GUI-Shift: Enhancing VLM-Based GUI Agents Through Self-supervised Reinforcement Learning

*Longxi Gao[1], Li Zhang[1], Pengzhi Gao[2] , Wei Liu[2], Jian Luan[2] , Mengwei Xu[1]**
*[1]Beijing University of Posts and Telecommunications [2]Independent Researcher *Corresponding Author*

Paper

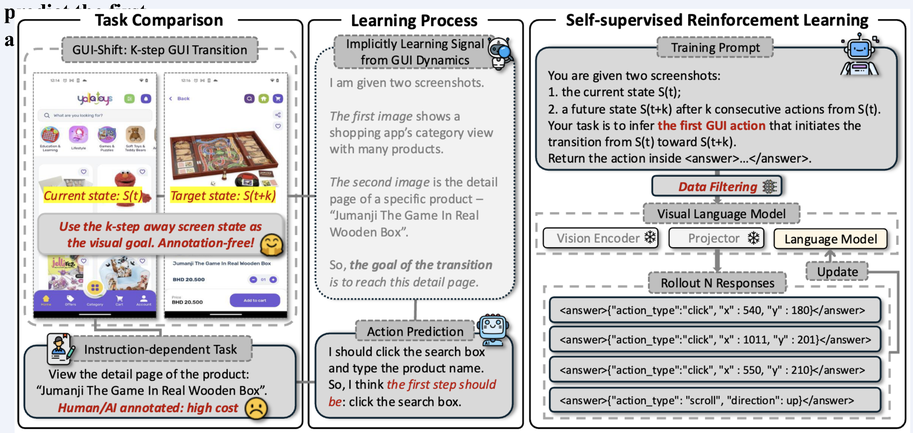
Introduction

- Mobile GUI agents interpret natural language instructions and perform actions (e.g., click, scroll) directly on smartphone screens. They can control diverse apps as a human would, improving accessibility for users who are visually impaired, elderly, or have their hands occupied. Breakthroughs of vision language models (VLMs) have reshaped the design paradigm of mobile GUI agents, transitioning from handcrafted heuristics to learned, vision-grounded policies. However, training effective VLMs-based GUI agents typically depends on large-scale annotated datasets, whose collection is both labor-intensive and error prone.
- This study aims to address a fundamental challenge: **how to train capable mobile GUI agents using large-scale, unlabeled GUI trajectories, rather than relying on costly human-annotated instructions.**
- This study introduce **K-step GUI Transition**, a self-supervised inverse dynamics task in which VLMs learn GUI dynamics by predicting the initial action that causes a transition between two GUI states. Building on this task, we propose **GUI-Shift**, a reinforcement learning (RL) framework that combines rule-based optimization with data filtering to improve VLM performance.

Methods

(1) K-step GUI Transition: a self-supervised training task.

Inspired by inverse-dynamics modeling in robotics and biomechanics, where a model predicts control commands linking two consecutive physical states, our task treats screenshots as states and GUI actions as commands. Each training sample in K-step GUI Transition consists of two screenshots, S(t) and S(t+k), where S(t+k) results from executing K actions starting from S(t). The VLM is trained to p a



(2) GUI-Shift: a self-supervised RL framework.

- **Left:** K-step GUI Transition replaces annotated instructions with the target state S(t+k), enabling scalable data construction through automated offline exploration.
- **Middle:** The model learns GUI dynamics by predicting the action that causes the transition.
- **Right:** GUI-Shift achieves self-supervised training by applying GRPO to GUI Transition.

Results

1. GUI Task Automation Evaluation

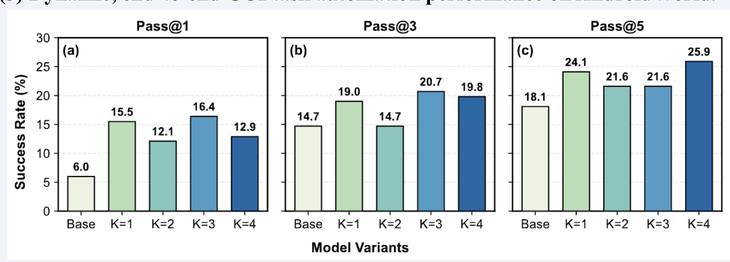
(1) Static, single step GUI task automation on AndroidControl and GUI Odyssey.

Model	# Training Samples	AC-Low		AC-High		GUI Odyssey	
		TM	EM	TM	EM	TM	EM
Proprietary models							
GPT-4o (OpenAI, 2024)	-	74.3	19.4	66.3	20.8	34.3	3.3
Models trained with annotations							
SeeClick (Cheng et al., 2024)	1M	93.0	75.0	82.9	59.1	71.0	53.9
OS-Atlas-7B (Wu et al., 2024b)	2.3M	93.6	85.2	85.2	71.2	-	62.0
Aguvis-7B (Xu et al., 2024)	1M	-	80.5	-	61.5	-	-
UI-TARS-7B (Qin et al., 2025)	-	98.0	90.8	83.7	72.5	**94.6**	**87.0**
UI-R1-3B (Lu et al., 2025)	136	94.3	88.5	57.9	45.4	52.2	32.5
GUI-R1-7B (Xia & Luo, 2025)	3K	85.2	66.5	71.6	51.7	65.5	38.8
InfiGUI-R1-3B (Liu et al., 2025)	32K	96.0	92.1	82.7	71.1	-	-
AgentCPM-GUI (Zhang et al., 2025)	470K	94.4	90.2	77.7	69.2	90.9	75.0
UI-Venus-Navi-7B (Gu et al., 2025)	350K	97.1	92.4	86.5	**76.1**	87.3	71.5
Ours: Qwen2.5-VL-7B as the base model							
Qwen2.5-VL-7B (Bai et al., 2025)	-	94.9	83.8	72.9	59.2	59.8	44.9
GUI-Shift-Qwen ($k = 1$)	2K	98.0↑3.1	90.6↑6.8	85.9↑13.0	70.4↑11.2	78.5↑18.7	54.8↑9.9
Ours: InternVL3-8B as the base model							
InternVL3-8B (Chen et al., 2024)	-	97.8	90.0	71.5	49.8	48.8	20.3
GUI-Shift-Intern ($k = 4$)	2K	97.3↓0.5	88.0↓2.0	78.5↑7.0	56.6↑6.8	59.6↑10.8	23.3↑3.0
Ours: Mimo-VL-7B-SFT as the base model							
Mimo-VL-7B-SFT (Xiaomi, 2025b)	-	90.8	85.7	75.2	63.1	86.9	62.0
GUI-Shift-Mimo-SFT ($k = 3$)	2K	98.6↑7.8	**93.2**↑7.5	**87.2**↑12.0	73.4↑10.3	86.1↓0.8	60.7↓1.3
Ours: Mimo-VL-7B-RL as the base model							
Mimo-VL-7B-RL (Xiaomi, 2025b)	-	91.8	87.2	76.5	64.6	87.2	63.1
GUI-Shift-Mimo-RL ($k = 1$)	2K	**98.9**↑7.1	**93.2**↑6.0	86.9↑10.4	71.7↑7.1	84.8↓2.4	59.5↓3.6

(2) Static, multi-step GUI task automation performance on AndroidControl.

Model	Qwen2.5-VL-7B		InternVL3-8B		Mimo-VL-7B-SFT		Mimo-VL-7B-RL	
	AC-Low	AC-High	AC-Low	AC-High	AC-Low	AC-High	AC-Low	AC-High
Base	50.2	22.4	68.3	15.4	48.4	16.4	53.3	18.0
K=1	67.5↑17.3	29.9↑7.5	60.7↓7.6	18.9↑3.5	72.4↑24.0	32.1↑15.7	**76.3**↑23.0	32.2↑14.2
K=2	65.1↑14.9	29.9↑7.5	60.3↓8.0	19.2↑3.8	73.9↑25.5	33.1↑16.7	71.8↑18.5	**34.6**↑16.6
K=3	**69.3**↑19.1	29.4↑7.0	61.0↓7.3	19.2↑3.8	**75.7**↑27.3	**34.1**↑17.7	71.0↑17.7	33.6↑15.6
K=4	67.6↑17.4	**28.7**↑6.3	**61.1**↓7.2	**21.9**↑6.5	74.2↑25.8	31.1↑14.7	69.5↑16.2	33.7↑15.7

(3) Dynamic, end-to-end GUI task automation performance on AndroidWorld.



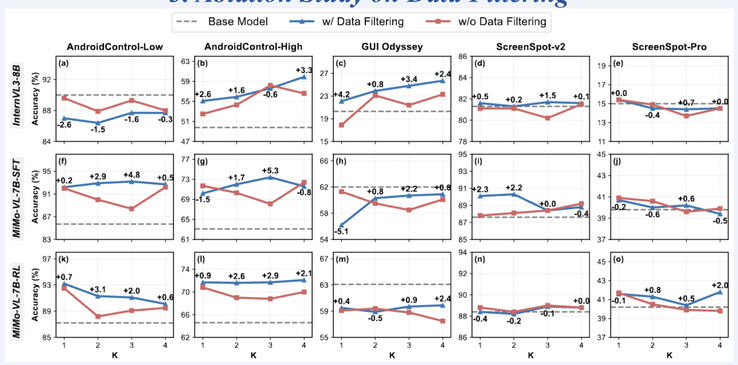
Test Model: Mimo-VL-7B-SFT and GUI-Shift-Mimo-SFT; Agent Setting: the original M3A protocol.

2. GUI Grounding Evaluation

Performance on GUI grounding benchmarks: ScreenSpot-v2 and ScreenSpot-Pro.

Model	# Training Samples	ScreenSpot-v2							-Pro
		Mobile		Desktop		Web		Avg.	Avg.
		Text	Icon	Text	Icon	Text	Icon		
Models trained with annotations									
CogAgent-18B (Wang et al., 2024)	-	-	-	-	-	-	-	-	7.7
SeeClick-9.6B (Cheng et al., 2024)	1M	78.4	50.7	70.1	29.3	55.2	32.5	55.1	1.1
UGround-7B (Gou et al., 2024)	1.3M	75.1	84.5	85.1	61.4	84.6	71.9	76.3	16.5
OS-Atlas-7B (Wu et al., 2024b)	2.3M	95.2	75.8	90.7	63.6	90.6	77.3	84.1	18.9
ShowUI-2B (Lin et al., 2024)	256K	-	-	-	-	-	-	-	7.7
UI-TARS-7B (Qin et al., 2025)	-	96.9	89.1	95.4	85.0	93.6	85.2	91.6	35.7
UI-R1-E-3B (Lu et al., 2025)	2K	83.0	**97.1**	85.0	**91.7**	77.9	**95.4**	89.5	33.5
InfiGUI-R1-3B (Liu et al., 2025)	32K	-	-	-	-	-	-	-	35.7
LPO (Tang et al., 2025)	-	97.9	82.9	**95.9**	86.4	95.6	84.2	90.5	-
UI-Venus-Ground-7B (Gu et al., 2025)	107K	**99.0**	90.0	97.0	90.7	**96.2**	88.7	**94.1**	**50.8**
Ours: Qwen2.5-VL-7B as the base model									
Qwen2.5-VL-7B (Bai et al., 2025)	-	98.3	86.3	88.7	67.1	92.7	81.8	87.7	26.4
GUI-Shift-Qwen ($k = 4$)	2K	98.6	89.6	86.1	75.0	92.7	82.8	89.0↑1.3	27.1↑0.7
Ours: InternVL3-8B as the base model									
InternVL3-8B (Chen et al., 2024)	-	93.4	81.5	80.4	52.1	91.0	73.4	81.3	15.0
GUI-Shift-Intern ($k = 1$)	2K	93.8	83.4	80.4	51.4	91.0	73.4	81.6↑0.3	15.4↑0.4
Ours: Mimo-VL-7B-SFT as the base model									
Mimo-VL-7B-SFT (Xiaomi, 2025b)	-	96.6	84.4	92.8	80.0	88.9	76.8	87.6	39.8
GUI-Shift-Mimo-SFT ($k = 1$)	2K	98.3	87.7	92.3	82.1	94.0	79.8	90.1↑2.5	40.7↑0.9
Ours: Mimo-VL-7B-RL as the base model									
Mimo-VL-7B-RL (Xiaomi, 2025b)	-	98.3	86.3	90.2	80.7	92.7	75.4	88.4	40.2
GUI-Shift-Mimo-RL ($k = 1$)	2K	**99.0**	87.7	91.2	83.6	89.7	72.9	88.4↑0.0	41.7↑1.5

3. Ablation Study on Data Filtering



- **The data filtering used in this work selects K-step GUI transitions with mixed correct and incorrect sampled responses, producing more informative training data.**
- **Filtered data consistently improves results on both GUI task automation and GUI grounding benchmarks.**
- **More ablations on CoT, training tasks and algorithms are available in the paper.**

Advantages

- Scalable data construction. GUI-Shift enables large-scale data filtering without annotated instructions, reducing annotation cost.
- Maximized data utilization. An N-image trajectory yields up to N-K training pairs for each K.

Contributions

- We introduce K-step GUI Transition, a training task that leverages abundant unlabeled GUI trajectories to enhance VLMs used in GUI agents.
- We propose GUI-Shift, a self-supervised RL framework that bridges the gap between GUI dynamics modeling and action-level GUI learning, mitigating the limitation of SFT in handling action multiplicity and poor generalization in GUI tasks.
- GUI-Shift improves performance across four VLMs and five benchmarks, with up to 11.2% gains.