Recovery-Aware HIL-SERL for Contact-Rich USB Insertion with Global State Encoding

Tsinghua University

Guanghui Shen, Peishi Yan, Chenyang Wang, Xinyi Xu, Junda Cao

GitHub https://sgh21.github.io/USBInsertionRL

Introduction

Background

We use real-world RL to automate **USB pick-and-insert** on a UR5e robot—a contact-rich assembly task where small errors and improper contact forces can cause jamming or damage—targeting **fast, safe,** and **robust** policy learning.

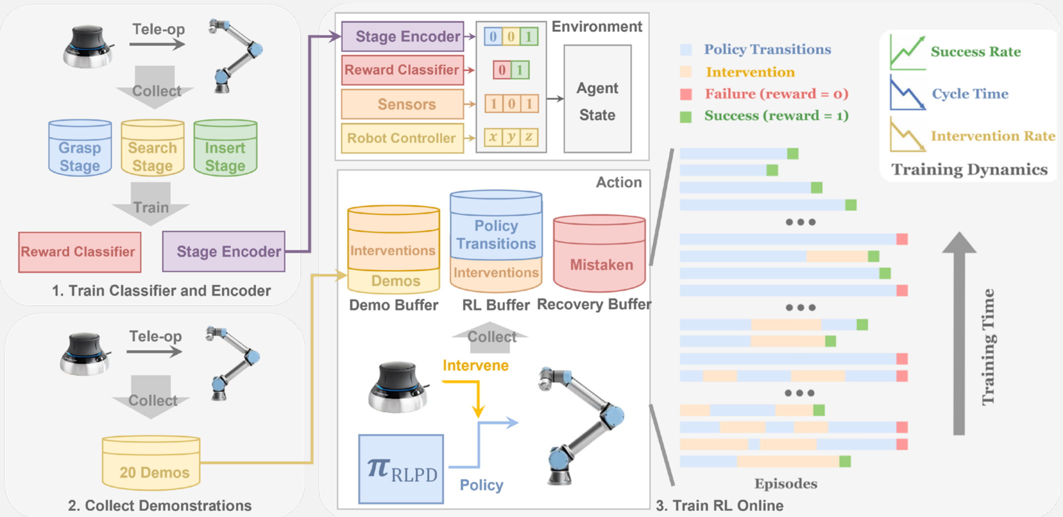
Motivation

- Simulation-based approaches often break in the real world due to the **sim-to-real gap**.
- In contrast, pure real-world RL can work but is **data-hungry, slow,** and **risky.**

Contributions

- **Safe compliant execution on UR5e:** High-rate control with **compliant force** regulation for reliable, **low-risk** in-contact training.
- **Global stage-aware state encoding:** Global observations to better **model multi-stage,** long-horizon insertion sequences, **speeding up convergence**.
- **Human-gated Recovery Buffer ("mistake notebook"):** Human-gated replay of **recovery-critical samples** improves **recovery skills** and **robustness** under failures.

Framework



An asynchronous **actor–learner loop** fuses **robot signals** with an **external-camera** global feature, executes the policy, and **stores transitions in replay buffers**. A **binary success classifier** provides **sparse episode-level reward**, and **off-policy replay** updates iteratively **improve the policy**.

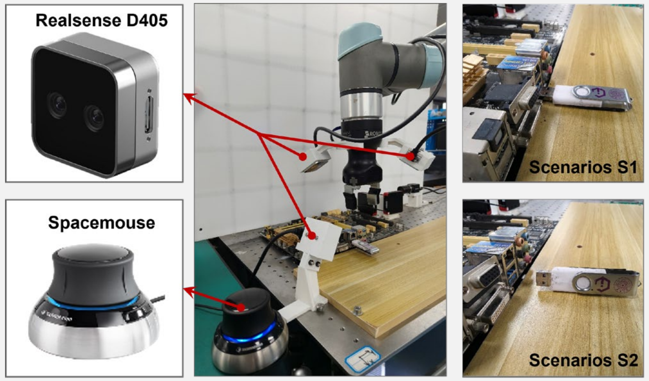
Method Details

- **Compliant Force Control:** Maintain a **desired pose updated** at up to **500 Hz,** applies **PD control** to generate **force commands** for compliant contact.

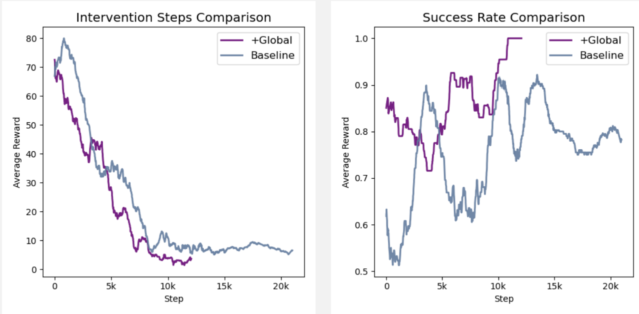
$$\delta x = \hat{x} - x \qquad \mathbf{F} = k_p \delta x + k_d \Delta(\delta x)$$

- **Global Observer:** An **external-camera encoder** is first trained for **stage classification**. During RL, the **encoder is frozen** and its **embedding is appended** to the **policy input** as a **compact stage-aware** global feature.
- **Recovery Buffer:** A **human supervisor gates** whether an episode is added to the **Recovery Buffer**. Each **policy parameter** update **samples** from the Demo / RL / Recovery buffers with a fixed **4:4:2 ratio**.

Experiments



Global Observer Convergence and Performance



Success Rate	6k	10k	12k
Baseline	0%	75%	90%
+Global	70%	95%	100%

Recovery Buffer Performance in Two Scenarios

Success Rate	S1: contact	S2: misoriented
Baseline	55%	0%
+Recovery	90%	70%