Efficient Learning of High Level Plans from Play
Núria Armengol-Urpi\textsuperscript{1}, Marco Bagatella\textsuperscript{2}, Otmar Hilliges\textsuperscript{1}, Georg Martius\textsuperscript{2}, Stelian Coros\textsuperscript{1}
\textsuperscript{1}Department of Computer Science, ETH Zurich; \textsuperscript{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 $\mathcal{D} = \{(s_1, a_1), \ldots, (s_N, a_N)\}$
- Learn a goal independent behavioral prior from play $\pi^\beta(\cdot | s)$ by minimizing:
  \[
  \mathcal{L}_{\text{NLL}} = \mathbb{E}_{s, a \sim \pi^\beta} \left[ -\frac{1}{B} \sum_{i=1}^{B} \log \pi^\beta(a|s) \right]
  \]
- Select feasible actions $\alpha(s) = \{a \in \mathcal{A} | \pi^\beta(a|s) > \rho\}$
- Learn in a reduced MDP $Q(s, a, g) \leftarrow (1 - \delta)Q(s, a, g) + \delta(r + \gamma \max_{a' \in \alpha(s)} Q(s', a', g))$

Experiments

Related Works