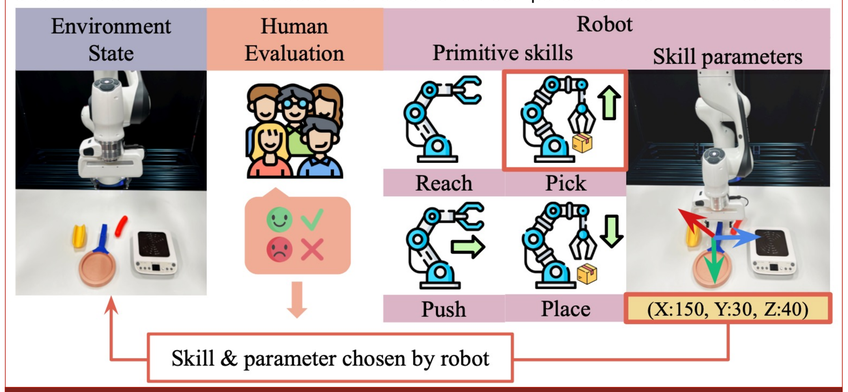
Primitive Skill-based Robot Learning from Human Evaluative Feedback

Ayano Hiranaka*, Minjune Hwang*, Sharon Lee, Chen Wang, Li Fei-Fei, Jiajun Wu, Ruohan Zhang
(*: equal contribution, alphabetically ordered)
Correspondence: {ayanoh, mjhwang, zharu} @ stanford.edu

Abstract

RL algorithms face significant challenges for long-horizon robot manipulation tasks in the real-world due to sample inefficiency and safety issues. To overcome such challenges, we propose a novel framework which combines RL from human feedback (RLHF) and learning with primitive skills. Our algorithm, **SEED,** reduces human effort, and its parameterized skills provide a clear view of the agent's high-level intentions, allowing humans to evaluate skill choices before execution in a safer and more efficient manner. **SEED** significantly outperforms state-of-the-art algorithms in sample efficiency and safety and exhibits a substantial reduction of human effort compared to other RLHF methods.

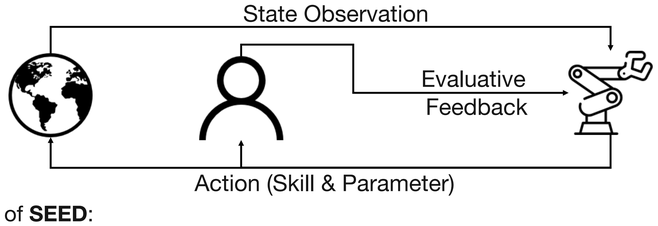


Introduction

Challenges of RL in real-world robotics:
sample inefficiency, safety concerns, and reward design

Our framework, **SEED**, integrates two approaches to overcome such issues:

1. **learning from human evaluative feedback**
2. **primitive skill-based motion control**.



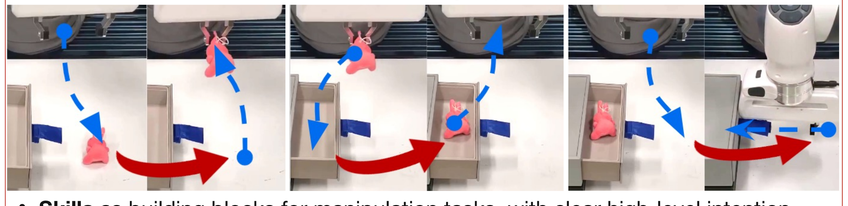
Benefits of **SEED**:

- Human feedback provide **dense training signals**.
- Skills represent robot's intent in an **intuitive** way.
- Evaluation without execution is **safe** & **efficient** at **reduced human effort**.

Parameterized Skills

Skills are implemented with Deoxys API operational space control for Franka Arm.

Pick(x, y, z) Place(x, y, z) Push(x, y, z, delta)

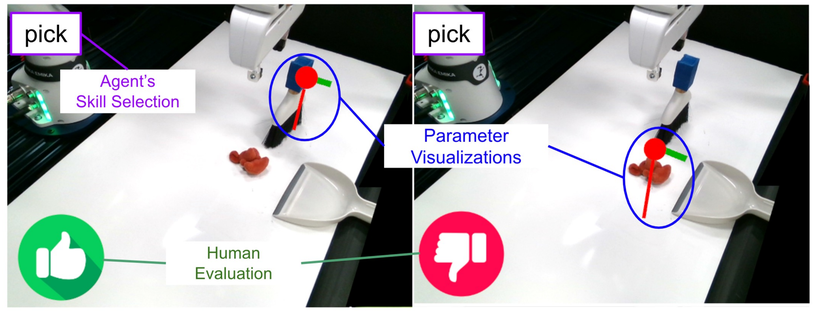


- **Skills** as building blocks for manipulation tasks, with clear high-level intention.
- **Parameters** with clear semantic meanings.

→ **Goal**: efficient learning without the burden of learning low-level control

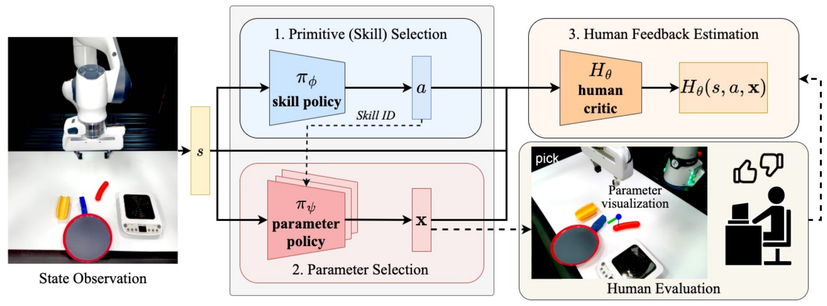
Evaluation without Execution

User Interface for Human Trainer



Intuitive Skill-Based Actions → Evaluation without Execution → More Safe and Efficient Learning

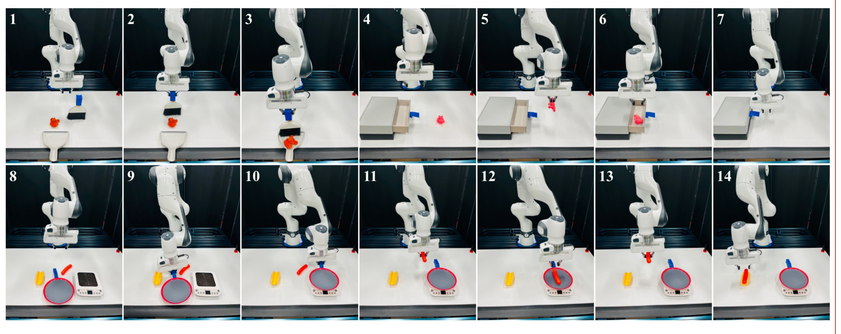
Model Architecture



- Hierarchical framework: skill policy and parameter policy
- Human feedback as reward: human evaluation instead of environment rewards
- Balanced replay buffer: equal number of good & bad samples in off-policy batch

Real-world Long-Horizon Manipulation Tasks

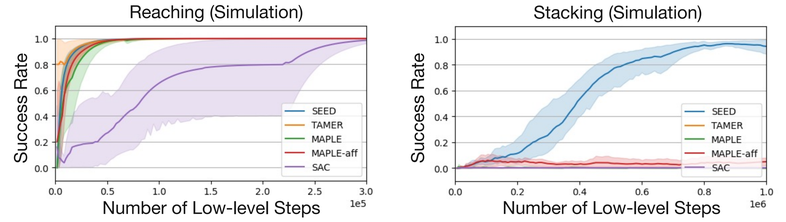
Visualization of real-world long-horizon manipulation tasks with intermediate steps.



- 1~3: **Sweeping** task
- 4~7: **Collecting-Toy** task
- 8~14: **Cooking-Hotdog** task (a task with the longest horizon)

Simulation Experiments

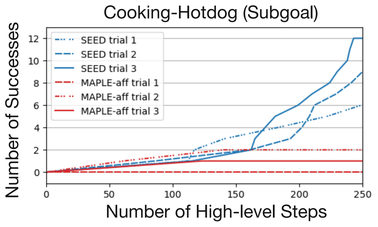
Simulation tasks in **Robosuite** [1] with synthetic feedback.



SEED is as efficient in a simple task, and far more successful in a complex task.

Real-world Experiments

For real-world robot experiment, a single human trains the agent by providing evaluation signals via a keyboard press.

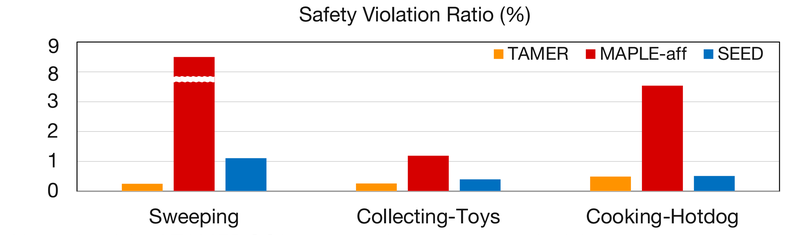


SEED is sample efficient.

- 10x times faster training than MAPLE [4]
- 9x times higher success rate than MAPLE

SEED ensures better safety.

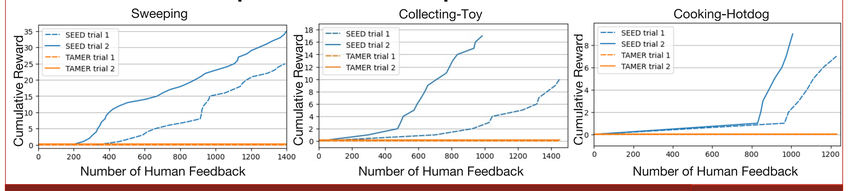
- SEED exhibited significantly lower safety violation: 3~7x lesser than baseline [4].



Note: the risk of TAMER [3] is underestimated as one decision step corresponds to one skill step, which involves around 100 low-level steps/feedback for TAMER.

SEED significantly reduces human effort.

- **SEED** succeeded within the fixed amount of feedback, while TAMER failed.
- Human trainers **adapted** and **learned to provide better feedback** as well.



Reference

[1] Zhu, Yuke, et al. “robosuite: A Modular Simulation Framework and Benchmark for Robot Learning.” arXiv 2018.
[2] Haarnoja, Tuomas, et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." ICML 2018.
[3] Knox, W. Bradley, et al. "TAMER: Training an Agent Manually via Evaluative Reinforcement." ICSR 2013.
[4] Nasiriany, Soroush, et al. " Augmenting Reinforcement Learning with Behavior Primitives for Diverse Manipulation Tasks." ICRA 2022.

Webpage

Find out more in our project website!