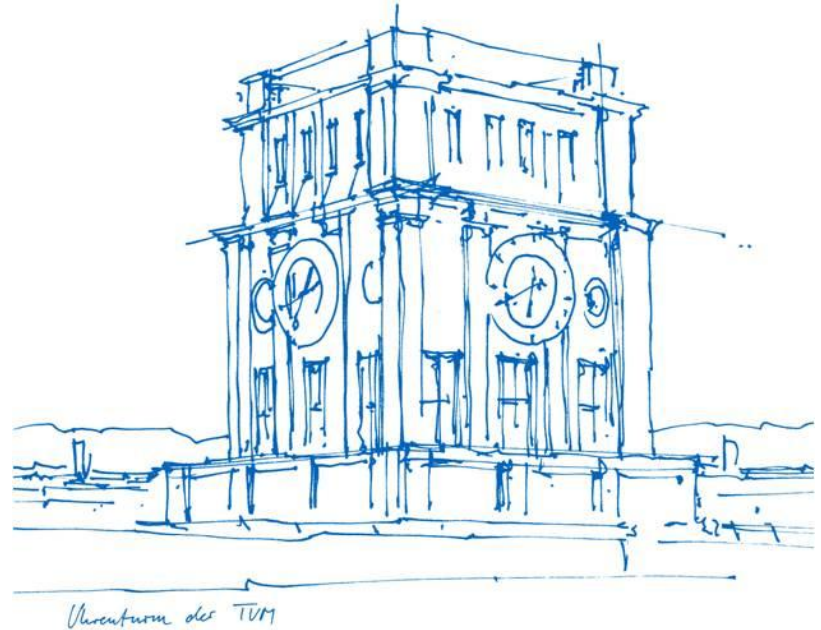


Reason-RFT: Reinforcement Fine-Tuning for Visual Reasoning

Technical University of Munich

Bivek Panthi

Munich, 16th July 2025



Overview

1 Background

2 Motivation

3 Methodology

4 Experimental Setup

5 Results

6 Training Insights

7 Advantages

Background: Vision Language Model (VLM)

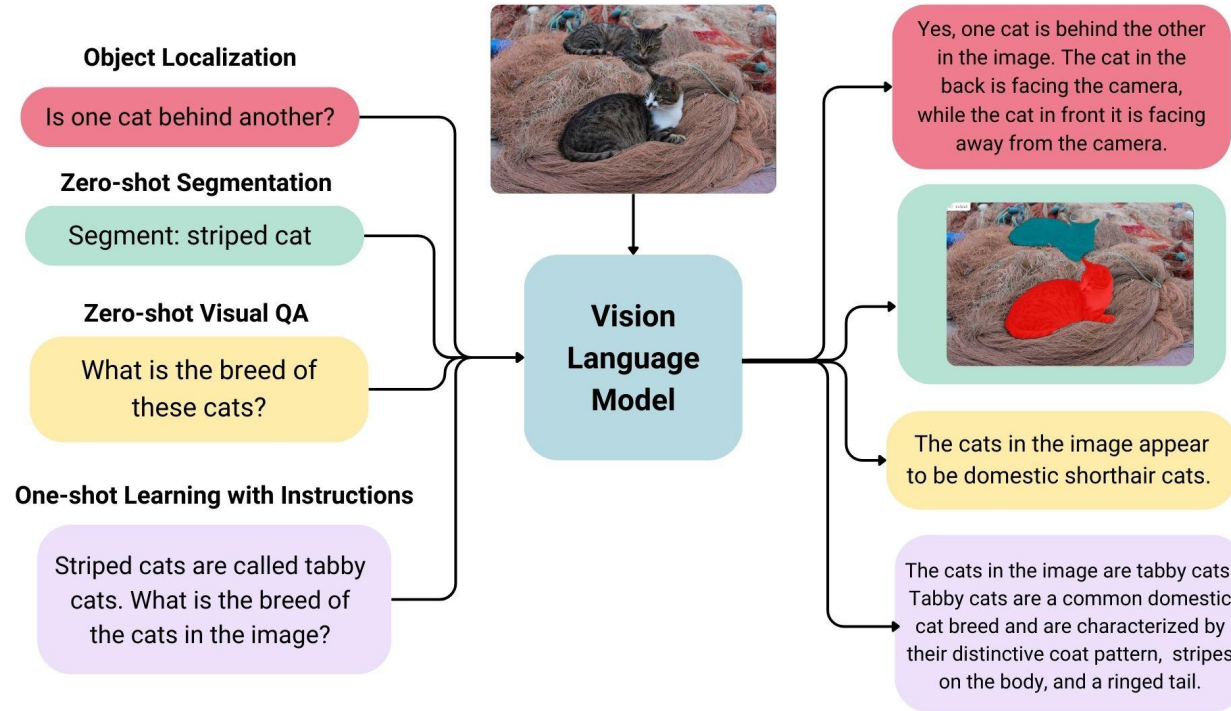


Fig: VLM [1]

Background: Vision Language Model (VLM)

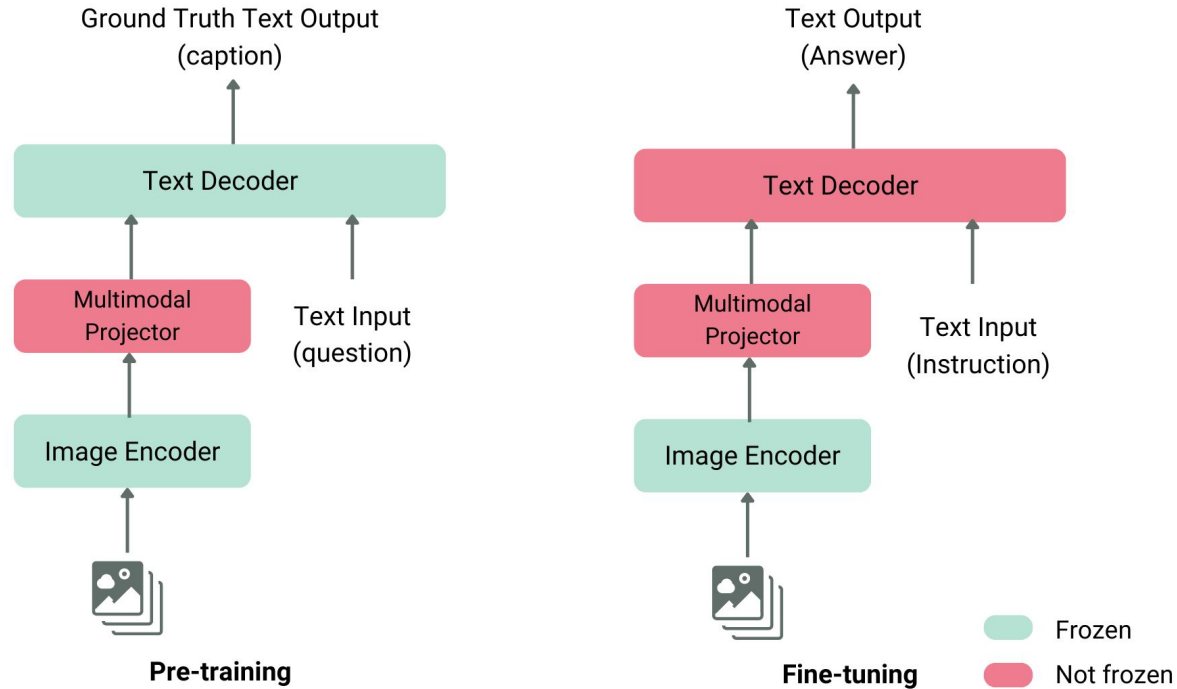


Fig: Pre-training vs Fine-tuning [1]

Background: Supervised Fine-Tuning with Chain-of-Thought data

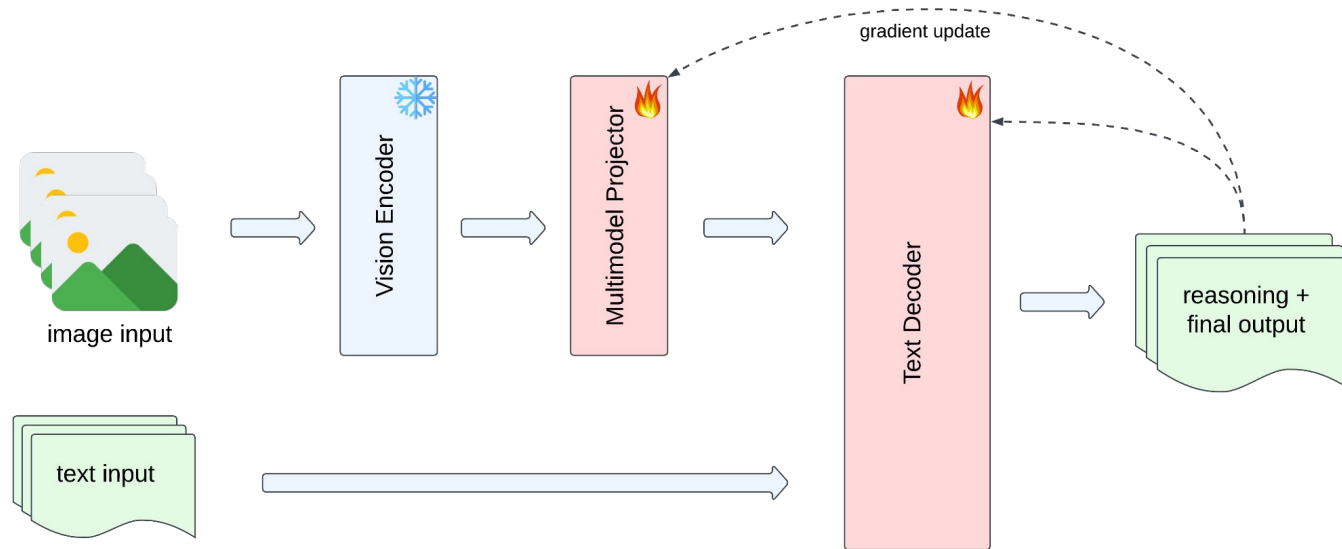


Fig. Supervised Fine-Tuning (SFT) with Chain-of-Thought (CoT) data

Background: Group Relative Policy Optimization

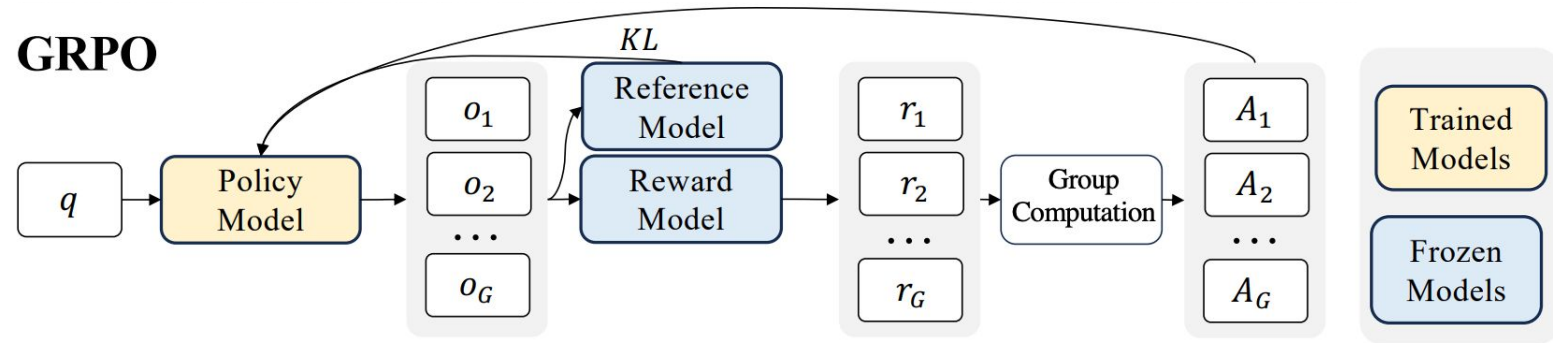
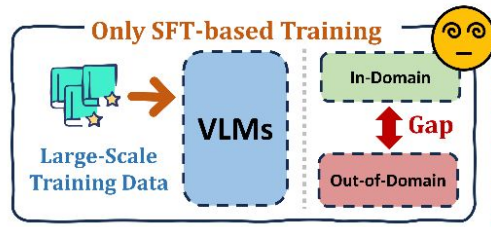


Fig: Group Relative Policy Optimization (GRPO) [2]

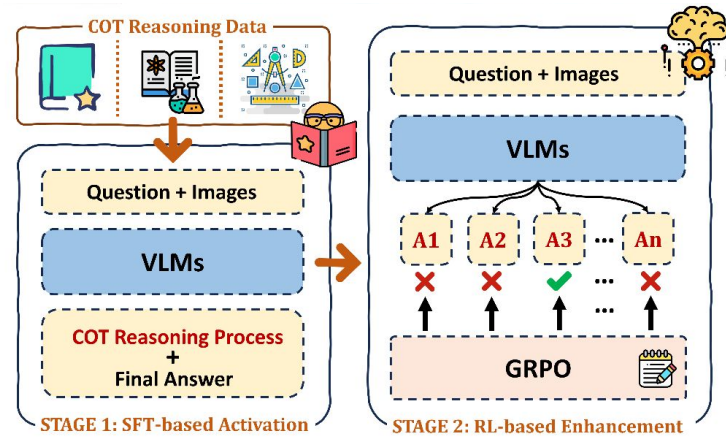
Motivation

Existing Approach



causes overfitting and cognitive rigidity,
limiting cross-domain generalization

Reason-RFT Proposal



CoT data in SFT to activate
reasoning capabilities

generate diverse reasoning-response
pairs, improving generalization.

Methodology : STAGE 1

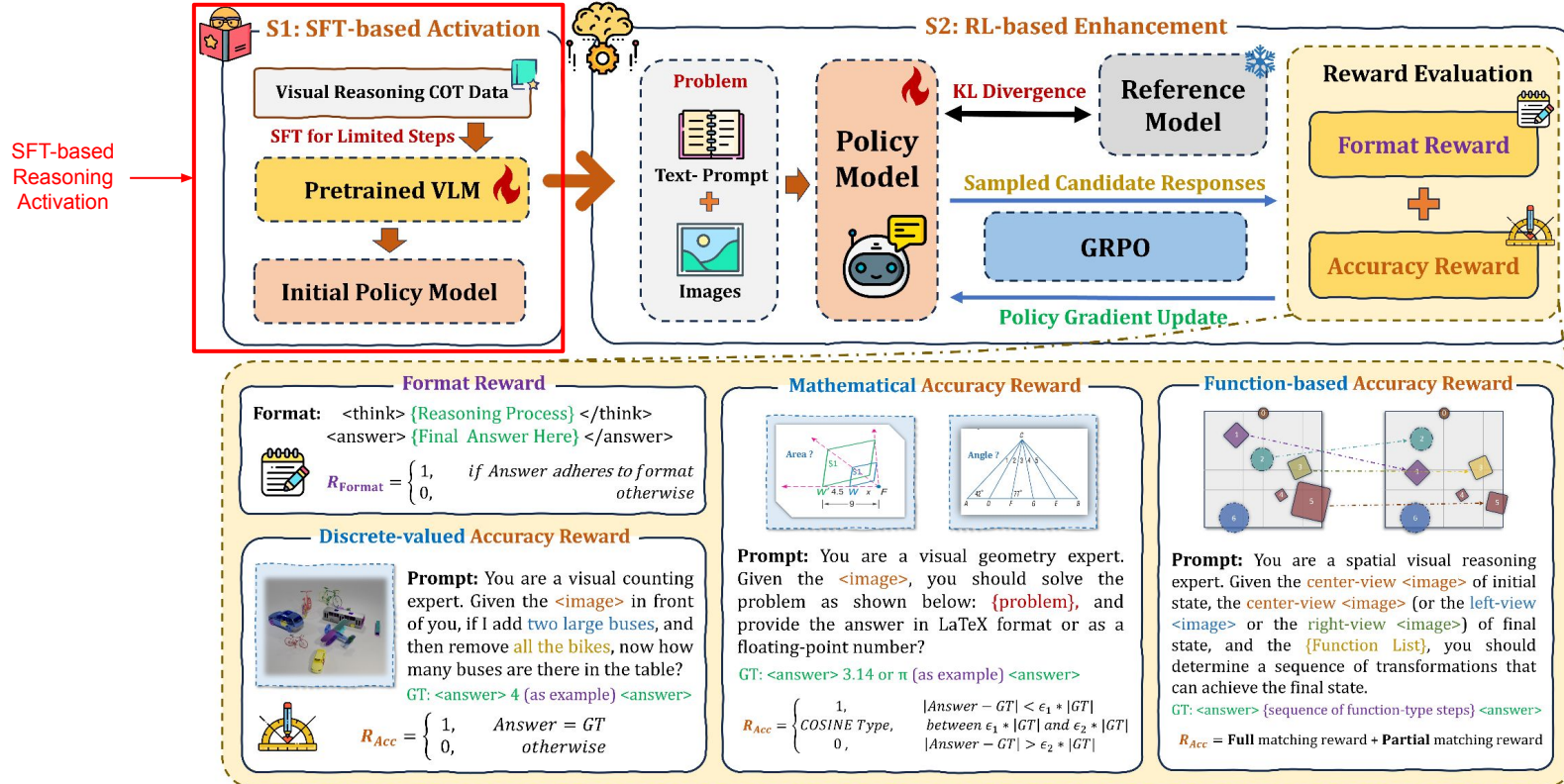


Fig: Framework of Reason-RFT [3]

Methodology : STAGE 2

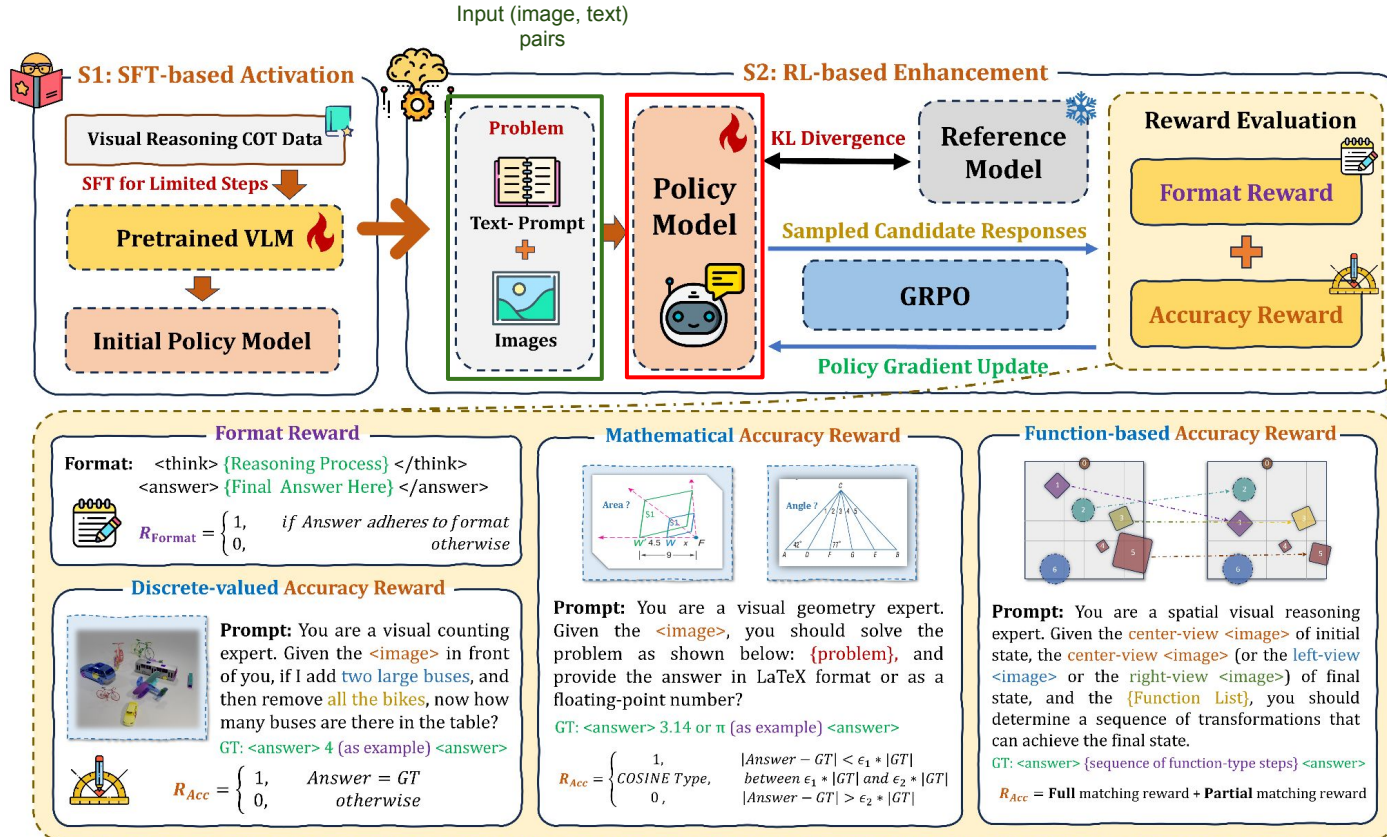


Fig: Framework of Reason-RFT [3]

Methodology : STAGE 2

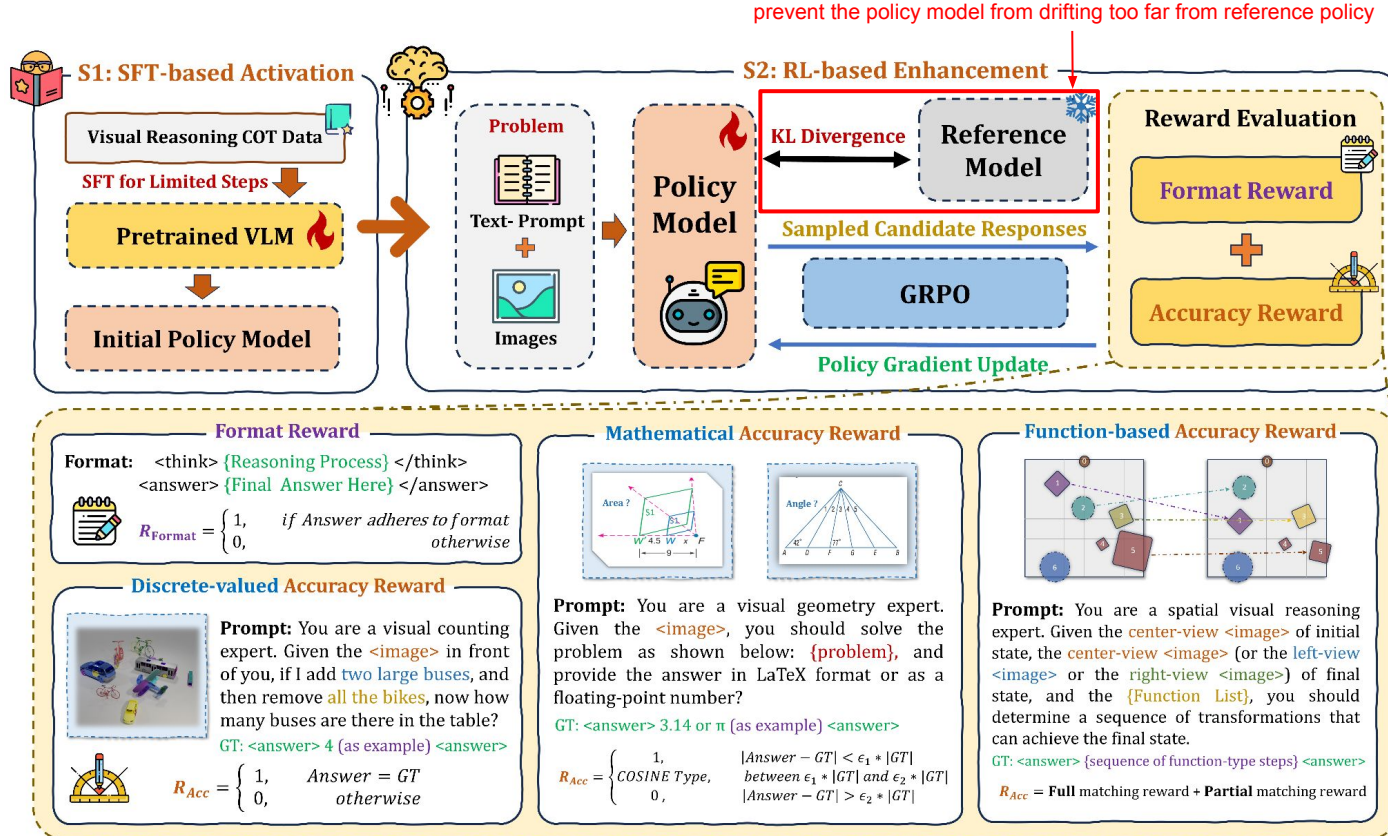


Fig: Framework of Reason-RFT [3]

Methodology : STAGE 2

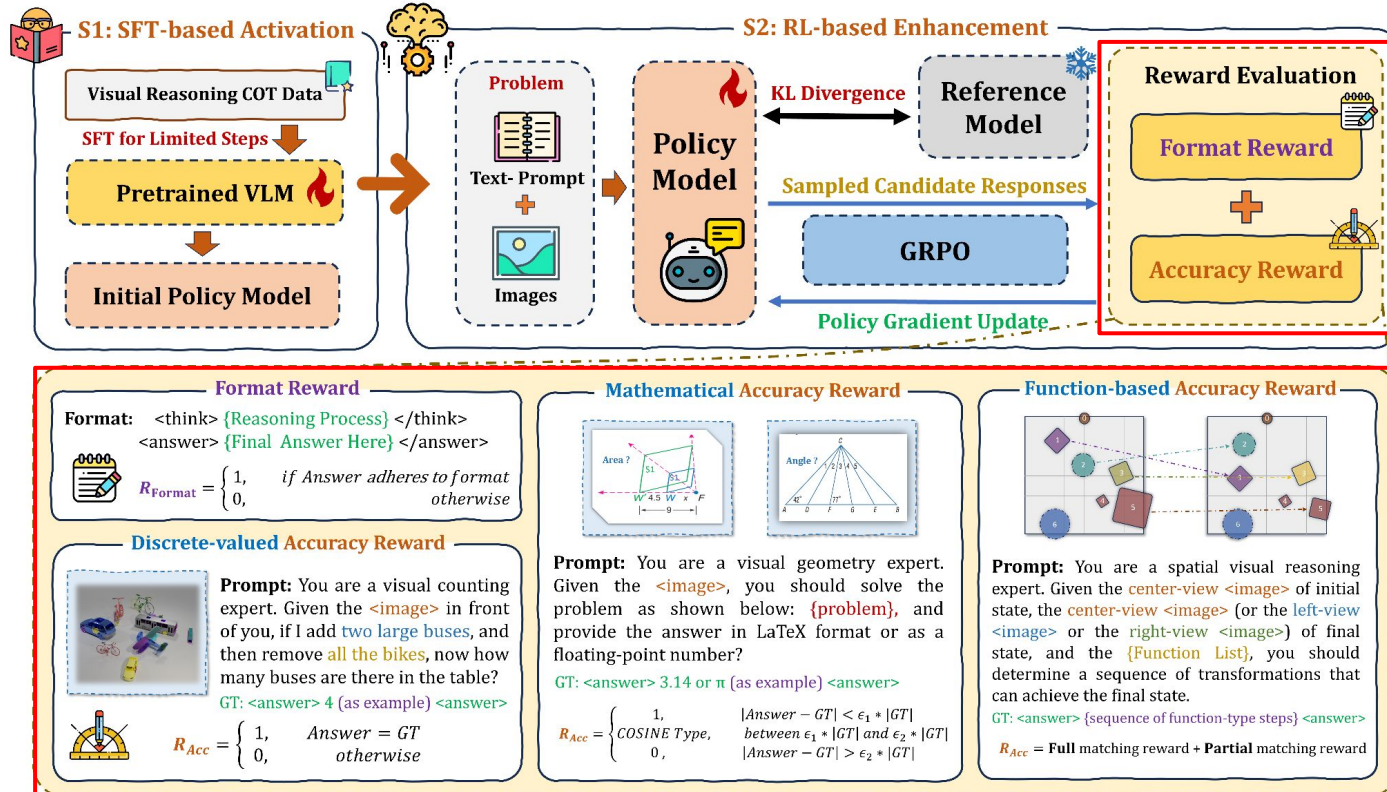


Fig: Framework of Reason-RFT [3]

Methodology : STAGE 2

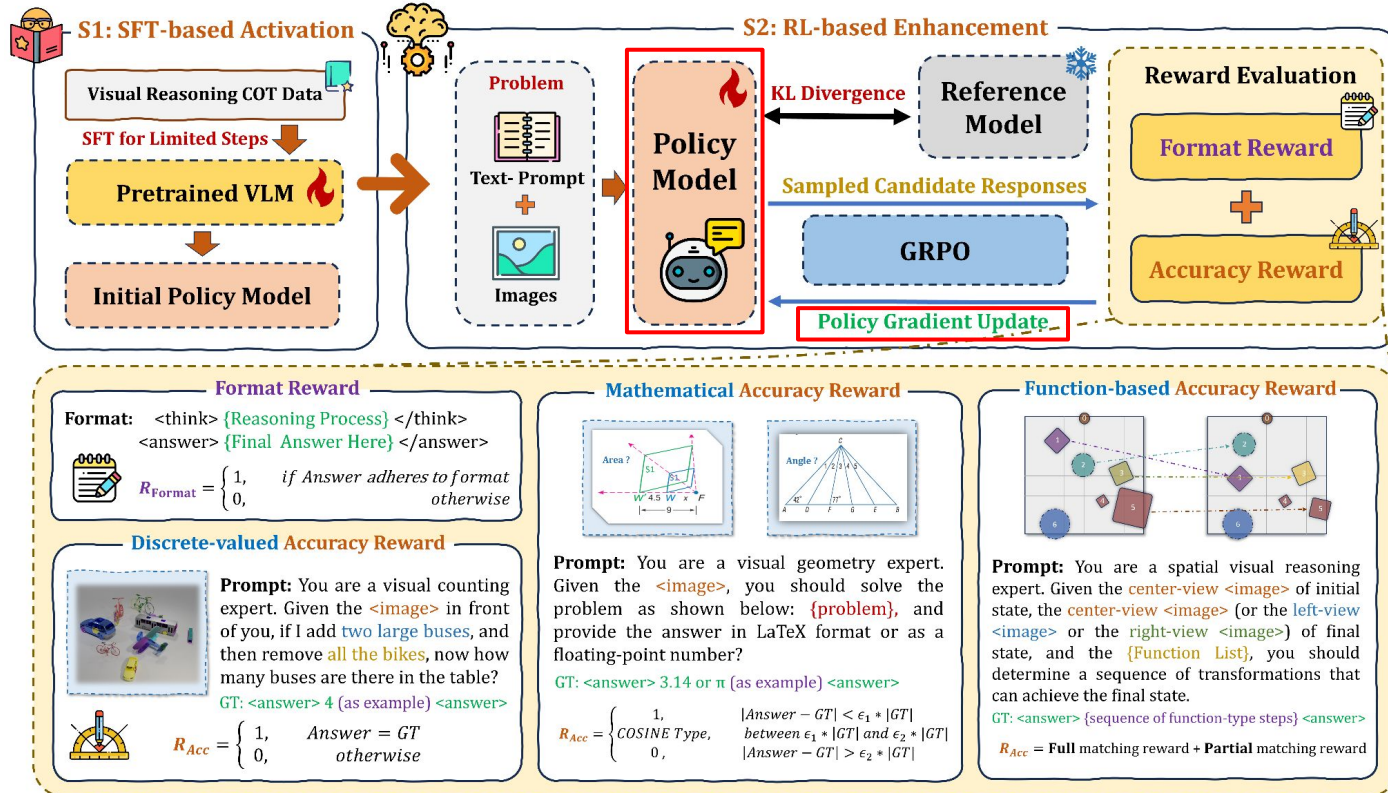
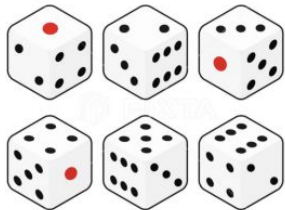


Fig: Framework of Reason-RFT [3]

Results



COT-SFT:

<think>Let me think about the properties of a standard six-sided die. On a six-sided die, each number from 1 to 6 is opposite to another number. The sum of the numbers on opposite sides is always 7. Number 4 is opposite to number 2 because $4 + 2 = 6$, and the sum of 6 and 1 is 7. Therefore, the number directly opposite to number 4 is 2.</think>

<answer>B</answer>

Question: Based on the image, determine which number is directly opposite to number 4 on the die depicted. Use reasoning to arrive at the correct answer from the choices provided.
Choices: A. 1 B. 2 C. 3 D. 6

Reason-RFT:

<think>Let's analyze the image. The numbers shown are 2, 3, and 4. To find the number opposite to 4, we need to consider the properties of a standard six-sided die. On a die, opposite numbers add up to 7. Since 4 is opposite to 3, we can conclude that the number opposite to 4 is 3.</think>

<answer>C</answer>

Fig: Visualization examples of Reason-RFT and CoT-SFT on Structure Perception task [3]

Experimental Setup : Task Categories

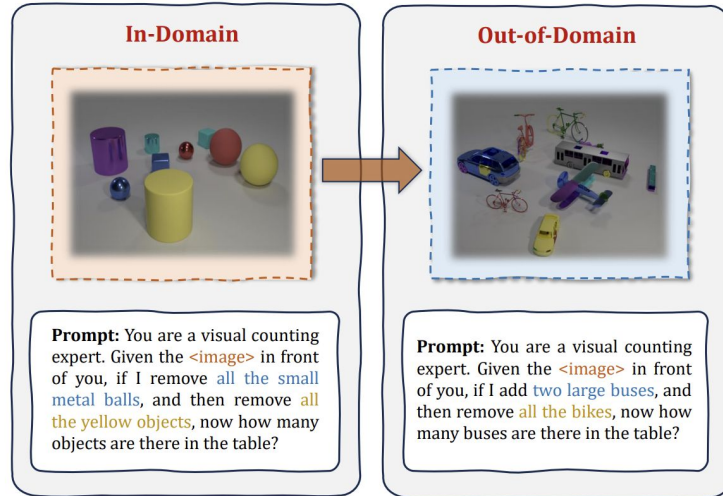


Fig: Example of Visual Counting [3]

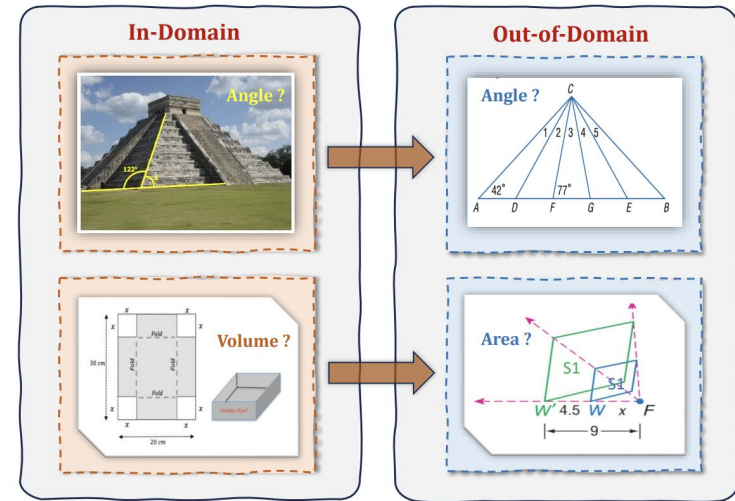


Fig: Example of Structure Perception [3]

Experimental Setup : Task Categories

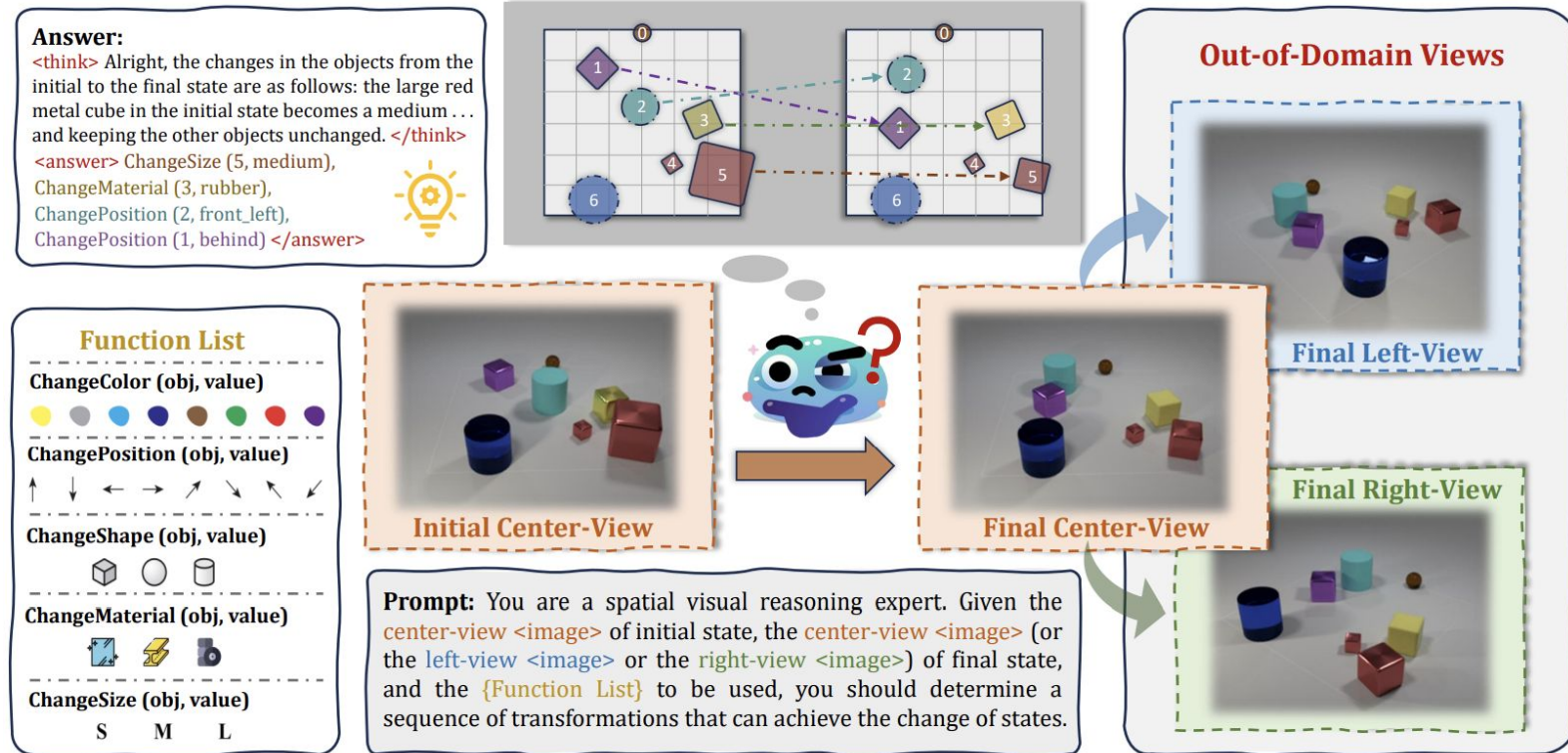


Fig: Example of Spatial Transformation [3]

Results

Method	Visual Counting			Structure Perception			Spatial Transformation			
	Clevr-Math ID	Super-Clevr OOD	AVG	GeoMath ID	Geometry3k OOD	AVG	TRANCE ID	TRANCE-L OOD	TRANCE-R OOD	AVG
Proprietary Models										
GPT-4o-2024-08-06 [26]	68.10	34.31	51.20	50.18	43.49	46.83	42.55	28.67	29.76	35.88
Gemini-1.5-Pro [55]	61.80	37.50	49.65	50.12	48.38	49.45	26.22	18.76	19.88	22.77
Open-Source Models										
Qwen2.5-VL-3B-Instruct [5]	75.90	39.30	57.60	36.75	37.44	37.09	8.57	8.26	8.31	8.42
Phi-3.5-Vision-4B-Instruct [1]	21.40	15.20	18.30	36.83	50.25	43.54	7.42	2.45	4.02	5.33
Llava-OneVision-7B [35]	69.70	29.10	49.40	77.63	43.66	60.64	10.00	8.33	8.74	9.27
Qwen2.5-VL-7B-Instruct [5]	74.60	35.20	54.90	44.00	45.61	44.80	19.63	13.12	13.42	16.45
InternVL-2.5-8B [6]	93.50	35.30	64.40	63.00	47.32	51.60	7.19	6.62	6.63	6.91
Llama-3.2-11B-Vision [44]	10.30	9.50	9.90	13.75	20.85	17.30	8.22	8.40	9.03	8.47
Pixtral-12B [3]	42.60	22.90	32.75	30.38	36.09	33.23	7.35	5.03	5.22	6.42
Qwen2VL-2B-Instruct										
Zero-Shot	82.40	32.00	57.20	25.86	20.63	23.25	3.78	4.60	4.67	4.35
+ ANS-SFT	96.20	39.20	67.70	51.34	22.50	36.92	77.39	49.24	50.33	58.99
+ CoT-SFT	85.50	46.50	66.00	43.05	25.25	34.15	64.37	43.19	42.86	50.14
+ Reason-RFT-Zero	98.40	44.80	71.60	47.68	32.50	40.09	42.13	34.07	33.41	33.74
+ Reason-RFT	96.80	51.20	74.00	49.03	33.13	41.08	74.61	64.05	64.08	67.58
Qwen2VL-7B-Instruct										
Zero-Shot	98.60	42.10	70.35	43.30	43.88	43.59	13.53	12.72	12.78	13.01
+ ANS-SFT	95.00	33.90	64.45	51.34	25.38	38.36	82.19	54.29	54.83	63.77
+ CoT-SFT	87.30	42.40	64.85	50.49	33.00	41.75	81.31	47.90	47.80	59.00
+ Reason-RFT-Zero	99.40	53.00	76.20	55.00	54.75	54.88	67.67	57.20	56.15	56.68
+ Reason-RFT	95.60	51.00	73.30	59.27	49.25	54.26	79.97	59.36	58.61	65.98

Fig: Results on three visual reasoning tasks [3]

Results

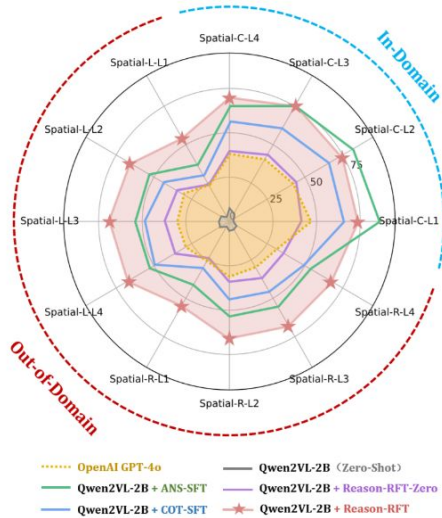
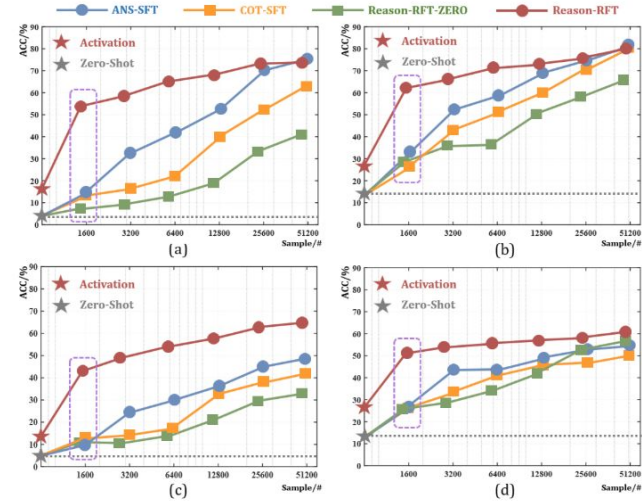


Fig: Out-of-Domain generalization for spatial transformation [3]

2B model

7B model

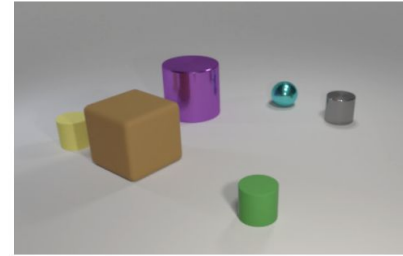
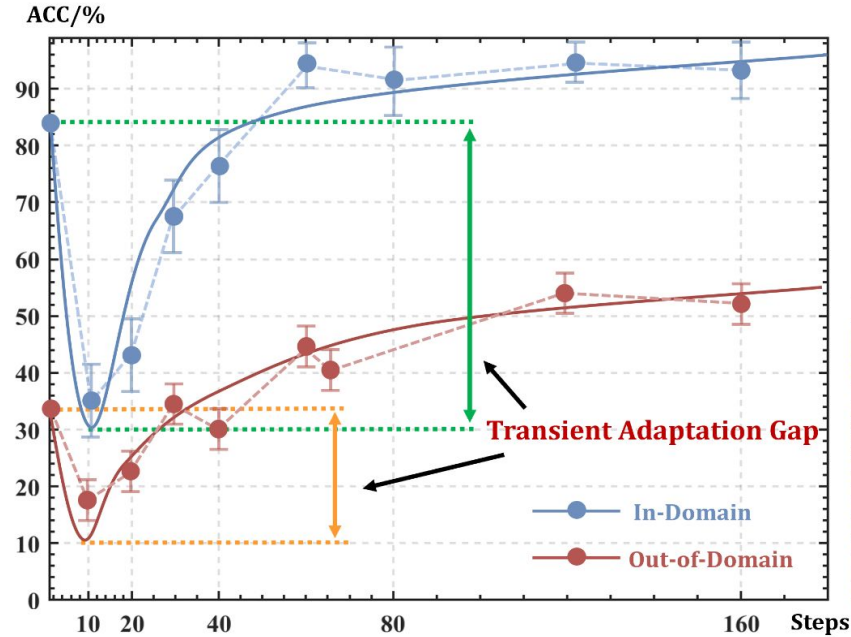


In-Domain

Out-of-Domain

Fig: more efficient utilization of training data [3]

Training Insights



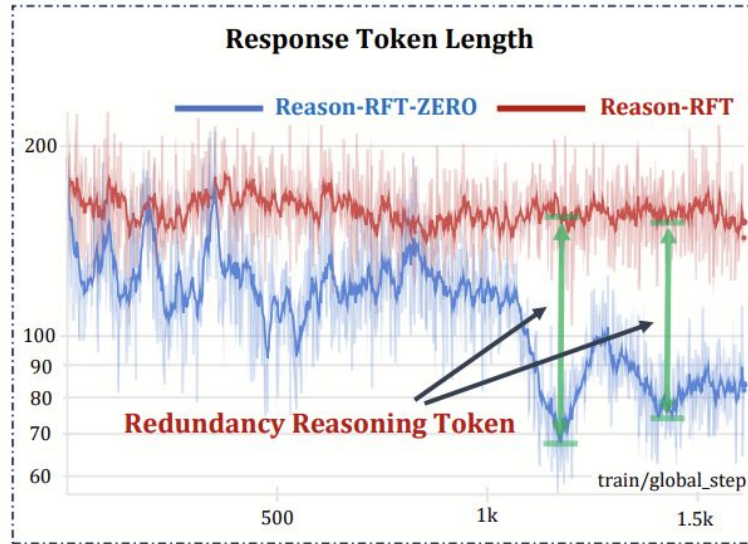
Prompt: You are a visual counting expert. Given the `<image>` in front of you, if I remove **all the blocks**, then how many objects are there in the table?

Direct Response: The final answer is 5.

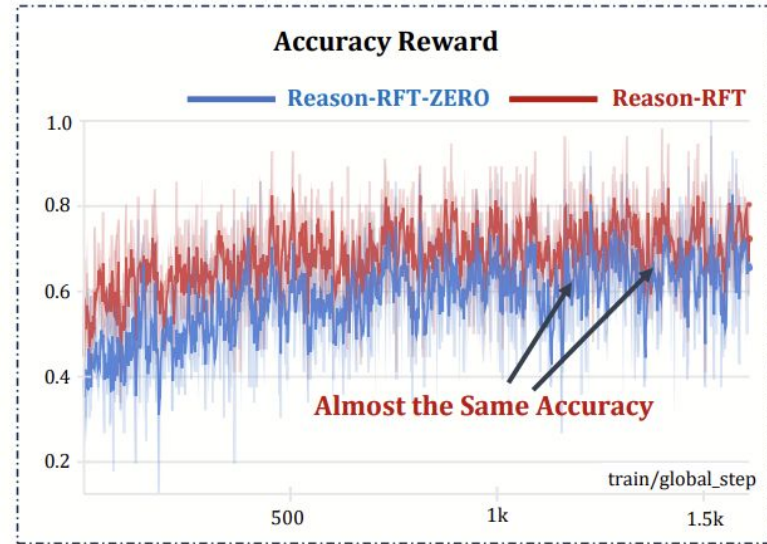
Reasoning Response: `<think>` Initially, there are 5 objects in the image: a yellow cylinder, a brown cube, a purple cylinder, a cyan sphere, and a gray cylinder. The brown cube is a block, so we need to remove it. **The final answer is $5 - 1 = 4$** `</think>`
`<answer> 4 </answer>`

Fig: Transient Adaptation Gap [3]

Training Insights



(a) reasoning token length during training on the Structure Perception task



(b) Accuracy Reward

Fig: Reasoning Redundancy [3]

Advantages

1

Performance Improvement: It significantly outperforms mainstream VLMs in tasks like visual counting, structure perception, and spatial transformation.

2

Enhanced Generalization: It consistently surpasses SFT-only and RL-only paradigms across diverse tasks, validated through multi-dimensional evaluations.

3

Data Efficiency: It achieves over 95% of the performance of SFT-only approaches using less than 20% of the data.

Thank You!

References

1. <https://huggingface.co/blog/vlms>
2. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models (<https://arxiv.org/abs/2402.03300>)
3. <https://arxiv.org/abs/2503.20752>