (b) Pick and place (c) Push (d) Kitchen

Figure 1: **Comparison with non-stationary reward function.** This figure illustrates the success rate comparison across four sparse-reward maze navigation and robotic manipulation tasks. We evaluate HPO against the hierarchical approach trained with non-stationary reward function. As seen from figure, HPO demonstrates strong performance and significantly outperforms the baseline.



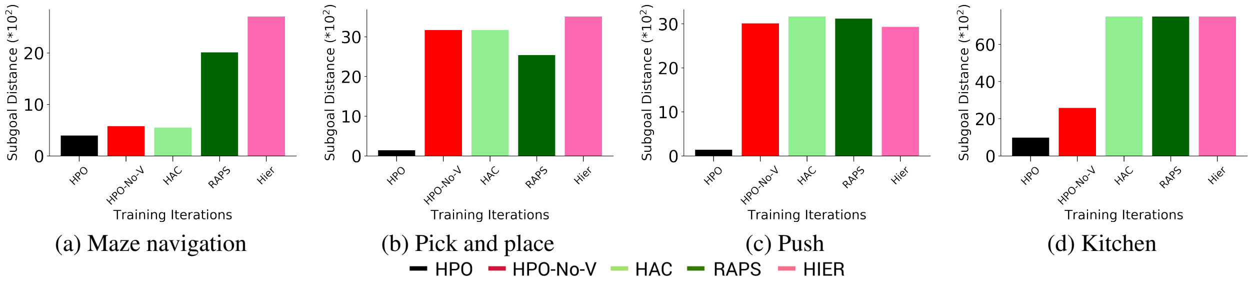
(a) Maze navigation (b) Pick and place (c) Push (d) Kitchen

Figure 2: **Comparison with PIPER.** This figure illustrates the success rate comparison across four sparse-reward maze navigation and robotic manipulation tasks. We evaluate HPO and HPO+HER (HPO with Hindsight Experience Replay) approach against PIPER and PIPER+HER baseline. As seen from figure, although PIPER outperforms HPO in maze, pick and place and push tasks, HPO+HER demonstrates impressive performance and outperforms the PIPER PIPER+HER baselines on all tasks.



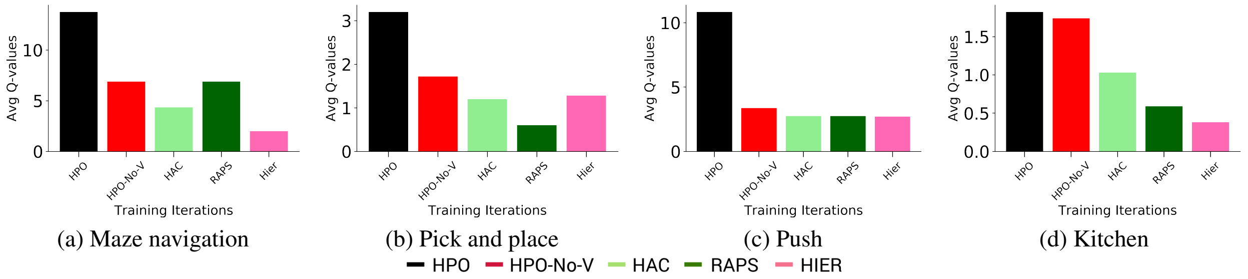
(a) Maze navigation (b) Pick and place (c) Push (d) Kitchen

Figure 3: **Comparison with SAC+HER.** This figure illustrates the success rate comparison across four sparse-reward maze navigation and robotic manipulation tasks. We evaluate HPO and HPO+HER (HPO with Hindsight Experience Replay) approach against SAC+HER. As seen from figure, although SAC+HER outperforms HPO in pick and place and push tasks, HPO+HER demonstrates impressive performance and outperforms the SAC+HER baseline on maze, push and kitchen tasks.



(a) Maze navigation (b) Pick and place (c) Push (d) Kitchen

Figure 4: **Non-stationarity ablation.** This figure compares `HPO` with `HPO-No-V, HAC, RAPS, HIER` baselines, based on average distance between subgoals predicted by the higher level policy and subgoals achieved by the lower level primitive. `HPO` consistently generates low average distance values, which implies that in `HPO`, the higher level policy generates achievable subgoals that induce optimal lower primitive goal reaching behavior. This shows that HPO is able to address non-stationary in HRL and generate feasible subgoals.



(a) Maze navigation (b) Pick and place (c) Push (d) Kitchen

Figure 5: **Q-value ablation.** This figure compares `HPO` with `HPO-No-V, HAC, RAPS, HIER` baselines, based on average lower level Q function values for the subgoals predicted by the higher level policy. `HPO` consistently large Q-function values, which implies that in `HPO`, the higher level policy generates achievable subgoals. This shows that HPO is able to address non-stationary in HRL and generate feasible subgoals.