MEDREFINEAGENT: RISK-AWARE MULTI-TURN MASK REFINEMENT FOR MEDICAL IMAGE SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Medical image segmentation has advanced along two largely separate lines: **prompt-free specialist models** that automatically produce complete masks, and **promptable foundation models** that refine regions through points, boxes, or other user prompts. In practical clinical pipelines, however, the input is often neither an empty image nor a manually prompted target, but an imperfect coarse mask produced by an automatic segmenter. This setting calls for a unified mechanism that can inherit the efficiency of prompt-free segmentation while exploiting the local correction ability of promptable models. In this paper, we propose **MedRefineAgent**, a risk-aware multi-turn agent for automatic coarse-to-fine medical mask refinement across both 2D images and 3D volumes. Specifically, **1)** a **prompt-free backbone** provides the initial coarse mask and an auxiliary reference for automatic risk localization; **2)** a **promptable local refiner** is invoked only on high-risk regions through a structured action space of region, repair primitive, and hyper-parameter; and **3) conservative repair control** coordinates gating, action pruning/reranking, and best-so-far rollback to reduce harmful edits. A teacher-distilled training pipeline further transfers search-based repair trajectories into lightweight test-time policies. Experiments across diverse 2D/3D medical segmentation settings show average improvements over the coarse prediction, while also exposing the limitations of disagreement-based localization and proxy rollback.

1 INTRODUCTION

Medical image segmentation has recently benefited from both strong specialist models such as nnU-Net Isensee et al. (2021) and promptable foundation models such as SAM, SAM 2, and MedSAM Kirillov et al. (2023); Ravi et al. (2024); Ma et al. (2024). Yet in many clinical pipelines the practical input is not an empty image but a *coarse* segmentation already produced by an automatic model. In this setting, the challenge is no longer how to segment from scratch, but how to identify local failure modes, choose suitable repair operations, and stop editing before useful structures are overwritten. We target this coarse-to-fine correction setting and model it as a multi-turn repair problem for both 2D and 3D medical segmentation.

In contrast to prior interactive or promptable segmentation work, our goal is not to request new human prompts for segmenting an object from scratch, but to automatically repair an existing coarse prediction by coordinating a promptable local refiner over multiple turns. In this sense, our framework is intended to bridge two previously separated paradigms: prompt-free automatic segmentation, which produces the initial coarse mask, and promptable interactive segmentation, which provides the local correction capability. The resulting problem is an underexplored autonomous medical mask-repair setting with explicit stopping and conservative control, rather than a claim that mask correction itself is new.

Our contributions are threefold:

1. We formulate an underexplored setting, *coarse-to-fine autonomous medical segmentation repair*, which connects prompt-free automatic segmentation and promptable interactive segmentation: an automatic model first generates a coarse mask, after which a promptable refiner performs autonomous correction and conservative stopping on high-risk disagreement regions.

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.

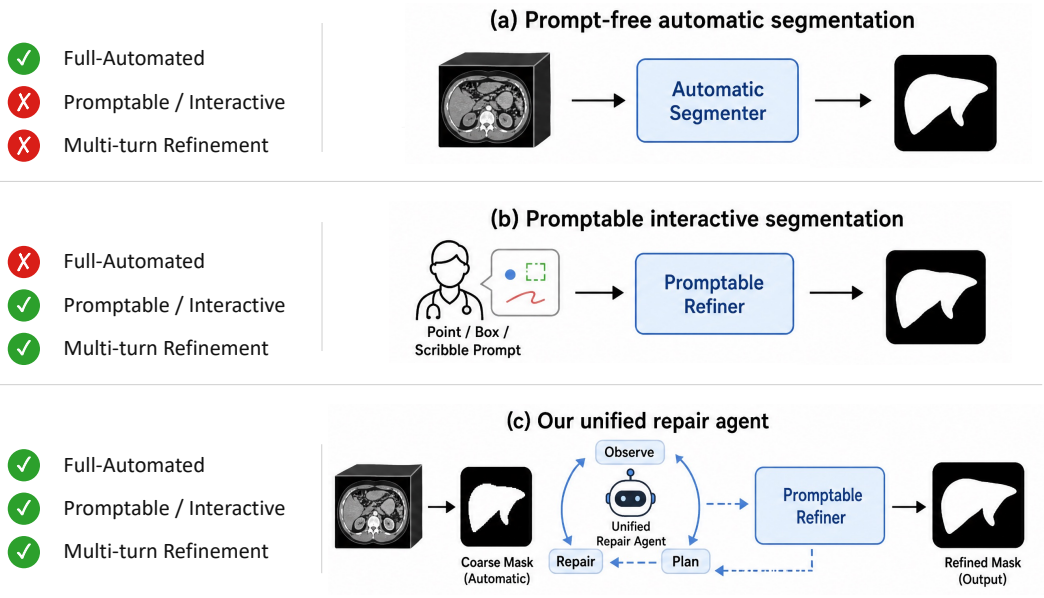

Figure 1: Overview of the proposed unified repair agent. The left panel shows prompt-free automatic segmentation producing a coarse mask, the center panel depicts our multi-turn risk-guided repair process, and the right panel illustrates promptable interactive segmentation.

2. We propose an interpretable structured repair policy that decomposes each repair decision into *region* × *primitive* × *hyper-parameter*. This explicit action design improves the transparency of the repair process and makes the agent's decisions more clinically inspectable than black-box mask prediction.
3. We develop a small-sample, training-light teacher-distilled learning pipeline. A stronger search-based teacher is first used to construct high-quality repair trajectories, and lightweight gate/ranker modules are then distilled from a limited amount of supervision to approximate the stronger teacher at substantially lower training cost.

2 Related Work

Medical Image Segmentation and Foundation Models

Medical segmentation spans convolutional encoder–decoders (U-Net, V-Net, U-Net++, and DeepLabV3+) Ronneberger et al. (2015); Milletari et al. (2016); Zhou et al. (2018); Chen et al. (2018), transformer hybrids (TransUNet, UNETR, Swin UNETR, CoTr, and nnFormer) Chen et al. (2021); Hatamizadeh et al. (2022); Tang et al. (2022); Xie et al. (2021); Zhou et al. (2021), and the self-configuring nnU-Net backbone used for our coarse masks Isensee et al. (2021). Promptable models such as SAM, SAM 2, and MedSAM Kirillov et al. (2023); Ravi et al. (2024); Ma et al. (2024) have also been adapted to 2D, 3D, and text-defined medical targets Cheng et al. (2023); Wang et al. (2023a); Du et al. (2024); Zhao et al. (2025); Zhu et al. (2024).

Automatic-prompt pipelines reduce target-image interaction under different assumptions. RFMedSAM 2 derives masks and boxes from a U-Net for two-stage SAM 2 refinement Xie et al. (2025); AutoMedSAM uses diffusion-based class prompts Huang et al. (2025); Sam2Rad learns prompts for musculoskeletal ultrasound Wahd et al. (2025); Proxy Prompt transfers cues from an annotated support pair Wang et al. (2025); and prompt-distillation methods improve SAM robustness under noisy prompts Kang et al. (2026). Unlike these predefined pipelines, MedRefineAgent re-estimates risk after each edit and jointly selects the region, action, hyper-parameter, and stopping point.

Segmentation Post-processing and Error Correction

Classical morphology and CRFs smooth predictions or sharpen boundaries Krähenbühl & Koltun (2011), with 3D CRF cleanup also used in DeepMedic Kamnitsas et al. (2017). Recent risk-aware segmentation also ranks errors from uncertainty or perturbation energy Fu et al. (2026). Model-agnostic refiners instead consume coarse masks: training-free SAMRefiner extracts noise-tolerant

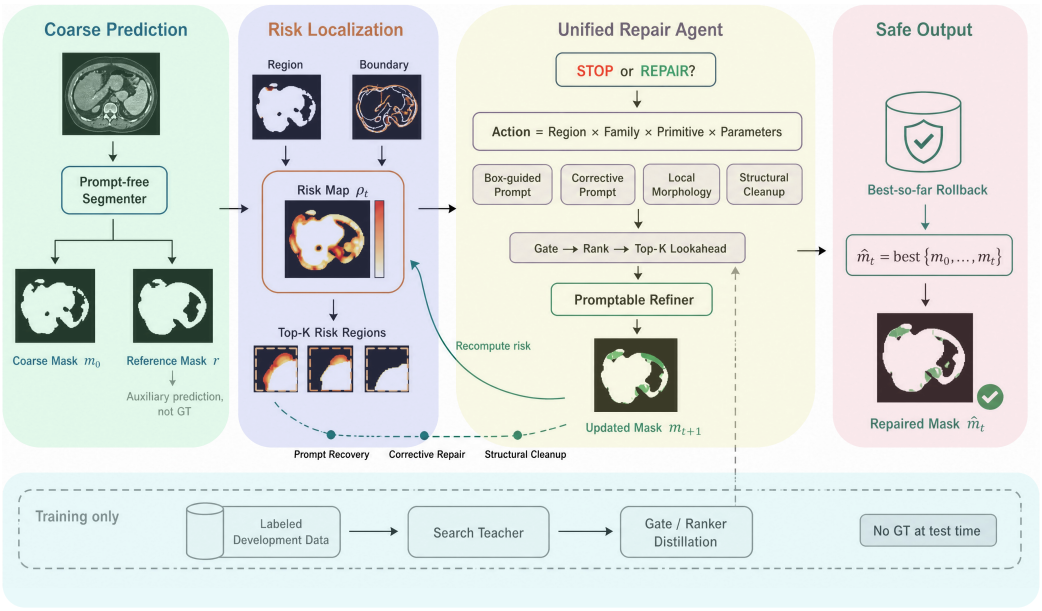

Figure 2: Full pipeline of the proposed risk-guided multi-turn repair framework. Stage 1 prepares the image, coarse mask, and auxiliary reference mask. Stage 2 constructs a fused risk map and extracts top-K_c candidate regions. Stage 3 applies structured family-aware action selection with progressive multi-turn scheduling. Stage 4 performs conservative proxy-based rollback and outputs the best-so-far mask, while teacher trajectories supervise lightweight gate and ranker distillation during training only.

point, box, and mask prompts Lin et al. (2025), whereas SegRefiner learns multi-step discrete-diffusion denoising Wang et al. (2023b). Both prescribe one refinement mechanism; our policy selects the repair region, operation, and stopping decision with conservative proxy control.

Interactive and Iterative Segmentation Refinement

Interactive methods refine masks through human clicks or boxes, including coarse-to-fine click learning Sun et al. (2024), FocalClick's localized correction and progressive merge Chen et al. (2022), efficient click-based mask correction Du et al. (2023), and interactive 3D refinement Cong et al. (2026). Test-time energy adaptation also iteratively adjusts predictions using learned shape energy Zhang et al. (2025b). These methods assume user-provided correction clicks, whereas we study autonomous medical repair in which risk localization, action selection, stopping, and proxy rollback are performed without test-time human clicks.

Agentic and Reinforcement-Learning-Based Segmentation

Reasoning-oriented models such as LISA, PixelLM, and GLaMM generate masks from language or grounded conversation Lai et al. (2024); Ren et al. (2024); Rasheed et al. (2024). SegAgent emits iterative clicks Zhu et al. (2025), while IBISAgent extends tool-using agents to biomedical referring and reasoning segmentation Jiang et al. (2026). Related sequential formulations use guideline-checking agents, contour-point cooperation, RL-assisted feature alignment, or multi-turn SAM prompting with agentic reinforcement learning Vats et al. (2026); Zhang et al. (2025a); Qu & Zhao (2026); Liu et al. (2026). MedRefineAgent instead distills a compact search-teacher policy for explicit risk-region repair of an available medical mask, with conservative stopping and rollback.

3 Method

Problem Formulation

Let Ω denote the spatial domain of either a native 2D image or a 3D volume. We process each case–label pair independently. We are given an image $I : \Omega \to \mathbb{R}$, a coarse mask $m_0 \in \{0,1\}^{|\Omega|}$ produced by one automatic segmentation predictor, and an auxiliary reference mask r obtained from

complementary predictors. We use e_{ens} for a direct ensemble or guarded-voting output when it is evaluated as a separate baseline, and $\hat{m}$ for the final MedRefineAgent output. The objective is not to segment from scratch, but to find a short repair trajectory

$$\tau = (a_0, a_1, \ldots, a_{T-1}) \tag{1}$$

that improves m_0 while avoiding harmful edits. We call our formulation *Risk-Ordered Structured Repair*: the agent first localizes where the coarse and reference masks disagree, then applies an interpretable repair operator, and finally decides whether further edits are still justified.

The reference mask is used as a low-cost auxiliary signal rather than as a replacement for the target prediction. In the reported protocol, m_0 is the fold-0 nnU-Net prediction and r is formed by guarded voting over the complementary folds; e_{ens} denotes a direct fusion baseline and is kept conceptually distinct from m_0. The main tables therefore use the explicitly named coarse predictor as the starting point, while reference-only and fusion-only scores should be read as separate information-budget controls rather than as the agent baseline.

Core insight. Coarse-mask errors are not homogeneous. Early errors are often region-level discrepancies that benefit from high-recall promptable edits, while later residuals are more often boundary, hole, or component-level errors. This motivates a progressive action design: prompt-based repair is used to recover missing or mislocalized regions, and structural operators are used later to consolidate local topology. This ordering is the main mechanism behind our multi-turn design rather than a flat enumeration of prompts. Empirically, strong medical segmentation backbones tend to leave a long-tailed error pattern: a few severe false-negative or false-positive regions dominate the Dice loss, while many remaining errors are small boundary leaks, holes, thin-structure breaks, or isolated components. The proposed turn schedule matches this hierarchy by moving from high-recall region recovery to boundary and topology consolidation.

Risk Map and Candidate Regions

At turn t, the agent observes $s_t = \{m_t, r, \rho_t, \mathcal{C}_t, h_t\}$, where m_t is the current mask, ρ_t is a dense risk map, $\mathcal{C}_t$ is a set of candidate regions, and h_t stores trajectory memory such as the turn index, previous family, previous region, and previous gain.

We compute ρ_t from two complementary disagreement cues. The first cue is the region-level mismatch between m_t and r. The second cue is boundary mismatch, obtained by extracting a width-δ boundary band from each mask using a signed distance transform. Denoting the two normalized disagreement maps by d_{reg} and d_{bd}, the fused risk map is

$$\rho_t(x) = \mathcal{N}\left(\alpha d_{\text{reg}}(x) + (1-\alpha) d_{\text{bd}}(x)\right), \tag{2}$$

where $\mathcal{N}(\cdot)$ normalizes values to $[0, 1]$ and α controls the region-boundary trade-off.

Candidate regions are then obtained by thresholding the risk map and running connected-component analysis. Components outside a valid area range are removed. Each retained component is summarized by its bounding box, anchor point or central slice, area, error type (false positive, false negative, or mixed), and a priority score computed from its average and peak risk. We keep the top K_c candidates, so subsequent decisions are made over a compact and interpretable set of risky regions rather than the full image grid.

This construction has an explicit blind spot. If $m_t = r$ at a location where both masks disagree with the ground truth, the disagreement risk can be low even though the prediction is wrong. Conversely, if r is worse than m_t, reference agreement can encourage an undesirable edit. We therefore interpret the risk map as a detector of recoverable disagreement errors, not as a complete error detector or a formal safety certificate.

Structured Repair Actions

Each turn selects either `stop` or a repair action. We factorize an action as

$$a_t = (u_t, c_t, g_t, p_t, \eta_t), \tag{3}$$

where $u_t \in \{\texttt{stop}, \texttt{repair}\}$, $c_t \in \mathcal{C}_t$ is the candidate region, g_t is an action family, p_t is a primitive inside that family, and η_t contains discrete hyper-parameters such as in-plane margin and, for 3D, a z-margin. Conceptually, the policy follows the conditional decomposition

$$\begin{aligned} \pi(a_t|s_t) = \pi(u_t|s_t)\pi(c_t|s_t, u_t) \\ \pi(g_t|s_t, c_t)\pi(p_t, \eta_t|s_t, c_t, g_t). \end{aligned} \tag{4}$$

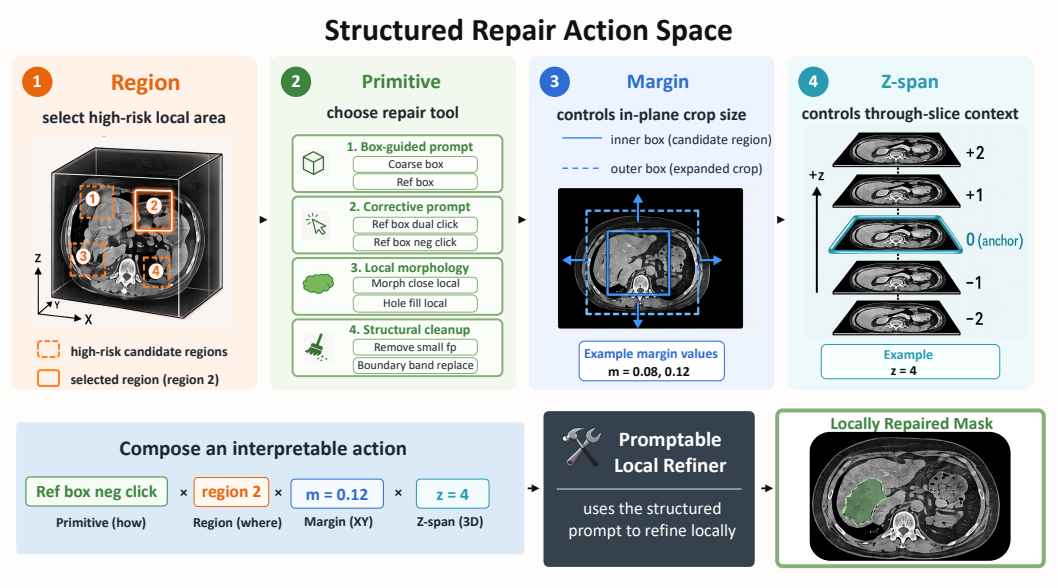

Figure 3: Structured repair action space. Each repair action is composed from a risky candidate region, a repair primitive, an in-plane margin, and, for 3D volumes, a through-slice z-span. This factorization makes the action space both interpretable and modular, while allowing the local promptable refiner to apply targeted edits around the selected region.

Algorithm 1 Risk-Ordered Structured Repair

Require: Image/volume I, coarse mask m_0, reference mask r, turn budget T
Ensure: Repaired mask $\hat{m}$
$m \leftarrow m_0$, $h \leftarrow \emptyset$, $\hat{m} \leftarrow m_0$
for $t = 0, 1, \ldots, T-1$ **do**
  Compute ρ_t and candidate regions $\mathcal{C}_t$
  Enumerate legal actions under the turn schedule
  Gate, prune, and rank actions with $Q_\theta(s_t, a)$
  Optionally select among top-K actions by proxy lookahead
  if the selected action is `stop` **then**
    break
  end if
  Execute local repair and fuse the edited crop into m
  Update memory h and best-so-far mask $\hat{m}$
end for
return $\hat{m}$

This decomposition is useful both computationally and clinically: it separates where to repair, what kind of edit to perform, and how aggressively to apply it.

The primitive vocabulary is grouped into four families: **box-guided prompt actions** for high-recall local refinement, **corrective prompt actions** for discrepancy-aware prompting, **local morphology actions** such as `morph_expand` for boundary thickness and gap correction, and **structural cleanup actions** for holes, false-positive components, boundary bands, and largest-component preservation. The progressive schedule is

Turn	Enabled families
Turn 1	box-guided prompt + corrective prompt
Turn 2	corrective prompt + local morphology
Turn 3+	local morphology + structural cleanup

Thus later turns are encouraged to perform complementary structural consolidation rather than repeatedly sampling near-duplicate prompt variants. For 2D datasets, z-margin is omitted; for 3D datasets, it controls the local slice context.

Policy Scoring, Proxy Lookahead, and Conservative Rollback

For each enumerated action, a lightweight scorer uses risk statistics, local shape and topology descriptors, trajectory history, primitive identity, and hyper-parameters as input features. At inference, the deployed policy has three explicit stages: a gate $G_\theta(s_t)$ decides whether repair is worth attempting, a ranker $Q_\theta(s_t, a)$ orders legal repair actions, and a stop action is selected when the gate rejects the state or no candidate has positive proxy gain. For an accepted repair state, a top-K proxy lookahead evaluates a small candidate set before committing:

$$a_t = \operatorname*{arg\,max}_{a \in \mathrm{TopK}(Q_\theta)} \Big(S(F(m_t, a); r, m_0) - S(m_t; r, m_0) \Big), \tag{5}$$

where $F(I, m_t, a)$ is the local executor, Q_θ is the learned action ranker, and S is a ground-truth-free proxy score. We define this proxy as a normalized linear score

$$\begin{aligned} S(m; r, m_0) = \lambda_r A_{\mathrm{reg}}(m, r) + \lambda_b A_{\mathrm{bd}}(m, r) \\ - \lambda_a D_{\mathrm{area}}(m, m_0) - \lambda_c C_{\mathrm{frag}}(m), \end{aligned} \tag{6}$$

where $A_{\mathrm{reg}}(m, r) = \mathrm{Dice}(m, r)$ measures whole-mask agreement with the auxiliary reference, $A_{\mathrm{bd}}(m, r) = \mathrm{Dice}(\partial_\delta m, \partial_\delta r)$ measures agreement between width-δ boundary bands, $D_{\mathrm{area}}(m, m_0) = \big||m| - |m_0|\big| / (|m_0| + \varepsilon)$ penalizes area drift from the coarse mask, and $C_{\mathrm{frag}}(m) = \max(0, N_{\mathrm{cc}}(m) - 1)$ penalizes fragmented foreground components after normalization by the candidate-crop size. Unless otherwise stated, we use $\lambda_r = 0.4$, $\lambda_b = 0.4$, $\lambda_a = 0.1$, and $\lambda_c = 0.1$; these weights are selected on the validation split and then fixed for test-time evaluation. In the reported teacher experiments, the trajectory reward is Dice gain; an action-cost penalty is optional and is set to zero unless explicitly stated. These terms are deliberately conservative: reference agreement can move a correct coarse region toward an imperfect reference, whereas area and fragmentation penalties discourage anatomically implausible expansion, shrinkage, or scattered components. The final prediction uses best-so-far proxy rollback:

$$\hat{m}_T = \arg \max_{m_j \in \{m_0, \ldots, m_T\}} S(m_j; r, m_0). \tag{7}$$

This allows non-greedy exploration and reduces proxy-scored harmful final edits, but it does not guarantee safety when m_t and r share an error or when proxy agreement is misaligned with the ground truth. The agent may explore intermediate edits, while the delivered mask is selected from the best proxy-scored state observed along the trajectory rather than blindly trusting the last action.

Teacher Construction and Distillation

On labeled development splits, a strong teacher evaluates candidate actions using their Dice gain. More generally, its trajectory reward can be written as

$$R(\tau) = \mathrm{Dice}(m_T^\tau, y) - \lambda_{\mathrm{cost}} \sum_{t=0}^{T-1} c(a_t), \tag{8}$$

where $c(a_t)$ is an action cost and $\lambda_{\mathrm{cost}} = 0$ for the reported tables unless otherwise stated. Beam search over the structured action space gives

$$\tau^\star = \arg \max_{\tau \in \mathcal{T}_K} \mathrm{Dice}(m_T^\tau, y), \tag{9}$$

which captures non-greedy trajectories that one-step imitation would miss.

The teacher supervises two lightweight modules. Stage 1 is a repair gate trained to predict whether the teacher trajectory improves the coarse mask by more than a small margin. Stage 2 is a family-first action ranker: actions whose reward is within ϵ_r of the best action at the same turn are treated as near-best positives, and the ranker is trained with pairwise/listwise ranking losses. The same construction is applied first at the family level and then within the selected family. This makes the learned policy match the structured action design rather than forcing a single flat classifier to choose among all primitives.

4 Experiments

Experiment Settings

We organize the evaluation into protocol definition, overall performance, and ablation analysis. The experiments are designed to answer four questions: whether the repair agent improves over coarse

masks and direct ensembling, whether it generalizes across 3D CT, 3D MRI, and native 2D images, whether teacher distillation remains effective with limited supervision, and which modules contribute most to reliable multi-turn refinement.

Dataset setup. We evaluate on five public benchmarks spanning volumetric and native 2D segmentation. For 3D evaluation, we use two representative multi-organ benchmarks covering abdominal and cardiac anatomy. AMOS22 Ji et al. (2022) provides separate independent CT and MRI cohorts with 15 abdominal organs and serves as the main cross-modality abdominal benchmark. MM-WHS Zhuang et al. (2019) provides disjoint CT/MRI cardiac volumes with seven whole-heart substructures for symmetric modality comparison. The 2D benchmarks include Kvasir-SEG, ISIC 2018 Task 1, and REFUGE, covering endoscopy, dermoscopy, and fundus imaging, respectively. We follow official splits when available. Hidden-test benchmarks are tuned only on validation data and evaluated following the official protocol. Each case-label pair is converted into a common repair tuple consisting of an image, a coarse mask, a reference mask, and a ground-truth mask for evaluation or teacher construction.

Baselines. We compare against three groups of baselines: prompt-free automatic segmentation, prompt-based foundation models and adaptation, and agent-style automatic prompting and refinement. Tab. 1 summarizes the independent-cohort 3D benchmarks on AMOS22 and MM-WHS, while Tab. 2 reports the common DSC metric on the native 2D datasets. The fold-0 coarse and MedRefineAgent rows are evaluated in our pipeline; comparison rows without an explicit same-protocol marker are literature-reported values and may use different splits, preprocessing, or prompt information. Manual-point, box, text-prompt, and automatic-prompt methods therefore do not share the same information budget.

Coarse and reference generation. For each target dataset, we train fold-specific nnU-Net v2 models on its official training split, using the 3D full-resolution pipeline for AMOS22 and MM-WHS and the native 2D pipeline for Kvasir-SEG, ISIC 2018, and REFUGE. Unless stated otherwise, the fold-0 prediction is fixed as the coarse mask m_0, and predictions from the remaining folds are fused into a GT-free auxiliary reference r by adaptive guarded voting. A reference-only prediction and a direct fusion output e_{ens} are separate controls and are not conflated with the fold-0 coarse row. The present headline tables report the coarse and agent rows; a standalone reference-only Dice must be reported separately before attributing gains to the controller. Ground-truth masks are used only for teacher construction on the development split and for final evaluation; they are never used to generate m_0 or r, localize risk, select actions, or stop a test-time trajectory.

Promptable local refiner. All prompt actions are executed by the public MedSAM2 checkpoint Zhu et al. (2024) with the `sam2.1_hiera_t512` configuration. Native 2D images are treated

Method	AMOS CT DSC↑	MMWHS CT DSC↑	AMOS MRI DSC↑	MMWHS MRI DSC↑
Prompt-free automatic segmentation				
U-Net	89.04	89.19	85.59	76.33
UNETR	79.43	88.96	75.30	74.72
Swin UNETR	86.32	90.48	75.70	80.46
CoTr	88.01	91.38	77.50	75.16
nnFormer	88.30	91.43	80.60	74.54
nnU-Net coarse (fold-0)	**86.33**	**90.05**	**87.10**	**80.71**
nnU-Net reference (4-fold)	**86.29**	**90.11**	**87.06**	**80.74**
Prompt-based foundation models and adaptation				
SAM-Med3D (1pt)	79.06	52.09	71.01	54.55
SAM-Med3D (10pt)	83.18	69.18	77.47	72.99
SegVol (box)	79.82	61.53	45.96	51.48
SAT-Pro (text)	86.37	89.97	78.90	86.70
Agent-style automatic prompting and refinement				
SAMRefiner	86.65	90.23	87.36	80.98
SegRefiner	84.47	88.89	85.47	79.36
Ours	**89.24**	**91.58**	**89.47**	**82.34**
Gain vs. coarse	+2.91	+1.53	+2.37	+1.63

Table 1: 3D comparison on AMOS and MM-WHS (DSC↑). The coarse row is the fold-0 nnU-Net prediction used as m_0; gains are absolute Dice-point improvements over this starting mask.

Method	Kvasir DSC↑	ISIC 2018 DSC↑	REFUGE Disc DSC↑	REFUGE Cup DSC↑
Prompt-free automatic segmentation				
U-Net	87.96	89.10	90.22	73.46
DeepLabV3+	87.07	85.70	85.66	71.71
TransUNet	89.21	89.30	95.27	88.66
nnU-Net coarse (fold-0)	**85.44**	**90.52**	**94.70**	**84.90**
nnU-Net reference (4-fold)	**85.51**	**90.48**	**94.73**	**84.87**
Prompt-based foundation models and adaptation				
SAM (training-free, 1-point)	68.0	66.0	39.10	33.50
SAM (training-free, box)	77.0	78.8	54.20	71.60
MedSAM (1-point)	68.0	86.8	92.90	82.10
SAM-Med2D (5-point)	82.20	88.40	92.10	82.00
Agent-style automatic prompting and refinement				
SAMRefiner	85.64	90.69	94.83	85.13
SegRefiner	83.51	89.40	94.43	83.74
Ours	**88.51**	**91.75**	**95.82**	**86.04**
Gain vs. coarse	+3.07	+1.23	+1.12	+1.14

Table 2: 2D comparison on Kvasir, ISIC 2018, and REFUGE (DSC↑). The coarse row is the fold-0 nnU-Net prediction used as m_0; gains are absolute Dice-point improvements over this starting mask.

as single-slice volumes. For 3D data, the selected axial subvolume is converted to three-channel 512×512 slices; the box, point, scribble, or mask prompt is applied on the anchor slice, and MedSAM2 propagates the prediction forward and backward only within the selected z-span. Multi-organ volumes are evaluated as case-label pairs, so the agent receives one binary coarse/reference pair at a time and no ground-truth information at inference.

Implementation details. For all datasets, the agent repairs a fixed coarse prediction rather than segmenting from scratch. Unless otherwise specified, the maximum turn budget is $T = 5$, the number of candidate regions is $K_c = 2$, in-plane margins are selected from $\{0.08, 0.12\}$, and 3D z-margins are selected from $\{4, 6\}$. The risk-map weight is set to $\alpha = 0.5$. The teacher uses beam size 8, near-best tolerance $\epsilon_r = 0.005$, and top-3 proxy lookahead. For 2D images, the z-margin is removed and the same action factorization reduces to region $\times$ primitive $\times$ margin. Hyper-parameters are selected on development splits and frozen for external testing.

Evaluation metrics. We report Dice as the main metric across all 2D and 3D benchmarks. Agent-specific metrics include mean gain over the coarse mask, improvement rate, degradation rate, and average turn count.

Overall Performance

3D experimental analysis. As shown in Tab. 1, the 3D suite evaluates whether structured repair transfers across abdominal multi-organ segmentation on AMOS22 and whole-heart segmentation on MM-WHS. Relative to the fold-0 coarse baseline, MedRefineAgent improves Dice by 2.91 and 1.53 points on AMOS22-CT and MM-WHS-CT, respectively, and by 2.37 and 1.63 points on AMOS22-MRI and MM-WHS-MRI. These average gains indicate improvement over the starting mask across CT and MRI, while direct fusion and reference-only controls remain necessary to isolate the controller contribution.

2D experimental analysis. As summarized in Tab. 2, the 2D experiments test whether the same repair abstraction transfers to native image domains without cross-slice propagation. Compared with the fold-0 coarse baseline, MedRefineAgent improves Dice by 3.07 points on Kvasir-SEG and 1.23 points on ISIC 2018, while yielding gains of 1.12 and 1.14 points for REFUGE optic-disc and optic-cup segmentation. These average improvements suggest that risk-guided refinement remains useful after removing the z-margin, but do not by themselves establish superiority over same-information fusion controls.

Ablation Studies

We report three complementary ablations in the main paper. The turn-budget study tests whether additional repair turns remain useful; the action-space study isolates the value of combining prompt families; and the risk-localization study evaluates whether the proposed disagreement cues identify true error regions. Increasing the budget from one to five turns improves Kvasir Dice from 0.8806 to 0.8851 and reduces degradation from 3.98% to 1.00%. The fused risk map also improves Precision@1 and Density Lift@1 over either cue alone, although it cannot resolve errors shared by the coarse and reference masks. Additional teacher–student and few-shot analyses are provided in Appendix B.

Max turns	Dice	Gain	Imp.	Deg.	Avg. T	Inference time (s)
$T = 0$ (coarse)	0.8544	0.0000	0.0000	0.0000	0.00	7.23
$T = 1$	0.8806	0.0262	0.4693	0.0398	1.00	15.41
$T = 2$	0.8828	0.0284	0.4795	0.0295	1.51	19.02
$T = 3$	0.8839	0.0295	0.5000	0.0100	2.02	23.76
$T = 4$	0.8845	0.0301	0.5000	0.0100	2.53	28.78
$T = 5$	0.8851	0.0307	0.5000	0.0100	3.04	33.57

Table 3: Turn-budget ablation on Kvasir-SEG. "Imp." and "Deg." denote improvement and degradation rates over the coarse mask. Inference time is the measured per-sample runtime in seconds, including the corresponding repair budget.

The first turn provides most of the improvement, while later turns give smaller gains for residual cleanup. The added refinement also increases runtime from 15.41 seconds at $T = 1$ to 33.57 seconds at $T = 5$.

Variant	Final Dice	Gain	Success rate
Box-guided prompts only	0.8558	0.0245	0.3500
Corrective prompts only	0.8575	0.0262	0.4700
Full hybrid action space	0.8579	0.0267	0.4700

Table 4: Oracle one-step ablation of prompt action-space capacity at the first repair turn. Candidate actions are selected by ground-truth Dice; the table measures the potential of each action vocabulary rather than the deployed student policy.

Corrective prompts outperform box-guided prompts alone, and the full hybrid vocabulary gives the best oracle result. The small hybrid gain suggests that action diversity is most useful when it complements discrepancy-aware prompting.

Risk variant	Hit@1	Precision@1	Density Lift@1
Region disagreement only	0.9900	0.1888	9.6984×
Boundary disagreement only	1.0000	0.1750	9.5127×
Fused risk map (ours)	1.0000	0.2439	16.2681×

Table 5: Risk-region localization ablation at turn 0. Higher is better for all three metrics.

The fused map preserves perfect Hit@1 while raising Precision@1 to 0.2439 and Density Lift@1 to 16.2681×. Combining region and boundary disagreement therefore concentrates candidates more effectively, although shared coarse/reference errors remain undetected.

5 Conclusion

In this work, we presented MedRefineAgent, a risk-aware multi-turn framework for refining coarse medical segmentation masks in both 2D images and 3D volumes. The framework connects prompt-free automatic segmentation with promptable local refinement through explicit risk localization, structured repair actions, learned gating and ranking, and conservative proxy rollback. Experiments on AMOS22, MM-WHS, Kvasir-SEG, ISIC 2018, and REFUGE show average Dice improvements across abdominal, cardiac, endoscopic, skin-lesion, and fundus segmentation. The method remains limited by consensus errors, imperfect references, and proxy misalignment, so the results should be interpreted as evidence for an autonomous repair protocol rather than a formal safety guarantee.

AI Assistance Disclosure

LLM-based tools were used to assist with drafting and editing parts of the manuscript, improving readability and structure, LaTeX formatting, bibliography verification, and wording the analysis of already obtained experimental results. The authors designed and implemented the method, designed and ran the experiments, and generated and verified the reported results without using LLMs to generate data or replace the experimental pipeline. The authors reviewed all AI-assisted material and take responsibility for the final content of this work, including all claims, analyses, and artifacts.

REFERENCES

Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision*, pp. 801–818, 2018.

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. doi: 10.1109/CVPR52688.2022.00136.

Junlong Cheng, Jin Ye, Zhongying Deng, Jianpin Chen, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyan Huang, Jilong Chen, Lei Jiang, Hui Sun, Junjun He, Shaoting Zhang, Min Zhu, and Yu Qiao. Sam-med2d. *arXiv preprint arXiv:2308.16184*, 2023.

Wei Cong, Yang Cong, Jiahua Dong, and Gan Sun. Towards efficient and effective interactive 3d segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 3425–3433, 2026.

Fei Du, Jianlong Yuan, Zhibin Wang, and Fan Wang. Efficient mask correction for click-based interactive image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22773–22782, 2023. doi: 10.1109/CVPR52729.2023.02181.

Yuxin Du, Fan Bai, Tiejun Huang, and Bo Zhao. Segvol: Universal and interactive volumetric medical image segmentation. In *Advances in Neural Information Processing Systems*, 2024.

Tianhao Fu, Austin Wang, Charles Chen, Roby Aldave-Garza, and Yucheng Chen. Segwithu: Uncertainty as perturbation energy for single-forward-pass risk-aware medical image segmentation. *arXiv preprint arXiv:2604.15271*, 2026.

Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R. Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 574–584, 2022.

Peng Huang, Shu Hu, Bo Peng, Xun Gong, Penghang Yin, Hongtu Zhu, Xi Wu, and Xin Wang. Diffusion-empowered AutoPrompt MedSAM. *arXiv preprint arXiv:2502.06817*, 2025.

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(2):203–211, 2021.

Yuanfeng Ji, Haotian Bai, Chongjian Ge, Jie Yang, Ye Zhu, Ruimao Zhang, Zhen Li, Lingyan Zhang, Wanling Ma, Xiang Wan, and Ping Luo. Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. In *Advances in Neural Information Processing Systems*, pp. 36722–36732, 2022. doi: 10.52202/068431-2661.

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. *arXiv preprint arXiv:2601.03054*, 2026.

Konstantinos Kamnitsas, Christian Ledig, Virginia F. J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78, 2017.

Jingxuan Kang, Ziqi Zhang, Shaoming Zheng, Shuang Li, Uday Bharat Patel, Alexander Harry Fitzhugh, Phillip Lung, and Yusuf Kiberu. Learning from noisy prompts: Saliency-guided prompt distillation for robust segmentation with sam. *arXiv preprint arXiv:2604.23314*, 2026.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.

Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in Neural Information Processing Systems*, 2011.

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.

Yuqi Lin, Hengjia Li, Wenqi Shao, Zheng Yang, Jun Zhao, Xiaofei He, Ping Luo, and Kaipeng Zhang. SAMRefiner: Taming segment anything model for universal mask refinement. *arXiv preprint arXiv:2502.06756*, 2025.

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.

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

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision*, pp. 565–571, 2016. doi: 10.1109/3DV.2016.79.

Jierui Qu and Jianchun Zhao. Rl-u2net: A dual-branch unet with reinforcement learning-assisted multimodal feature fusion for accurate 3d whole-heart segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 8640–8648, 2026. doi: 10.1609/aaai.v40i11.37816.

Hanoona Rasheed, Muhammad Maaz, Sahal Shaji Mullappilly, Abdelrahman Shaker, Salman Khan, Hisham Cholakkal, Rao M. Anwer, Erix 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.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 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*, pp. 234–241, 2015.

Hao Sun, Mingyao Zhou, Wenjing Chen, and Wei Xie. Cfr-icl: Cascade-forward refinement with iterative click loss for interactive image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 4998–5007, 2024.

Yucheng Tang, Dong Yang, Wenqi Li, Holger R. Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20730–20740, 2022.

Vanshika Vats, Ashwani Rathee, and James Davis. Guideline-consistent segmentation via multi-agent refinement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 9612–9620, 2026. doi: 10.1609/aaai.v40i12.37923.

Assefa Seyoum Wahd, Banafshe Felfeliyan, Yuyue Zhou, Shrimanti Ghosh, Adam McArthur, Jiechen Zhang, Jacob L. Jaremko, and Abhilash Hareendranathan. Sam2Rad: A segmentation model for medical images with learnable prompts. *Computers in Biology and Medicine*, 187: 109725, 2025. doi: 10.1016/j.compbiomed.2025.109725.

Haoyu Wang, Sizheng Guo, Jin Ye, Zhongying Deng, Junlong Cheng, Tianbin Li, Jianpin Chen, Yanzhou Su, Ziyan Huang, Yiqing Shen, Bin Fu, Shaoting Zhang, Junjun He, and Yu Qiao. Sam-med3d: Towards general-purpose segmentation models for volumetric medical images. *arXiv preprint arXiv:2310.15161*, 2023a.

Mengyu Wang, Henghui Ding, Jun Hao Liew, Jiajun Liu, Yao Zhao, and Yunchao Wei. SegRefiner: Towards model-agnostic segmentation refinement with discrete diffusion process. In *Advances in Neural Information Processing Systems*, volume 36, 2023b.

Xinyi Wang, Hongyu Kang, Peishan Wei, Shuai Li, Yu Sun, Sai Kit Lam, and Yongping Zheng. Proxy prompt: Endowing sam and sam 2 with auto-interactive-prompt for medical segmentation. *arXiv preprint arXiv:2502.03501*, 2025.

Bin Xie, Hao Tang, Yan Yan, and Gady Agam. RFMedSAM 2: Automatic prompt refinement for enhanced volumetric medical image segmentation with SAM 2. *arXiv preprint arXiv:2502.02741*, 2025.

Yutong Xie, Jianpeng Zhang, Chunhua Shen, and Yong Xia. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. *arXiv preprint arXiv:2103.03024*, 2021.

Ruicheng Zhang, Yu Sun, Zeyu Zhang, Jinai Li, Xiaofan Liu, Au Hoi Fan, Haowei Guo, and Puxin Yan. Marl-mambacontour: Unleashing multi-agent deep reinforcement learning for active contour optimization in medical image segmentation. *arXiv preprint arXiv:2506.18679*, 2025a.

Xiaoran Zhang, Byung-Woo Hong, Hyoungseob Park, Daniel H. Pak, Anne-Marie Rickmann, Lawrence H. Staib, James S. Duncan, and Alex Wong. Progressive test time energy adaptation for medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22338–22348, 2025b. doi: 10.1109/ICCV51701.2025.02074.

Ziheng Zhao, Yao Zhang, Chaoyi Wu, Xiaoman Zhang, Xiao Zhou, Ya Zhang, Yanfeng Wang, and Weidi Xie. Large-vocabulary segmentation for medical images with text prompts. *npj Digital Medicine*, 2025. doi: 10.1038/s41746-025-01964-w.

Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*, 2021.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11, 2018.

Jiayuan Zhu, Abdullah Hamdi, Yunli Qi, Yueming Jin, and Junde Wu. Medical sam 2: Segment medical images as video via segment anything model 2. *arXiv preprint arXiv:2408.00874*, 2024.

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.

Xiahai Zhuang, Lei Li, Christian Payer, Darko Štern, Martin Urschler, Mattias P. Heinrich, Julien Oster, Chunliang Wang, Örjan Smedby, Cheng Bian, Xin Yang, Pheng-Ann Heng, Aliasghar Mortazi, Ulas Bagci, Guanyu Yang, Chenchen Sun, Gaetan Galisot, Jean-Yves Ramel, Thierry Brouard, Qianqian Tong, Weixin Si, Xiangyun Liao, Guodong Zeng, Zenglin Shi, Guoyan Zheng, Chengjia Wang, Tom MacGillivray, David Newby, Kawal Rhode, Sebastien Ourselin, Raad Mohiaddin, Jennifer Keegan, David Firmin, and Guang Yang. Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. *Medical Image Analysis*, 58:101537, 2019. doi: 10.1016/j.media.2019.101537.

A Additional method and protocol details

State Features and Candidate Filtering

The lightweight gate and ranker operate on structured, ground-truth-free features rather than additional deep segmentation networks. The feature groups include risk statistics (mean and peak risk, region risk, boundary risk, and risk density), geometry (box area, aspect ratio, center, and component compactness), topology (hole count, component count, and boundary-band ratio), trajectory state (turn index, previous family, previous gain, and repeated-region flags), and action identity (primitive family, margin, z-span, and prompt type). Missing values are median-imputed before fitting the tabular controllers.

Candidate regions are extracted by thresholding ρ_t and applying connected-component analysis. Components are filtered by minimum area and compactness constraints to remove tiny noise islands and overly large regions that are unlikely to be useful local repair targets. Each retained component is summarized by its bounding box, center point, area, average risk, peak risk, disagreement type, and boundary-risk fraction. Only the top K_c candidates are retained at each turn.

Feature group	**Representative features**
Risk statistics	Mean/maximum risk, region risk, boundary risk, risk density
Geometry	Box area, aspect ratio, center, component compactness
Topology	Hole count, component count, boundary-band ratio
Trajectory	Turn index, previous family, previous gain, repeated-region flag
Action identity	Primitive family, margin, 3D z-span, prompt type

Table 6: Feature groups used by the lightweight repair gate and action ranker.

Structured Action Library

The action generator enumerates prompt actions and explicit local mask operators using the same candidate region and margin parameters. Prompt actions use the selected region box; mask and scribble variants are constructed from current/reference disagreement inside that crop. For a selected component with bounding box $b = (x_1, y_1, x_2, y_2)$, the crop expands by $m = \mathrm{round}(\gamma\ \mathrm{size}(b))$ pixels on each side, where γ is the selected in-plane margin. The crop is clamped to the image boundary. In 3D, the selected slice interval is additionally expanded by the selected z-span and the anchor-slice prediction is propagated through the local volume. If a prompt degenerates to an empty mask, the executor falls back to the corresponding box-only action and records the event.

Family	**Primitive**	**Prompt or local operation**
Box-guided prompts	`coarse_box`	Current-mask box and local mask prompt
	`ref_box`	Auxiliary-reference box prompt
Corrective prompts	`ref_box_dual_click`	Reference box with positive and negative clicks
	`dense_delta_mask`	Dense current/reference disagreement mask
	`ref_pos_scribble_mask`	Positive reference-supported scribble mask
Local morphology	`morph_expand`	One local dilation step around the selected crop
	`morph_shrink`	One local erosion step around the selected crop
	`morph_close_local`	Local gap closing
	`morph_open_local`	Local noise opening
Structural cleanup	`hole_fill_local`	Fill internal holes inside the selected crop
	`remove_small_fp`	Remove unsupported foreground components
	`boundary_band_replace_local`	Replace the local boundary band with the reference
	`keep_largest_component_local`	Keep the dominant connected component locally

Table 7: Action primitives implemented by the repair executor.

Teacher, Controller, and Trace Details

The teacher evaluates candidate actions on labeled development data using Dice gain minus an optional action cost. Beam search uses width 8, and actions within $\epsilon_r = 0.005$ of the best action at the

same turn are retained as near-best positives. The gate label is one when the teacher action improves the coarse mask by more than 0.005. The stage-2 ranker uses standardized features and grouped five-fold cross-validation; actions within 0.01 of the best teacher action are treated as positives and up to five hard negatives are sampled per positive. The default nonlinear gate uses histogram gradient boosting with maximum depth 3, learning rate 0.05, 200 iterations, and minimum leaf size 4. The ranker uses maximum depth 4, learning rate 0.05, 300 iterations, and minimum leaf size 4.

At test time, an action is accepted only when its decision metric exceeds the continuation threshold. The standard launch configuration uses a minimum continuation reward of 10^{-4}; a long-turn 2D preset uses -10^{-3} to avoid premature termination. Best-so-far rollback is applied after every accepted action. Each rollout records the selected region box, crop box, z-range, maximum and mean risk, reference-agreement gain, changed pixels/voxels, fallback usage, selection source, accepted-turn count, and stop reason. These trace fields support per-dataset turn-count, degradation, and failure-case analysis without exposing ground truth at test time.

Dataset and Input Protocol

The benchmark suite covers AMOS22 and MM-WHS for 3D repair and Kvasir-SEG, ISIC 2018, and REFUGE for native 2D repair. The reported pipeline separates the prompt-free segmentation stage from local repair: the designated nnU-Net fold produces m_0, while complementary fold predictions are fused into the auxiliary reference r. Native 2D images are represented as single-slice pseudo-volumes, so the through-slice parameter is zero. For 3D local refinement, each slice in a selected crop is converted to three-channel input and resized to 512×512 before normalization. The ISIC resize preset additionally caps the long side at 1280 pixels before local refinement.

Unless a dataset-specific launch script overrides the threshold, risk regions are obtained from positive entries of the fused risk map using the 85th percentile of positive risk values. The default minimum connected-component area is 16 pixels/voxels, and the top two regions are retained. The selected region uses an in-plane margin from $\{0.08, 0.12\}$. For 3D runs, the through-slice context is swept over $\{4, 6\}$; for 2D runs it is fixed to zero.

Hyper-parameter	**Default value**
Maximum turns T	5
Candidate regions K_c	2
Risk fusion weight α	0.5
In-plane margins	$\{0.08, 0.12\}$
3D z-spans	$\{4, 6\}$
Teacher beam size	8
Near-best tolerance ϵ_r	0.005
Proxy lookahead candidates	3

Table 8: Default hyper-parameters used by MedRefineAgent.

B Additional analyses

Few-shot distillation

To evaluate the data efficiency of the teacher-distilled controller, we vary the number of teacher trajectories used for training, i.e., 10, 25, 50, 100, and 880 cases. We report final Dice, improvement rate, and repairable-case recall in Tab. 9. This setting tests whether lightweight gate/ranker modules can approximate a strong teacher under limited supervision, rather than relying on large-scale reinforcement learning.

Teacher cases	Final Dice	Improvement rate	Repairable recall
10	87.25	27.50%	45.83%
25	87.47	30.83%	51.67%
50	87.63	34.17%	56.67%
100	87.98	39.20%	65.80%
880	**88.51**	**46.67%**	**79.17%**

Table 9: Few-shot distillation on the Kvasir-SEG development split with varying numbers of teacher trajectories. Repairable recall is the fraction of development cases for which the teacher identifies at least one action with positive Dice gain over m_0; values are from the reported single split.

QUALITATIVE TRAJECTORY

Figure 4 illustrates one Kvasir-SEG 2D repair trajectory from coarse mask to final mask across five turns.

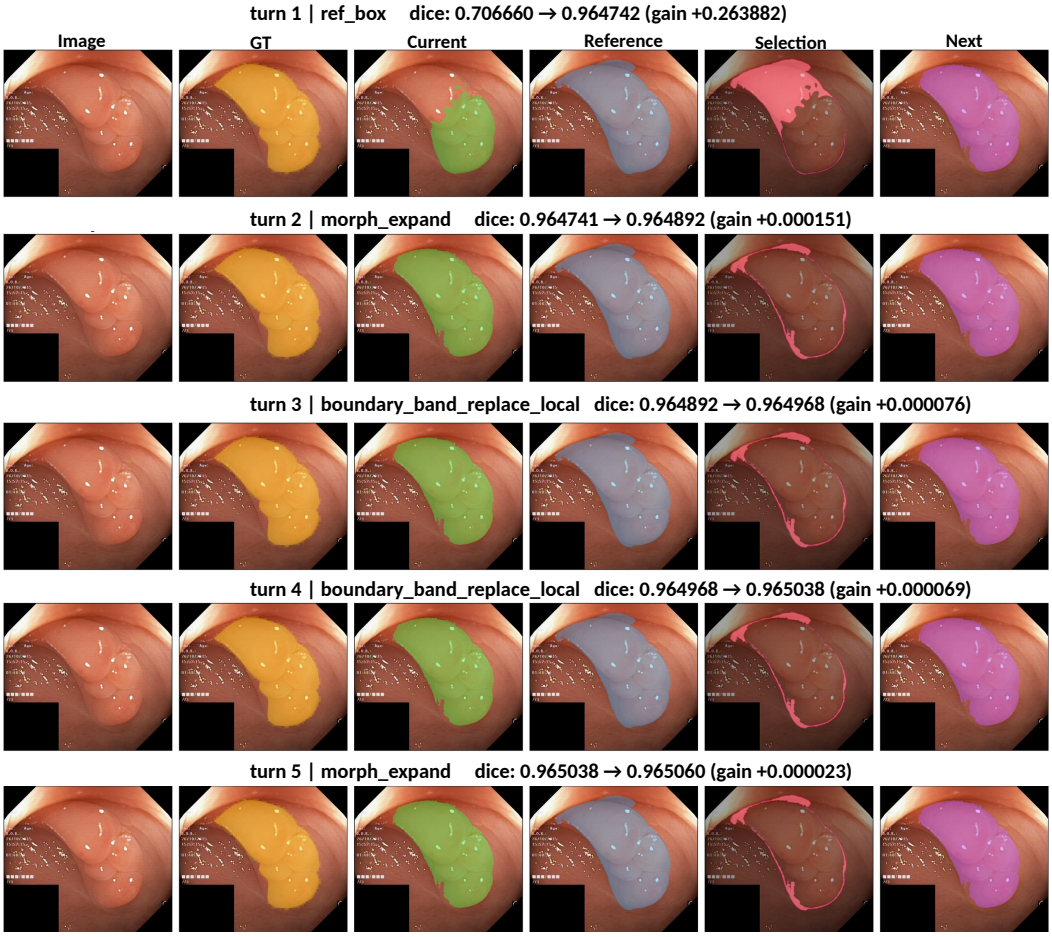

Figure 4: Qualitative multi-turn repair on one Kvasir-SEG 2D example. Each row reports one turn and shows the image, ground truth, current mask, auxiliary reference, selected repair region, and updated mask. The example is illustrative rather than a representative 2D/3D trajectory summary.

TEACHER–STUDENT COMPARISON

Tab. 10 contrasts the coarse mask, the deployed distilled student, and the oracle teacher. The student is GT-free at inference, while the oracle uses ground-truth Dice only for trajectory selection.

Policy variant	Dice	Gain	Imp.	Deg.
Coarse mask	85.44	–	–	–
Distilled student	88.51	3.07	50.00%	2.05%
Oracle teacher	89.13	3.69	52.05%	0%

Table 10: Teacher–student policy comparison. “Imp.” and “Deg.” denote improvement and degradation rates over the coarse mask. The distilled student is the deployed ground-truth-free policy, while the oracle teacher uses ground-truth Dice for trajectory selection.