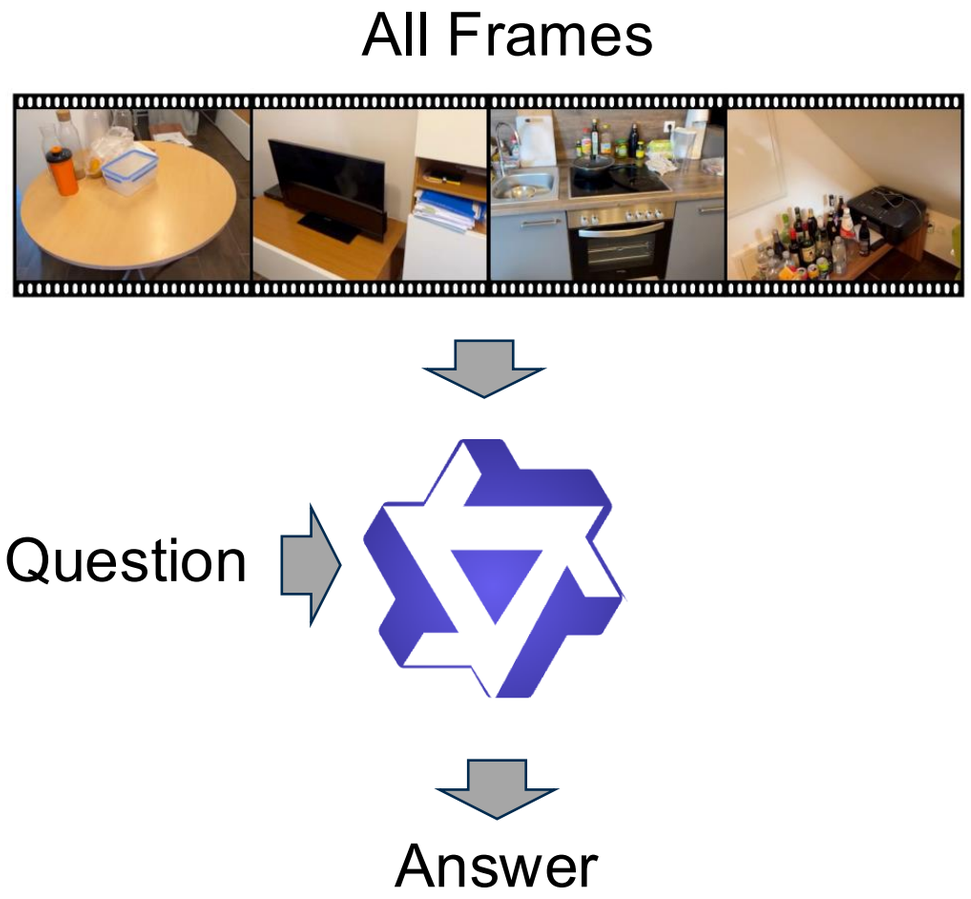
Active Exploring like a Pigeon: Reinforcing Spatial Reasoning via Agentic Vision-Language Models

Wei Deng[12], Xianlin Zhang[23], Mengshi Qi*[12]

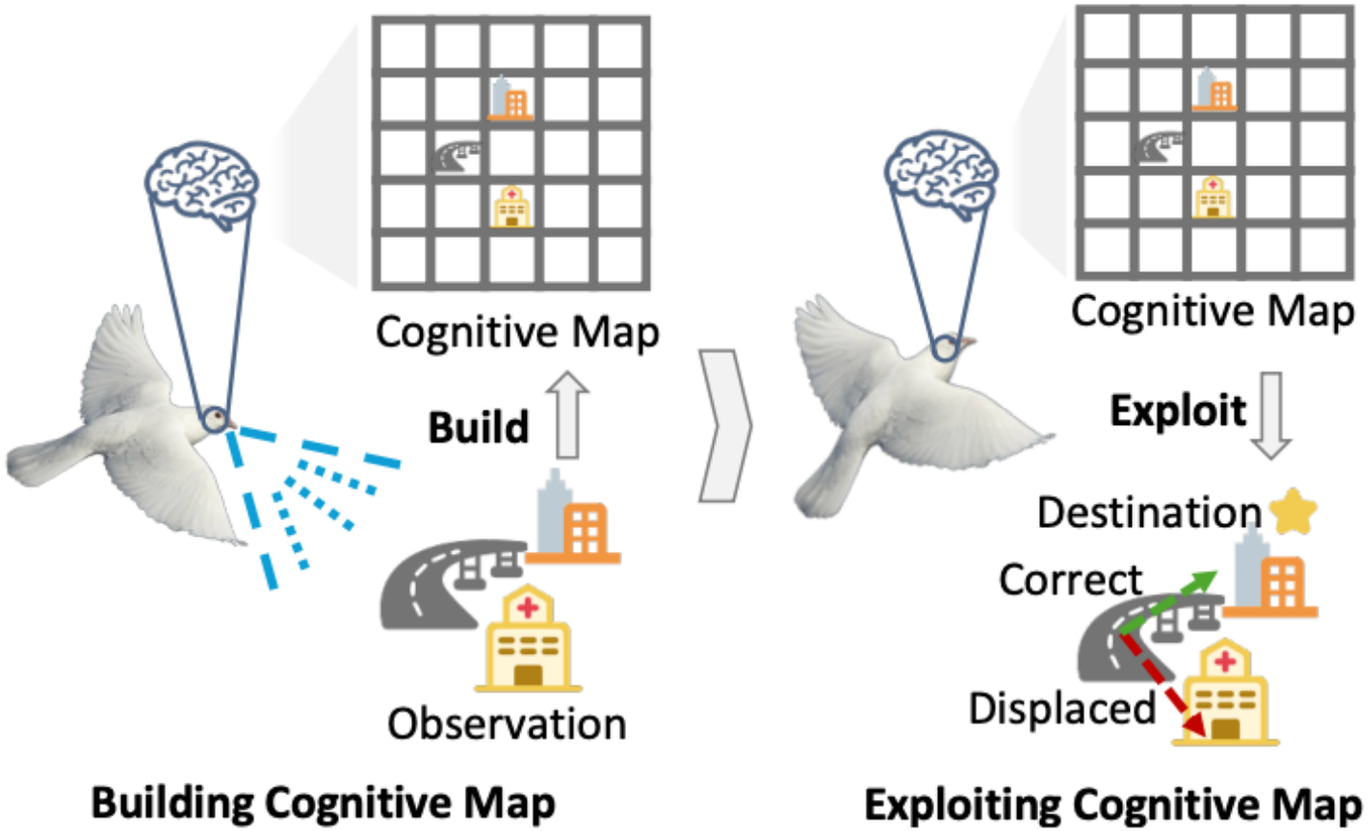
43nd International Conference on Machine Learning (ICML), 2026

[1]State Key Laboratory of Networking and Switching Technology
[2]School of Digital Media & Design Arts
[3]Beijing University of Posts and Telecommunications, China
*Corresponding author

Introduction & Motivation



VLM serves as a passive observer for spatial reasoning



Homing pigeons build and exploiting cognitive maps for navigation

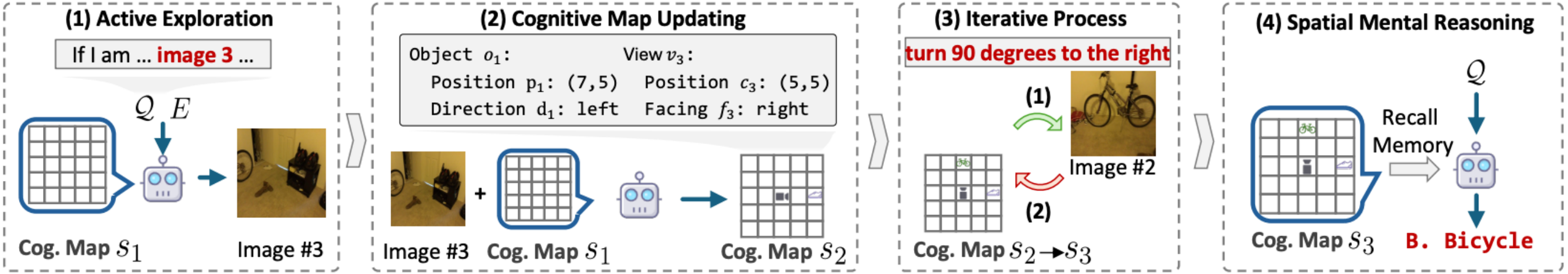
The Core Idea: Active Perception *vs* Passive Perception

Previous: Passive perception

Ours: Active perception

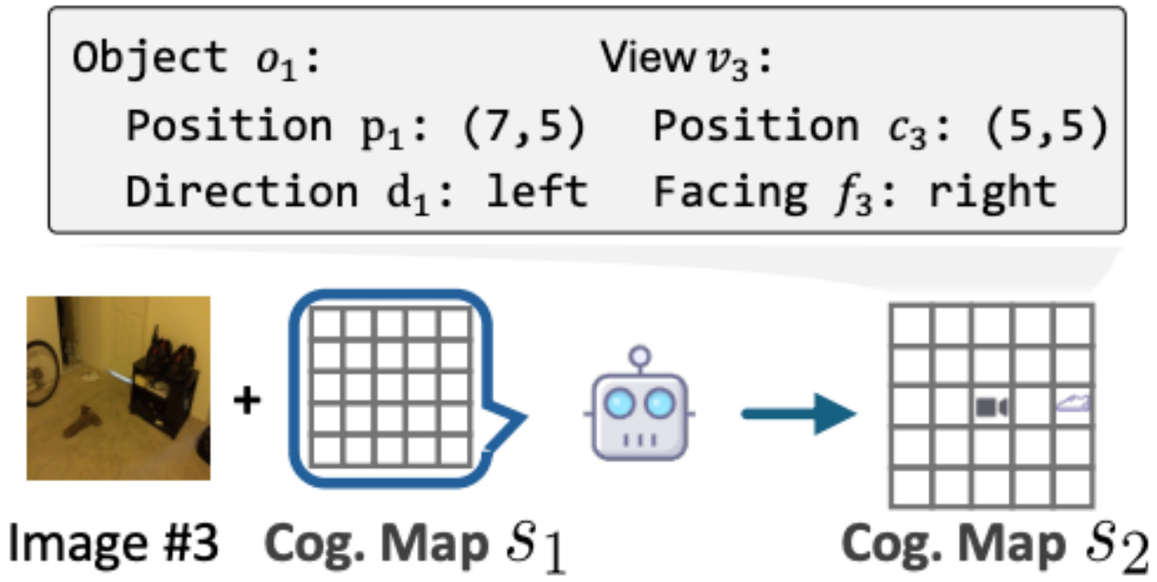
Methodology: Agentic Pipeline

1. Active exploration
2. Cognitive map updating
3. Iterative process
4. Spatial mental reasoning



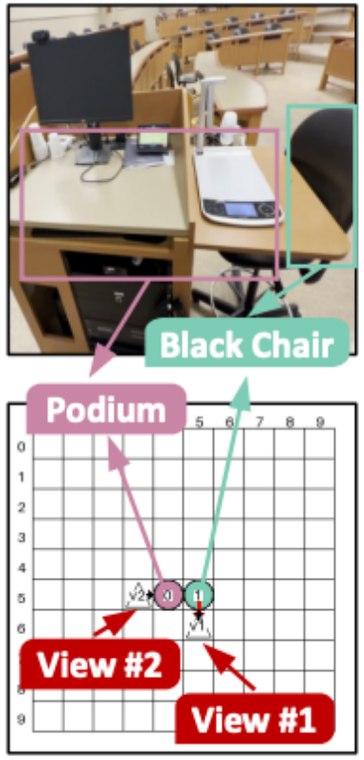
Methodology: Dynamic Cognitive Map

- A unified top-down coordinate system
- Storing objects with their positions and orientations, and camera viewpoints
- A persistent memory



Methodology: Spatial Assertion Codes (SAC)

- SAC is an executable Python expression
- SAC describe spatial relationships
- Example: object_1 on the left of object_2 from the perspective of view 4 -> obj1 in obj0.left(view=v4)

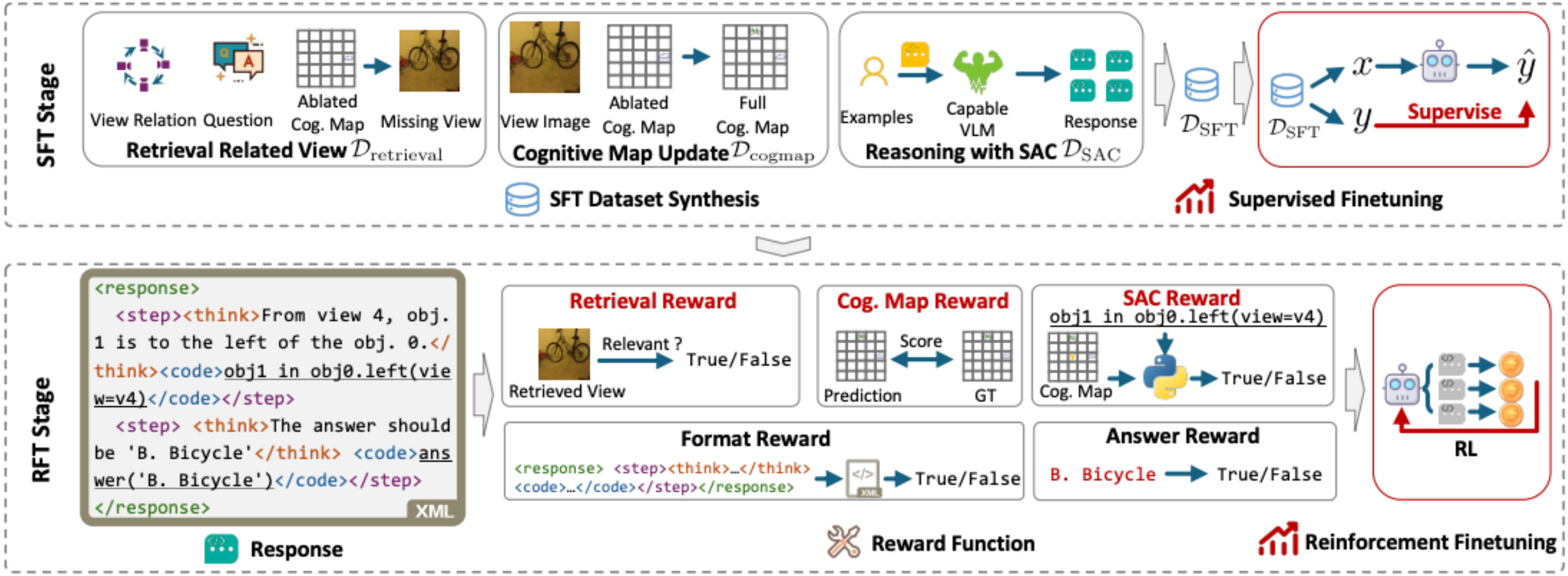



```xml
<step>
 <think>According to the cognitive map, the first view
 (v1) is at `(5, 6)` and the second view (v2) is at `(
 3, 5)`. The "podium" (obj0) is at `(4, 5)` and the "b
 lack chair" (obj1) is at `(5, 5)`. The second view is
 in the forward direction relative to the first view.<
 /think>
 <code>v2 in v1.front(view=v1) and v2 in v1.left(view=
 v1)</code>
</step>
<step>
 <think>So, the answer is B.</think>
 <code>agent.answer('B')</code>
</step>
```

XML

Methodology: Post-Training Process

- SFT: training with view retrieval, cognitive map updating, and spatial reasoning with SAC
- RFT: Finetuning with GRPO and dense reward functions



Experimental Results: MindCube Performance

- We achieve the best performance on MindCube
- We achieve a significant improvement on the Rotation subset, which is important for EAI

Methods	MindCube-Tiny Benchmark				
	Features	Overall ↑	Rotation ↑	Among ↑	Around ↑
Baseline					
Random (chance)	–	32.35	36.36	32.29	30.66
Random (frequency)	–	33.02	38.30	32.66	35.79
Open-Source VLMs					
Qwen2.5-VL-7B-Instruct (Bai et al., 2025)	Passive	29.26	38.76	29.50	21.35
Qwen2.5-VL-3B-Instruct (Bai et al., 2025)	Passive	33.21	37.37	33.26	30.34
Proprietary Models					
GPT-4o (OpenAI, 2024)	Passive	38.81	32.65	40.17	29.16
Claude-4-Sonnet-20250514 (Anthropic, 2024)	Passive	44.75	48.42	44.21	47.62
Spatial VLMs					
Spatial-MLLM (Wu et al., 2025a)	Passive, 2D+3D	32.06	38.39	20.92	32.82
Space-Qwen (Chen et al., 2024a)	Passive	33.28	38.02	33.71	26.32
$\text{MindCube}_{\text{Qwen2.5-VL-3B}}$ (Yin et al., 2025)	Passive, Cog. Map	70.7	48.0	79.2	68.4
$\text{3DThinker}_{\text{Qwen2.5-VL-3B}}$ (Chen et al., 2026)	Passive, 3D Rec.	75.2	55.5	**81.8**	75.2
Ours	Active, Cog. Map	**80.5**↑5.3	**85.0**↑29.5	81.0↓0.8	**75.6**↑0.4

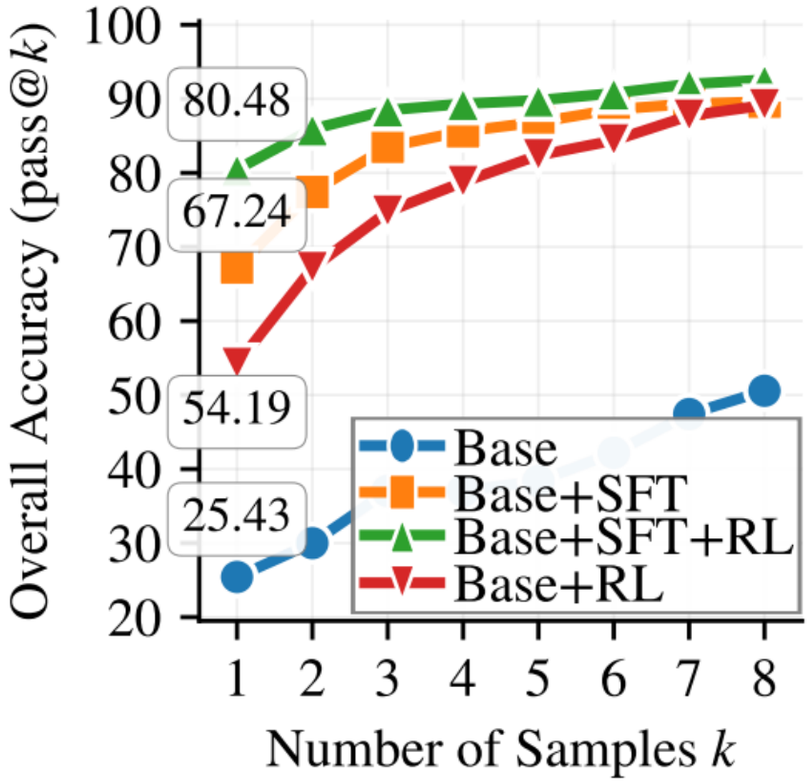
Experimental Results: Ablation Studies

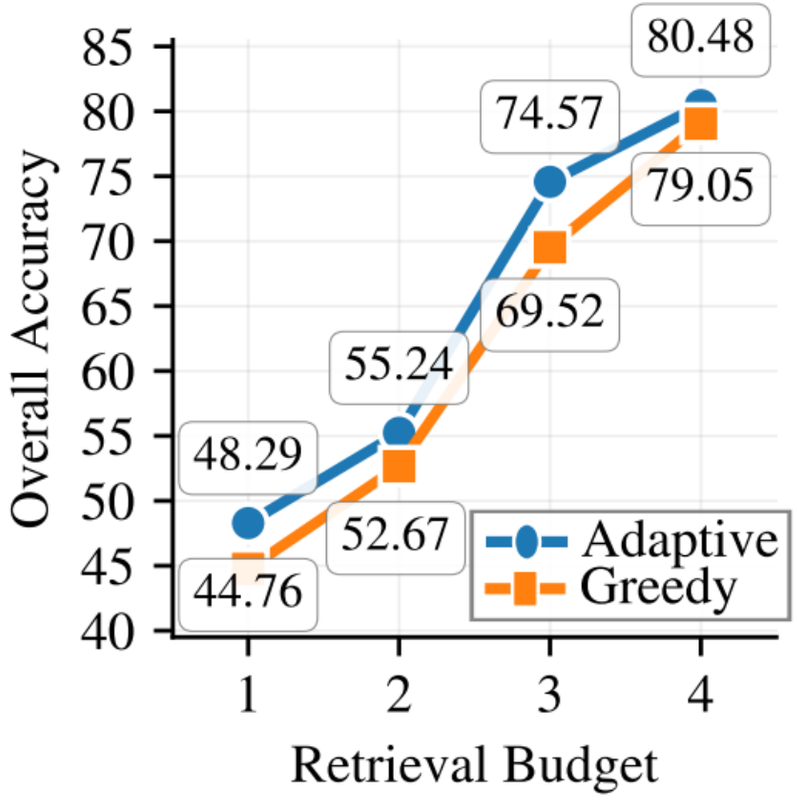
- Active perception pipeline outperforms passive
- All reward components contribute to the final performance
- The cognitive map provides clear gains over plain context

Setting	Acc. ↑	Pre. ↑	Rec. ↑	F1 ↑
Study of Perception Pipeline				
Passive	27.5	26.7	27.3	26.6
Active (Ours)	**38.5**↑11	**30.6**↑3.9	**29.2**↑1.9	**29.0**↑2.4
Reward Component Ablation				
Ours Full	**80.4**	**79.2**	**79.3**	**79.2**
Ours w/o $R_{\text{retrieval}}$	72.5↓7.9	57.4↓21.8	57.0↓22.3	57.1↓22.1
Ours w/o R_{cogmap}	72.6↓7.8	57.3↓21.9	57.0↓22.3	57.1↓22.1
Ours w/o R_{SAC}	70.2↓10.2	55.5↓23.7	54.5↓24.8	54.9↓24.3
Study of Memory Mechanism				
Context	50.9	36.2	34.8	34.8
Cog. Map (Ours)	**54.2**↑3.3	**40.6**↑4.4	**39.4**↑4.6	**39.5**↑4.7

Experimental Results: Ablation Studies

- Both of SFT and RFT post-training stages are essential
- Our adaptive retrieval strategy always surpasses the greedy one.





Core Contribution

- Re-formulate spatial reasoning and propose an agentic pipeline
- Propose SAC for RLVR
- Achieve state-of-the-art performance on MindCube

ICML
International Conference
On Machine Learning

Thank You!