SLAC: Simulation-Pretrained Latent Action Space for Whole-Body Real-World RL

Jiaheng Hu, Peter Stone, Roberto Martin-Martin **The University of Texas at Austin**

Motivation

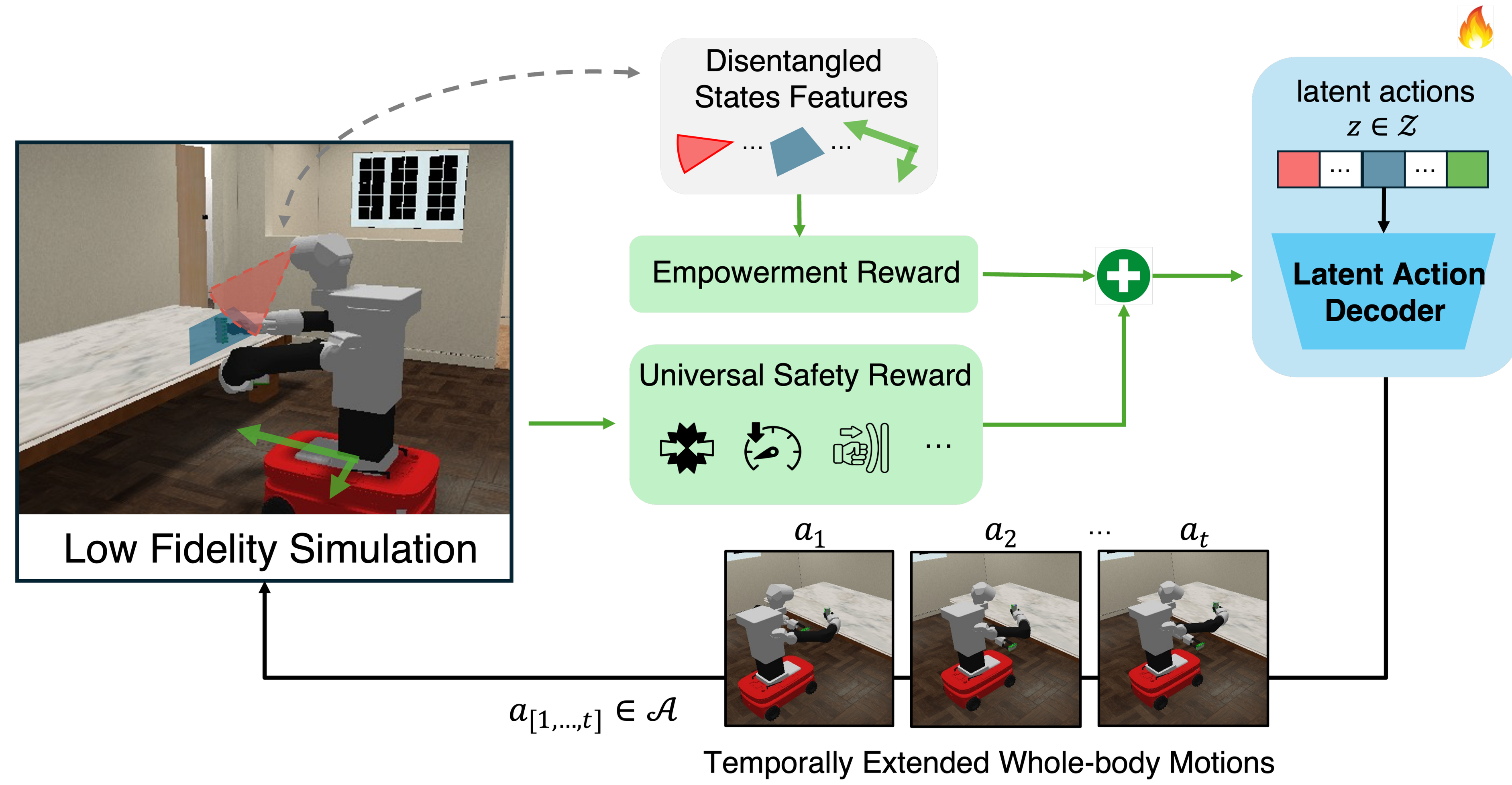
Why Real-World Reinforcement Learning?

- Enables Robots to learn **autonomously** from direct **physical interactions**
- Opens possibilities for self-improvement

Why is it hard?

- RL can take millions of **samples** to learn
- The robot can easily induce **unsafe** behavior and break itself during training

Step 1: Latent Action Space Learning



Unsupervised Skill Discovery:

- Learn diverse behaviors from **reward-free interactions** with an environment
- Often done via maximizing the **empowerments** of latent skills

Objective Breakdown:

Disentangled MI: $\mathcal{J}(\theta) = \sum_{i=1}^{N} I(S^i; Z^i) - \lambda I(S^{\neg i}; Z^i)$

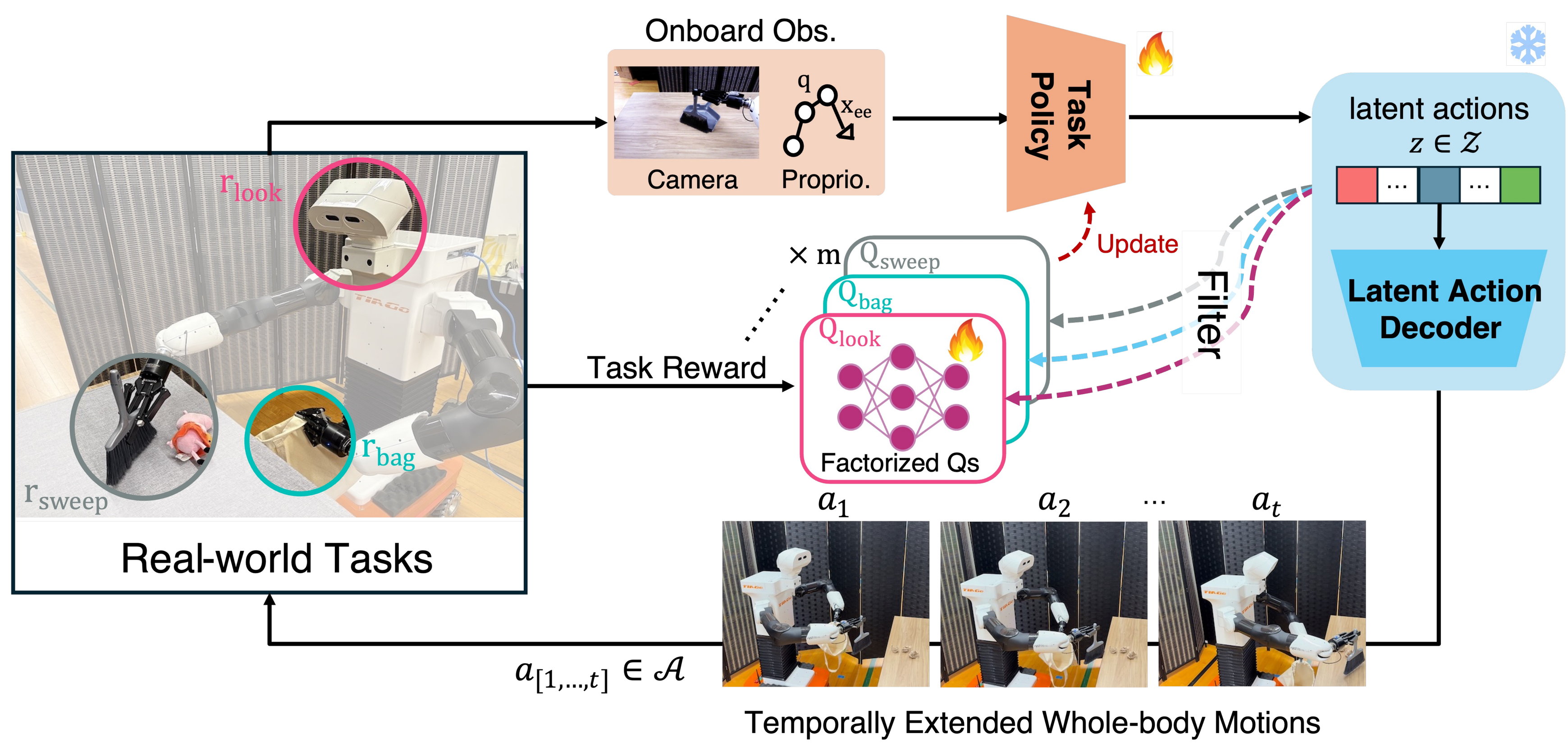
Tractable Intrinsic Reward: $r_{skill}(s, a) \triangleq \sum_{i=1}^{N} q_\phi^i(z^i|s^i) - \lambda q_\psi^i(z^i|s^{\neg i})$

Safety-aware skill learning: $r_{latent} = r_{skill} + r_{safe}$

How can we make **Real-world Reinforcement Learning** work on **Mobile Manipulators**?

Our key idea: boost sample efficiency and ensure safety by operating in a **good action space**, one that can be pre-learned entirely in a **low-fidelity simulator**.

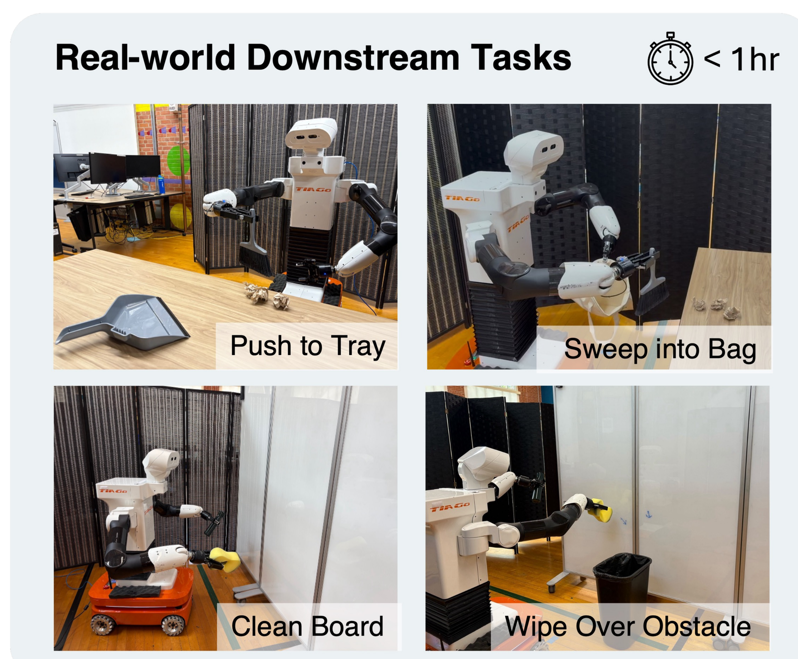
Step 2: Downstream Task Learning via Real-World RL

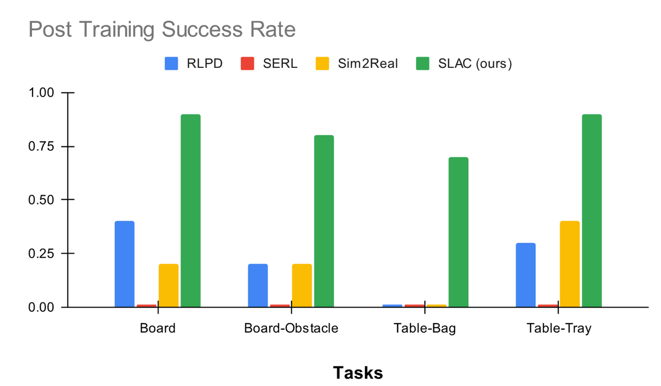


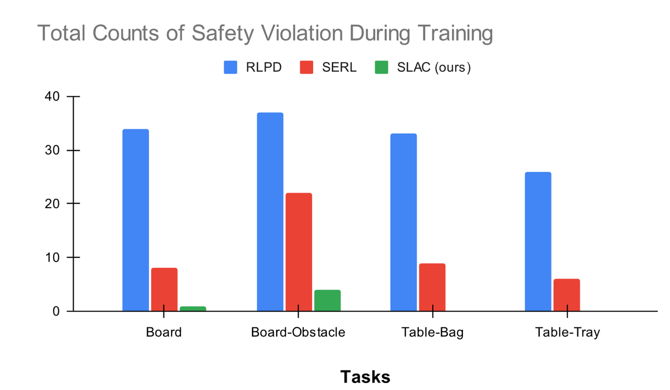
$$\pi(a|o) = \int_z \pi_{dec}(a|o_{dec}, z)\, \pi_{task}(z|o)\, dz \qquad Q_\pi(s, z) = \sum_{i=1}^{m} Q_\pi^i(s, z)$$

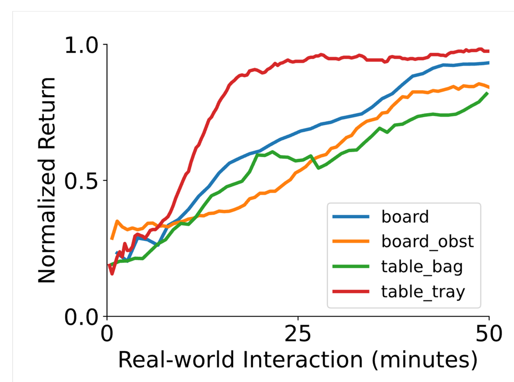
Result

SLAC can learn challenging tasks within an hour of interactions!









Check out our website for videos!

website & code