Website

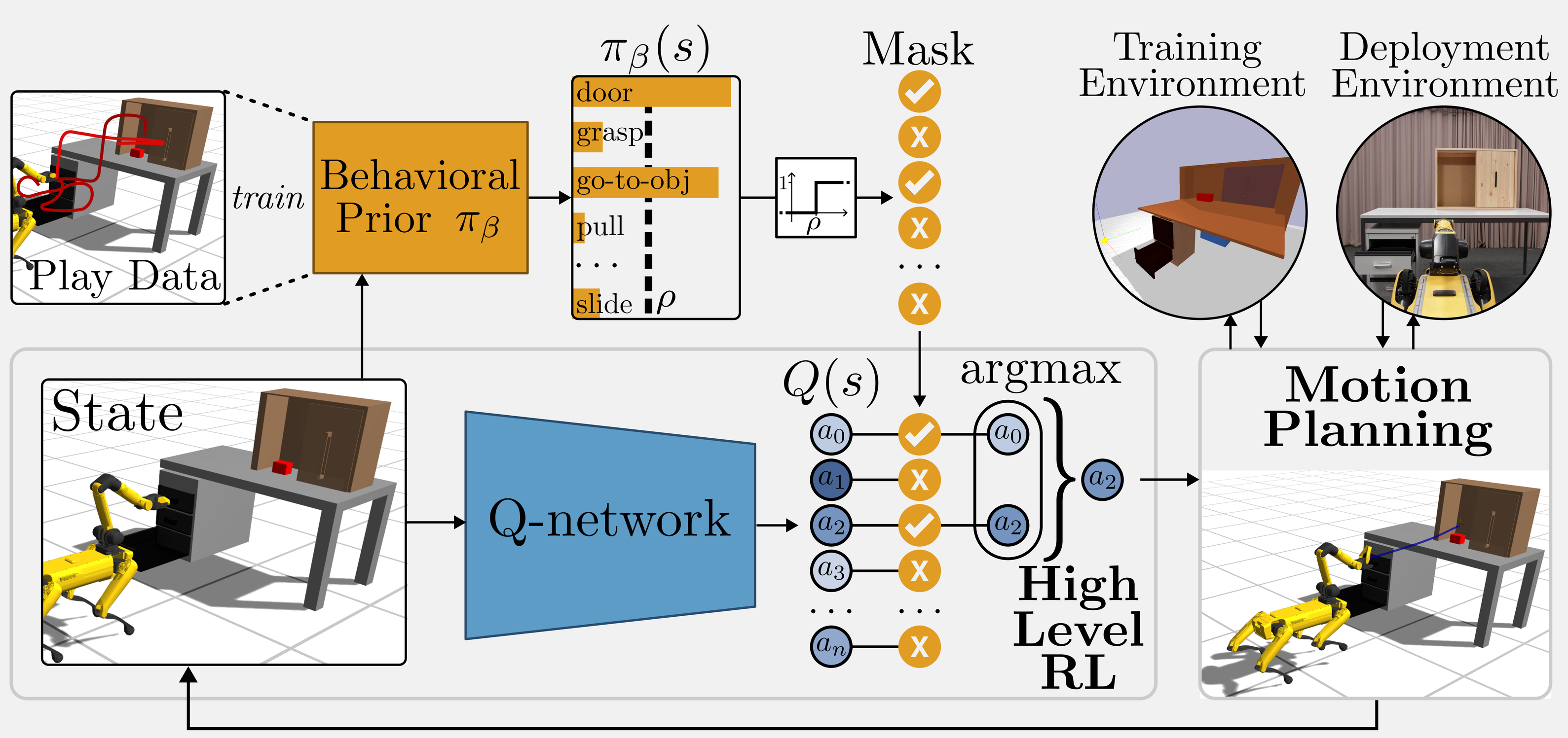
Efficient Learning of High Level Plans from Play

Núria Armengol-Urpí[1], Marco Bagatella[2], Otmar Hilliges[1], Georg Martius[2], Stelian Coros[1]

[1]Department of Computer Science, ETH Zurich; [2]Max Planck Institute for Intelligent Systems Tübingen, Germany

Contributions

- We explore the **integration of RL with classic motion planning algorithms**
- We highlight how the **exploration problem** can be addressed by introducing a **prior** model learned from **play data**
- We show how learned policies can be deployed to **physical hardware**



Method Overview

- Collect play data $\mathcal{D} = \{(s_1, a_1), ...(s_N, a_N)\}$
- Learn a goal independent behavioral prior from play $\pi^{\beta}(\cdot|s)$ by minimizing:

$$\mathcal{L}_{NLL} = \mathbb{E}_{B\sim\mathcal{D}}\left[\frac{1}{|B|}\sum_{(s,a)\in B} -\log \pi^{\beta}(a|s)\right]$$

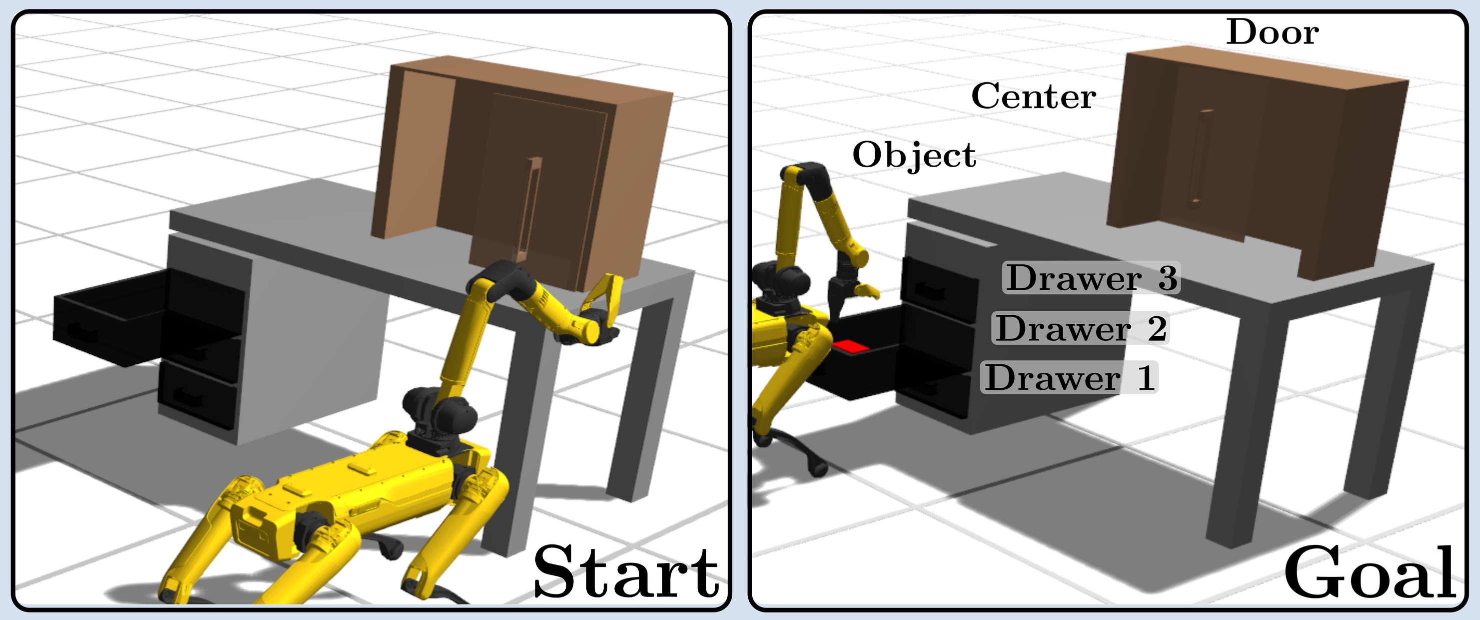
- Select feasible actions

$$\alpha(s) = \{a \in \mathcal{A} \mid \pi_{\beta}(a|s) > \rho\}$$

- Learn in a reduced MDP

$$Q(s,a,g) \leftarrow (1-\delta)Q(s,a,g) + \delta(r + \gamma \max_{a' \in \alpha(s)} Q(s', a', g))$$

Experimental Setup



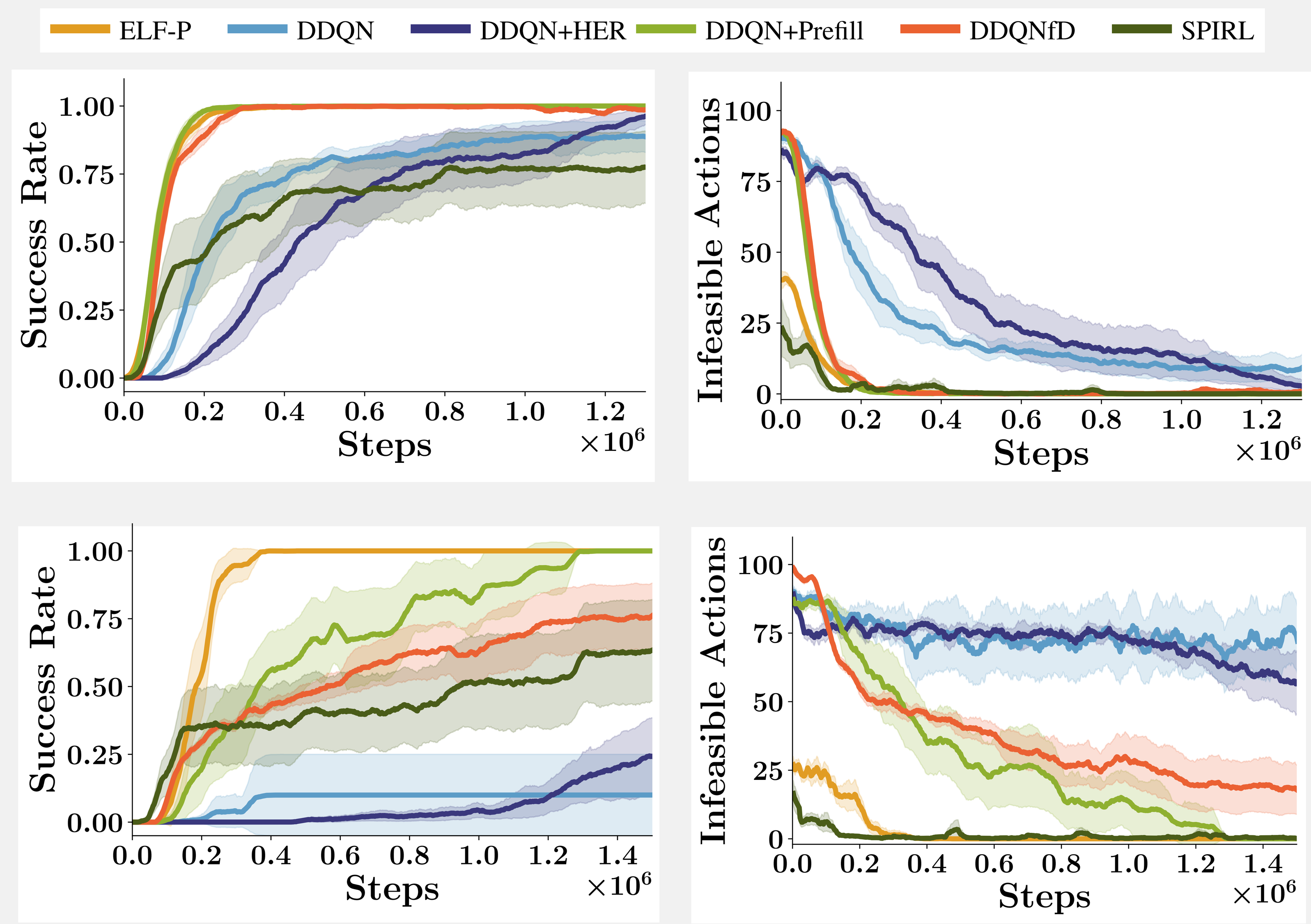
Start: object inside cabinet **Goal**: block inside mid drawer

Plan:

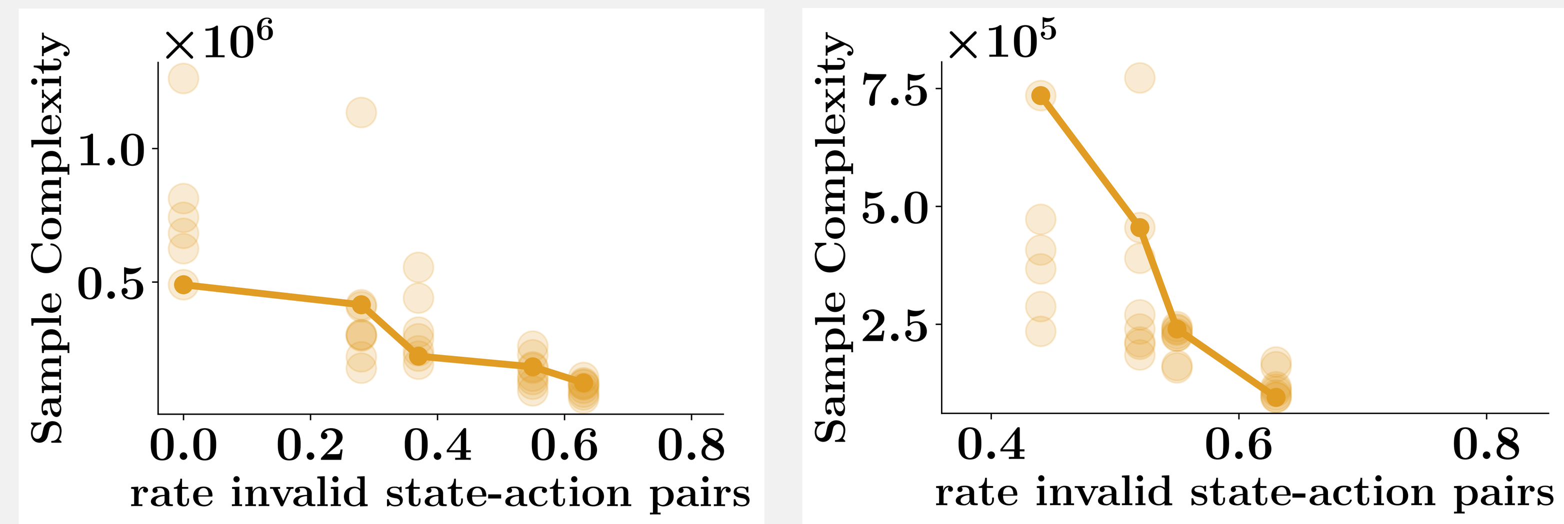
1. center	6. drawer2	11. door	16. grasp
2. drawer3	7. grasp/release	12. grasp/release	17. center
3. grasp/release	8. push/pull	13. slide	18. go-to-goal
4. push/pull	9. grasp/release	14. grasp/release	19. release
5. grasp/release	10. center	15. object	

Experiments

Performance



Sample complexity



Related Works

K. Pertsch, Y. Lee, and J. J. Lim, "Accelerating reinforcement learning with learned skill priors," in Conference on Robot Learning (CoRL), 2020.

C. Lynch, M. Khansari, T. Xiao, V. Kumar, J. Tompson, S. Levine, and P. Sermanet, "Learning latent plans from play," Conference on Robot Learning (CoRL), 2019.