浙江大學 ZHEJIANG UNIVERSITY

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

CVPR JUNE 3-7, 2026 DENVER COLORADO

IBISAgent: Reinforcing Pixel-Level Visual Reasoning in MLLMs for Universal Biomedical Object Referring and Segmentation

Yankai Jiang, Qiaoru Li, Binlu Xu, Haoran Sun, Chao Ding, Junting Dong, Yuxiang Cai, Xuhong Zhang, Jianwei Yin

Motivation: Why pixel-level reasoning matters?

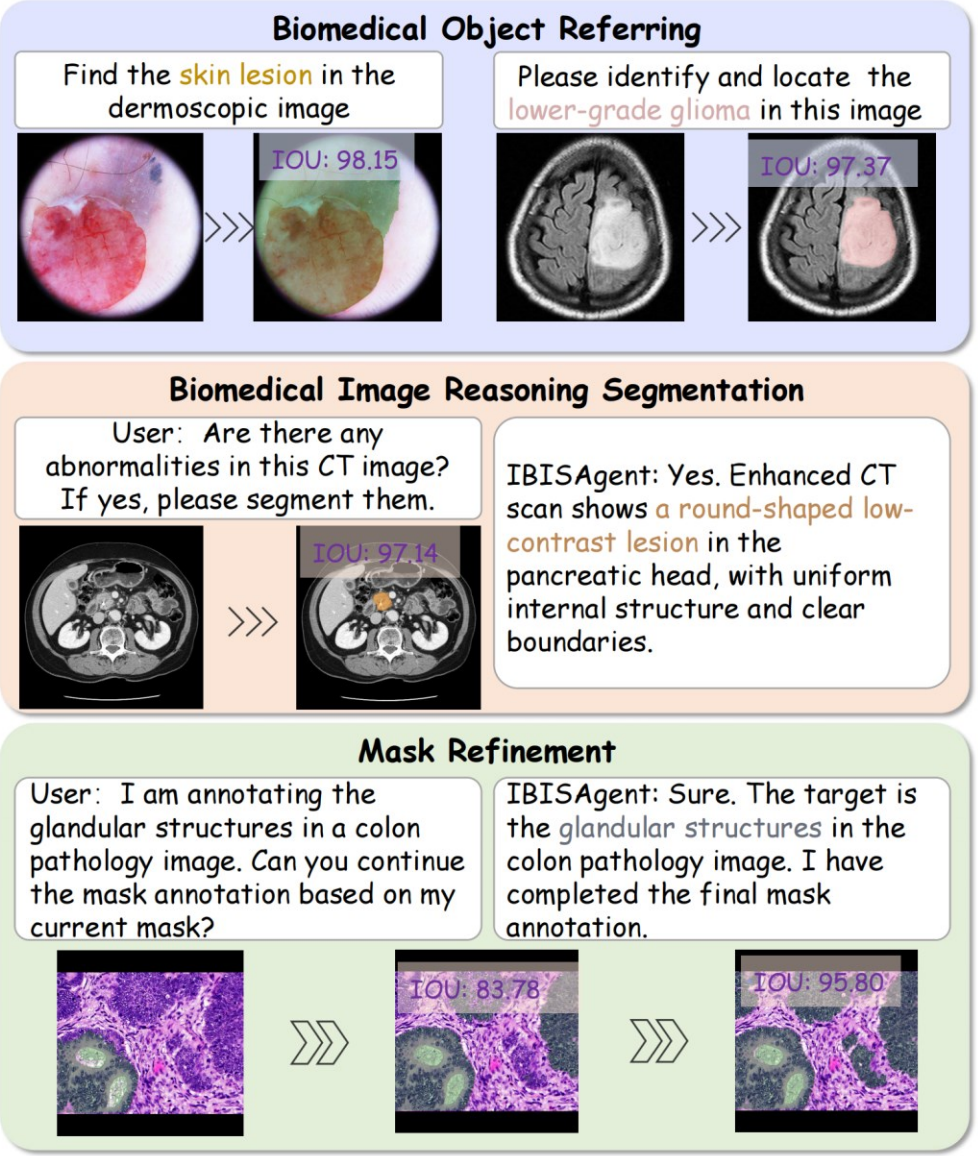
a. Medical MLLMs are strong at image-level QA, but **weak at pixel-level understanding**.

b. **Real clinical analysis is not a single QA step**. It requires localization, segmentation, and iterative verification.

c. **Biomedical images are especially challenging** because targets can be tiny, subtle, ambiguous, and visually heterogeneous.

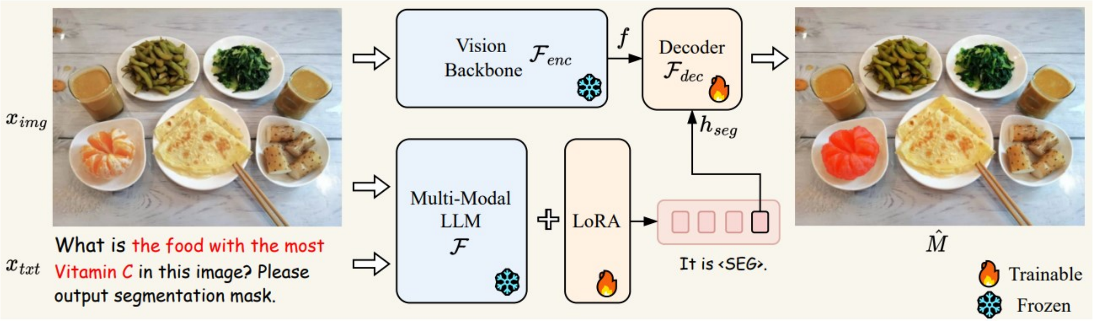
d. **Human annotators work interactively**: observe the current mask, place positive/negative clicks, and refine step by step.

Question: Can an MLLM behave like an annotator and segment through reasoning-driven interaction?

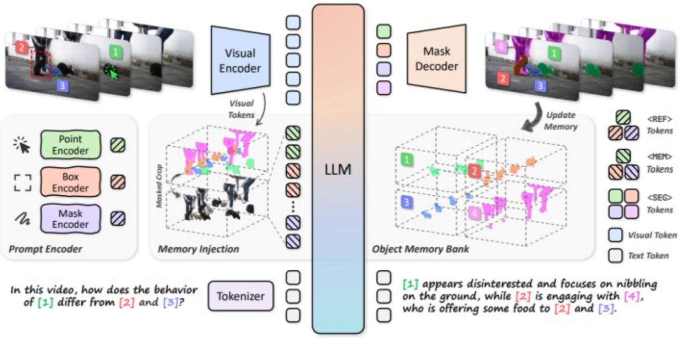


Limitations of existing methods

1. Rely on **special implicit segmentation tokens** (e.g. <seg>)

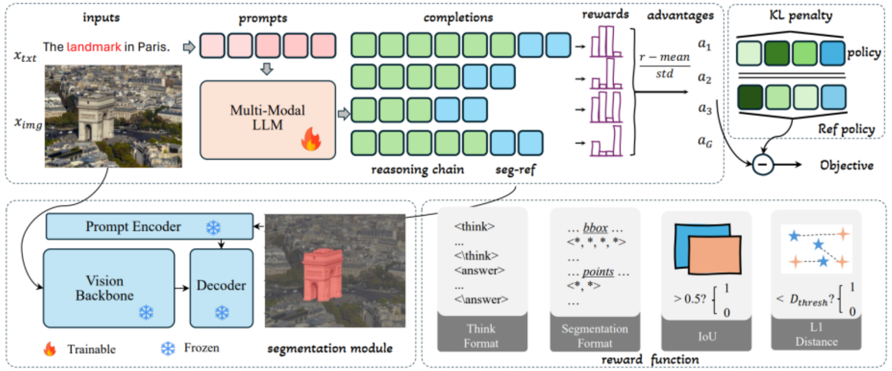


LISA (CVPR 2024)

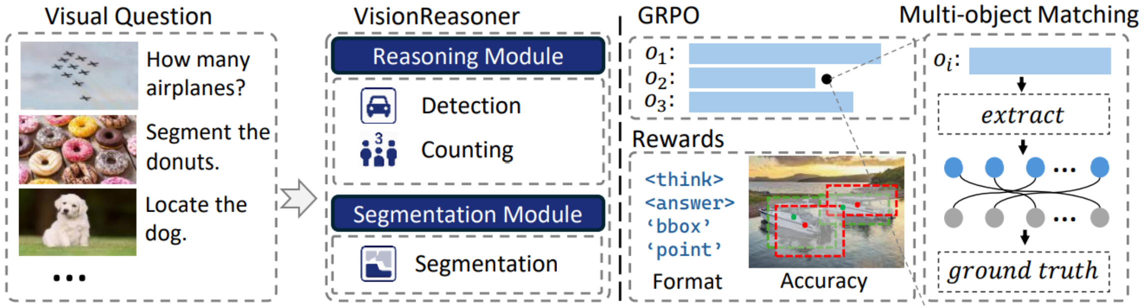


UniPixel (NeurIPS 2025)

2. Are **single-turn** and cannot autonomously revisit or refine



Seg-Zero (arxiv)



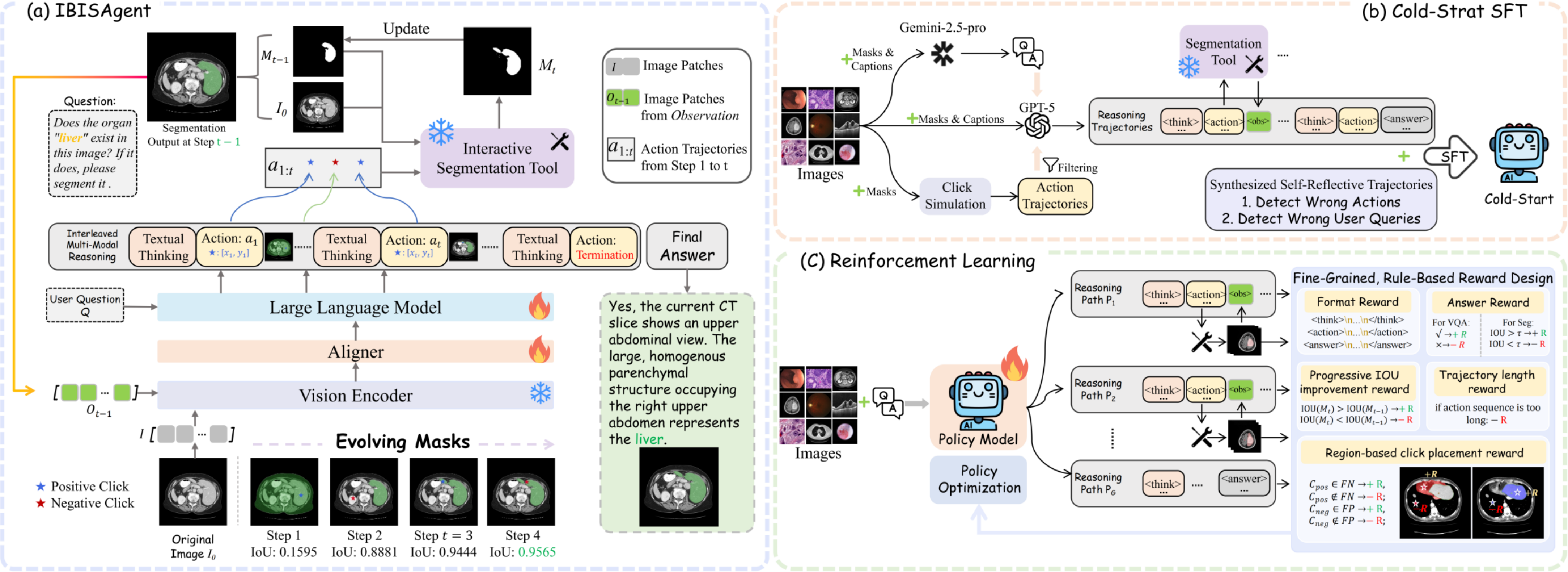
(a) Pipeline of VisionReasoner (b) Training of VisionReasoner

VisionReasoner (ICLR 2026)

Core idea of IBISAgent

IBISAgent models segmentation as a **Markov Decision Process.** At each step, the agent performs a **thought → action → observation** loop:

1. **Thought:** reason about what is wrong with the current mask
2. **Action:** generate positive or negative click commands with normalized coordinates
3. **Observation:** invoke a segmentation tool and observe the updated mask



Key Feature: Agentic RL with fine-grained rewards

ZHEJIANG UNIVERSITY

- After cold-start SFT, we further optimize the agent with **reinforcement learning**.
- The RL set contains 888K VQA samples, encouraging the model to explore beyond supervised imitation.
- We design five rewards:
  a. **Format reward** for valid structured outputs
  b. **Final-answer reward** for QA correctness or final segmentation quality
  c. **Region-based click placement reward** for clicking inside the right FP/FN regions
  d. **Progressive segmentation improvement reward** for step-by-step IoU improvement
  e. **Trajectory length reward** for efficient interaction

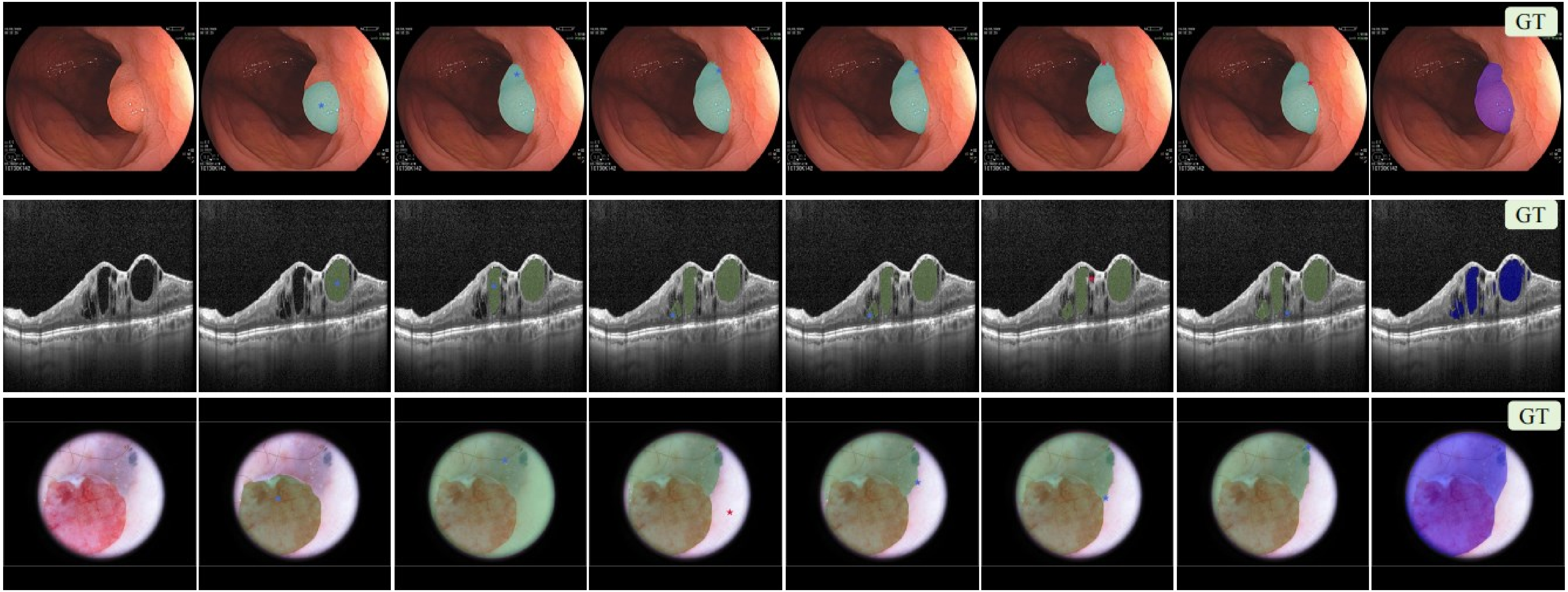
- **Key intuition:** reward not only the final mask, but also whether each intermediate action is meaningful.

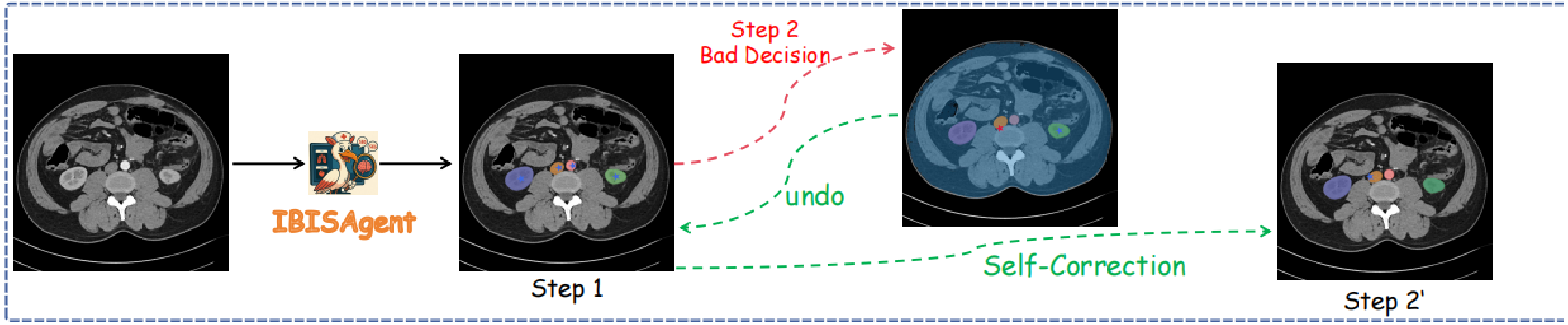
Results

ZHEJIANG UNIVERSITY

Benchmarks \ Methods	General-Purpose MLLMs with Segmentation Capability				Medical MLLMs with Segmentation Capability				Ours
	LISA [18]	LISA++ [45]	SAM4MLLM [5]	VisionReasoner [25]	MedPLIB [13]	Citrus-V [40]	UniBiomed [41]	MMedAgent [20]	IBISAgent
In-domain testset $\mathcal{D}_{\text{test}}$									
IOU ↑	9.44 (20.46)	9.49 (20.76)	15.85 (27.84)	16.11 (29.11)	22.29	30.61	50.74	36.13	**85.58**
DSC ↑	14.11 (25.73)	14.30 (25.94)	21.16 (33.04)	22.05 (35.50)	27.35	37.63	58.31	42.85	**92.21**
F1-score ↑	20.18 (32.15)	20.75 (32.34)	32.53 (42.75)	34.78 (46.72)	38.94	53.75	69.22	56.64	**96.39**
Out-of-domain testset MeCOVQA-G+									
IOU ↑	10.07 (15.24)	9.87 (15.01)	16.99 (21.19)	18.27 (24.46)	33.36	46.54	24.88	26.54	**80.63**
DSC ↑	15.44 (21.30)	14.70 (21.26)	21.85 (26.35)	25.08 (30.24)	41.19	52.65	31.74	33.81	**89.27**
F1-score ↑	21.69 (28.04)	21.25 (27.96)	32.94 (38.57)	37.83 (42.08)	53.47	69.84	43.63	44.17	**95.24**
Held-out in-house testset									
IOU ↑	5.23 (9.12)	5.46 (9.45)	8.28 (14.00)	10.10 (17.66)	20.12	32.08	35.62	27.39	**72.09**
DSC ↑	9.58 (14.33)	9.69 (14.80)	13.59 (18.04)	15.88 (24.57)	27.80	38.63	41.55	34.26	**83.78**
F1-score ↑	13.03 (17.15)	13.17 (17.72)	19.07 (25.26)	22.49 (30.07)	39.42	50.76	54.97	45.88	**91.76**

Methods	MeCOVQA-G+			In-House Testset		
	IoU ↑	DSC ↑	F1 ↑	IoU ↑	DSC ↑	F1 ↑
GPT-4o [15] + MedSAM2 [27]	11.75	17.39	22.42	7.23	10.40	15.16
LLaVA-Med [21] + MedSAM2 [27]	24.54	31.38	35.70	20.03	26.94	37.75
HuatuoGPT-Vision [4] + MedSAM2 [27]	35.86	43.41	54.79	30.25	36.72	52.28
Lingshu [43] + MedSAM2 [27]	39.63	47.18	60.31	31.19	37.55	54.02
Chiron [37] + MedSAM2 [27]	39.58	47.16	60.25	30.75	36.90	53.37
IBISAgent	**80.63**	**89.27**	**95.24**	**72.09**	**83.78**	**91.76**


GT
GT
GT


Step 2
Bad Decision
undo
IBISAgent
Step 1
Self-Correction
Step 2'

浙江大学 ZHEJIANG UNIVERSITY

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

CVPR JUNE 3-7, 2026 DENVER COLORADO

Thanks for watching !