Visual Agentic Reinforcement Fine-Tuning

Ziyu Liu[1,2] **Yuhang Zang**[2✉] **Yushan Zou**[4] **Zijian Liang**[1]
Xiaoyi Dong[2,3] **Yuhang Cao**[2] **Haodong Duan**[2] **Dahua Lin**[2,3] **Jiaqi Wang**[2✉]
[1]Shanghai Jiaotong University [2]Shanghai Artificial Intelligence Laboratory
[3]The Chinese University of Hong Kong [4]Wuhan University
liuziyu77@sjtu.edu.cn, {zangyuhang, wangjiaqi}@pjlab.org.cn

简介

如何让开源多模态大模型也能像 GPT-4o 一样既能推理也能用工具?

本文提出了Visual-ARFT: 一种基于可验证奖励（RLVR）的多模态 Agent 强化训练方法。
本文创建了 **MAT Benchmark**：覆盖图像搜索与图像编程两类工具使用场景，用于评估多模态Agent

背景

越来越多的商业大模型（比如 OpenAI 的 o3）已经不是单纯地回答问题，而是可以：

- 主动思考（规划任务、分解子任务）
- 使用工具（比如：搜索引擎、Python 代码）
- 联动多模态（图像+文字）解决真实问题

但目前开源模型普遍缺乏这种能力，尤其在“视觉 + 工具”的任务上。

VISUAL-ARFT

- 实验对象：Qwen2.5-VL 3B/7B
- 强化学习算法：GRPO
- Reward设计

$$R_{\text{total}}(q, o) = R_{\text{format}}(o) + R_{\text{acc}}(q, o)$$

框架结构

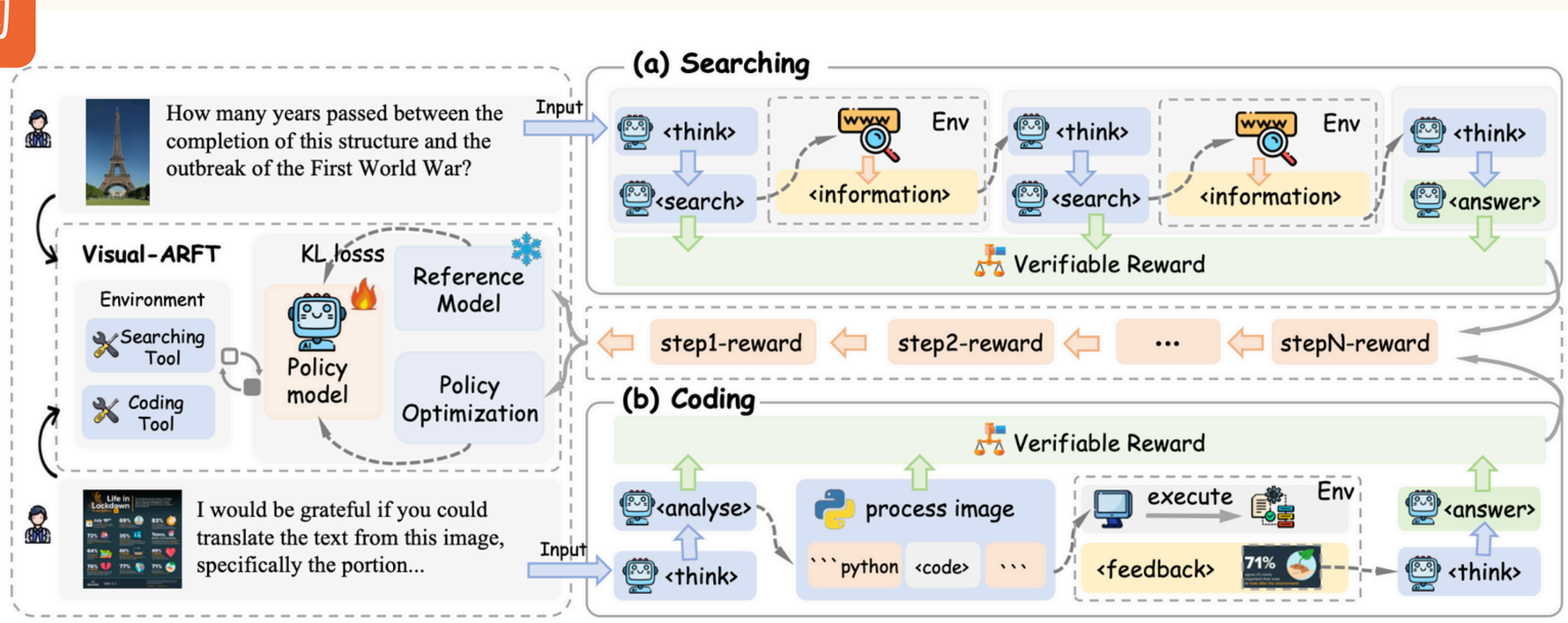

Figure 2: **Overview of Visual-ARFT.** We successfully empower LVLMs with multimodal agentic capabilities, including (a) agentic search and (b) agentic coding, enabling them to solve complex multimodal tasks through reasoning, decomposition, and tool interaction.

实验结果

Models	Reasoning with Tools	MAT-Coding						MAT-Search					
		Simple		Hard		Avg		Simple		Hard		Avg	
		F1	*EM*	*F1*	*EM*	*F1*	*EM*	*F1*	*EM*	*F1*	*EM*	*F1*	*EM*
GPT-4o [10]	✗	47.12	38.57	27.57	15.38	34.41	23.5	68.55	61.33	53.61	42.67	61.08	52.00
OpenAI-o3 [29]	✓	70.38	65.38	75.00	70.59	72.99	68.33	79.72	70.67	63.74	52.00	71.73	61.33
LLaVa-v1.5-7B [24]	✗	19.50	12.86	9.30	5.38	12.87	8.00	56.55	52.00	30.32	25.33	43.44	38.67
LLaVa-Next-7B [17]	✗	30.78	17.14	17.11	10.00	21.89	12.5	63.27	56.00	38.75	29.33	51.01	42.67
LLaVa-OneVision-7B [16]	✗	39.86	28.57	16.05	11.54	24.38	17.5	61.78	54.67	31.66	26.67	46.72	40.67
Xcomposer2.5 [51]	✗	36.06	22.86	19.90	10.77	25.56	15.0	60.16	54.67	31.93	28.00	46.04	41.33
InternVL2.5-8B [3]	✗	39.48	28.57	26.62	13.85	31.12	19.00	61.72	53.33	41.69	33.33	51.70	43.33
Qwen2.5-VL-3B [1]	✗	46.29	35.71	17.98	13.85	27.89	21.50	57.54	50.67	33.11	26.67	45.32	38.67
+ Visual-ARFT	✓	49.78	40.00	28.42	13.08	35.90	22.50	56.41	50.67	45.55	36.00	50.98	43.33
Δ	–	**+3.49**	**+4.29**	**+10.44**	**-0.78**	**+8.01**	**+1.0**	**-1.13**	**+0.0**	**+12.44**	**+9.33**	**+5.66**	**+4.66**
Qwen2.5-VL-7B [1]	✗	55.23	40.00	19.67	11.54	32.12	21.50	67.40	61.33	39.59	32.00	53.49	46.67
+ Visual-ARFT	✓	60.10	51.43	45.60	25.38	50.68	34.50	71.78	66.67	55.77	44.00	63.77	55.33
Δ	–	**+4.87**	**+11.43**	**+25.93**	**+13.84**	**+18.56**	**+13.00**	**+4.38**	**+5.37**	**+16.18**	**+12.00**	**+10.28**	**+8.66**

CONCLUSION

从文本LLM为核心的reasoning, agent发展到如何以多模态大模型为核心做reasoning, agent