MedVR: Annotation-Free Medical Visual Reasoning via Agentic Reinforcement Learning

Zheng Jiang*, Heng Guo*, Chengyu Fang*, Changchen Xiao, Xinyang Hu, Lifeng Sun†, Minfeng Xu†

Tsinghua University · DAMO ACADEMY · ZHEJIANG UNIVERSITY

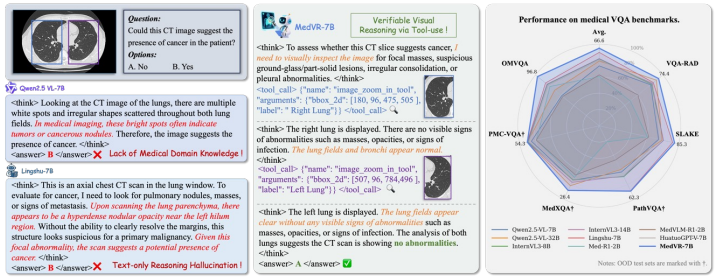
ICLR

Paper · WeChat · Github

Motivation

Medical VLMs need reliable reasoning over visual evidence, but existing approaches still rely heavily on text-only deliberation.

- Text-only reasoning misses fine-grained clinical details.
- It increases the risk of language shortcuts and visual hallucination.
- Clinical reasoning should be interactive and visually verifiable.



Challenges

Enabling medical visual reasoning is difficult because explicit grounding supervision is scarce and expensive.

- Intermediate ROI annotations are difficult to obtain.
- Final-answer rewards are too weak for visual credit assignment.
- Naive tool exploration is inefficient and unstable.

Contributions

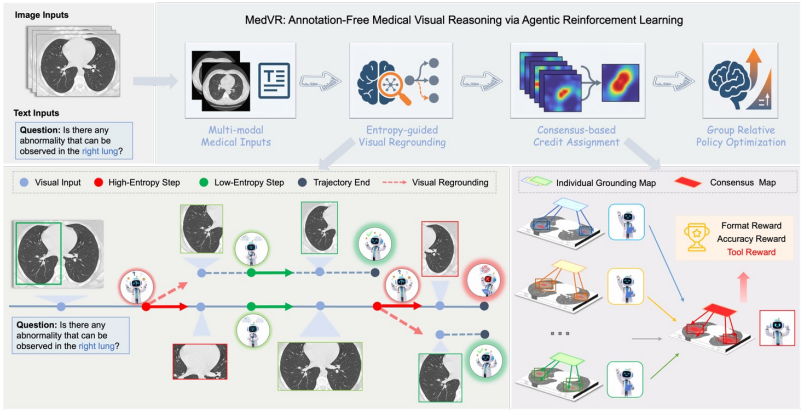
We propose MedVR, the first annotation-free RL framework for grounded medical visual reasoning.

- Interleaves reasoning with image manipulation tools.
- EVR enables uncertainty-aware visual exploration.
- CCA provides consensus-based pseudo supervision.
- Achieves SOTA on diverse medical VQA benchmarks.

Methods

MedVR trains a medical VLM to reason through an agentic loop of deliberation, tool use, observation, and reward-driven refinement.

- EVR monitors token entropy during tool generation and branches when uncertainty is high, exploring alternative ROIs.
- CCA aggregates successful trajectories into a consensus mask and rewards tool use that aligns with this shared visual strategy.



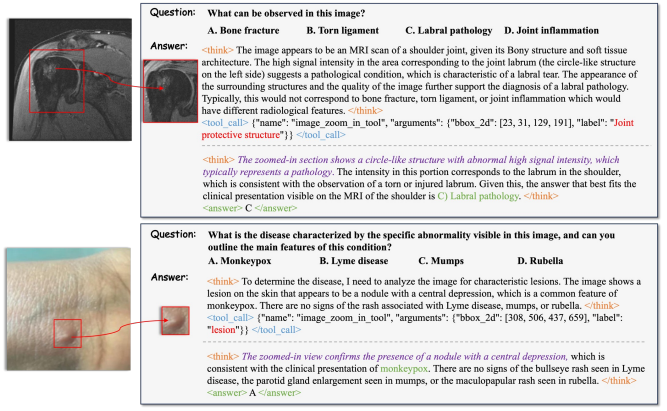
Main Results

MedVR sets new SOTA on six public medical VQA benchmarks.

Settings	General-Domain VQA (Multi-choice)				Modality-Specific VQA (Free-text)			
Tasks	OMVQA	PMC-VQA°	MedXQA°	Avg.	VQA-RAD	SLAKE	PathVQA°	Avg.
General-Purpose VLMs								
Qwen2.5-VL-7B	59.0	51.2	22.3	44.2	64.5	67.2	44.1	58.6
Qwen2.5-VL-32B	69.9	54.1	24.5	49.5	71.8	71.2	41.9	61.6
InternVL2.5-8B	83.9	51.3	21.0	52.1	59.4	69.0	42.1	56.8
InternVL3-8B	81.4	53.8	22.1	52.4	65.4	72.8	48.6	62.3
InternVL3-14B	81.9	54.1	23.1	53.0	66.3	72.8	48.0	62.4
Medical-Specific VLMs								
Med-R1-2B	77.3	47.4	21.1	48.6	39.0	54.5	15.3	36.3
MedVLM-R1-2B	57.4	47.5	20.1	41.7	48.6	56.0	32.5	45.7
MedGemma-4B	70.5	49.9	15.4	45.3	72.5	76.4	48.8	65.9
LLaVA-Med-7B	29.3	30.5	20.3	26.7	53.7	48.0	38.8	46.8
HuatuoGPT-V-7B	80.1	53.3	22.3	51.2	67.0	67.8	48.0	61.6
Lingshu-7B	84.2	**54.3**	**26.5**	55.0	67.9	83.1	61.9	70.3
MedVR (Ours)	**96.8**	**54.3**	26.4	**59.2**	**74.4**	**85.3**	**62.3**	**74.0**

Case Study

MedVR grounds its reasoning by iteratively inspecting clinically relevant image regions.



Ablation Study

Both EVR and CCA are critical to the final gains.

- The Zoom-in tool alone is insufficient for improving performance.
- EVR enhances exploration and improves OOD generalization.
- CCA provides more reliable supervision for visual grounding.

Methods	Components			Performance		
	Zoom-in	EVR	CCA	OmniMedVQA	PMC-VQA	MedXpertQA
Textual RL	—	—	—	94.50	53.40	21.38
MedVR	✓	—	—	94.31	52.62	22.26
	✓	✓	—	95.38	53.81	24.73
	✓	—	✓	96.55	53.30	23.09
	✓	✓	✓	**96.77**	**54.31**	**26.38**