E-MRL: Cross-view Aligned Evidence-driven Multimodal Reinforcement Learning for Reliable 3D Tumor Analysis

Sijing Li[1,2], Zhongwei Qiu[2,3*], Zhuoya Wang[4], Boxiang Yun[5], Zhenyu Yi[6], Jianwei Xu[2], Wenqiao Zhang[1*], Yingda Xia[2], and Ling Zhang[2]

[1] Zhejiang University
[2] DAMO Academy, Alibaba Group
[3] Hupan Lab
[4] Huazhong University of Science and Technology
[5] East China Normal University
[6] Shanghai Jiao Tong University

Abstract. While Vision-Language Models (VLMs) show great promise in volumetric medical report generation, they frequently suffer from visual hallucinations and a lack of grounding in 3D CT data. Current Supervised Fine-Tuning (SFT) and Reinforcement Learning (RL) strategies typically optimize text fidelity alone, essentially rewarding correct diagnoses derived from language priors rather than genuine visual perception. To address this, we propose cross-view aligned **E**vidence-driven **M**ultimodal **R**einforcement **L**earning (Evidence-MRL, noted as **E-MRL**), a reliable RL reasoning framework that formulates the generation process as a Markov Decision Process of "diagnosis-localization-verification". Unlike standard approaches, our model is explicitly trained to identify a "key evidence slice" alongside the global diagnostic report, grounding its findings in verifiable visual evidence. Crucially, we introduce a novel cross-view consistency reward, which validates the semantic alignment between the golden-standard report and a local visual re-query of the selected key slice, providing additional rewards for correctly-localized reasoning. Experiments on large-scale 3D CT tumor datasets demonstrate that E-MRL significantly reduces hallucinations and improves diagnostic accuracy compared to SFT and RL baselines, offering a clinically interpretable solution for visually-grounded and tumor analysis.

Keywords: Medical VLM · RL · Tumor Analysis · Interpretability

1 Introduction

The rapid advancement of Vision-Language Models (VLMs) has opened new frontiers in automated medical reasoning tasks [28,25,22,27], transitioning from

* Correspondence to: Wenqiao Zhang (wenqiaozhang@zju.edu.cn) and Zhongwei Qiu (qiuzhongwei.qzw@alibaba-inc.com).

2D image analysis to complex 3D volumetric interpretation (e.g., CT and MRI scans) [1,3,20]. However, the application of VLMs to 3D medical imaging is severely compromised by "visual hallucinations" [13], a phenomenon in which models generate clinically plausible but factually incorrect reports. Due to the excessive length of 3D sequences and the sparsity of pathological features, models often struggle to maintain attention on critical visual cues. Consequently, they tend to rely on language priors rather than genuine visual perception, fabricating lesion details that are inconsistent with or entirely absent from the input volume. Such ungrounded reasoning poses significant risks in clinical diagnostics.

Existing alignment strategies, such as Supervised Fine-Tuning (SFT) and text-based Reinforcement Learning from Human Feedback (RLHF) [33,18], primarily optimize the linguistic coherence and factual correctness of reports at text level. However, these approaches share a fundamental limitation: they provide no process-level visual supervision over where the model looks during the reasoning process. If a model draws correct conclusions based on irrelevant slices (e.g., misidentifying a vessel cross-section as a tumor), standard loss functions fail to penalize this visual grounding error. The core challenge, therefore, is not merely improving textual quality, but establishing a verifiable correspondence between generated report content and its visual evidence in the 3D volume.

To address this challenge, inspired by the radiologist's diagnostic workflow of "global scanning followed by local details verification", we propose E-MRL, the first cross-view aligned **E**vidence-driven **M**ultimodal **R**einforcement **L**earning framework for reliable tumor analysis in 3D CT scans. Unlike conventional methods that rely solely on textual rewards, E-MRL frames report generation as a "diagnosis-localization-verification" Markov Decision Process and introduces multi-granularity textual and visual evidence-driven rewards based on a novel cross-view (global and local) consistency mechanism. While pioneers like RadGPT [2] rely on complex multi-stage, segmentation-assisted pipelines to surpass standard medical VLMs which often lack sufficient visual evidence, E-MRL represents the first end-to-end framework to surpass such paradigms, proving that its cross-view verification RL mechanism is a more effective strategy for ensuring diagnostic reliability than traditional segmentation-based assistance. Specifically, E-MRL not only generates a global diagnostic report, but also explicitly localizes a "key evidence slice"—the most informative frame within the 3D CT volume to support its diagnosis. The RL environment extracts this slice for an independent local re-query. If the model describes a tumor in the global report and confirms it in slice verification, our RL mechanism grants an extra reward beyond base reward for lesion attributes. This mechanism compels the model to seek visual evidence for local verification in the global reasoning results before generating descriptions, thereby significantly mitigating hallucinations.

Our main contributions are as follows:

• **RL Framework:** We propose a novel "Diagnosis-Localization-Verification" RL framework, marking the first end-to-end VLM to surpass complex multi-stage segmentation-assisted paradigms and demonstrating that fine-grained evidence-driven reasoning can substantially elevate clinical diagnostic precision.

• **Multi-Granularity Multimodal Reward:** We design a comprehensive reward mechanism that integrates textual diagnostic correctness with visual evidence grounding. By providing dense, multimodal feedback, this approach effectively mitigates hallucinations at both the reasoning and perception levels.

• **Cross-view Consistency Alignment:** We develop a global-to-local alignment mechanism that synchronizes global diagnostic findings with local slice-level evidence. This ensures rigorous consistency between the generated text and the underlying visual cues.

• **Substantial Performances and Insights:** Extensive experiments demonstrate that E-MRL achieves synergistic improvements in both diagnostic accuracy and visual evidence localization. Furthermore, our framework and evidence-driven dataset for E-MRL will be open-sourced to foster the development of reliable and trustworthy medical VLMs.

2 Related Work

Medical Vision-Language Models. Recent Med-LVLMs [25,17,30], including LLaVA-Med [16] and HealthGPT [21], employ supervised fine-tuning (SFT) for aligning vision and text modalities [9], while M3D [1], CT-Chat [8], and Merlin [3] extend SFT to 3D images. However, SFT alone can be limited in complex reasoning. Recent works, such as MedVLThinker [10] and MedVLM-R1 [23], incorporate reinforcement learning (RL) with Chain-of-Thought [18,5] or multi-agent [29,6] strategies to address these challenges. Overall, SFT and RL have become common and effective post-training approaches in Med-LVLMs.

Reinforcement Learning Strategies and Interpretability. Most medical RL methods, such as MedVLM-R1 [23], Med3D-R1 [14], and VALOR [**?**], lack explicit logical grounding, while MedReason-R1 [19] relies on pre-processed suspicious regions as auxiliary inputs. Existing text-based rewards [15,24] provide sparse supervision and lead to limited visual grounding and insufficient interpretability. Although prior visual rewards mainly target 2D image localization [**?**,4], clinical 3D visual rewards remain absent. To address this gap, we enable interpretable tumor reasoning via 3D evidence-guided multimodal RL.

3 Method

3.1 Overview: The Diagnosis-Localization-Verification Framework

To address the "black-box" reasoning and visual hallucination issues inherent in medical VLMs, we propose **Evidence-MRL**, an interpretable reinforcement learning framework with a cross-view consistency verification mechanism. Our method mimics the radiologists' workflow, formalizing the reasoning process into a closed-loop "Diagnosis-Location-Verification" paradigm: **1) Global Diagnosis:** Analyze the entire 3D CT volume and generate a global report with lesion descriptions. **2) Evidence Localization:** Identify and output a "Key-Slice" index as visual evidence supporting the diagnosis, ensuring traceability. **3) Verification:** Extract the selected key slice, generate a local description, and verify reliability by checking the correspondence of slice indices and lesion attributes.

This decomposition turns end-to-end generation into grounded reasoning, but selecting a discrete slice index is non-differentiable. To enable end-to-end training and visual verification under a unified objective, we formulate the process as a Markov Decision Process (MDP), defined by the tuple $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}\}$. **State Space ($\mathcal{S}$)**: At time step t, the state $s_t = [I_{ct}, y_{<t}]$ consists of the input 3D CT volume I_{ct} and the history of generated text tokens $y_{<t}$. **Action Space ($\mathcal{A}$)**: To enable interpretability, we define a hybrid action space. An action a_t can be either a text token from the vocabulary (for constructing the report) or a special pointing action $k \in \{1, ..., N\}$, representing the selection of the k-th slice as the key visual evidence. **Policy ($\mathcal{P} = \pi_\theta$)**: We denote a pre-trained VLM (Qwen3-VL) fine-tuned via SFT as the policy model to learn the distribution $\pi_\theta(a_t|s_t)$. **Reward ($\mathcal{R}$)**: We design a multi-granularity reward, including rewards of Tumor Existence r_e, Fine-Grained Information r_f, and Cross-View Consistency r_c.

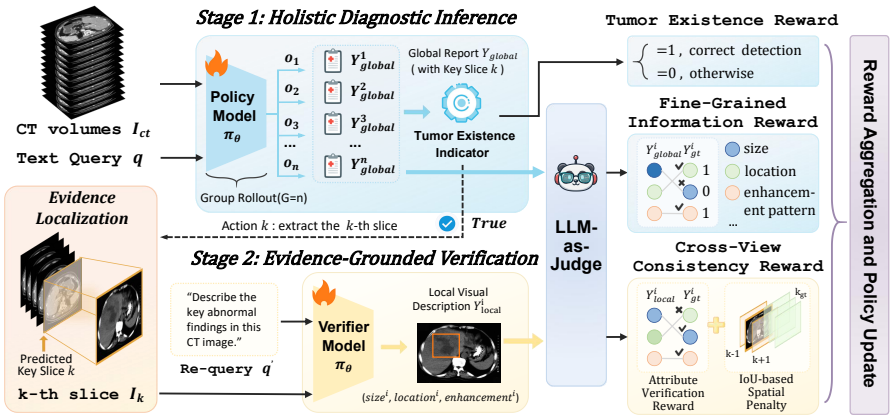

Fig. 1. The Diagnosis-Localization-Verification framework of Evidence-MRL.

3.2 Evidence-Driven Decomposed Reasoning

As illustrated in Fig. 1, our framework operates in two distinct stages.

Stage 1: Holistic Diagnostic Inference. The policy model π_θ processes the entire CT sequence to generate a comprehensive diagnostic report Y_{global} and explicitly predicts a key slice index k. The output is structured as follows:

`<findings>` f `</findings>` || `<impression>` i `</impression>` || (opt.) The key lesion is in image `<key_slice>` k `</key_slice>`, `<bbox>` $[x_1, y_1, x_2, y_2]$ `</bbox>`

Stage 2: Evidence-Grounded Verification. This acts as an environmental feedback step. Once the action k is selected, the environment extracts the k-th slice, I_k. Subsequently, I_k is fed into a Verifier (a weight-shared copy of the VLM) with a fixed prompt (e.g., "Describe the abnormality in this image. Should a tumor be identified, please describe its size, location and enhancement pattern.") to perform a single-image query, yielding a local visual description Y_{local}. This step mimics the radiologist's process of "double-checking the slice".

3.3 Multi-Granularity Reward

To drive evidence-based tumor analysis, we formulate a composite reward, deriving textual and visual modalities, with three components: $R_{\text{total}} = \alpha r_e + \beta r_f + \gamma r_c$, where α, β, γ are empirically tuned coefficients, finally set to (0.3, 0.5, 0.2).
Tumor Existence Reward ($r_\mathbf{e}$): This reward measures binary correctness for tumor existence detection. It is computed against the ground truth labels to ensure baseline clinical safety: $r_e = \mathbb{I}\left[\mathcal{E}(Y_{\text{global}}) = \mathcal{E}(Y_{\text{gt}})\right]$, where $\mathcal{E}(\cdot)$ denotes the tumor existence mapping, yielding 1 if a tumor is present and 0 otherwise, and $\mathbb{I}[\cdot]$ is the indicator function.
Fine-Grained Information Reward ($r_\mathbf{f}$): To promote clinical characterization of tumor lesions, we leverage a powerful LLM (Qwen-235B) to systematically extract structured entities and evaluate three key tumor attributes: size, location, and enhancement pattern. The reward is formulated into a unified metric weighted by clinical significance (coefficients: 0.3, 0.4, and 0.3, respectively):

$$r_{\text{f}} = \sum_{att \in \Lambda} w_{att} \cdot \mathcal{S}_{att}(Y_{\text{global}}, Y_{\text{gt}}) \tag{1}$$

where $\Lambda = \{size, location, enhancement\}$ denotes the set of key attributes and w_{att} reflects the clinical significance of current attribute. Attribute-level LLM scores can be defined as:

$$\mathcal{S}_{att}(Y_{\text{global}}, Y_{\text{gt}}) = \begin{cases} 1, & \text{if prediction for attribute matches gold answer} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Cross-View Consistency Reward ($r_\mathbf{c}$): This core contribution is designed to validate the accuracy and interpretability of visual evidence from a local slice perspective. Specifically, we focus on assessing the consistency between the attributes of the largest tumor region in the local slice and those documented in the ground-truth report. It aggregates attribute-level scores from an LLM-as-judge and imposes a spatial penalty based on intersection-over-union (IoU):

$$r_c = \sum_{att \in \Lambda} w_{att} \cdot \mathcal{S}_{att}(Y_{\text{local}}, Y_{gt}) - \lambda \cdot \mathcal{L}_{\text{dist}}(k, k_{gt}) \tag{3}$$

where

$$\mathcal{L}_{\text{dist}}(k, k_{gt}) = 1 - \text{IoU}([k-1, k+1], [k_{gt}-1, k_{gt}+1]) \tag{4}$$

Here, k and k_{gt} denote the slice number with the maximum tumor cross-sectional area for the prediction and ground truth. To account for potential annotation noise and segmentation inaccuracies, we expand the specific slice index into $[k-1, k+1]$ to calculate IoU. $\lambda = 0.3$ is a tunable penalty coefficient.

3.4 Policy Optimization via GRPO

To efficiently optimize the policy model π_θ without the overhead of a value network, we employ Group Relative Policy Optimization (GRPO). For each input I_v, we sample a group of G outputs $\{o_i\}_{i=1}^{G}$ from the policy π_θ, where $o_i = (Y^i_{\text{global}}, k^i, Y^i_{local})$. The advantage A_i is estimated by normalizing rewards within the group:

$$A_i = \frac{R(o_i) - \text{mean}(\{R(o_j)\}_{j=1}^{G})}{\text{std}(\{R(o_j)\}_{j=1}^{G}) + \epsilon} \tag{5}$$

The policy is updated by maximizing the surrogate objective with a KL-divergence penalty to maintain linguistic stability:

$$\mathcal{J}_{\mathrm{MRL}}(\theta) = \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\left(\min\left(\rho_i A_i, \mathrm{clip}(\rho_i, 1-\varepsilon, 1+\varepsilon)A_i\right) - \beta\mathbb{D}_{\mathrm{KL}}(\pi_\theta||\pi_{\mathrm{ref}})\right)\right] \tag{6}$$

where $\rho_i = \frac{\pi_\theta(o_i|I_v)}{\pi_{\theta_{\mathrm{old}}}(o_i|I_v)}$ is the probability ratio, and π_{ref} is the reference SFT model used to prevent reward hacking.

4 Experiments

4.1 Experimental Setup

Data Preparation. We focus on evidence-driven tumor analysis on a public tumor dataset **AbdomenAtlas3.0** [2]. All samples are cleaned and formatted uniformly and reviewed by experienced radiologists. Based on the tumor segmentation mask, we output key slices as visual evidence in the reports. AbdomenAtlas3.0 contains 9,262 CT images with 1,451 liver, 1,057 pancreas, and 2,049 kidney lesions. We organize each case into a single-organ report, and follows the train-test splitation as [2]. Our model is trained on both SFT and RL data derived from the training set and evaluated on test set. To construct a challenging RL dataset, we identify about 2,000 samples with erroneous predictions of tumor attributes or key slice in the SFT outputs for further RL training.

Implementation Details. Qwen-3-VL-8B [32] is used as base model of E-MRL. The model first undergoes SFT to acquire essential tumor analysis abilities, during which only the encoder and projection layer parameters are updated. During RL, both the projection layer and LLM parameters are updated and the learning rate is reduced to 5×10^{-6} to ensure stable convergence.

Baselines. We compare E-MRL with multiple general and medical models, including two Commercial LVLMs (GPT-4o [11] and Gemini 3 [26]), three 2D Medical Models (HuluMed [12], Lingshu [31] and MedVLM-R1 [23]), four 3D Medical Models (CT-Chat [8], CT2Rep [7], Merlin [3] and M3D [1]) and RadGPT [2].

Metrics. Following previous approaches [2], we use sensitivity and specificity as clinical metrics. To reflect overall clinical ability, we report balanced-accuracy (**B-Acc**), defined as the mean of sensitivity and specificity. For fine-grained tumor characterization, we assess accuracy for tumor size, location, enhancement pattern, and the hit rate of key slice (**K-Hit**) using Qwen-235B. K-Hit evalute the visual evidence reliability by calculating the hit rate based on whether the generated key slice index fall within the range of $[k_{gt} - 1,\ k_{gt} + 1]$.

4.2 Results

Diagnosis Performance. Table 1 demonstrates that our approach consistently surpasses end-to-end commercial solutions, as well as specialized 2D and 3D medical VLM models, across three major tumor types in terms of balanced accuracy (B-Acc). The suboptimal tumor detection accuracy (sensitivity and specificity) observed in most baselines underscores the challenges faced by current LVLMs

Table 1. Clinical Results (Sen., Spe., and Balanced-Acc.) on AbdomenAtlas3.0 benchmark. **Bold** denotes the best. Underline denotes the second highest scores.

Model	Pancreatic (%)				Kidney (%)				Liver (%)				Avg. B-Acc.
	Sen. (≤2cm)	Sen. (>2cm)	Spec.	B-Acc.	Sen. (≤2cm)	Sen. (>2cm)	Spec.	B-Acc.	Sen. (≤2cm)	Sen. (>2cm)	Spec.	B-Acc.	
Commercial VLM Models													
GPT-4o-mini	57.58	68.12	35.06	37.28	27.85	55.47	66.11	55.51	79.03	95.52	31.55	59.57	50.79
Gemini 3	96.67	98.41	9.78	54.38	94.29	100.00	8.89	53.35	87.50	98.33	13.68	53.40	53.71
2D Medical VLM Models													
HuluMed-7B	96.97	97.10	6.49	51.77	73.42	74.22	34.30	54.11	46.77	77.61	60.11	61.45	55.78
Lingshu-7B	78.79	88.41	19.82	51.92	68.35	87.50	36.84	58.51	50.00	76.12	49.02	56.29	55.57
MedVLM-R1-2B	46.88	57.35	37.82	45.19	55.70	62.20	57.89	58.80	37.78	70.37	30.89	42.80	48.93
3D Medical VLM Models (w/ SFT)													
CT-Chat	66.70	51.90	61.20	58.40	31.10	32.80	74.20	53.03	5.70	3.20	94.70	49.52	53.65
CT2Rep	0.00	0.00	92.50	46.25	36.50	39.30	70.40	54.08	35.80	49.20	70.40	56.74	52.36
Merlin	33.30	51.90	71.80	59.52	28.40	45.90	86.60	61.45	30.20	41.30	95.90	66.06	62.34
M3D	8.18	18.84	51.17	32.57	11.39	16.41	53.36	33.68	22.58	26.87	61.72	43.47	36.57
Previous SOTA: Segmentation-assisted Report Generation Model													
RadGPT	66.70	81.50	93.20	**85.50**	54.80	93.30	51.80	62.00	39.60	96.80	64.40	67.53	71.68
Our Models (Backbone: Qwen3-VL-8B)													
Zero-shot Inference	6.06	15.94	90.27	50.85	15.19	34.38	88.02	57.53	18.03	16.18	87.48	52.27	53.55
SFT	66.67	71.01	82.74	75.88	79.75	92.97	84.57	86.24	86.57	87.10	90.31	88.58	83.57
SFT+GRPO ($w/\ r_e$)	60.61	78.34	85.40	77.82	77.22	92.19	89.42	<u>87.94</u>	83.51	89.55	95.47	91.06	85.61
SFT+GRPO ($w/\ r_e, r_f$)	75.76	84.06	83.82	<u>82.04</u>	82.28	91.41	85.08	86.50	88.52	93.03	93.78	<u>92.32</u>	<u>86.95</u>
SFT+E-MRL	78.79	84.06	83.21	**82.43**	89.74	94.44	83.82	**88.23**	90.32	92.54	94.81	**93.14**	**87.93**
Δ vs SFT model	↑ 12.12	↑ 13.05	↑ 0.47	↑ 6.55	↑ 9.99	↑ 1.47	↓ 0.75	↑ 1.99	↑ 3.75	↑ 5.44	↑ 4.50	↑ 4.56	↑ 4.36

in generating CT reports for clinically significant tasks. To ensure a fair comparison, all competing 3D medical models [8,7,3,1] were supervised fine-tuned on our curated dataset. Notably, our SFT variant yielded substantial performance gains; we attribute this to our strategy of fine-tuning the image encoder exclusively while treating CT slices as video sequences. This design effectively retains the foundational representation capabilities of Qwen3-VL-8B. Furthermore, ablation studies(see Fig. 2a) validate the efficacy of our proposed RL rewards: consistent improvements were observed, with the full "Evidence-MRL" configuration (utilizing all three reward functions) achieving state-of-the-art (SOTA) performance of 87.93% Avg. B-Acc across three types tumor.

Notably, our approach even outperforms the previous SOTA RadGPT [2] by over 16% in Avg. B-Acc. RadGPT leverages expert tumor segmentation models for fine-grained report generation. Our model is by far the first end-to-end VLM solution to achieve such clinically applicable performance for tumor-related report generation.

Attribute Reliability and Visual Grounding Abilities. Table 2 reveals that both commercial and specialized medical models exhibit limited accuracy in fine-grained tumor grounding. Key attributes, including tumor size, location, relative enhancement, and key slice localization, represent critical characteristics prioritized by radiologists for interpretation, yet remain challenging for automated report generation systems. In contrast, our approach demonstrates consistent performance gains across training stages: from supervised fine-tuning to

Table 2. Lesion Attributes and visual evidence evaluation (size, location, enhancement and K-Hit) on AbdomenAtlas3.0 benchmark. "–" denotes that the model does not have the capability to generate key slice descriptions.

Model	Pancreatic Tumor (%)				Kidney Tumor (%)				Liver Tumor (%)				Avg.
	Size	Loc.	Enhan.	K-Hit	Size	Loc.	Enhan.	K-Hit	Size	Loc.	Enhan.	K-Hit	
Baseline Models													
GPT-4o-mini	20.59	50.00	45.10	27.45	11.11	41.06	54.59	14.49	44.96	20.93	77.52	23.26	35.92
Gemini 3	31.96	43.30	51.55	30.52	56.57	63.74	77.80	28.28	66.67	36.43	82.95	27.13	49.74
HuluMed-7B	30.39	12.75	25.49	9.49	37.20	36.71	51.69	11.32	37.98	28.68	40.31	7.85	27.49
Lingshu-7B	26.47	20.59	26.47	19.79	28.23	11.21	15.80	13.21	38.08	8.72	12.21	10.91	19.31
MedVLM-R1-2B	21.77	14.92	12.79	1.94	38.65	31.88	25.12	2.83	11.34	1.06	2.47	0.77	13.80
M3D	6.86	5.88	17.45	–	3.01	11.11	18.02	–	3.18	1.55	13.18	–	8.92
Our Models (Backbone: Qwen3-VL-8B)													
Zero-shot Inference	4.90	15.69	32.35	1.96	9.18	32.85	46.38	2.42	10.85	3.10	37.21	1.83	16.56
SFT	50.98	53.92	49.02	48.04	64.73	73.43	70.53	56.04	65.12	25.58	74.42	58.10	57.49
SFT+GRPO ($w/$ r_e)	46.27	46.08	45.20	44.12	66.18	70.46	71.50	60.80	65.12	26.36	72.87	52.64	55.63
SFT+GRPO ($w/$ r_e, r_f)	50.00	64.71	50.98	53.92	71.50	74.88	77.29	63.77	76.77	46.51	86.40	60.36	64.76
SFT+E-MRL	**51.96**	**66.67**	**52.94**	**54.90**	**73.91**	**76.33**	**79.71**	**71.01**	**79.84**	**49.61**	**88.37**	**65.12**	**67.53**

GRPO optimization incorporating the tumor existence reward (r_e), fine-grained information reward (r_f), and cross-view consistency reward (r_c). These results underscore our model's superiority in generating grounded evidence and enhancing interpretability. Note that our models cannot be compared with RadGPT because we used RadGPT-generated fine-grained reports as references to evaluate the metrics in Table 2.

Cases Study. Fig. 2(b) illustrates the comparison between the SFT model and the E-MRL model in liver CT report generation. The SFT model missed both tumor detection and attribute analysis. In contrast, E-MRL model accurately detected the small tumor and correctly reported its location and enhancement pattern. The key slice was also identified within the tolerance range, showing a high consistency with the ground truth. This case demonstrates the significant advantages of our proposed E-MRL method in improving tumor detection rates and enriching structured details in radiology reports.

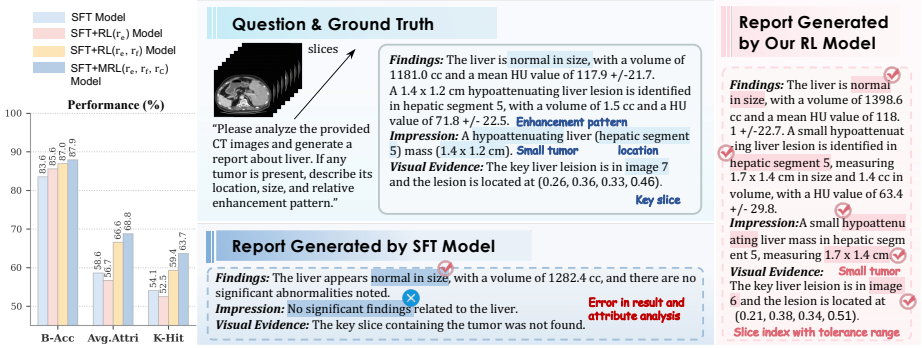

Fig. 2. (a) **Ablation** study on three kinds of multi-granularity rewards. (b) A **case analysis** of performance enhancement through E-MRL built upon the SFT model.

5 Conclusion

In this work, we introduced Evidence-MRL, a multimodal reinforcement learning framework designed to address visual hallucinations in 3D CT tumor analysis. By formulating report generation as a diagnosis-localization-verification process, our method enforces explicit visual grounding through a Cross-View Consistency Reward, ensuring that global diagnostic conclusions are substantiated by local slice evidence. Extensive experiments demonstrate that Evidence-MRL significantly outperforms existing baselines in both diagnostic accuracy and fine-grained attribute grounding. This work sheds light on developing trustworthy and interpretable end-to-end LVLM systems capable of supporting reliable clinical decision-making for tumor analysis.

Disclosure of Interest

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements

This work has been supported in part by the NSFC (No. 62436007), the China Postdoctoral Science Foundation under Grant Number 2024M752794, the ZJNSF (No. LZ25F020004), the Key Research and Development Projects in Zhejiang Province (No. 2025C01128, 2025C01030, 2025C02156), Ningbo Yongjiang Talent Introduction Programme (2023A400-G).

References

1. Bai, F., Du, Y., Huang, T., Meng, M.Q.H., Zhao, B.: M3d: Advancing 3d medical image analysis with multi-modal large language models. arXiv preprint arXiv:2404.00578 (2024)
2. Bassi, P.R., Yavuz, M.C., Hamamci, I.E., Er, S., Chen, X., Li, W., Menze, B., Decherchi, S., Cavalli, A., Wang, K., et al.: Radgpt: Constructing 3d image-text tumor datasets. In: ICCV. pp. 23720–23730 (2025)
3. Blankemeier, L., Cohen, J.P., Kumar, A., Van Veen, D., Gardezi, S.J.S., Paschali, M., Chen, Z., Delbrouck, J.B., Reis, E., Truyts, C., et al.: Merlin: A vision language foundation model for 3d computed tomography. Research Square pp. rs–3 (2024)
4. Deria, A., Kumar, K., Dukre, A.M., Segal, E., Khan, S., Razzak, I.: MedMO: Grounding and understanding multimodal large language model for medical images. arXiv preprint arXiv:2602.06965 (2026)
5. Fan, Z., Cao, J., Dai, Y., Lv, Z., Zhang, W., Xie, Z., LU, P., Ooi, B.C.: Ctrlcot: Dual-granularity chain-of-thought compression for controllable reasoning. arXiv preprint arXiv:2601.20467 (2026)
6. Feng, Y., Wang, J., Zhou, L., Lei, Z., Li, Y.: Doctoragent-rl: A multi-agent collaborative reinforcement learning system for multi-turn clinical dialogue. arXiv preprint arXiv:2505.19630 (2025)
7. Hamamci, I.E., Er, S., Menze, B.: Ct2rep: Automated radiology report generation for 3d medical imaging. In: MICCAI. pp. 476–486. Springer (2024)
8. Hamamci, I.E., Er, S., Wang, C., Almas, F., Simsek, A.G., Esirgun, S.N., Dogan, I., Durugol, O.F., Hou, B., Shit, S., et al.: Generalist foundation models from a multimodal dataset for 3d computed tomography. Nature Biomedical Engineering pp. 1–19 (2026)
9. Hu, Q., Yi, Z., Zhou, Y., Peng, F., Liu, M., Li, Q., Wang, Z.: Sali: Short-term alignment and long-term interaction network for colonoscopy video polyp segmentation. In: MICCAI. pp. 531–541. Springer (2024)
10. Huang, X., Wu, J., Liu, H., Tang, X., Zhou, Y.: Medvlthinker: Simple baselines for multimodal medical reasoning. arXiv preprint arXiv:2508.02669 (2025)
11. Hurst, A., Lerer, A., Goucher, A.P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., et al.: Gpt-4o system card. arXiv preprint arXiv:2410.21276 (2024)
12. Jiang, S., Wang, Y., Song, S., Hu, T., Zhou, C., Pu, B., Zhang, Y., Yang, Z., Feng, Y., Zhou, J.T., et al.: Hulu-med: A transparent generalist model towards holistic medical vision-language understanding. arXiv preprint arXiv:2510.08668 (2025)
13. Jiang, Z., Guo, H., Fang, C., Xiao, C., Hu, X., Sun, L., Xu, M.: MedVR: Annotation-free medical visual reasoning via agentic reinforcement learning. In: ICLR (2026)
14. Lai, H., Jiang, Z., Zhang, K., Yao, Q., Wang, R., He, Z., Tao, X., Wei, W., Zhou, S.K.: Med3D-R1: Incentivizing clinical reasoning in 3d medical vision-language models for abnormality diagnosis. arXiv preprint arXiv:2602.01200 (2026)
15. Lai, Y., Zhong, J., Li, M., Zhao, S., Li, Y., Psounis, K., Yang, X.: Med-r1: Reinforcement learning for generalizable medical reasoning in vision-language models. IEEE Transactions on Medical Imaging (2026)
16. Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.: Llava-med: Training a large language-and-vision assistant for biomedicine in one day. Advances in NeurIPS **36**, 28541–28564 (2023)

17. Li, S., Lin, T., Lin, L., Zhang, W., Liu, J., Yang, X., Li, J., He, Y., Song, X., Xiao, J., et al.: Eyecaregpt: Boosting comprehensive ophthalmology understanding with tailored dataset, benchmark and model. In: ACM MM. pp. 3893–3902 (2025)
18. Li, S., Qiu, Z., Liu, J., Zhang, W., Lin, T., Xie, Y., An, J., Yun, B., Yang, C., Xiao, J., et al.: Tumorchain: Interleaved multimodal chain-of-thought reasoning for traceable clinical tumor analysis. In: ICLR
19. Li, Y., Tang, F., Li, Y., Zhou, S.K.: Medreason-r1: Learning to reason for ct diagnosis with reinforcement learning and local zoom. In: ISBI. pp. 1–5. IEEE (2026)
20. Lin, T., Qiu, Z., Zhang, W., Liu, J., Xie, Y., Gao, M., Fan, Z., Li, Z., Li, S., Xie, Z., et al.: Omnict: Towards a unified slice-volume lvlm for comprehensive ct analysis. arXiv preprint arXiv:2602.16110 (2026)
21. Lin, T., Zhang, W., Li, S., Yuan, Y., Yu, B., Li, H., He, W., Jiang, H., Li, M., Song, X., et al.: Healthgpt: A medical large vision-language model for unifying comprehension and generation via heterogeneous knowledge adaptation. arXiv preprint arXiv:2502.09838 (2025)
22. Liu, C., Wu, J., Xie, Z., Zhang, W., Zheng, K., Zhu, J., Cai, Q., Anne, O.G., Tan, M.C.J., Yin, J., et al.: Diyhealth suite: Dataset, model, and benchmark for health management at home. arXiv preprint arXiv:2606.07542 (2026)
23. Pan, J., Liu, C., Wu, J., Liu, F., Zhu, J., Li, H.B., Chen, C., Ouyang, C., Rueckert, D.: Medvlm-r1: Incentivizing medical reasoning capability of vision-language models (vlms) via reinforcement learning. In: MICCAI. pp. 337–347. Springer (2025)
24. Pellegrini, C., Özsoy, E., Busam, B., Wiestler, B., Navab, N., Keicher, M.: Radialog: Large vision-language models for x-ray reporting and dialog-driven assistance. In: Medical Imaging with Deep Learning (2025)
25. Sellergren, A., Kazemzadeh, S., Jaroensri, T., Kiraly, A., Traverse, M., Kohlberger, T., Xu, S., Jamil, F., Hughes, C., Lau, C., et al.: Medgemma technical report. arXiv preprint arXiv:2507.05201 (2025)
26. Team, G., Anil, R., Borgeaud, S., Alayrac, J.B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A.M., Hauth, A., Millican, K., et al.: Gemini: a family of highly capable multimodal models. arXiv preprint arXiv:2312.11805 (2023)
27. Tu, T., Azizi, S., Driess, D., Schaekermann, M., Amin, M., Chang, P.C., Carroll, A., Lau, C., Tanno, R., Ktena, I., et al.: Towards generalist biomedical ai. Nejm Ai **1**(3), AIoa2300138 (2024)
28. Wu, C., Zhang, X., Zhang, Y., Hui, H., Wang, Y., Xie, W.: Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data. Nature Communications **16**(1), 7866 (2025)
29. Xia, P., Wang, J., Peng, Y., Zeng, K., Wu, X., Tang, X., Zhu, H., Li, Y., Liu, S., Lu, Y., et al.: Mmedagent-rl: Optimizing multi-agent collaboration for multimodal medical reasoning. arXiv preprint arXiv:2506.00555 (2025)
30. Xie, Y., Li, S., Lin, T., Wang, Z., Yang, C., Zhong, Y., Zhang, W., Li, H., Jiang, H., Zhang, F., et al.: Heartcare suite: Multi-dimensional understanding of ecg with raw multi-lead signal modeling. arXiv e-prints pp. arXiv–2506 (2025)
31. Xu, W., Chan, H.P., Li, L., Aljunied, M., Yuan, R., Wang, J., Xiao, C., Chen, G., Liu, C., Li, Z., et al.: Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning. arXiv preprint arXiv:2506.07044 (2025)
32. Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al.: Qwen3 technical report. arXiv preprint arXiv:2505.09388 (2025)
33. Yu, T., Yao, Y., Zhang, H., He, T., Han, Y., Cui, G., Hu, J., Liu, Z., Zheng, H.T., Sun, M., et al.: Rlhf-v: Towards trustworthy mllms via behavior alignment from fine-grained correctional human feedback. In: CVPR. pp. 13807–13816 (2024)