MedReasoner: Reinforcement Learning Drives Reasoning Grounding from Clinical Thought to Pixel-Level Precision

Zhonghao Yan[1*], Muxi Diao[12*], Yuxuan Yang[1], Ruoyan Jing[1], Jiayuan Xu[1], Kaizhou Zhang[1], Lele Yang[1], Yanxi Liu[3], Kongming Liang[1†], Zhanyu Ma[1]

[1]Beijing University of Posts and Telecommunications
[2]Zhongguancun Academy
[3]Beijing Information Science and Technology University

Abstract

Accurately grounding regions of interest (ROIs) is critical for diagnosis and treatment planning in medical imaging. While multimodal large language models (MLLMs) combine visual perception with natural language, current medical-grounding pipelines still rely on supervised fine-tuning with explicit spatial hints, making them ill-equipped to handle the implicit queries common in clinical practice. This work makes three core contributions. We first define **Unified Medical Reasoning Grounding (UMRG)**, a novel vision–language task that demands clinical reasoning and pixel-level grounding. Second, we release **U-MRG-14K**, a dataset of 14K samples featuring pixel-level masks alongside implicit clinical queries and reasoning traces, spanning 10 modalities, 15 super-categories, and 108 specific categories. Finally, we introduce **MedReasoner**, a modular framework that distinctly separates reasoning from segmentation: an MLLM reasoner is optimized with reinforcement learning, while a frozen segmentation expert converts spatial prompts into masks, with alignment achieved through format and accuracy rewards. MedReasoner achieves state-of-the-art performance on U-MRG-14K and demonstrates strong generalization to unseen clinical queries, underscoring the significant promise of reinforcement learning for interpretable medical grounding.

1 Introduction

Medical imaging plays a central role in modern healthcare, where clinicians routinely examine regions of interest (ROIs) within these images to assess the health of organs and tissues (Cheng et al. 2023; Lin et al. 2024; Yan et al. 2025). Consequently, precise object detection and image segmentation (often called **grounding**) are essential for tasks such as disease diagnosis and treatment planning (Chen et al. 2021; Ma et al. 2024). To further enhance diagnostic efficiency and interpretability, medical Multimodal Large Language Models (MLLMs) have recently emerged (Li et al. 2023; Chen et al. 2024a; Xu et al. 2025). These models integrate visual perception with language interaction, allowing them to accept free-form language queries, generate high-quality responses, and even identify queried ROIs.

Despite these significant advances, a crucial limitation persists: **MLLM outputs remain at the image level**. To translate reasoning into visual outputs, every language reference must be grounded to a spatial location. However, while expert models (Cheng et al. 2023; Yue et al. 2024) achieve high grounding accuracy, they rely on precise spatial prompts (e.g., *bounding boxes* and *points*). Such detailed annotations are rarely provided by clinicians in real workflows (see Fig. 1 for an example query). Recent MLLMs attempt to move beyond handcrafted prompts by coupling rich visual components (Da et al. 2024; Huang et al. 2025b). However, existing medical grounding pipelines are still trained in a fully supervised manner on explicitly phrased referring expressions (e.g., "*segment the left lung*") (Liu et al. 2023; Koleilat et al. 2024). Collecting such finely annotated data is costly and, more importantly, misaligned with real clinical queries, which are often **implicit** (e.g., "*What can be inferred from the irregular shadow?*"). Although some models can name anatomical structures, they often fail to ground them (see Fig. 1). Therefore, we need models with reasoning that can turn implicit clinical phrases into explicit spatial targets for grounding in clinical scenarios.

Existing medical visual–question answering (VQA) datasets (Lau et al. 2018; He et al. 2020; Liu et al. 2021) evaluate semantic understanding with image-level question–answer pairs but lack spatial labels. Conversely, large-scale segmentation datasets (Ye et al. 2023; Zhao et al. 2024; Li et al. 2024) provide pixel-accurate masks yet omit language annotations. Neither class of dataset addresses the implicit queries that arise in real clinical practice. **We have no principled way to measure whether a framework can translate implicit clinical queries into precise spatial grounding**. Here, we are particularly interested in two research questions that must be addressed before implicit clinical queries can be grounded reliably:

- ***RQ1:*** *How can we create data that mirrors clinicians' implicit query patterns while still providing the pixel-level annotations needed for training and evaluation?*
- ***RQ2:*** *How can we enable models to interpret implicit clinical queries and accurately ground the corresponding image regions without handcrafted spatial prompts?*

Guided by the research questions above, we formally introduce the **Unified Medical Reasoning Grounding**

[*]These authors contributed equally.
[†]Corresponding Author

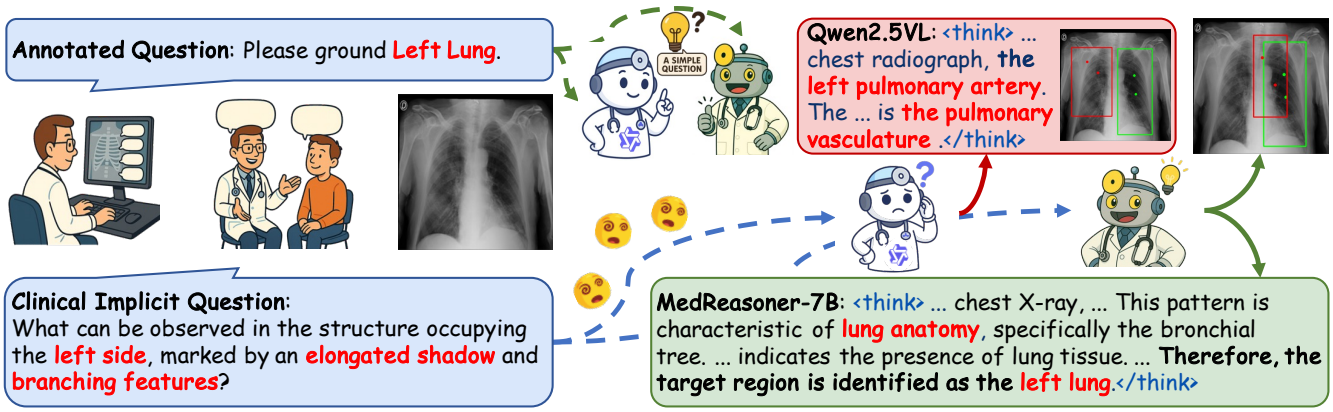

Figure 1: Comparison of annotated question and implicit clinical question. The ground-truth bounding box is green, and models' predicted box is red. **MedReasoner** precisely identifies the target with the reasoning trace and achieves accurate grounding.

(UMRG) task. UMRG demands a framework that integrates linguistic reasoning with spatial grounding. To succeed, a framework need: (1) interpret the implicit query, (2) reason over visual cues and anatomical priors to infer the latent target, and (3) generate the accurate pixel-level grounding of that ROIs. This three-stage process mirrors how clinicians inspect images, reflect, and mark ROIs. Full task specifications are given in Section 4.1.

In response to ***RQ1***, we propose **U-MRG-14K**, a rigorously curated dataset of 14K high-quality samples tailored to the UMRG task. U-MRG-14K is constructed from three open-source datasets. To generate semantically rich and clinically meaningful supervision, we employ GPT-4o (OpenAI 2024) as a simulator of clinician behavior. And we design a three-stage prompting pipeline that yields high-quality QA pairs, including implicit queries, chain-of-thought (CoT) reasoning traces, and final grounding steps for each target region. Further construction details appear in Section 3.

In response to ***RQ2***, we present **MedReasoner**, a reinforcement learning (RL) framework for medical reasoning and grounding. MedReasoner is decoupled into two plug-and-play components: **Clinical Reasoning Module (CRM),** any MLLM that reasons over implicit queries and generates lightweight spatial prompts (a bounding box plus two key points); **Anatomical Segmentation Module (ASM),** any model that accepts these prompts and returns a pixel-level mask. Because CRM and ASM exchange minimal geometric cues, they can be upgraded without retraining the other. Most existing grounding pipelines rely on supervised fine-tuning (SFT) with special tokens (Lai et al. 2024; Tong et al. 2025). This approach suffers from: **(1) annotation hunger,** it requires large, heavily annotated datasets and CoT traces are especially costly; and **(2) phrase overfitting,** it encourages models to echo explicit referring phrases and fails to develop genuine reasoning ability. MedReasoner solve these weaknesses through a rule-based RL training scheme that optimizes only the CRM. In each step, the CRM produces a *think* trace and an *answer* containing spatial prompts, and the frozen ASM renders a mask. Rewards for output format and spatial accuracy drive exploration, gradually aligning reasoning with precise grounding and achieving state-of-the-art performance on U-MRG-14K. As shown in Fig. 1, the RL-driven MedReasoner yields sharper grounding and more coherent reasoning than an instruction-tuned baseline, demonstrating its superiority on implicit-query grounding.

To summarize, our contributions are as follows:

- We formulate the **UMRG** task and propose **U-MRG-14K**. U-MRG-14K pairs implicit clinical queries with pixel-level masks and includes CoT traces to improve the interpretability of grounding.
- We present **MedReasoner**, an RL-driven, plug-and-play framework in which the CRM and the ASM are fully decoupled. This design enables easy substitution and extension to future models and clinical modalities.
- We demonstrate through extensive empirical evaluations the effectiveness of our proven MedReasoner framework. We will release the code, and dataset for future research.

2 Related Work

2.1 MLLMs for Medical Image Analysis

Recent advancements in MLLMs have significantly enhanced their capabilities for medical image analysis, with contributions from visual-language alignment techniques (Zhu et al. 2025; Wang et al. 2024; Yuan et al. 2024; Bai et al. 2025; Guo et al. 2025) . These progressions have been further extended to various medical applications, including the integration of visual expert modules into pre-trained language models (Li et al. 2023; Sellergren et al. 2025), and the unification of medical understanding and generation through heterogeneous knowledge adaptation and general foundation models (Chen et al. 2024a; Lin et al. 2025; Xu et al. 2025). However, significant gaps remain in their handling of clinical complexities and crucial clinical grounding tasks, which have seen limited exploration.

2.2 Visual Grounding with Medical Reasoning

Recent MLLMs have demonstrated powerful reasoning capabilities (OpenAI 2024; Guo et al. 2025; Liu et al. 2025; Zhu et al. 2025; Bai et al. 2025). For visual grounding in

general-purpose images, these models often leverage segmentation tools like SAM (Kirillov et al. 2023), with methods ranging from training new tokens (Lai et al. 2024; Ren et al. 2024) to prompting for geometric outputs (Chen et al. 2024b; Uesato et al. 2022). However, direct application in medical scenarios is challenging due to opaque reasoning and noisy data. While some specialized works have attempted to address this (Huang et al. 2025b; Trinh et al. 2025; Luo et al. 2024; Li et al. 2025), they often struggle with the natural language found in clinical practice. Inspired by Seg-Zero (Uesato et al. 2022), we employ reinforcement learning to generate an explicit CoT (Wei et al. 2022). This approach enhances medical visual grounding performance while offering a transparent reasoning process, thereby increasing trust in clinical applications.

3 U-MRG-14K Dataset

3.1 Data Generation

Most existing medical imaging datasets treat visual-grounding and VQA as separate tasks. As a result, some models support natural-language interaction without pixel-level analysis, whereas the accuracy of mainstream segmentation models hinges on the precision of supplied visual prompts. MoCoVQA (Huang et al. 2025a) attempts to unify the two tasks, yet its questions use explicit phrasing that fails to reflect the ambiguity common in routine clinical practice.

To address this gap, we construct **U-MRG-14K**, a medical grounding dataset centered on implicit referential expressions. U-MRG-14K is generated with GPT-4o (OpenAI 2024) through carefully designed prompts. As shown in Fig. 2, its generation process has three stages.

Stage 1: Dataset Preprocessing. We collect 14K image–mask pairs from SA-Med2D-20M (Ye et al. 2023), BiomedParse (Zhao et al. 2024), and IMIS-Bench (Cheng et al. 2025). We then standardize and complete the *super-category* labels (coarse anatomical regions) and *category* labels (specific organs or lesions) from the source datasets, producing a consistent and reliable taxonomy. The dataset comprises 15 super-categories and 108 categories. Table 1 provides a systematic comparison showing the advantages of U-MRG-14K over existing datasets.

Stage 2: Descriptions & QA Formats Generation. To facilitate the creation of high-quality QA pairs, we perform two preparatory steps. First, for each image, we generate two complementary descriptions: (i) a **short description** capturing the visual appearance of the region in plain and intuitive language, and (ii) a **long description** providing a medically precise interpretation of the target area. Second, we use GPT-4o to design a set of QA formats for each super-category. The *questions* mimic realistic clinical queries with vague or implicit references, while the *answers* provide a step-by-step, clinical reasoning path for correct grounding. On average, we create about 20 formats per super-category, with the exact number manually adjusted for class diversity.

Stage 3: QA Pairs Construction. Using the per-image descriptions and super-category QA formats from Stage 2, we prompt GPT-4o to synthesize instance-level QA pairs.

Dataset	# Prompts	QAs	Sup.	Cat.	CoT
SA-Med2D	20M	✗	-	219	✗
BioMedParse	1.1M	✗	3	82	✗
IMED	361M	✗	6	204	✗
MoCoVQA	100K	✓	-	-	✗
U-MRG-14K	14K	✓	15	108	✓

Table 1: Comparison of U-MRG-14K with existing medical vision–language datasets. **Sup.** and **Cat.** denote the numbers of *super-categories* and fine-grained *categories*, respectively. U-MRG-14K supplies customized QA templates for each category, and is the only dataset that includes CoT annotations for reasoning-aware evaluation.

Each answer contains an explicit, step-by-step reasoning trace guiding the model from an under-specified query to the correct spatial grounding, thereby enhancing interpretability and enabling manual verification. Prompts are iteratively refined, and all generated QA pairs undergo manual screening to remove factual inconsistencies or misaligned reasoning. U-MRG-14K is the first medical-image grounding dataset that includes both pixel-level annotations and complete CoT reasoning traces, providing a valuable resource for reasoning-based grounding and implicit-query QA tasks.

3.2 Dataset statistics

U-MRG-14K contains 14K image-mask pairs from ten imaging modalities. The dataset is organized into 15 super-categories covering frequent anatomical regions and pathology-oriented classes, providing broad clinical coverage. Within these, 108 fine-grained categories denote specific structures, reflecting hierarchical structure of anatomy. For instance, *left lung* and *right lung* are separate categories nested under the *lung* super-category. Beyond pixel-level masks, every sample includes a CoT reasoning trace. These annotations make the reasoning process verifiable, allowing researchers to inspect the model's decision path.

4 MedReasoner

4.1 Task Definition

Given a medical image $\mathcal{I}$ and a clinical query $\mathcal{Q}$ with implicit referring expressions, the model $\mathbf{G}$ outputs a bounding box $\mathcal{B}$, two semantic key points $\mathcal{P}_1$ and $\mathcal{P}_2$, and a pixel-level segmentation mask $\mathcal{M}$. The process can be formulated as:

$$\{\mathcal{T}, \mathcal{B}, \mathcal{P}_1, \mathcal{P}_2, \mathcal{M}\} = \mathbf{G}(\mathcal{I}, \mathcal{Q}). \quad (1)$$

where $\mathcal{T}$ is an optional CoT trace that records the model's intermediate reasoning, analogous to how a clinicians infers the target from implicit linguistic cues.

4.2 Model Architecture

Enabling native pixel-level segmentation in an MLLM usually requires custom `[MASK]` tokens, multi-head decoders, and large collections of manual mask annotations (Pi et al. 2024; Lai et al. 2024; Tong et al. 2025). However, MedSAM2 (Ma et al. 2025) already yields modality-agnostic

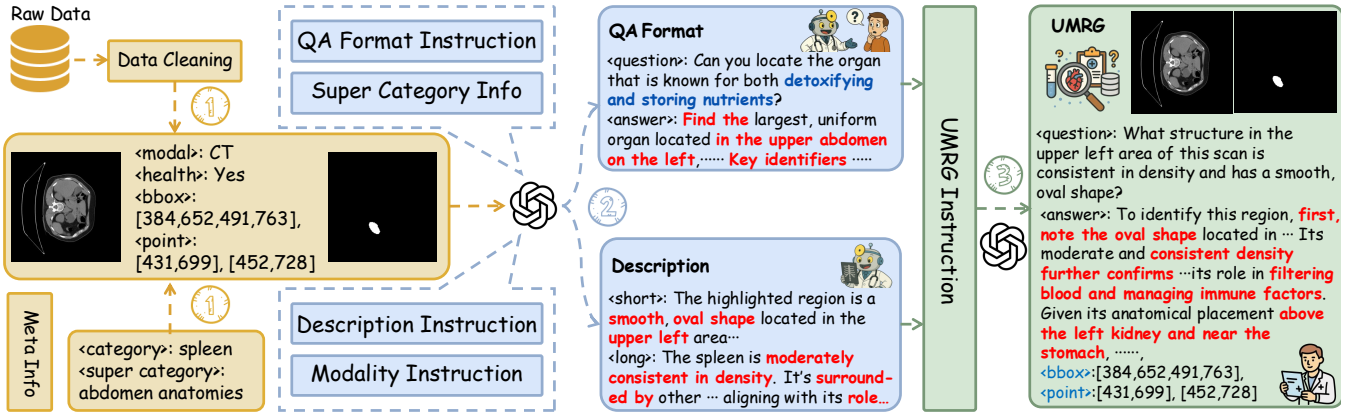

Figure 2: Overview of the **U-MRG-14K** construction pipeline: (1) Data cleaning and metadata organization manually, (2) Description and QA format generation via GPT-4o, (3) QA pair generation with GPT-4o and human verification.

masks out of the box. As shown in Fig. 3, **MedReasoner** decouples language reasoning from pixel-level grounding, thereby (1) mitigating phrase over-fitting to enable authentic reasoning, and (2) treating MedSAM-family models as plug-and-play components controllable by language.

Clinical Reasoning Module. We employ Lingshu (Xu et al. 2025) as our **Clinical Reasoning Module (CRM)** $\mathbf{F}_{reason}$. Given $(\mathcal{I}, \mathcal{Q})$, CRM outputs a structured tuple `<think>···<think><answer>`$\mathcal{B}, \mathcal{P}_1, \mathcal{P}_2$`<answer>`. A bounding box $\mathcal{B}$ is often inadequate in medical images: boxes may enclose multiple organs or lesions, and their corners lack semantics for SAM-style prompters. We therefore add two key points $\mathcal{P}_1, \mathcal{P}_2$ on visually distinctive regions. These enrich spatial cues at low annotation cost. To learn reliable cues without compromising linguistic competence, we train $\mathbf{F}_{\text{reason}}$ with **Group Relative Policy Optimization (GRPO)** (Shao et al. 2024), using: (1) **format rewards** enforcing the output schema, and (2) **accuracy rewards** measuring spatial correctness.

Anatomical Segmentation Module. We instantiate the Anatomical Segmentation Module (ASM) with a frozen MedSAM2 (Ma et al. 2025), denoted as $\mathbf{F}_{\text{seg}}$. The tuple $(\mathcal{B}, \mathcal{P}_1, \mathcal{P}_2)$ produced by the CRM is fed to $\mathbf{F}_{\text{seg}}$, which transforms these coarse prompts into a high-resolution mask $\mathcal{M}$ without any task-specific fine-tuning. Freezing $\mathbf{F}_{\text{seg}}$ preserves MedSAM2's strong zero-shot delineation ability, while allowing $\mathbf{F}_{\text{reason}}$ to concentrate on language understanding and spatial reasoning.

4.3 Reward Functions

Reward functions in RL guide a model toward the behaviors we desire. For UMRG, we introduce three rewards that first prompt the model to reason about the implicit target and then to predict the bounding box and key points.

Reasoning Formats Reward. This reward evaluates the structural validity of the model's output, focusing on the formatting of the reasoning and answer components. It assigns $\mathbb{R}_{\text{think}}$ to assess whether the model produces a well-structured `<think>` block, and $\mathbb{R}_{\text{answer}}$ to verify whether the `<answer>` block is a valid JSON object containing the required fields: `bbox`, `points_1`, and `points_2`. These rewards do not evaluate the correctness or quality of the reasoning content itself, but rather the presence and structural completeness of the expected formats. Both $\mathbb{R}_{\text{think}}$ and $\mathbb{R}_{\text{answer}}$ are assigned discrete values in the range $[0, 1]$.

Grounding Box Reward. This reward evaluates the quality of the predicted bounding box $\mathcal{B}_p = [x_1^p, y_1^p, x_2^p, y_2^p]$ against the ground-truth box $\mathcal{B}_g = [x_1^g, y_1^g, x_2^g, y_2^g]$, where all coordinates are normalized to $[0, 1]$. First, the **IoU reward** measures the spatial overlap between two boxes:

$$\mathbb{R}_{\text{iou}} = \frac{\text{Area}(\mathcal{B}_p \cap \mathcal{B}_g)}{\text{Area}(\mathcal{B}_p \cup \mathcal{B}_g)}. \tag{2}$$

Second, the **Alignment reward** computes the average L1 distance between corresponding corner coordinates, normalized by the diagonal of $\mathcal{B}_g$:

$$\mathbb{R}_{\text{align}} = \frac{1}{4} \sum_{i=1}^{4} \left| \mathcal{B}_p^{(i)} - \mathcal{B}_g^{(i)} \right|. \tag{3}$$

Third, the **Scale reward** captures shape consistency in terms of area and aspect ratio. Specifically, we compute the logarithmic difference in box area and aspect ratio, and define:

$$\mathbb{R}_{\text{scale}} = \sqrt{(\Delta \log A)^2 + (\Delta \log R)^2}, \tag{4}$$

where A denotes box area and R the aspect ratio. Smaller values indicate better structural alignment.

Grounding Points Reward. This reward evaluates the quality of the predicted key point pair $\mathcal{P}_p = \{\mathbf{p}_1^p, \mathbf{p}_2^p\}$ against the ground-truth pair $\mathcal{P}_g = \{\mathbf{p}_1^g, \mathbf{p}_2^g\}$, where each point $\mathbf{p} = (x, y)$ is normalized to $[0, 1]$. First, the **pDice reward** quantifies the spatial overlap between circles, where each circle is defined by a point pair serving as its diameter:

$$\mathbb{R}_{\text{pdice}} = \frac{2 \cdot \text{Area}(O_p \cap O_g)}{\text{Area}(O_p) + \text{Area}(O_g)}, \tag{5}$$

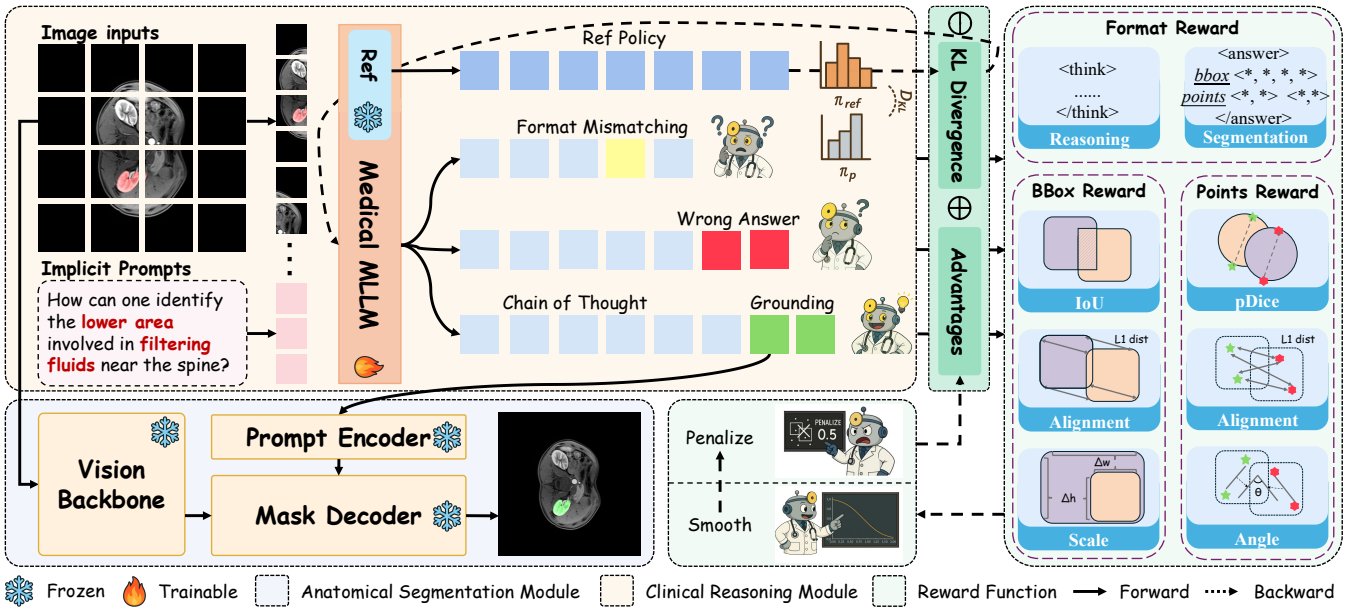

Figure 3: Overview of the **MedReasoner** framework. MedReasoner transforms implicit clinical prompts into pixel-level grounding via a two-stage process. The **CRM** first generates intermediate reasoning and grounding outputs (*CoT*, *bounding box*, and *key points*). Then, the **ASM** converts the grounded outputs into final segmentation masks.

where O_p and O_g are the circles constructed from $\mathcal{P}_p$ and $\mathcal{P}_g$, respectively. Second, the **Alignment reward** computes the mean absolute error between corresponding points:

$$\mathbb{R}_{\text{align}} = \frac{1}{2}\sum_{i=1}^{2}\left(|x_i^p - x_i^g| + |y_i^p - y_i^g|\right). \tag{6}$$

Third, the **Angle reward** measures the cosine similarity between the predicted and ground-truth direction vectors, capturing angular consistency:

$$\mathbb{R}_{\text{angle}} = |\cos(\theta)| = \left|\frac{\langle \mathbf{v}_p, \mathbf{v}_g\rangle}{\|\mathbf{v}_p\|_2 \cdot \|\mathbf{v}_g\|_2}\right|, \tag{7}$$

where $\mathbf{v}_p = \mathbf{p}_2^p - \mathbf{p}_1^p$ and $\mathbf{v}_g = \mathbf{p}_2^g - \mathbf{p}_1^g$.

Smoothing and Penalization. To enhance training stability and differentiate prediction quality, we apply smoothing functions to all reward components. For the $\mathbb{R}_{\text{iou}}$, $\mathbb{R}_{\text{pdice}}$ and $\mathbb{R}_{\text{angle}}$ rewards, we use logarithmic smoothing:

$$\mathcal{S}_{\log}(r;k) = \frac{\log(kr+1)}{\log(k+1)}, \tag{8}$$

where $r \in [0,1]$ is the raw reward and k is a scaling factor (default $k = 3$). For the $\mathbb{R}_{\text{align}}$ and $\mathbb{R}_{\text{scale}}$ rewards, we use exponential smoothing:

$$\mathcal{S}_{\exp}(d;k,c) = \frac{1}{1+e^{k(d-c)}}, \tag{9}$$

where $d \in [0,2]$ is the normalized distance, and c is the target center (default $c = 1$).

After smoothing, we apply a penalization function $\mathcal{N}(\cdot)$ to softly down-weight unreliable predictions. For each reward, two validity scores are computed to reflect the spatial plausibility of the output. The final reward is adjusted as:

$$\mathcal{N}(r;v_1,v_2) = \lambda r + (1-\lambda)r \cdot \frac{v_1+v_2}{2}, \tag{10}$$

where r is the smoothed reward, v_1 and v_2 are the two validity scores, and $\lambda = 0.7$ by default.

5 Experiments

5.1 Experimental Settings

Models. We conduct a comprehensive comparison across a wide range of models. For general MLLMs, we utilized GPT-4o (OpenAI 2024), Gemini-2.5-flash (Google 2025), Qwen2.5VL-7B/72B (Bai et al. 2025) and InternVL3-8B/78B (Zhu et al. 2025). For medical-specific MLLMs, we selected MedR1-2B (Lai et al. 2025), MiniInternVL-4B (Gao et al. 2024), MedGamma-4B (Sellergren et al. 2025), HuatuoGPT-7B-Qwen2.5VL (Chen et al. 2024a), Lingshu-7B (Xu et al. 2025), and Chiron-o1-8B (Sun et al. 2025). For segmentation models, we chose MedSAM (Ma et al. 2024), SAM-Med2D (Cheng et al. 2023) and MedSAM2 (Ma et al. 2025). For grounding-specific models, we included SAM4MLLM (Chen et al. 2024b), VLMR1-REC-3B (Shen et al. 2025) and SegZero-7B (Liu et al. 2025).

Datasets. We train MedReasoner on U-MRG-14K, using the data preparation strategy mentioned in Section 3.1. We randomly hold out 2.5K samples as a test set, and use the remaining data for training. All quantitative results reported in this paper are obtained on the test set.

Implementation Details. We adopt Lingshu-7B as our default CRM and default ASM to MedSAM2.

Method	IoU↑	pDice↑	Dice↑	Super-Categories (IoU↑)									
				Abd.	Brain	Eye	Heart	Hist.	Lung	Ves.	Neo.	N-Neo.	Inf.
General MLLMs													
GPT-4o	2.65	1.12	4.72	0.92	0.91	3.29	0.36	2.8	11.70	1.83	1.01	4.16	6.37
Gemini-2.5-flash	7.86	3.24	14.29	3.99	5.69	6.39	7.77	6.63	16.37	9.08	7.15	13.91	11.4
Qwen2.5VL-7B	12.61	7.14	22.73	6.84	23.97	29.35	8.37	9.22	20.79	20.46	8.00	24.97	19.4
InternVL3-8B	5.70	2.46	9.23	3.72	6.54	2.02	3.67	5.56	14.44	7.88	3.78	8.71	9.00
Qwen2.5-VL-72B	18.32	12.39	29.71	13.60	20.06	38.3	15.51	8.74	35.25	20.64	20.69	30.19	16.92
InternVL3-78B	4.02	1.55	7.23	2.04	2.95	2.33	2.12	6.12	12.21	4.19	1.33	8.19	5.62
Medical-Specific MLLMs													
MedR1-2B	8.18	3.60	14.73	3.53	12.55	1.10	3.53	8.14	25.58	8.81	4.39	13.57	17.35
MiniInternVL-4B	2.88	0.85	4.76	1.88	2.67	0.68	1.60	3.45	7.99	3.59	1.56	3.76	6.59
MedGamma-4B	5.39	1.90	8.90	4.23	6.92	1.28	3.41	4.78	17.22	6.92	3.17	3.90	10.04
HuatuoGPT-7B	10.13	5.23	19.76	5.88	18.16	3.88	6.63	9.56	22.94	15.58	8.25	16.12	15.87
Lingshu-7B	8.19	3.73	16.48	4.03	15.72	6.97	6.27	8.06	19.77	8.63	6.34	13.31	11.99
Chiron-o1-8B	6.40	2.46	10.05	3.82	6.90	4.29	4.20	5.99	12.86	9.50	5.53	11.31	10.86
Grounding-Specific MLLMs													
VLMR1-REC-3B	13.96	-	22.19	8.64	21.81	25.09	8.19	10.69	29.77	21.35	8.76	26.59	21.41
SegZero-7B	16.14	5.23	26.05	11.66	23.37	40.23	13.12	9.35	22.18	20.68	12.58	29.46	21.93
SAM4MLLM-8B	7.94	-	16.49	6.30	14.69	5.09	5.81	7.46	12.61	11.99	6.24	11.96	12.40
MedReasoner-7B	**32.42**	**26.55**	**37.78**	**30.27**	**32.81**	**51.50**	**34.72**	**11.66**	**50.75**	**29.91**	**33.58**	**37.19**	**30.48**

Table 2: Results on the **U-MRG-14K** test set under the **MedReasoner** paradigm. Each candidate uses one medical MLLM as the **CRM** to output a bounding box and two key points; the **ASM** is fixed to *MedSAM2*. **Bold** numbers denote the best score in each column, and underlined numbers denote the second best.

Evaluation Metrics. We compute three evaluation metrics: **IoU**, **pDice**, and **Dice** to assess model performance. **IoU** measures the bounding box localization accuracy predicted by MLLMs. **pDice** quantifies keypoint pair semantic alignment by evaluating the overlap of circles formed by predicted endpoints (formally defined in Section 4.3). **Dice** assesses segmentation quality based on masks generated by downstream models conditioned on MLLM outputs.

5.2 Medical Reasoning Grounding Results

For fair comparison, we evaluated models under the MedReasoner paradigm, using a single MLLM as CRM to return bounding box and key point, with MedSAM2 fixed as ASM. All MLLMs are driven by the same user prompt. As shown in Table 2, MedReasoner-7B achieved superior overall performance, significantly leading the second-best Qwen2.5VL-72B by 14.10 in IoU, 14.16 in pDice, and 8.07 in Dice. This highlights its precise spatial prompting capability. While General MLLMs, such as GPT-4o (IoU 2.65), and Medical-Specific MLLMs, like HuatuoGPT-7B (IoU 10.13), demonstrated cross-modal understanding or domain benefits, they consistently lacked the fine-grained precision required for UMRG. In contrast, MedReasoner-7B established a substantial lead among Grounding-Specific MLLMs, surpassing SegZero-7B by over 16 IoU points, validating our RL-driven grounding strategy for accurate regional prompt translation. This superiority extended across super-categories, with MedReasoner-7B leading in most (e.g., Lung's IoU was 50.75), though all models, including ours, faced challenges in complex categories like Histology.

Method	IoU↑	pDice↑	Dice↑	# Ref.↓
Lingshu	8.19	3.73	16.51	2
Lingshu w/ SFT	9.15	2.88	15.22	2
Lingshu w/ RL(Base)	15.85	8.29	28.79	0
Lingshu w/ RL(Hard)	31.69	24.36	33.51	0
Lingshu w/ RL(Soft)	**32.42**	**26.55**	**37.78**	0

Table 3: Comparison of the **SFT** baseline with three RL variants: **Base**, **Hard**, and **Soft** on U-MRG-14K. **# Ref.** denotes refusals to ground answers with the reasoning prompt.

5.3 Ablation Study

Ablations verify our proposed design's effectiveness, all trained on U-MRG-14K and using the same user prompt.

Effect of Reward Types. This ablation study investigated how reward design influences RL training for CRM. Our SFT baseline had a low IoU of 9.15 and 2 query refusals (as Table 3 shows). RL fine-tuning drastically improved performance, eliminating all refusals. We evaluated three reward variants: **Base** (a hard-threshold scheme (Liu et al. 2025)), **Hard** (our full reward), and **Soft** (IoU and pDice only). While Base removed refusals, its IoU of 15.85 was considerably lower. Our Hard reward significantly outperformed Base, increasing IoU by 15.84 points. The Soft reward variant achieved the best overall IoU of 32.42, surpassing Hard by 0.73 points, suggesting that less strict alignment fosters better exploration and more accurate grounding solutions.

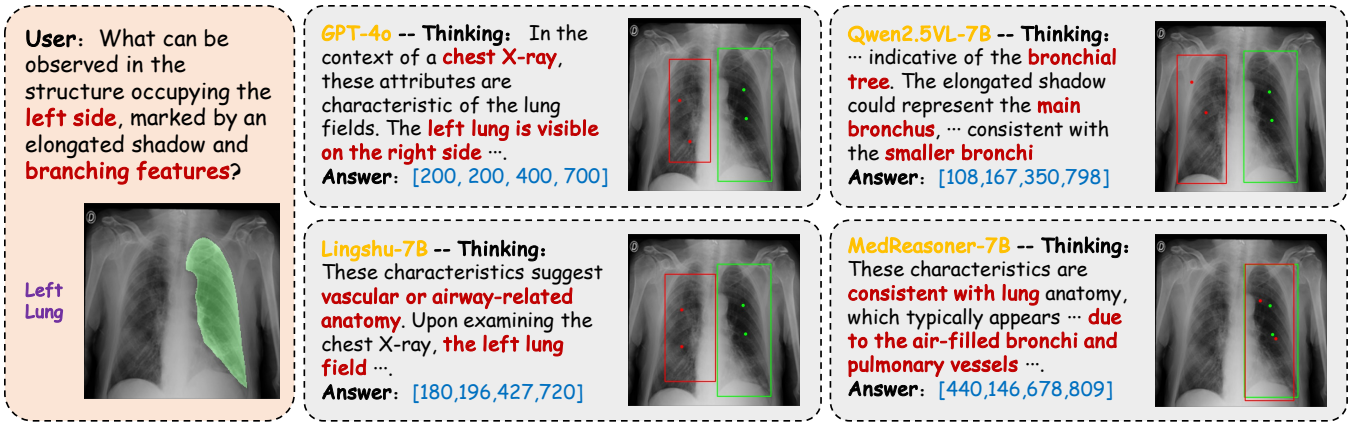

Figure 4: Open-source (**Qwen2.5VL-7B**), closed-source (**GPT-4o**), and medically post-trained (**Lingshu-7B**) models reason and specify referents within CoT processes, a characteristic tied to their respective training methodologies. Our **MedReasoner-7B** integrates grounding information during training, thereby aligning pixel-level grounding with semantic-level understanding.

Method	Dice↑		
	w/ Points	w/ BBox	BBox & Points
MedSAM	5.67	28.39	19.00
SAM-Med2D	33.23	35.03	36.48
MedSAM2	**34.86**	**37.15**	**37.78**

Table 4: Dice scores for three segmentation backbones under three prompt types: **key points only**, **bounding box only**, and the **combined bounding box & points**.

Method	Reason	IoU↑	pDice↑	# Ref.↓
Qwen2.5VL-7B	✗	14.57	8.14	13
Qwen2.5VL-7B	✓	12.61	7.14	0
Lingshu-7B	✗	9.35	2.40	4
Lingshu-7B	✓	8.19	3.73	2
MedReasoner-7B	✗	30.29	25.82	12
MedReasoner-7B	✓	**32.42**	**26.55**	0

Table 5: Impact of adding an explicit reasoning step vs. a direct prompt for three CRMs. **Reason** indicates whether the model is prompted to reason first (✓) or respond directly (✗).

Effect of Segmentation Backbones. This ablation assesses the ASM. Table 4 reports results for three medical SAM variants: MedSAM, SAM-Med2D, and MedSAM2. To investigate prompt influence, we evaluated three input formats per backbone: points only, bounding box only, and the combined bounding box and points. The combination consistently yielded the best Dice (37.78), with MedSAM2 achieving the highest performance across all configurations.

Effect of Reasoning Strategies. This ablation tests whether prompting the model to reason before grounding helps when answering implicit queries. We designed two prompts: **Direct** asks the CRM to output the spatial prompt immediately, whereas **Reasoning** instructs it to first generate a brief CoT. As Table 5 shows, the Reasoning prompt significantly reduces refusal rates compared to the Direct prompt. While base Qwen2.5VL and Lingshu exhibit a slight performance drop due to their limited inherent reasoning capabilities, this is expected. However, after CRM training within the MedReasoner framework, the Reasoning strategy clearly outperforms the Direct one. This confirms that an explicit reasoning phase is valuable for implicit-query grounding.

5.4 Qualitative Results

Figure 4 compares four MLLMs' predictions on a chest X-ray query requiring implicit reasoning. **GPT-4o** produces a coherent CoT and an accurate image-level answer, but its spatial output is wrong: the bounding box is misplaced and coordinates are rounded, indicating limited fine-grained grounding. **Qwen2.5VL-7B** fails at the reasoning stage, resulting in an incorrect diagnosis and an irrelevant box. **Lingshu-7B** correctly identifies the *left lung* but misplaces the box, demonstrating that it alone doesn't guarantee accurate localization. Only **MedReasoner-7B** precisely identifies and pinpoints the target; its box tightly encloses the bronchial tree of the left lung, with key points aligning to it. These observations highlight the necessity of explicit RL-based grounding. It preserves the reasoning quality of large models while enforcing the spatial precision for UMRG.

6 Conclusion

We present the **UMRG** task, which challenges models to transform implicit clinical queries into precise pixel-level grounding. To support this, we introduce **U-MRG-14K**, a large-scale dataset featuring rich annotations and reasoning traces. To solve UMRG, we propose **MedReasoner**, a modular framework that decouples reasoning from segmentation and leverages RL to align linguistic reasoning with spatial precision. Extensive experiments demonstrate that MedReasoner consistently outperforms existing models in accuracy. We believe this framework offers a promising step toward trustworthy and generalizable medical grounding systems.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (Grant 62476029, 62225601, U23B2052), funded by the Fundamental Research Funds for the Beijing University of Posts and Telecommunications under Grant 2025TSQY08, and sponsored by Beijing Nova Program.

References

Bai, S.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; Song, S.; Dang, K.; Wang, P.; Wang, S.; Tang, J.; Zhong, H.; Zhu, Y.; Yang, M.; Li, Z.; Wan, J.; Wang, P.; Ding, W.; Fu, Z.; Xu, Y.; Ye, J.; Zhang, X.; Xie, T.; Cheng, Z.; Zhang, H.; Yang, Z.; Xu, H.; and Lin, J. 2025. Qwen2.5-VL Technical Report. *arXiv preprint arXiv:2502.13923*.

Chen, J.; Gui, C.; Ouyang, R.; Gao, A.; Chen, S.; Chen, G. H.; Wang, X.; Zhang, R.; Cai, Z.; Ji, K.; et al. 2024a. Huatuogpt-vision, towards injecting medical visual knowledge into multimodal llms at scale. *arXiv preprint arXiv:2406.19280*.

Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A. L.; and Zhou, Y. 2021. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

Chen, Y.-C.; Li, W.-H.; Sun, C.; Wang, Y.-C. F.; and Chen, C.-S. 2024b. Sam4mllm: Enhance multi-modal large language model for referring expression segmentation. In *European Conference on Computer Vision*, 323–340. Springer.

Cheng, J.; Fu, B.; Ye, J.; Wang, G.; Li, T.; Wang, H.; Li, R.; Yao, H.; Cheng, J.; Li, J.; et al. 2025. Interactive medical image segmentation: A benchmark dataset and baseline. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 20841–20851.

Cheng, J.; Ye, J.; Deng, Z.; Chen, J.; Li, T.; Wang, H.; Su, Y.; Huang, Z.; Chen, J.; Jiang, L.; Sun, H.; He, J.; Zhang, S.; Zhu, M.; and Qiao, Y. 2023. SAM-Med2D. arXiv:2308.16184.

Da, L.; Wang, R.; Xu, X.; Bhatia, P.; Kass-Hout, T.; Wei, H.; and Xiao, C. 2024. Segment as You Wish–Free-Form Language-Based Segmentation for Medical Images. *arXiv preprint arXiv:2410.12831*.

Gao, Z.; Chen, Z.; Cui, E.; Ren, Y.; Wang, W.; Zhu, J.; Tian, H.; Ye, S.; He, J.; Zhu, X.; et al. 2024. Mini-internvl: A flexible-transfer pocket multimodal model with 5% parameters and 90% performance. *arXiv preprint arXiv:2410.16261*.

Google. 2025. Gemini-2.5-Flash. Online; GA release.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

He, X.; Zhang, Y.; Mou, L.; Xing, E.; and Xie, P. 2020. Pathvqa: 30000+ questions for medical visual question answering. *arXiv preprint arXiv:2003.10286*.

Huang, X.; Shen, L.; Liu, J.; Shang, F.; Li, H.; Huang, H.; and Yang, Y. 2025a. Towards a multimodal large language model with pixel-level insight for biomedicine. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 3779–3787.

Huang, Y.; Peng, Z.; Zhao, Y.; Yang, P.; Yang, X.; and Shen, W. 2025b. MedSeg-R: Reasoning Segmentation in Medical Images with Multimodal Large Language Models. *arXiv preprint arXiv:2506.10465*.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4015–4026.

Koleilat, T.; Asgariandehkordi, H.; Rivaz, H.; and Xiao, Y. 2024. Medclip-sam: Bridging text and image towards universal medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, 643–653. Springer.

Lai, X.; Tian, Z.; Chen, Y.; Li, Y.; Yuan, Y.; Liu, S.; and Jia, J. 2024. Lisa: Reasoning segmentation via large language model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9579–9589.

Lai, Y.; Zhong, J.; Li, M.; Zhao, S.; and Yang, X. 2025. Med-r1: Reinforcement learning for generalizable medical reasoning in vision-language models. *arXiv preprint arXiv:2503.13939*.

Lau, J. J.; Gayen, S.; Ben Abacha, A.; and Demner-Fushman, D. 2018. A dataset of clinically generated visual questions and answers about radiology images. *Scientific data*, 5(1): 1–10.

Li, C.; Wong, C.; Zhang, S.; Usuyama, N.; Liu, H.; Yang, J.; Naumann, T.; Poon, H.; and Gao, J. 2023. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems*, 36: 28541–28564.

Li, J.; Liu, C.; Bai, W.; Arcucci, R.; Bercea, C. I.; and Schnabel, J. A. 2025. Enhancing Abnormality Grounding for Vision Language Models with Knowledge Descriptions. *arXiv preprint arXiv:2503.03278*.

Li, W.; Qu, C.; Chen, X.; Bassi, P. R.; Shi, Y.; Lai, Y.; Yu, Q.; Xue, H.; Chen, Y.; Lin, X.; et al. 2024. Abdomenatlas: A large-scale, detailed-annotated, & multi-center dataset for efficient transfer learning and open algorithmic benchmarking. *Medical Image Analysis*, 97: 103285.

Lin, T.; Chen, Z.; Yan, Z.; Yu, W.; and Zheng, F. 2024. Stable diffusion segmentation for biomedical images with single-step reverse process. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 656–666. Springer.

Lin, T.; Zhang, W.; Li, S.; Yuan, Y.; Yu, B.; Li, H.; He, W.; Jiang, H.; Li, M.; Song, X.; et al. 2025. Healthgpt: A medical large vision-language model for unifying comprehension and generation via heterogeneous knowledge adaptation. *arXiv preprint arXiv:2502.09838*.

Liu, B.; Zhan, L.-M.; Xu, L.; Ma, L.; Yang, Y.; and Wu, X.-M. 2021. Slake: A semantically-labeled knowledge-enhanced dataset for medical visual question answering.

In *2021 IEEE 18th international symposium on biomedical imaging (ISBI)*, 1650–1654. IEEE.

Liu, J.; Zhang, Y.; Chen, J.-N.; Xiao, J.; Lu, Y.; A Landman, B.; Yuan, Y.; Yuille, A.; Tang, Y.; and Zhou, Z. 2023. Clip-driven universal model for organ segmentation and tumor detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 21152–21164.

Liu, Y.; Peng, B.; Zhong, Z.; Yue, Z.; Lu, F.; Yu, B.; and Jia, J. 2025. Seg-zero: Reasoning-chain guided segmentation via cognitive reinforcement. *arXiv preprint arXiv:2503.06520*.

Luo, L.; Tang, B.; Chen, X.; Han, R.; and Chen, T. 2024. Vividmed: Vision language model with versatile visual grounding for medicine. *arXiv preprint arXiv:2410.12694*.

Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; and Wang, B. 2024. Segment anything in medical images. *Nature Communications*, 15(1): 654.

Ma, J.; Yang, Z.; Kim, S.; Chen, B.; Baharoon, M.; Fallahpour, A.; Asakereh, R.; Lyu, H.; and Wang, B. 2025. Medsam2: Segment anything in 3d medical images and videos. *arXiv preprint arXiv:2504.03600*.

OpenAI. 2024. GPT-4o (GPT-4 Omni). Online.

OpenAI. 2024. OpenAI o1. https://openai.com/o1/.

Pi, R.; Yao, L.; Gao, J.; Zhang, J.; and Zhang, T. 2024. Perceptiongpt: Effectively fusing visual perception into llm. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 27124–27133.

Ren, Z.; Huang, Z.; Wei, Y.; Zhao, Y.; Fu, D.; Feng, J.; and Jin, X. 2024. Pixellm: Pixel reasoning with large multimodal model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 26374–26383.

Sellergren, A.; Kazemzadeh, S.; Jaroensri, T.; Kiraly, A.; Traverse, M.; Kohlberger, T.; Xu, S.; Jamil, F.; Hughes, C.; Lau, C.; et al. 2025. MedGemma Technical Report. *arXiv preprint arXiv:2507.05201*.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y.; Wu, Y.; et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Shen, H.; Liu, P.; Li, J.; Fang, C.; Ma, Y.; Liao, J.; Shen, Q.; Zhang, Z.; Zhao, K.; Zhang, Q.; et al. 2025. Vlm-r1: A stable and generalizable r1-style large vision-language model. *arXiv preprint arXiv:2504.07615*.

Sun, H.; Jiang, Y.; Lou, W.; Zhang, Y.; Li, W.; Wang, L.; Liu, M.; Liu, L.; and Wang, X. 2025. Enhancing Step-by-Step and Verifiable Medical Reasoning in MLLMs. *arXiv preprint arXiv:2506.16962*.

Tong, Q.; Lu, Z.; Liu, J.; Zheng, Y.; and Lu, Z. 2025. MediSee: Reasoning-based Pixel-level Perception in Medical Images. *arXiv preprint arXiv:2504.11008*.

Trinh, Q.-H.; Nguyen, M.-V.; Peng, J.; Bagci, U.; and Jha, D. 2025. PRS-Med: Position Reasoning Segmentation with Vision-Language Model in Medical Imaging. *arXiv preprint arXiv:2505.11872*.

Uesato, J.; Kushman, N.; Kumar, R.; Song, F.; Siegel, N.; Wang, L.; Creswell, A.; Irving, G.; and Higgins, I. 2022. Solving math word problems with process-and outcome-based feedback. *arXiv preprint arXiv:2211.14275*.

Wang, W.; Lv, Q.; Yu, W.; Hong, W.; Qi, J.; Wang, Y.; Ji, J.; Yang, Z.; Zhao, L.; XiXuan, S.; et al. 2024. Cogvlm: Visual expert for pretrained language models. *Advances in Neural Information Processing Systems*, 37: 121475–121499.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Xu, W.; Chan, H. P.; Li, L.; Aljunied, M.; Yuan, R.; Wang, J.; Xiao, C.; Chen, G.; Liu, C.; Li, Z.; et al. 2025. Lingshu: A Generalist Foundation Model for Unified Multimodal Medical Understanding and Reasoning. *arXiv preprint arXiv:2506.07044*.

Yan, Z.; Yin, Z.; Lin, T.; Zeng, X.; Liang, K.; and Ma, Z. 2025. PGP-SAM: Prototype-Guided Prompt Learning for Efficient Few-Shot Medical Image Segmentation. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, 1–5. IEEE.

Ye, J.; Cheng, J.; Chen, J.; Deng, Z.; Li, T.; Wang, H.; Su, Y.; Huang, Z.; Chen, J.; Jiang, L.; et al. 2023. Sa-med2d-20m dataset: Segment anything in 2d medical imaging with 20 million masks. *arXiv preprint arXiv:2311.11969*.

Yuan, Y.; Tang, H.; Wang, C.; Zheng, Y.; and Hao, J. 2024. ED2: Environment Dynamics Decomposition World Models for Continuous Control. *Visual Intelligence*, 3: Article 23.

Yue, W.; Zhang, J.; Hu, K.; Xia, Y.; Luo, J.; and Wang, Z. 2024. SurgicalSAM: Efficient class promptable surgical instrument segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 6890–6898.

Zhao, T.; Gu, Y.; Yang, J.; Usuyama, N.; Lee, H. H.; Naumann, T.; Gao, J.; Crabtree, A.; Abel, J.; Moung-Wen, C.; et al. 2024. BiomedParse: a biomedical foundation model for image parsing of everything everywhere all at once. *arXiv preprint arXiv:2405.12971*.

Zhu, J.; Wang, W.; Chen, Z.; Liu, Z.; Ye, S.; Gu, L.; Tian, H.; Duan, Y.; Su, W.; Shao, J.; et al. 2025. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*.