REVIVE: COUNTERFACTUAL ON-POLICY DIAGNOSIS AND RECOVERY FOR INTERACTIVE ULTRASOUND SEGMENTATION AGENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent MLLM-based segmentation agents automate multi-round interactive segmentation by generating corrective prompts for SAM-based models. However, they primarily emphasize action imitation while lacking explicit diagnosis of the current error state before action selection. Moreover, trajectory-level reinforcement learning optimizes sampled actions through sparse reward signals but does not explicitly compare alternative corrections at the same on-policy state, limiting recovery from suboptimal decisions and allowing errors to accumulate across interaction rounds. In clinical settings, human annotators follow a "diagnose-act-verify" self-correction loop: they need to diagnose the current mask errors before acting and then verify their effectiveness from the updated mask. We propose **REVIVE**, a self-improving segmentation agent that explicitly diagnoses the residual-error state before action, which further translates residual error evidence into action-specific representations. Then, the Counterfactual On-Policy Action Ranking (COAR) is proposed to execute alternative corrective actions at on-policy states and align the action likelihoods with the resulting segmentation outcomes, providing dense state-wise supervision for corrective action-type selection, while its integration with trajectory-level GRPO improves multi-round interaction behavior. Experiments demonstrate that our REVIVE consistently improves segmentation quality and interaction reliability, while reducing error-action misalignment and improving recovery from self-induced errors.

1 INTRODUCTION

Medical image segmentation plays a crucial role in computer-assisted diagnosis, treatment, and image-guided intervention (Ronneberger et al., 2015; Deng et al., 2024; Li et al., 2022). Promptable segmentation foundation models (Ma et al., 2024) have substantially advanced this paradigm by allowing segmentation masks to be iteratively refined through sparse user prompts. More recently, MLLM-based segmentation agents (Liu et al., 2026; Jiang et al., 2026) have further automated this process by autonomously generating corrective clicks and invoking SAM-based segmentation tools (Ma et al., 2024) over multiple interaction rounds. These methods are typically optimized through supervised fine-tuning (SFT) on expert trajectories and reinforcement learning (RL) with trajectory-level rewards.

Despite different training objectives, existing agents largely learn an interactive action implicitly, without explicitly modeling the residual error state that should justify the action. This limitation can lead to error-action misalignment, where the selected correction is inconsistent with the dominant residual error, e.g., issuing a positive click under severe false-positive errors or stopping despite substantial false-negative regions. In multi-round interaction, such suboptimal decisions further alter subsequent states, causing the trajectory to drift from expert supervision and errors to accumulate across rounds.

Existing segmentation agents are primarily optimized through supervised fine-tuning (SFT) and reinforcement learning (RL). SFT minimizes token-level cross-entropy on expert trajectories, teaching the policy to imitate demonstrated actions but providing limited supervision for recovery at self-induced states. Meanwhile, RL signals (Guo et al., 2025; Yu et al., 2026) in interactive segmen-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

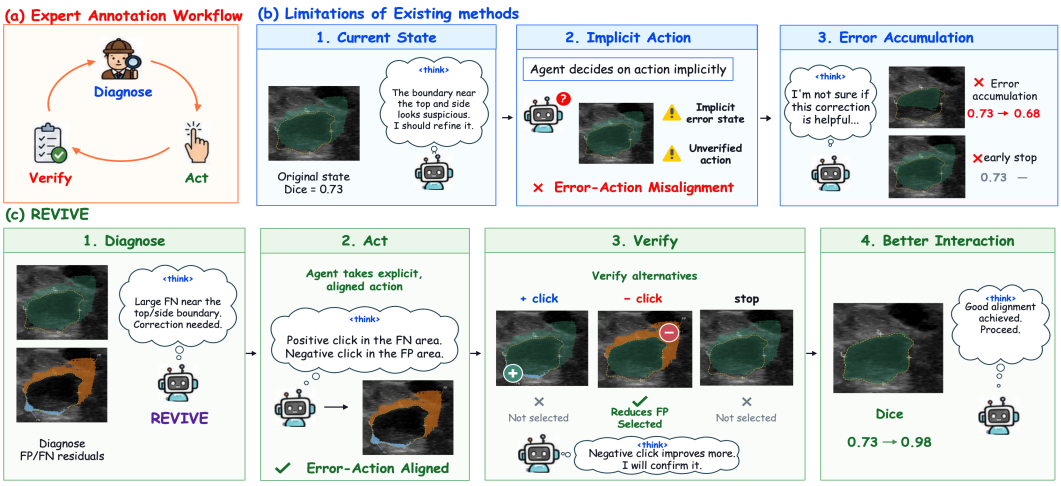

Figure 1: Existing agents directly predict corrective actions without explicit error diagnosis, leading to error accumulation. REVIVE follows a "diagnose-act-verify" paradigm by diagnosing FN/FP errors and counterfactually evaluating alternative actions for robust recovery.

tation are typically sparse and do not directly evaluate alternative actions at each interaction round. Consequently, the agent cannot determine whether an unchosen correction at the same state would have yielded a better segmentation result.

These limitations are especially problematic in ultrasound imaging, where low tissue contrast, ambiguous boundaries, and severe speckle artifacts make segmentation errors difficult to identify. Existing agents typically predict corrective actions directly from image-mask observations. As a result, visual ambiguity and action selection become entangled within the policy itself, leaving the agent with insufficient structured guidance about what aspect of the mask actually requires correction. Therefore, reliable interactive ultrasound segmentation requires explicit residual-error diagnosis before action selection and direct verification of alternative corrections for effective policy recovery.

In clinical settings, human annotators naturally address these challenges through a closed-loop self-correction process. Before acting, they first inspect the current mask and diagnose the dominant error regions. After each corrective action, the updated mask provides direct verification of whether the action was effective. This motivates a straightforward "diagnose-act-verify" principle:

$$\text{Error Diagnosis} \rightarrow \text{Corrective Action} \rightarrow \text{Action Verification}. \tag{1}$$

Following this principle, we propose **REVIVE**, a self-improving interactive ultrasound segmentation agent that explicitly incorporates error-state diagnosis and action verification into policy learning. First, **Explicit Error-State Diagnosis** learns a lightweight Residual-Error Verifier to estimate false-negative (FN) and false-positive (FP) error states from inference-available MedSAM signals. Then the diagnosed error states are explicitly aligned with corrective action semantics, providing state-dependent evidence for positive-click, negative-click, and stop decisions. Second, we introduce **Counterfactual On-Policy Recovery Learning** to improve the MLLM policy directly from segmentation states induced by its own interaction behavior. At each on-policy state, REVIVE executes alternative positive-click, negative-click, and stop actions with the frozen segmentation tool and evaluates their actual segmentation outcomes. Based on these outcomes, the **Counterfactual On-Policy Action Ranking (COAR)** loss is proposed to align the MLLM action preferences with the state-wise ordering of alternative corrections, enabling the segmentation agent to improve both local corrective decisions and long-horizon interaction behavior.

Our contributions are summarized as follows:

- We formulate interactive segmentation as an explicit diagnose-act-verify process, revealing the missing explicit residual-error diagnosis before action selection and state-wise evaluation of alternative corrections during policy optimization.

- We propose **REVIVE**, which explicitly estimates FN/FP residual states and grounds the resulting diagnostic evidence into action-specific representations for positive-click, negative-click, and stop decisions.
- We introduce **COAR**, which counterfactually evaluates alternative corrections at on-policy states and converts their actual outcomes into state-wise action ranking for robust multi-round interaction.

2 Related Work

2.1 Promptable Ultrasound Image Segmentation

Medical image segmentation has traditionally been formulated as a static dense prediction problem, where a task-specific network maps an input image directly to a pixel-wise mask. U-Net and its variants have achieved strong performance in fully supervised medical image segmentation (Ronneberger et al., 2015; Isensee et al., 2021). However, these static models produce masks in a single pass and cannot actively diagnose or correct residual under- and over-segmentation errors during inference (Chen et al., 2022). Promptable segmentation foundation models, represented by SAM and its medical variants such as MedSAM, change this paradigm by generating masks from sparse prompts including points and boxes (Kirillov et al., 2023; Ma et al., 2024; Wong et al., 2024). Compared with static segmenters, MedSAM-like models provide a more flexible interface for interactive refinement and can serve as strong medical segmentation tools (Lin et al., 2024). But they are still prompt-conditioned executors: they can respond to a given prompt, but cannot autonomously determine whether the current mask is reliable or which corrective click should be issued.

2.2 MLLM-based Medical Segmentation Agents

Recent MLLM-based segmentation methods combine language reasoning with segmentation modules to enable pixel-level grounding from complex instructions. General-domain methods such as LISA, PixelLM, and GLaMM (Lai et al., 2024; Ren et al., 2024; Rasheed et al., 2024), as well as medical models such as MedSeg-R and UniBioMed (Huang et al., 2025b; Wu et al., 2025), mainly focus on single-pass grounding or mask generation. More recent works formulate segmentation as a multi-round agentic process. SegAgent (Zhu et al., 2025), IBISAgent (Jiang et al., 2026), and MedSAM-Agent (Liu et al., 2026) iteratively generate corrective prompts and invoke external segmentation tools for mask refinement. However, these agents still learn the evidence-to-action mapping largely implicitly from trajectories or reward signals, without explicitly modeling the residual-error state that should justify the next correction. As a result, visual-action mismatch may propagate across interaction rounds (Guan et al., 2024). Our method instead explicitly diagnoses residual errors and verifies alternative corrections at on-policy states.

2.3 Policy Learning for Interactive Segmentation Agents

Existing segmentation agents are commonly optimized through supervised fine-tuning (SFT) (Ouyang et al., 2022) and reinforcement learning (RL) (Schulman et al., 2017). Although such SFT behavior cloning provides effective action initialization, the policy may encounter self-induced states that are not covered by the expert demonstrations during multi-round interaction (Ross et al., 2011). Trajectory-level RL can improve overall interaction outcomes, but its delayed rewards provide limited state-wise supervision for distinguishing among alternative corrections. In contrast, preference-based objectives directly compare candidate outputs and have been widely used to align language-model policies (Rafailov et al., 2023). However, their preferences are typically constructed from offline annotations or complete responses. Accordingly, we propose Counterfactual On-Policy Action Ranking (COAR), which executes alternative corrective action types from the same on-policy state and ranks them according to the resulting mask improvements. COAR thus provides dense, executor-aware supervision at each state, complementing sparse trajectory-level reinforcement learning.

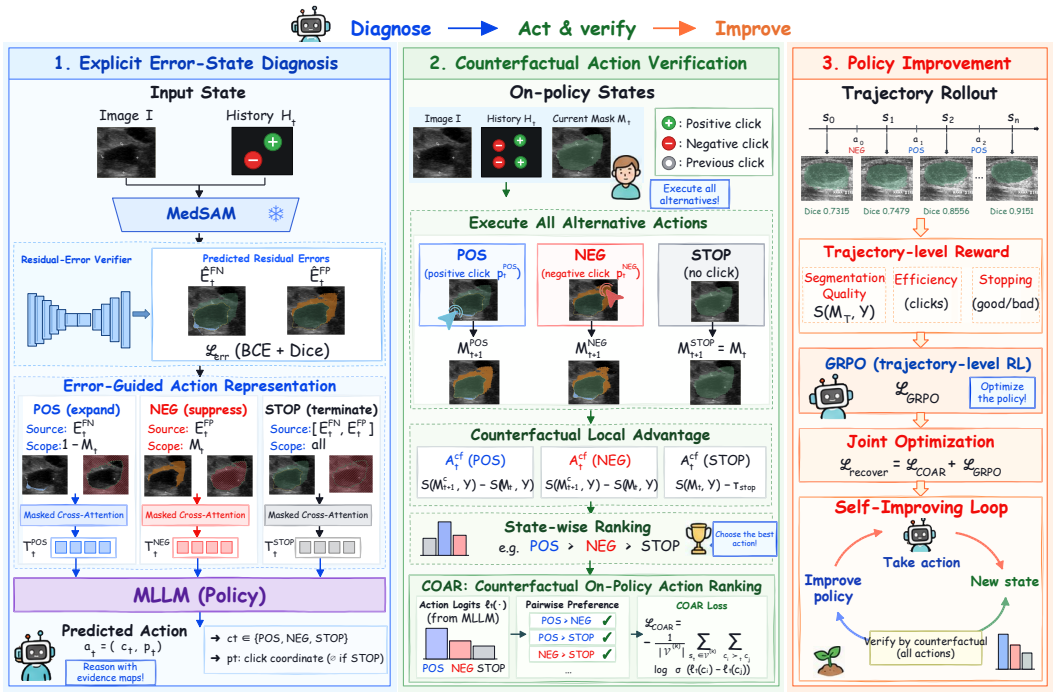

Figure 2: REVIVE follows a diagnose-verify-improve pipeline for interactive segmentation. First, the Residual-Error Verifier explicitly estimates FN/FP errors to construct action-specific representations for alternative actions. Second, alternative actions are counterfactually executed at each on-policy state, and their segmentation outcomes are ranked by COAR to provide state-wise recovery supervision, enabling the policy to progressively improve its corrective decisions during interaction. Third, COAR is jointly optimized with trajectory-level GRPO, combining state-wise action ranking with rewards for segmentation quality and interaction efficiency to improve the policy's ability to recover from its own suboptimal corrections.

3 METHOD

3.1 PROBLEM FORMULATION

Following the *diagnose-act-verify* clinical behavior, we formulate multi-round interactive segmentation as a sequential decision-making problem between an agent and a frozen segmentation tool. Given an image I, a task instruction q, and the interaction history H_t, the frozen MedSAM (Ma et al., 2024) produces the current segmentation mask M_t, mask logits L_t, and intermediate visual features F_t:

$$(M_t, L_t, F_t) = \mathcal{T}(I, H_t), \tag{2}$$

where $\mathcal{T}$ denotes the frozen MedSAM. The interaction state is therefore defined as:

$$s_t = (I, q, H_t, M_t, L_t, F_t). \tag{3}$$

At each interaction round, the agent predicts a corrective action $a_t = (c_t, p_t)$, where $c_t \in \mathcal{C} = \{\text{POS}, \text{NEG}, \text{STOP}\}$ denotes the action type and p_t is the normalized click coordinate, with $p_t = \varnothing$ for STOP.

Existing MLLM-based segmentation agents (Wang et al., 2025; Liu et al., 2026; Jiang et al., 2026) are primarily trained with SFT (Ouyang et al., 2022) on expert trajectories and RL with trajectory-level rewards (Guo et al., 2025; Yu et al., 2026). However, these objectives directly optimize action generation without explicitly modeling the residual-error state that supports each correction. SFT supervises only the demonstrated GT action, while trajectory-level RL evaluates sampled actions through sparse rewards; neither explicitly identifies the remaining error at each state nor compares alternative corrections at the same state. Consequently, suboptimal actions may shift the interaction trajectory away from expert states and propagate errors across subsequent rounds.

To address this limitation, we first explicitly diagnose the residual-error state and then improve state-wise decision making through counterfactual action verification, following:

$$\text{Error Diagnosis} \rightarrow \text{Corrective Action} \rightarrow \text{Action Verification}. \tag{4}$$

3.2 Explicit Error-State Diagnosis

In clinical annotation, reliable corrective actions should be guided by the residual errors in the current segmentation mask. However, existing MLLM-based agents (Wang et al., 2025; Liu et al., 2026; Jiang et al., 2026) often fail to explicitly perceive this error state when making corrective decisions. During training, we therefore construct explicit false-negative (FN) and false-positive (FP) residual states from the GT mask Y and the binarized MedSAM prediction:

$$P_t = \sigma(L_t), \quad \hat{M}_t = \mathbb{I}[P_t > 0.5]. \tag{5}$$

Then the residual errors of the current mask can be decomposed as:

$$E_t^{FN} = Y \odot (1 - \hat{M}_t), \quad E_t^{FP} = (1 - Y) \odot \hat{M}_t. \tag{6}$$

where E_t^{FN} and E_t^{FP} denote under- and over-segmented regions, respectively.

Hence, the teacher action c_t^* during SFT can be determined by the dominant residual-error type:

$$c_t^* = \begin{cases} \text{STOP}, & \mathcal{S}(\hat{M}_t, Y) \geq \tau_{\text{stop}}, \\ \text{POS}, & |E_t^{FN}| \geq |E_t^{FP}|, \\ \text{NEG}, & |E_t^{FN}| < |E_t^{FP}|, \end{cases} \tag{7}$$

where $\mathcal{S}(\cdot, \cdot)$ is the segmentation-quality measure and τ_{stop} is the stopping threshold. The click coordinate p_t^* is placed at the center of the dominant residual component, yielding the expert action $a_t^* = (c_t^*, p_t^*)$, with $p_t^* = \emptyset$ for STOP.

The SFT objective therefore supervises both categorical action-type selection and type-conditioned coordinate generation.

Since Y is unavailable at inference, we learn a lightweight Residual-Error Verifier V_ϕ from inference-available MedSAM signals:

$$(\widetilde{E}_t^{FN}, \widetilde{E}_t^{FP}) = V_\phi \left(\text{Concat} \left[\hat{M}_t, L_t, F_t, U_t \right] \right), \tag{8}$$

where U_t is the prediction uncertainty map.

During the training stage, the verifier V_ϕ is supervised to recover the GT-derived residual-error maps:

$$\mathcal{L}_{\text{err}} = \sum_{r \in \{FN, FP\}} \left[\text{BCE}(\widetilde{E}_t^r, E_t^r) + \text{DiceLoss}(\widetilde{E}_t^r, E_t^r) \right]. \tag{9}$$

Although the predicted FN/FP residual maps $\widetilde{E}_t^{FN}$ and $\widetilde{E}_t^{FP}$ provide pixel-level diagnostic evidence, such visual cues are not explicitly represented in the MLLM action space. We therefore introduce the learnable token for each candidate action and ground its representation in the corresponding visual error state, establishing an explicit state-dependent association between residual errors and corrective actions. Specifically, for each action $c \in \mathcal{C} = \{\text{POS}, \text{NEG}, \text{STOP}\}$, we associate an action-specific residual-error source $\mathbf{R}_t^c$ with a spatial scope $\mathcal{M}_t^c$:

$$(\mathbf{R}_t^c, \mathcal{M}_t^c) = \begin{cases} (\widetilde{E}_t^{FN},\, 1 - \hat{M}_t), & c = \text{POS}, \\ (\widetilde{E}_t^{FP},\, \hat{M}_t), & c = \text{NEG}, \\ (\text{Concat}[\widetilde{E}_t^{FN}, \widetilde{E}_t^{FP}],\, \mathbf{1}), & c = \text{STOP}. \end{cases} \tag{10}$$

This design grounds positive and negative actions in mask-exterior FN evidence and mask-interior FP evidence, respectively, while allowing the stop action to assess the global residual-error state.

Then an action-specific learnable query $\mathbf{q}_c$ extracts the corresponding diagnostic evidence through masked cross-attention:

$$T_t^c = \left[\text{MaskedCrossAttn}\left(\mathbf{q}_c, \text{Proj}(\mathbf{R}_t^c); \mathcal{M}_t^c\right); \mathbf{e}_c^{\text{sem}} \right], \quad c \in \mathcal{C}, \tag{11}$$

where $\mathbf{e}_c^{\text{sem}}$ denotes the learnable semantic embedding of action c. The resulting tokens are inserted into the MLLM context for action prediction. Hence, we optimize the diagnosis stage with $\mathcal{L}_{diag} = \mathcal{L}_{SFT} + \mathcal{L}_{err}$, where $\mathcal{L}_{\text{SFT}}$ denotes the standard SFT objective, and the GT mask is only used during training. At inference, V_ϕ estimates the residual state from inference-available MedSAM signals.

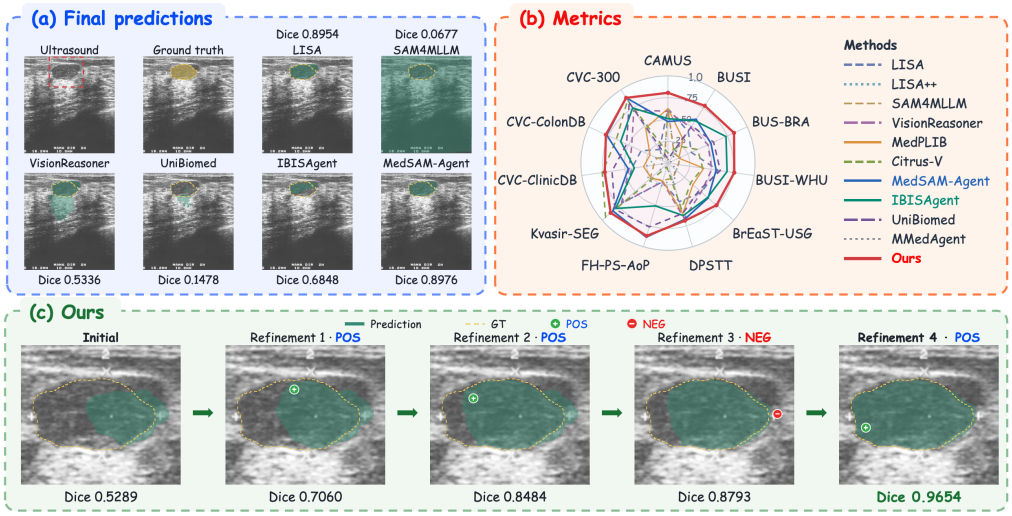

Figure 3: Comparison with recent general- and medical-purpose segmentation methods. (a) Qualitative final predictions with recent SOTAs; REVIVE yields masks closest to the ground truth. (b) Per-dataset metrics across these benchmark datasets. (c) REVIVE's multi-round refinement: POS/NEG corrective clicks raise the Dice score from 0.5289 to 0.9654.

3.3 Counterfactual On-Policy Recovery Learning

Explicit diagnosis teaches the agent how residual errors correspond to corrective actions, but it does not yet endow the MLLM with the ability to recover from states caused by its own suboptimal corrections. Once a suboptimal action is taken by the MLLM, the resulting trajectory may enter states that are poorly covered by offline supervision: SFT is largely restricted to expert-induced states, whereas on-policy RL optimizes sampled actions through rollout rewards without explicitly comparing alternative corrections.

Therefore, we introduce Counterfactual On-policy Recovery Learning, which evaluates all alternative actions $c \in \{POS, NEG, STOP\}$ at each on-policy state to align the MLLM action preferences with their actual segmentation benefits:

$$M_{t+1}^{c} = \begin{cases} \mathcal{T}\left(I, H_t \cup \{a_t^c\}\right), & c \in \{\text{POS}, \text{NEG}\}, \\ M_t, & c = \text{STOP}. \end{cases} \tag{12}$$

We then quantify the segmentation benefits of each alternative action c:

$$A_t^{\text{cf}}(c) = \begin{cases} \mathcal{S}(M_{t+1}^{c}, Y) - \mathcal{S}(M_t, Y), & c \in \{\text{POS}, \text{NEG}\}, \\ \mathcal{S}(M_t, Y) - \tau_{\text{stop}}, & c = \text{STOP}. \end{cases} \tag{13}$$

where $\mathcal{S}(\cdot, \cdot)$ denotes the segmentation-quality measure. We refer to $A_t^{\text{cf}}(c)$ as the *counterfactual local advantage*, as it measures the immediate segmentation state under action c. And STOP uses the signed margin to the stopping threshold.

We then rank actions by their counterfactual advantages, $c_i \succ_t c_j$ iff $A_t^{cf}(c_i) > A_t^{cf}(c_j)$. The proposed Counterfactual On-Policy Action Ranking (COAR) objective can be formulated as:

$$\mathcal{L}_{\text{COAR}} = -\frac{1}{|\mathcal{V}^{(k)}|} \sum_{s_t \in \mathcal{V}^{(k)}} \sum_{c_i \succ_t c_j} \log \sigma\left(\ell_t(c_i) - \ell_t(c_j)\right). \tag{14}$$

Hence, COAR converts the actual outcomes of each alternative action into dense state-wise recovery supervision, increasing the likelihood of actions that would have produced better segmentation outcomes and suppressing other inferior actions.

Table 1: Quantitative comparison with different methods on the ultrasound image datasets.

Benchmarks \ Methods	General-Purpose MLLMs				Medical-Purpose MLLMs						**Ours**
	LISA	LISA++	SAM4MLLM	VisionReasoner	MedPLIB	Citrus-V	MedSAM-Agent	IBISAgent	UniBiomed	MMedAgent	REVIVE
BUSI (Al-Dhabyani et al., 2020)											
DSC ↑	0.4307	0.1379	0.1307	0.4850	0.3220	0.2315	0.5980	0.5767	0.5075	0.2667	0.7818
IOU ↑	0.3561	0.0835	0.0774	0.4013	0.2338	0.1729	0.5126	0.5148	0.4328	0.2431	0.7036
Prec ↑	0.5450	0.0838	0.0774	0.4889	0.2988	0.3038	0.6598	0.6180	0.5678	0.2837	0.8419
Rec ↑	0.4369	0.8023	0.8333	0.5939	0.4707	0.2641	0.5723	0.5912	0.4958	0.2937	0.7520
BUSBRA (Gómez-Flores et al., 2024)											
DSC ↑	0.2951	0.1548	0.1524	0.5052	0.1364	0.3742	0.5431	0.7329	0.5768	0.0354	0.8342
IOU ↑	0.2481	0.0890	0.0872	0.4137	0.0914	0.2928	0.4224	0.6486	0.4718	0.0257	0.7467
Prec ↑	0.3941	0.0890	0.0872	0.4971	0.1464	0.4545	0.7015	0.7277	0.6046	0.0438	0.8458
Rec ↑	0.2899	0.9982	1.0000	0.6601	0.1922	0.4032	0.5034	0.7917	0.6379	0.0665	0.8645
CAMUS (Leclerc et al., 2019)											
DSC ↑	0.6237	0.6199	0.6132	0.5933	0.6260	0.1144	0.4760	0.5095	0.0472	0.0382	0.8062
IOU ↑	0.4585	0.4545	0.4476	0.4353	0.4841	0.0633	0.3235	0.3477	0.0274	0.0248	0.6954
Prec ↑	0.4664	0.4611	0.4531	0.4996	0.7558	0.7015	0.9160	0.9783	0.4159	0.2952	0.8327
Rec ↑	0.9689	0.9738	0.9771	0.8830	0.5735	0.0648	0.3270	0.3490	0.0276	0.0282	0.8079
BUSI_WHU (Huang et al., 2023)											
DSC ↑	0.5826	0.1231	0.1194	0.5073	0.4149	0.4026	0.5587	0.6857	0.6071	0.0581	0.7696
IOU ↑	0.4959	0.0702	0.0673	0.4158	0.3029	0.3082	0.4466	0.6000	0.5139	0.0475	0.6868
Prec ↑	0.6299	0.0704	0.0673	0.4777	0.3671	0.4574	0.6358	0.7251	0.6366	0.0589	0.7797
Rec ↑	0.6012	0.9866	1.0000	0.7061	0.5963	0.4531	0.5679	0.7018	0.6490	0.0883	0.7991
BrEaST_USG (Pawłowska et al., 2024)											
DSC ↑	0.5066	0.1264	0.1256	0.5059	0.3784	0.3443	0.6106	0.6027	0.4742	0.1010	0.7377
IOU ↑	0.4148	0.0729	0.0720	0.4047	0.2726	0.2527	0.5155	0.5122	0.3866	0.0825	0.6363
Prec ↑	0.6049	0.0729	0.0720	0.4799	0.3335	0.4521	0.6399	0.6750	0.5160	0.1094	0.8314
Rec ↑	0.5213	0.9943	0.9998	0.7066	0.5894	0.3592	0.6623	0.6189	0.5102	0.1709	0.7126
DPSTT (Li et al., 2022)											
DSC ↑	0.4611	0.1881	0.1834	0.6619	0.6044	0.5496	0.6734	0.6272	0.6018	0.2733	0.6886
IOU ↑	0.3445	0.1080	0.1047	0.5552	0.4802	0.4586	0.5541	0.5229	0.4809	0.2322	0.5962
Prec ↑	0.3631	0.1080	0.1047	0.6500	0.5119	0.6057	0.7649	0.7685	0.7188	0.3094	0.6960
Rec ↑	0.9369	1.0000	1.0000	0.7778	0.9255	0.5639	0.6779	0.6025	0.5977	0.2834	0.7747
FH–PS–AoP (Jieyun & ZhanHong, 2023)											
DSC ↑	0.2229	0.2731	0.2045	0.2240	0.2136	0.0467	0.8852	0.5116	0.7631	0.0471	0.8702
IOU ↑	0.1553	0.1714	0.1277	0.1554	0.1366	0.0297	0.8230	0.4499	0.6419	0.0340	0.7749
Prec ↑	0.1823	0.1779	0.1556	0.2002	0.2652	0.1575	0.8961	0.5978	0.8046	0.0776	0.8863
Rec ↑	0.3204	0.7431	0.6410	0.4490	0.2307	0.0309	0.8819	0.4884	0.7666	0.0370	0.8616

COAR and GRPO (Guo et al., 2025) address two complementary levels of policy improvement. COAR resolves which action is preferable at a particular self-induced state, whereas GRPO optimizes the quality of the complete multi-round interaction trajectory. Following MedSAM-Agent (Liu et al., 2026), we retain its multi-turn GRPO training and reward design to optimize trajectory-level segmentation quality and interaction efficiency:

$$\mathcal{L}_{\text{recover}} = \mathcal{L}_{\text{COAR}} + \mathcal{L}_{\text{GRPO}}. \tag{15}$$

4 Experiments

4.1 Experimental Setup & Implementation

We conduct an extensive evaluation on the datasets spanning ultrasound and endoscopic imaging, covering diverse anatomical structures and lesion types in both static images and videos. Our primary evaluation comprises following public ultrasound datasets: CAMUS (Leclerc et al., 2019), BUSI (Al-Dhabyani et al., 2020), BUSBRA (Gómez-Flores et al., 2024), BUSI-WHU (Huang et al., 2023), BrEaST-USG (Pawłowska et al., 2024), and DPSTT (Li et al., 2022), and FH-PS-AoP (Jieyun & ZhanHong, 2023) . Following the data partition of PraNet (Fan et al., 2020), we conduct experiments on 900 images from Kvasir-SEG (Jha et al., 2019) and 550 images from CVC-ClinicDB (Bernal et al., 2015). We then evaluate the segmentation performance on the held-out Kvasir-SEG (Jha et al., 2019) and CVC-ClinicDB (Bernal et al., 2015) test sets for in-dataset performance, and on CVC-ColonDB (Bernal et al., 2012) and CVC-300 (Vázquez et al., 2017) for out-of-distribution cross-dataset generalization. We conduct experiments on 2 Nvidia A100 GPUs. The training stage is performed for 8 epochs using LoRA with rank 16 and a scaling factor of 32. We used AdamW with a learning rate of 1.5e-5. The batch size is set to 2 with 4 gradient accumulations for each GPU, resulting in an effective global batch size of 16.

4.2 Comparison with Previous Methods

Following the evaluation protocols of recent medical segmentation agents (Liu et al., 2026; Jiang et al., 2026), we conduct a comprehensive comparison with state-of-the-art MLLM-based segmentation methods. These baselines cover two categories: (1) general-purpose MLLMs that extend visual-

Table 2: Quantitative comparison on colorectal polyp segmentation datasets following the data partition of PraNet (Fan et al., 2020). Kvasir-SEG and CVC-ClinicDB are held-out test sets from the source datasets, while CVC-ColonDB and CVC-300 are used for cross-dataset evaluation.

Metrics \ Methods	General-Purpose MLLMs				Medical-Purpose MLLMs						**Ours**
	LISA	LISA++	SAM4MLLM	VisionReasoner	MedPLIB	Citrus-V	MedSAM-Agent	IBISAgent	UniBiomed	MMedAgent	REVIVE
Held-out Test Sets (In-Dataset Evaluation)											
Kvasir-SEG (Jha et al., 2019)											
DSC ↑	0.6958	0.2987	0.2700	0.7085	0.2579	0.9480	0.8380	0.7949	0.8070	0.1159	0.8716
IoU ↑	0.6089	0.1945	0.1732	0.6256	0.1847	0.9066	0.7590	0.7274	0.7262	0.1030	0.8040
Prec ↑	0.6579	0.1977	0.1734	0.7497	0.4156	0.9773	0.9679	0.9244	0.9513	0.1535	0.9471
Rec ↑	0.8455	0.9920	0.9989	0.7800	0.2431	0.9254	0.7683	0.7441	0.7489	0.1042	0.8409
CVC-ClinicDB (Bernal et al., 2015)											
DSC ↑	0.4196	0.2449	0.1651	0.4506	0.2842	0.7634	0.4608	0.3904	0.6493	0.0716	0.7329
IoU ↑	0.3270	0.1681	0.1017	0.3708	0.2022	0.7018	0.3518	0.3431	0.5780	0.0617	0.6511
Prec ↑	0.3683	0.1684	0.1020	0.4544	0.2851	0.9037	0.7092	0.4747	0.7669	0.0818	0.8426
Rec ↑	0.7871	0.9746	0.9995	0.7497	0.3908	0.7382	0.3709	0.3985	0.6239	0.0692	0.7226
Unseen Test Sets (OOD Cross-Dataset Evaluation)											
CVC-ColonDB (Bernal et al., 2012)											
DSC ↑	0.3398	0.1531	0.1394	0.4587	0.2258	0.6524	0.7729	0.5969	0.5229	0.0587	0.7845
IoU ↑	0.2500	0.0935	0.0842	0.3888	0.1654	0.5875	0.6821	0.5377	0.4504	0.0464	0.7088
Prec ↑	0.2789	0.0937	0.0842	0.4814	0.2695	0.7800	0.8822	0.6883	0.6934	0.0768	0.8462
Rec ↑	0.7525	0.9914	0.9997	0.7038	0.2971	0.6242	0.7208	0.5751	0.4848	0.0553	0.7769
CVC-300 (Vázquez et al., 2017)											
DSC ↑	0.1846	0.0759	0.0730	0.7255	0.4738	0.8576	0.8911	0.7510	0.8241	0.1725	0.8893
IoU ↑	0.1104	0.0403	0.0386	0.6669	0.3817	0.7981	0.8176	0.6901	0.7348	0.1410	0.8164
Prec ↑	0.1122	0.0403	0.0386	0.7299	0.4816	0.8579	0.9382	0.7765	0.8999	0.2141	0.8954
Rec ↑	0.8953	0.9999	1.0000	0.8342	0.5908	0.8662	0.8694	0.7534	0.7938	0.1560	0.9079

language reasoning to pixel-level segmentation, including LISA (Lai et al., 2024), LISA++ (Yang et al., 2023), SAM4MLLM (Chen et al., 2024), and VisionReasoner (Liu et al., 2025); and (2) medical-domain MLLMs and segmentation agents, including MedPLIB (Huang et al., 2025a), Citrus-V (Wang et al., 2025), MedSAM-Agent (Liu et al., 2026), IBISAgent (Jiang et al., 2026), UniBiomed (Wu et al., 2025), and MMedAgent (Li et al., 2024). Together, these baselines cover direct mask generation, segmentation-token-based grounding, and iterative segmentation-tool interaction.

Quantitative Segmentation Metrics Compared to general-domain MLLMs, our method achieves higher performance across all benchmarks. Our method outperforms these models, demonstrating that its superiority stems primarily from our novel and efficient design, which simulates the annotation process of human experts: performing actions based on the current error region and validating the effectiveness of the actions through multiple rounds of inference. Compared with existing medical-specific MLLMs, our method achieves better overall segmentation performance. Despite being trained on much larger datasets, Citrus-V (Wang et al., 2025) and UniBiomed (Wu et al., 2025) are still outperformed by our method, highlighting the effectiveness of our proposed components rather than dataset scale. Moreover, our method surpasses MedSAM-Agent (Liu et al., 2026) and IBISAgent (Jiang et al., 2026), which also leverage MedSAM, demonstrating the benefit of more accurate target localization and iterative mask refinement beyond simply integrating segmentation tools.

Qualitative Segmentation Performance

Fig. 3 visualizes the trends in Table 1, and the visual evidence mirrors the failure modes that motivate our design. Figure (a) compares final masks on representative ultrasound cases with per-method Dice annotated: general-purpose reasoners either collapse to near-empty masks (SAM4MLLM (Chen et al., 2024), Dice 0.0677) or capture only a coarse part of the target (VisionReasoner Liu et al. (2025), 0.5336), while medical-purpose agents localize the right region but, lacking an explicit residual-error estimate, mis-place their boundaries (IBISAgent (Jiang et al., 2026), 0.6848; MedSAM-Agent (Liu et al., 2026), 0.8976). REVIVE instead achieves the tightest alignment with the ground truth, confirming that conditioning each action on the diagnosed residual-error state rather than on the visual cue alone. Figure (b) shows that this advantage is not case-specific, holding consistently across these benchmark datasets. Figure (c) traces REVIVE's own trajectory on a low-contrast case: from an initial mask (Dice 0.5289), the agent diagnoses the dominant residual error each round and issues the matching action POS clicks to absorb false negatives, NEG clicks to suppress false positives.

Table 3: Ablation study of the proposed components on the BUS-BRA dataset.

Methods	GRPO	Diagnosis	COAR	Dice↑	Jaccard↑	Precision↑	Recall↑
M1	√	–	–	0.7657	0.6739	0.7542	0.8189
M2	√	√	–	0.7801	0.7033	0.7804	0.8176
M3	√	–	√	0.8086	0.7139	0.7960	0.8671
Ours	√	√	√	**0.8342**	**0.7467**	**0.8458**	**0.8645**

Table 4: Statistics of error recovery ability on ultrasound datasets during the inference stage.

Methods	BBox↑	Point↑	Error↓	Recover@ALL↑	Recover@1/2+↑	Fail↓
MedSAM-Agent	100%	24.3%	50.4%	25.4%	24.9% / 0.5%	74.6%
IBISAgent	0.0%	100%	22.3%	41.3%	32.2% / 9.1%	58.7%
Ours	100%	29.1%	22.5%	88.0%	85.8% / 2.2%	11.9%

4.3 Ablation Studies

Effectiveness of Individual Components. Table 3 evaluates the contributions of Explicit Diagnosis and COAR on top of the same RL baseline model. We start from a common RL baseline and build three configurations. M1 is trained only by the sparse trajectory-level GRPO reward, without error diagnosis or action comparison. M2 adds Explicit Error-State Diagnosis, injecting error guidance from MedSAM residual maps. M3 adds COAR, ranking counterfactual action executions into dense state-wise supervision without diagnosis. Ours (REVIVE) combines the GRPO backbone with both modules, realizing the full "diagnose-act-verify" loop. Adding Explicit Diagnosis improves Dice from 0.7657 to 0.7801, while introducing COAR achieves a larger improvement to 0.8086. Combining both components further increases Dice to 0.8342, with consistent improvements in Jaccard and Precision. These results demonstrate that Explicit Diagnosis and COAR provide complementary benefits for interactive segmentation.

Error Recovery Analysis. We further compare the error-recovery behavior of interactive segmentation agents in Table 4. Although all existing methods may produce suboptimal actions during interaction, REVIVE recovers from 85.8% of such errors within the next interaction, substantially outperforming MedSAM-Agent (24.9%) and IBISAgent (32.2%). Consequently, its unrecovered failure rate is reduced to 11.9%, compared with 74.6% and 58.7% for the two baselines. These results indicate that our proposed REVIVE is substantially more capable of correcting self-induced errors during multi-round interaction, please see the Proportion of single-round revisions versus revisions involving two or more rounds (Recover@ALL and Fail score).

5 Conclusion

In this work, we presented REVIVE, a self-improving agent for interactive ultrasound segmentation. Unlike existing agents that directly map image–mask feedback to corrective actions, REVIVE follows an explicit diagnose–act–verify process. A lightweight Residual-Error Verifier estimates FN and FP regions from inference-available MedSAM signals, while action-specific tokens ground POS, NEG, and STOP decisions in the corresponding diagnostic evidence. To improve recovery at self-induced states, COAR evaluates alternative action types instantiated with conditionally generated coordinates and aligns their categorical likelihoods with the resulting segmentation outcomes. This state-wise supervision complements trajectory-level GRPO, which optimizes overall multi-round segmentation quality and interaction efficiency. Experiments across multiple ultrasound benchmarks demonstrate improved segmentation accuracy, reduced error-action misalignment, and more reliable recovery from suboptimal corrections. These results highlight the value of explicit error diagnosis and counterfactual action verification for robust interactive segmentation agents.

6 Statement

In this work, we used Al tools to assist with benchmark construction and language editing, including grammar checking, sentence refinement, and improving the clarity and readability of the manuscript. We also used generative Al tools to assist with parts of the code implementation and debugging.

REFERENCES

Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020.

Jorge Bernal, Javier Sánchez, and Fernando Vilarino. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, 45(9):3166–3182, 2012.

Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015.

Xi Chen, Zhiyan Zhao, Yilei Zhang, Manni Duan, Donglian Qi, and Hengshuang Zhao. Focalclick: Towards practical interactive image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1300–1309, 2022.

Yi-Chia Chen, Wei-Hua Li, Cheng Sun, Yu-Chiang Frank Wang, and Chu-Song Chen. Sam4mllm: Enhance multi-modal large language model for referring expression segmentation. In *European Conference on Computer Vision*, pp. 323–340. Springer, 2024.

Xiaolong Deng, Huisi Wu, Runhao Zeng, and Jing Qin. Memsam: Taming segment anything model for echocardiography video segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9622–9631, 2024.

Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *International conference on medical image computing and computer-assisted intervention*, pp. 263–273. Springer, 2020.

Wilfrido Gómez-Flores, Maria Julia Gregorio-Calas, and Wagner Coelho de Albuquerque Pereira. BUS-BRA: A breast ultrasound dataset for assessing computer-aided diagnosis systems. *Medical Physics*, 51(4):3110–3123, 2024. doi: 10.1002/mp.16812.

Tianrui Guan, Fuxiao Liu, Xiyang Wu, Ruiqi Xian, Zongxia Li, Xiaoyu Liu, Xijun Wang, Lichang Chen, Furong Huang, Yaser Yacoob, Dinesh Manocha, and Tianyi Zhou. Hallusionbench: An advanced diagnostic suite for entangled language hallucination and visual illusion in large vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14375–14385, 2024. doi: 10.1109/CVPR52733.2024.01363.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

J Huang, J Zhang, Y Zhang, X Li, X Ma, J Deng, H Shen, D Wang, L Mei, and C Lei. Busi_whu: Breast cancer ultrasound image dataset. *Mendeley Data*, 3:2025, 2023.

Xiaoshuang Huang, Lingdong Shen, Jia Liu, Fangxin Shang, Hongxiang Li, Haifeng Huang, and Yehui Yang. Towards a multimodal large language model with pixel-level insight for biomedicine. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3779–3787, 2025a.

Yu Huang, Zelin Peng, Yichen Zhao, Piao Yang, Xiaokang Yang, and Wei Shen. Medseg-r: Reasoning segmentation in medical images with multimodal large language models. *arXiv preprint arXiv:2506.10465*, 2025b.

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18:203–211, 2021. doi: 10.1038/s41592-020-01008-z.

Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas De Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *International conference on multimedia modeling*, pp. 451–462. Springer, 2019.

Yankai Jiang, Qiaoru Li, Binlu Xu, Haoran Sun, Chao Ding, Junting Dong, Yuxiang Cai, Xuhong Zhang, and Jianwei Yin. Ibisagent: Reinforcing pixel-level visual reasoning in mllms for universal biomedical object referring and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20996–21005, 2026.

Bai Jieyun and Ou ZhanHong. Pubic symphysis-fetal head segmentation and angle of progression, 2023.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026, 2023.

Xin Lai, Zhuotao Tian, Yukang Chen, Yanwei Li, Yuhui Yuan, Shu Liu, and Jiaya Jia. Lisa: Reasoning segmentation via large language model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9579–9589, 2024.

Sarah Leclerc, Erik Smistad, João Pedrosa, Andreas Ostvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, Carole Lartizien, Jan Dhooge, Lasse Lovstakken, and Olivier Bernard. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE Transactions on Medical Imaging*, 38(9):2198–2210, 2019. doi: 10.1109/TMI.2019.2900516.

Binxu Li, Tiankai Yan, Yuanting Pan, Jie Luo, Ruiyang Ji, Jiayuan Ding, Zhe Xu, Shilong Liu, Haoyu Dong, Zihao Lin, et al. Mmedagent: Learning to use medical tools with multi-modal agent. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 8745–8760, 2024.

Jialu Li, Qingqing Zheng, Mingshuang Li, Ping Liu, Qiong Wang, Litao Sun, and Lei Zhu. Rethinking breast lesion segmentation in ultrasound: A new video dataset and a baseline network. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*, pp. 391–400. Springer, 2022.

Xian Lin, Yangyang Xiang, Li Yu, and Zengqiang Yan. Beyond adapting sam: Towards end-to-end ultrasound image segmentation via auto prompting. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume 15008, pp. 24–34. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-72111-3_3.

Shengyuan Liu, Liuxin Bao, Qi Yang, Wanting Geng, Boyun Zheng, Chenxin Li, Wenting Chen, Houwen Peng, and Yixuan Yuan. Medsam-agent: Empowering interactive medical image segmentation with multi-turn agentic reinforcement learning. *arXiv preprint arXiv:2602.03320*, 2026.

Yuqi Liu, Tianyuan Qu, Zhisheng Zhong, Bohao Peng, Shu Liu, Bei Yu, and Jiaya Jia. Visionreasoner: Unified visual perception and reasoning via reinforcement learning. *arXiv e-prints*, pp. arXiv–2505, 2025.

Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15:654, 2024.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27730–27744, 2022.

Anna Pawłowska, Anna Ćwierz-Pieńkowska, Agnieszka Domalik, Dominika Jaguś, Piotr Kasprzak, Rafał Matkowski, Łukasz Fura, Andrzej Nowicki, and Norbert Żołek. Curated benchmark dataset for ultrasound based breast lesion analysis. *Scientific Data*, 11:148, 2024. doi: 10.1038/s41597-024-02984-z.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pp. 53728–53741, 2023.

Hanoona Rasheed, Muhammad Maaz, Sahal Shaji, Abdelrahman Shaker, Salman Khan, Hisham Cholakkal, Rao M. Anwer, Eric Xing, Ming-Hsuan Yang, and Fahad S. Khan. Glamm: Pixel grounding large multimodal model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13009–13018, 2024.

Zhongwei Ren, Zhicheng Huang, Yunchao Wei, Yao Zhao, Dongmei Fu, Jiashi Feng, and Xiaojie Jin. Pixellm: Pixel reasoning with large multimodal model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26374–26383, 2024.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pp. 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.

Stephane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pp. 627–635. PMLR, 2011.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

David Vázquez, Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Antonio M López, Adriana Romero, Michal Drozdzal, and Aaron Courville. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of healthcare engineering*, 2017(1):4037190, 2017.

Guoxin Wang, Jun Zhao, Xinyi Liu, Yanbo Liu, Xuyang Cao, Chao Li, Zhuoyun Liu, Qintian Sun, Fangru Zhou, Haoqiang Xing, et al. Citrus-v: Advancing medical foundation models with unified medical image grounding for clinical reasoning. *arXiv preprint arXiv:2509.19090*, 2025.

Hallee E. Wong, Marianne Rakic, John Guttag, and Adrian V. Dalca. Scribbleprompt: Fast and flexible interactive segmentation for any biomedical image. In *Computer Vision – ECCV 2024*, pp. 207–229. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-73661-2_12.

Linshan Wu, Yuxiang Nie, Sunan He, Jiaxin Zhuang, Luyang Luo, Tao Li, Zhuoyao Xie, Dexuan Chen, Yinghua Zhao, Neeraj Mahboobani, et al. Unibiomed: A universal foundation model for grounded biomedical image interpretation. *arXiv preprint arXiv:2504.21336*, 2025.

Senqiao Yang, Tianyuan Qu, Xin Lai, Zhuotao Tian, Bohao Peng, Shu Liu, and Jiaya Jia. Lisa++: An improved baseline for reasoning segmentation with large language model. *arXiv preprint arXiv:2312.17240*, 2023.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, et al. Dapo: An open-source llm reinforcement learning system at scale. *Advances in Neural Information Processing Systems*, 38:113222–113244, 2026.

Muzhi Zhu, Yuzhuo Tian, Hao Chen, Chunluan Zhou, Qingpei Guo, Yang Liu, Ming Yang, and Chunhua Shen. Segagent: Exploring pixel understanding capabilities in mllms by imitating human annotator trajectories. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3686–3696, 2025.