Phoenix: A Motion-based Self-Reflection Framework for Fine-grained Robotic Action Correction

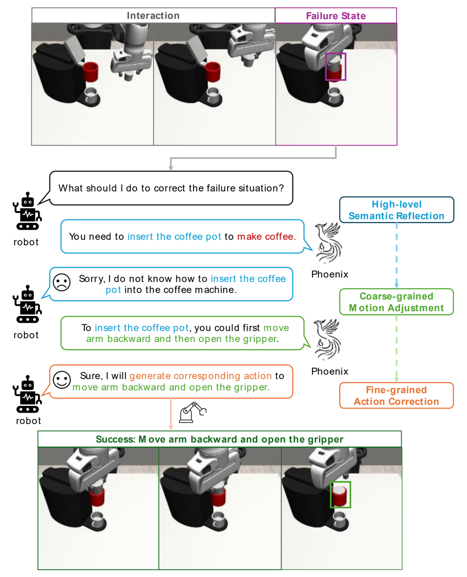
Wenke Xia[1,2], Ruoxuan Feng[1], Dong Wang[2], Di Hu[1,†]

[1] Gaoling School of Artificial Intelligence, Renmin University of China, [2] Shanghai AI Laboratory

github

arxiv

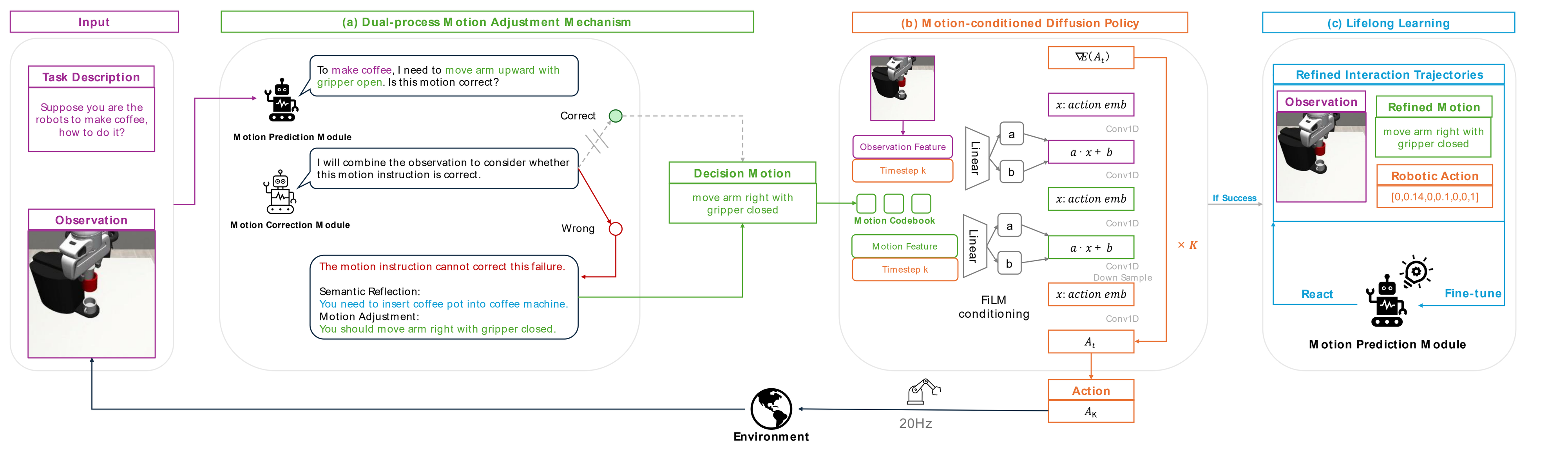
Motion-based Self-Reflection Framework



How to build a **generalizable self-reflection framework** for robots to learn recover from failures:

- **Coarse-grained motion instruction** as bridge to utilize the generalization ability of Multimodal Large Language Models (MLLMs) from high-level semantic reflection to fine-grained robotic action correction.
- **Motion-conditioned diffusion policy** convert motion instruction into high-frequency, low-level robotic action correction.
- **Lifelong learning process** for iterative self-improvement through interactions.

Motion-based Self-Reflection Framework



- **Dual-process motion adjustment mechanism** to ensure efficient motion prediction and comprehensive failure correction.
- **Motion-conditioned diffusion policy** for low-level action correction.

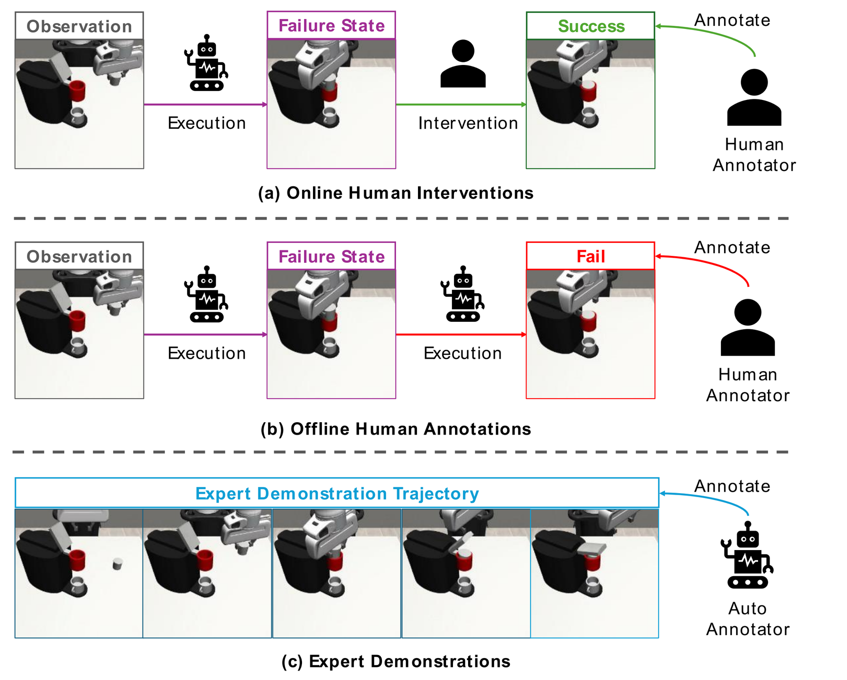
Dual-process Motion Adjustment Mechanism

To equip MCM with capabilities for failure detection and correction, we construct a comprehensive correction dataset:

Algorithm 1 Self-Reflection w/. Dual-Process Motion Adj.

Require: Task description T, Observation O, Environment E, Motion prediction module MPM, Motion correction module MCM, Motion-conditioned diffusion policy π, Exploration timestep K,

```
O_1 ← E.reset()
for k = 1 to K do
    m_i ← MPM(O_k, T)
    failure_flag, semanic_info ← MCM(O_k, m_i)
    if failure_flag is true then
        m_a ← MCM(O_k, sematic_info)
        m_d ← m_a
    else
        m_d ← m_i
    end if
    a ← π(O, m_d)
    O_{k+1} ← E.step(a)
end for
```

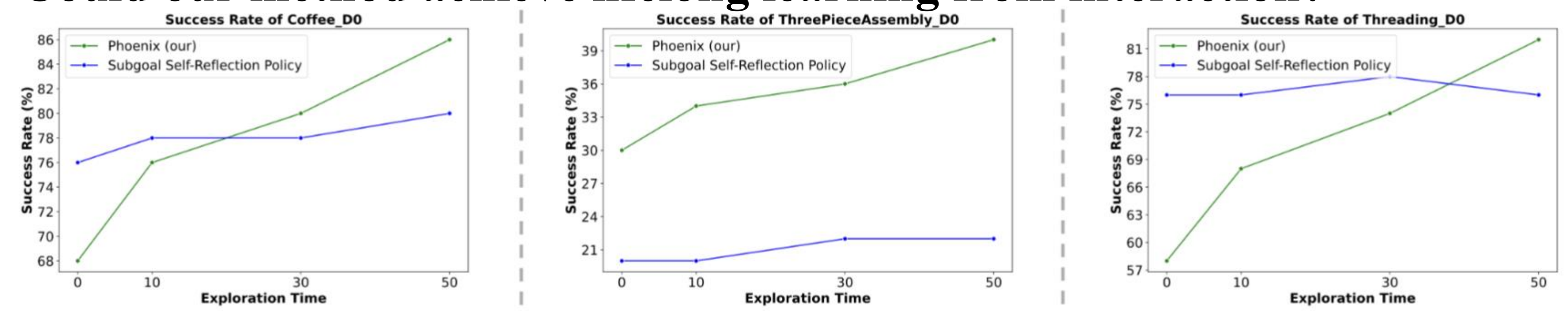


Experiment Results

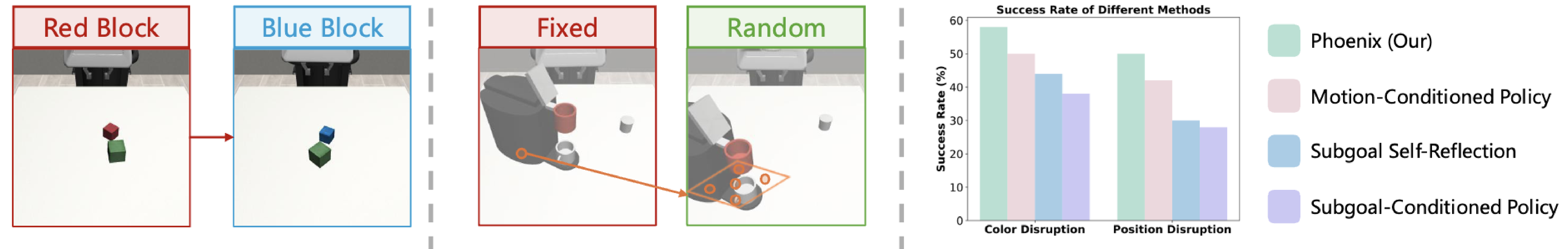
➢ Does motion-based self-reflection method enhance action correction?

Methods	Coffee_D0	Coffee_D1	Stack_D0	Stack_D1	StackThree_D0	StackThree_D1	Threading_D0	ThreePieceAssembly_D0	ThreePieceAssembly_D1	Mean
OpenVLA [11]	42%	18%	84%	**86%**	36%	**20%**	20%	28%	**8%**	38.0%
Task-conditioned	66%	24%	88%	68%	30%	6%	74%	20%	0%	41.8%
Subgoal-conditioned	76%	26%	88%	74%	24%	6%	**78%**	20%	2%	43.8%
Motion-conditioned	68%	32%	92%	84%	38%	16%	58%	30%	4%	46.9%
Subgoal Self-reflection	80%	32%	88%	78%	32%	6%	80%	34%	2%	48.0%
Phoenix (Ours)	**94%**	**48%**	**96%**	**86%**	**50%**	**20%**	68%	**52%**	6%	**57.8%**
Human Intervention (Oracle)	100%	100%	100%	90%	70%	40%	100%	70%	40%	78.9%

➢ Could our method achieve lifelong learning from interaction?

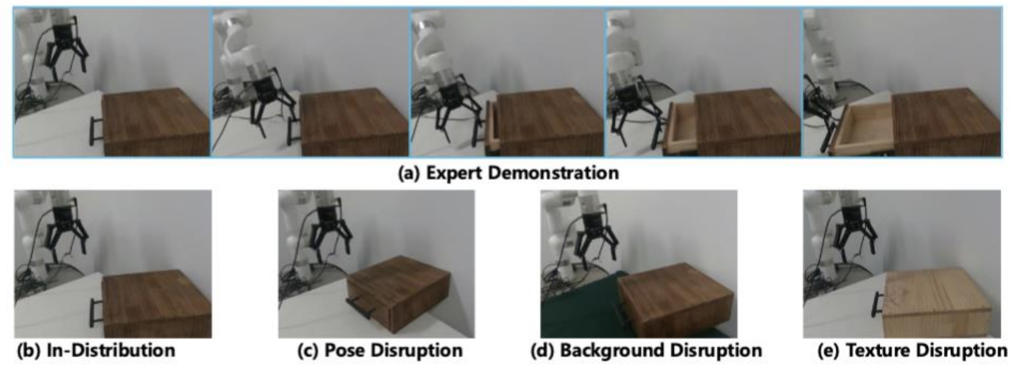


➢ Could our method generalize across novel scenarios?



(a) Color Disruption in Stack_D0 (b) Position Disruption in Coffee_D0 (c) Performance on Novel Tasks

➢ Could our framework ensure robustness in real-world scenarios?



Model	In-Dis.	Pose Dis.	Bg.	Tex.
OpenVLA	55%	30%	35%	45%
Task	60%	25%	25%	45%
Motion	60%	35%	30%	50%
Ours	75%	55%	45%	65%

Motion-conditioned Diffusion Policy

We propose the diffusion policy for high-frequency robotic action prediction:

$$\mathcal{L} = \mathrm{MSE}(\mathcal{E}^k, \pi(\mathcal{O}, \mathcal{M}, \mathcal{A}^0 + \mathcal{E}^k, k))$$

where π denote the diffusion policy, $\mathcal{M}$ is discriminative motion instruction feature, $\mathcal{O}$ is the observation feature, $\mathcal{A}^0$ represents the ground truth action.

Action Correction for Lifelong Learning

We equips MPM with motion prediction and failure correction capabilities through learning from refined interaction trajectory.

This process enhances our model to react quickly to the environment without human intervention and complex chain-of-thought.

Conclusion

- **Motion-based self-reflection framework** to convert the semantic reflection of MLLMs into fine-grained robotic action correction.
- Propose **dual-process motion adjustment mechanism** and **motion-conditioned diffusion policy** for action correction with motion instruction.
- Lifelong Learning for iterative self-improvement from interaction.