MedVol-R1: Reward-Driven Evidence Grounding for Volumetric Reasoning Segmentation

Zichun Wang[1†], Hairong Shi[2†], Bingzheng Wei[3], Yan Xu[1](✉), and Zihua Wang[4](✉)

[1] School of Biological Science and Medical Engineering, Beihang University, Beijing, China
xuyan04@gmail.com
[2] Center for Information and Computer Science, School of Science for Open and Environmental Systems, Graduate School of Science and Technology, Keio University, Kanagawa, Japan
[3] Bytedance Inc., China
[4] Tsinghua University, Beijing, China
wangzihua07@126.com

Abstract. Volumetric Reasoning Segmentation (VRS) aims to segment a target region in a 3D medical scan from a free-form clinical query, where the referent is often implicit and requires both medical knowledge and volume-grounded reasoning. Existing methods typically rely on specialized segmentation tokens to connect language with mask decoding, but this coupling collapses the decision process into opaque latent representations, limiting interpretability and generalization to diverse narrative expressions. In this paper, we present **MedVol-R1**, a reinforcement learning-based framework for VRS that explicitly decouples *evidence grounding* from *volumetric delineation*: the LVLM grounds clinical reasoning to a verifiable 2D evidence anchor (key axial slice and 2D bounding boxes), which is then propagated into a coherent 3D mask by a frozen MedSAM2 model. We train MedVol-R1 with cold-start supervised fine-tuning followed by GRPO, guided by a multi-component reward that encourages informative evidence selection, accurate 2D spatial grounding, and cross-slice volumetric coherence, without requiring costly chain-of-thought annotations. Experiments on CT-ORG, AbdomenCT-1K, and KiTS23 from the M3D-Seg benchmark demonstrate that MedVol-R1 consistently outperforms strong baselines and achieves state-of-the-art performance, with reinforcement learning providing clear gains over pure supervised fine-tuning. Code is available here.

Keywords: Volumetric Reasoning Segmentation · Medical Image Segmentation · Group Relative Policy Optimization.

† Equal contribution: Zichun Wang and Hairong Shi.
✉ Corresponding authors: Yan Xu and Zihua Wang.

1 Introduction

Volumetric Reasoning Segmentation (VRS) is an important capability for computer-aided diagnosis and clinician–AI interaction [1, 3, 8, 13, 15]. Given a 3D medical scan and a free-form clinical query, the goal is to produce a voxel-level binary mask for the referred target. Unlike label-based prompts, VRS queries often refer to targets implicitly, requiring either medical world knowledge (e.g., the organ that secretes bile), or volume-grounded reasoning (e.g., the kidney that contains the tumor) to resolve the referent [1]. This requires the model to not only parse language, but also resolve the hidden clinical reference by integrating external medical knowledge with volumetric visual evidence, and then translate this reasoning into reliable 3D localization with spatially coherent voxel masks.

Following the success of reasoning-based visual perception in the general domain [6, 7, 17], recent works have extended 3D medical Large Vision Language Models (LVLMs) from coarse semantic interpretation to dense voxel-level segmentation [1, 14]. In parallel, prompt-driven and interactive SAM-style medical segmentors have also advanced rapid 2D/3D annotation and refinement [2, 11, 12, 16], but they typically assume an explicit user prompt and do not address implicit narrative reasoning in 3D volumes. Among such LVLM-based methods, a common strategy is to introduce dedicated segmentation tokens (e.g., `<SEG>`) as latent semantic carriers that connect language-conditioned volumetric representations to a downstream mask decoder for voxel-level prediction. While effective for queries with explicit terminology, this paradigm suffers from two major limitations when confronting implicit, reasoning-dependent clinical narratives. First, by collapsing the entire decision process into implicit token representations, these methods provide no explicit reasoning chain from query interpretation to spatial grounding, offering neither verifiable evidence for target selection nor interpretable rationale for clinical decision support. Second, such imitation-based supervision tends to encourage surface-level pattern mimicry rather than genuine clinical understanding, leading to poor generalization beyond training-time phrasings.

Recent advancements in Group Relative Policy Optimization (GRPO) [5], originally proposed for large language model alignment, have been successfully extended to vision-language tasks [4, 9, 10, 18–20], enabling models to discover structured reasoning patterns without relying on expensive, manually-annotated Chain-of-Thought (CoT) sequences. Building on this insight, we propose MedVol-R1, the first reinforcement learning-based framework for VRS. MedVol-R1 employs GRPO initialized with a cold-start fine-tuning phase to autonomously develop structured reasoning capabilities. To further guide the model in the complex 3D clinical setting, we design a multi-dimensional reward mechanism that balances format constraints, hierarchical localization, and cross-slice consistency, ensuring segmentation results that are both logically grounded and anatomically coherent.

Our primary contributions are summarized as follows: (1) We propose MedVol-R1, an RL-based framework for VRS that applies GRPO on top of cold-start SFT, without requiring CoT annotations. (2) We design a multi-component re-

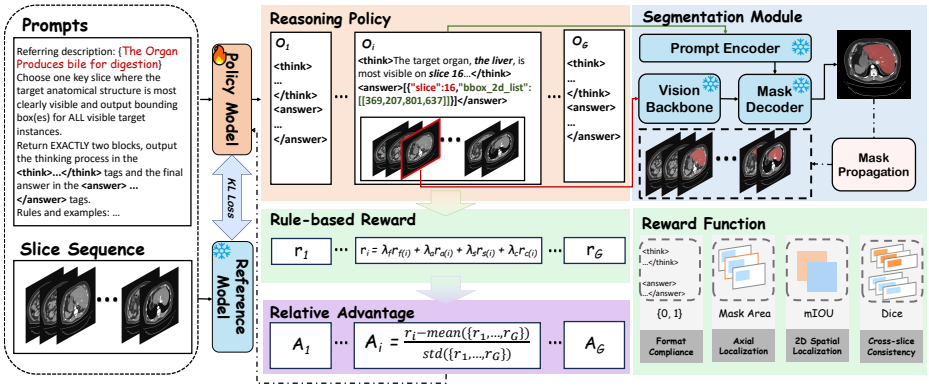

Fig. 1. Overall pipeline of MedVol-R1.

ward that enforces format validity, axial evidence selection, 2D box localization, and cross-slice volumetric consistency. (3) Extensive experiments on three M3D-Seg CT subsets show state-of-the-art performance.

2 Method

2.1 Task Formulation

We consider **VRS** on CT volumes. Given a CT volume $\mathbf{V} \in \mathbb{R}^{H\times W\times D}$ and a free-form implicit clinical query q, the objective is to produce a voxel-level binary mask of the referred target:

$$\Phi_\theta : \ (\mathbf{V}, q) \ \mapsto \ \hat{\mathbf{Y}} \in \{0,1\}^{H\times W\times D}, \tag{1}$$

where $\hat{\mathbf{Y}}$ is the predicted volumetric mask.

Unlike text-prompted volumetric segmentation [3, 21, 22], VRS queries are often implicit, requiring medical knowledge and volume-grounded reasoning to resolve the referent before producing a coherent 3D mask.

2.2 Preliminary: Group Relative Policy Optimization

Group Relative Policy Optimization [5] is a reinforcement learning algorithm that removes the need for an explicit value function by estimating advantages from *relative rewards* within a group of samples. Given a prompt p, GRPO draws a group of G responses $\{o_i\}_{i=1}^{G}$ from the current policy $\pi_\theta(\cdot \mid p)$. A reward function $R(\cdot,\cdot)$ assigns a scalar score to each response, producing $r_i = R(o_i, p)$ for $i = 1, \ldots, G$. The relative advantage A_i is then computed via group-wise normalization over $\{r_1, \ldots, r_G\}$:

$$A_i = \frac{r_i - \text{mean}(\{r_1, r_2, \ldots, r_G\})}{\text{std}(\{r_1, r_2, \ldots, r_G\})}. \tag{2}$$

GRPO optimizes a clipped policy-gradient objective with a Kullback–Leibler penalty to stabilize training:

$$\mathcal{L}_{\text{GRPO}}(\theta) = -\frac{1}{G}\sum_{i=1}^{G}\left[\min(\rho_i A_i,\ \text{clip}(\rho_i, 1-\epsilon, 1+\epsilon)\, A_i) - \beta\, D_{\text{KL}}(\pi_\theta \,\|\, \pi_{\text{ref}})\right], \tag{3}$$

where $\rho_i = \frac{\pi_\theta(o_i|p)}{\pi_{\theta_{\text{old}}}(o_i|p)}$, π_{ref} is the reference policy , and β is the regularization coefficient.

2.3 Pipeline

Unlike prior approaches [1] that encapsulate 3D semantics into opaque specialized tokens, MedVol-R1 explicitly decouples volumetric reasoning from geometric delineation. As illustrated in Fig. 1, the framework operates in two stages: (1) *Evidence Grounding*: given a multi-slice sequence $\mathcal{S}$ and a clinical query q, the LVLM identifies a key slice I_k and predicts 2D bounding boxes $\mathcal{B}_k = \{b_i\}_{i=1}^{N_k}$ on that slice; (2) *Volumetric Propagation*: a frozen MedSAM2 [12] lifts this 2D anchor $(I_k, \mathcal{B}_k)$ into a volumetric mask $\hat{\mathbf{Y}}$ via slice-to-slice memory propagation. We optimize the LVLM through cold-start SFT (Sec. 2.4) followed by GRPO (Sec. 2.5).

2.4 Stage I: Supervised Fine-tuning

We begin with SFT to obtain a *cold-start* policy that mitigates the sparse-reward problem in RL [4,5,9]. Each axial slice in $\mathcal{S}$ is prefixed with a discrete identifier (e.g., `<slice 1>`). Given a clinical query q, the model is trained to predict, within `<answer>` tags, the key-slice index k and the corresponding bounding boxes $\mathcal{B}_k$.

Supervision targets are derived from $\mathbf{Y}^*$ by Top-K visibility sampling (rank slices by foreground area, sample one from the Top-K), and 2D boxes are obtained by connected-component decomposition on the selected slice mask.

2.5 Stage II: Reward-Driven Policy Optimization via GRPO

Building on the SFT-initialized policy, we optimize the LVLM with GRPO [5]. For each query q, the policy π_θ samples a group of G outputs $\{o_i\}_{i=1}^{G}$ under a fixed `<think>`–`<answer>` format, where `<answer>` specifies an evidence anchor (key-slice index $\hat{k}$ and 2D boxes $\mathcal{B}_{\hat{k}}$). Each output receives a composite reward $r_i = R_{\text{total}}(o_i, q, \mathbf{Y}^*)$ that combines format validity, axial evidence selection, 2D box grounding, and cross-slice consistency. GRPO updates π_θ using within-group standardized advantages and a KL penalty to a frozen reference model π_{ref} [5].

Reward Function. Reward shaping is critical for effective policy optimization under GRPO. We therefore design a multi-dimensional reward mechanism $\mathcal{R}$ that evaluates each sampled output from complementary perspectives: (i) format compliance for reliable parsing, (ii) axial evidence selection for informative

key-slice choice, (iii) 2D spatial localization for accurate bounding-box grounding, and (iv) cross-slice consistency for coherent volumetric delineation under through-plane variations.

Format Compliance Reward. R_f enforces a strict output schema: $R_f = 1$ if the response contains valid `<think>...</think>` and `<answer>...</answer>` blocks whose `<answer>` content parses into the required schema (key-slice index and `bbox_2d_list`), and $R_f = 0$ otherwise.

Axial Localization Reward. R_a encourages selecting slices where the target is maximally visible. Letting F_t denote the ground-truth foreground area on slice t, we define $R_a = F_{\hat{k}} / \max_t F_t$, which provides graded credit proportional to the target visibility on the predicted key slice $\hat{k}$.

2D Spatial Localization Reward. R_s measures the spatial precision of predicted bounding boxes on the selected key slice. Let $\hat{\mathcal{B}} = \{\hat{b}_i\}_{i=1}^{N}$ denote the N predicted boxes and $\mathcal{B}^* = \{b_j^*\}_{j=1}^{N^*}$ the N^* ground-truth boxes on the predicted key slice $I_{\hat{k}}$. We construct an IoU-based cost matrix $C \in \mathbb{R}^{N \times N^*}$:

$$C_{i,j} = 1 - \mathrm{IoU}(\hat{b}_i, b_j^*),$$

and apply the Hungarian algorithm to obtain an optimal one-to-one matching set $\mathcal{M}$. The reward is defined as the normalized sum of matched IoUs:

$$R_s = \frac{1}{\max(N, N^*)} \sum_{(i,j)\in\mathcal{M}} \mathrm{IoU}(\hat{b}_i, b_j^*) \in [0, 1],$$

where the normalization naturally penalizes missed targets and spurious predictions.

Cross-slice Consistency Reward. To encourage volumetric coherence beyond a single slice, we introduce a cross-slice consistency reward R_c based on mask propagation. After obtaining the matched boxes on the predicted key slice (via Hungarian matching), we use them as prompts to an off-the-shelf MedSAM2 model and propagate slice-wise to obtain a volumetric prediction $\hat{\mathbf{Y}}$. We then compute the Dice score on a local axial neighborhood consisting of the key slice $\hat{k}$ and its adjacent slices:

$$R_c = \mathrm{Dice}\Big(\hat{\mathbf{Y}}_{\mathcal{N}(\hat{k})}, \mathbf{Y}^*_{\mathcal{N}(\hat{k})}\Big) \in [0, 1],$$

where $\mathcal{N}(\hat{k}) = \{t \mid |t - \hat{k}| \le \Delta\}$ denotes the set of $\pm\Delta$ adjacent slices around $\hat{k}$, and $\hat{\mathbf{Y}}_{\mathcal{N}(\hat{k})}$ and $\mathbf{Y}^*_{\mathcal{N}(\hat{k})}$ are the corresponding masks restricted to this neighborhood. This term directly evaluates whether the predicted evidence anchor can serve as an effective prompt for coherent volumetric delineation, while keeping RL rollouts computationally tractable.

Overall Reward. We combine the above components as a weighted sum:

$$R_{\text{total}} = \lambda_f R_f + \lambda_a R_a + \lambda_s R_s + \lambda_c R_c,$$

where $\lambda_f, \lambda_a, \lambda_s, \lambda_c$ are fixed weights and we set $\lambda_f = \lambda_a = \lambda_s = \lambda_c = 1.0$ in all experiments.

3 Experiments

3.1 Experimental Settings

Datasets and Evaluation Metrics We evaluate our framework on three CT sub-datasets from the M3D-Seg [1] benchmark: CT-ORG, AbdomenCT-1K, and KiTS23. Following the M3D-Seg protocol, all samples are organized as image–mask–text triplets, where fixed anatomical labels are replaced with free-form linguistic descriptions to support VRS.

To further assess the model's capacity for grounded reasoning over 3D volumes, we paraphrase the original semantic descriptions of all KiTS23 test queries into more diverse expressions. Unlike fixed category names, these queries couple lesion attributes with anatomical spatial relations (e.g., "the kidney containing an obvious fluid-filled sac"), requiring the model to jointly interpret semantic cues and 3D structural context. We report standard voxel-level segmentation metrics, including Dice Similarity Coefficient (DSC) and IoU.

Compared Methods We compare our method with two paradigms of baselines. First, we include two specialized medical parsers, SAT [22] and Biomed-ParseV2 [21], which are promptable medical segmentation models; we fine-tune them on the M3D-Seg training split using the same free-form referring queries to assess their adaptability to open-ended descriptions. Second, we compare with M3D, a native volumetric vision-language baseline for 3D medical image-text understanding.

Implementation Details We instantiate our framework with Qwen3-VL-4B, which provides a practical balance between vision-language reasoning capability and computational feasibility for GRPO-based volumetric evidence grounding. Each 3D volume is represented as a uniformly sampled sequence of 64 axial slices resized to 256×256. Training follows a two-stage pipeline. **(i) SFT.** We fine-tune the model for 1 epoch using LoRA (rank 128, all-linear) with a learning rate of 2×10^{-5} and a per-GPU batch size of 2. **(ii) RL.** We then optimize the policy for 3 epochs with GRPO (G=4, β=0.01), using a learning rate of 2×10^{-6} and a per-GPU batch size of 1. Both stages use AdamW with a 0.1 warmup ratio and a cosine learning-rate schedule. All experiments are run on NVIDIA RTX A6000 GPUs with bfloat16 precision and gradient checkpointing. The maximum generation length is 256 tokens.

3.2 Comparisons with Other Methods

Table 1 compares MedVol-R1 with representative baselines on AbdomenCT-1K, CT-ORG, and KiTS23. Our full model (SFT+RL) achieves the highest DSC and IoU on all three benchmarks, with particularly large margins on AbdomenCT-1K (89.86 vs. 73.63 DSC) and KiTS23 (45.46 vs. 30.69 DSC). Among the baselines, M3D remains competitive on CT-ORG where queries use standard terminology, but degrades sharply on KiTS23 whose paraphrased queries require

Table 1. Quantitative comparison with state-of-the-art methods.

Method	AbdomenCT-1K		CT-ORG		KiTS23	
	DSC	IoU	DSC	IoU	DSC	IoU
SAT [22]	58.44	41.28	48.64	32.14	28.70	16.75
BiomedParseV2 [21]	68.78	52.42	40.23	25.18	32.94	19.72
M3D [1]	73.63	58.27	83.49	71.66	30.69	18.13
Ours (Pure SFT)	85.52	74.70	83.23	71.28	36.21	22.11
Ours (SFT+RL)	**89.86**	**81.59**	**85.43**	**74.57**	**45.46**	**29.42**

(a) Effect of Reward Components

Setting	DSC	IoU
RL only (w/o SFT)	–	–
SFT only	85.52	74.70
w/o R_c	88.61	79.55
w/o R_s	87.58	77.90
w/o R_a	89.17	80.46
w/o R_f	88.98	80.53
Full	**89.86**	**81.59**

(b) Effect of Δ

Δ	DSC	IoU
2	84.41	73.03
5	**85.43**	**74.57**
10	85.16	74.16
15	83.99	72.40

Table 2. Ablation studies. (a) Effect of reward components on AbdomenCT-1K. "w/o" denotes removing one reward term from the full GRPO objective. (b) Effect of the local axial window size Δ used in the cross-slice consistency reward R_c on CT-ORG.

joint interpretation of lesion attributes and spatial context. SAT and BiomedParseV2, designed for explicit category prompts, fail to generalize to free-form narrative queries despite fine-tuning. GRPO brings consistent gains over pure SFT (+4.34 DSC on AbdomenCT-1K, +2.20 on CT-ORG, +9.25 on KiTS23), with the largest improvement on the most reasoning-heavy subset, confirming that reward-driven optimization is especially effective when queries demand non-trivial clinical inference.

3.3 Qualitative Results

Fig. 2 presents qualitative comparisons on four representative samples. Across all cases, MedVol-R1 produces segmentation masks that more closely match the ground-truth (GT) than the strongest baseline M3D. In particular, our results show cleaner boundaries and reduced leakage to adjacent structures, while preserving more complete target coverage. This indicates that the RL-enhanced policy provides more reliable evidence anchors for downstream MedSAM2 propagation, consistent with the quantitative gains in Table 1.

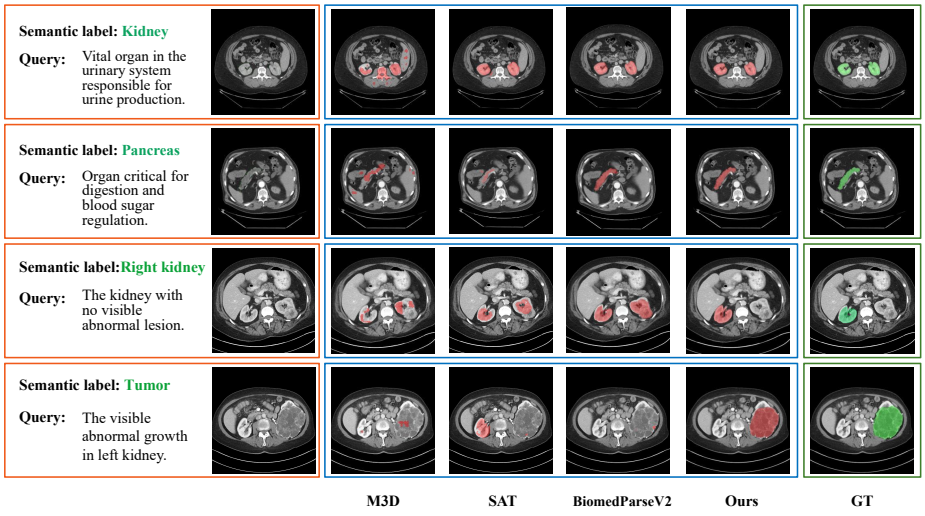

Fig. 2. Qualitative comparisons on four representative VRS samples.

3.4 Ablation Study

Table 2(a) ablates key design choices of our training pipeline on AbdomenCT-1K. RL without SFT warm-start fails to converge, confirming that cold-start initialization is essential for stable policy optimization. Compared with SFT-only, the full model improves DSC from 85.52 to 89.86, showing the benefit of GRPO over pure imitation learning. Removing any single reward component degrades performance, validating their complementarity. Among them, removing R_s causes the largest decline (DSC 87.58, IoU 77.90), indicating that accurate 2D spatial grounding is critical for downstream volumetric propagation. Removing R_c yields the second-largest drop, confirming the importance of cross-slice consistency.

Table 2(b) studies the neighborhood size Δ in R_c on CT-ORG. $\Delta = 5$ achieves the best performance, while smaller windows provide limited volumetric supervision and larger windows may introduce noisy signals from distant slices with weak or ambiguous target visibility. Thus, we use $\Delta = 5$ by default.

4 Conclusion

We presented MedVol-R1, a reinforcement learning-based framework for Volumetric Reasoning Segmentation that explicitly decouples evidence grounding from volumetric delineation. The LVLM was trained to predict a verifiable 2D evidence anchor (key axial slice and bounding boxes), which a frozen MedSAM2

propagated into a coherent 3D mask. A two-stage pipeline of cold-start SFT followed by GRPO, guided by a multi-component reward covering format validity, axial evidence selection, spatial grounding, and cross-slice consistency, avoids the need for manually annotated chain-of-thought supervision. Experiments on three M3D-Seg CT subsets showed consistent improvements over strong baselines, with GRPO providing clear gains beyond pure supervised fine-tuning. Future work will extend MedVol-R1 to multi-modal volumetric data, such as MRI and ultrasound volumes and explore richer evidence representations beyond single-slice anchors to further improve volumetric coherence.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grants 62371016 and U23B2063, the Fundamental Research Funds for the Central Universities from the State Key Laboratory of Software Development Environment at Beihang University, the 111 Project under Grant B13003, and the high-performance computing (HPC) resources at Beihang University.

Disclosure of Interests. We have no conflicts of interest to disclose.

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. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. arXiv preprint arXiv:2308.16184 (2023)
3. Du, Y., Bai, F., Huang, T., Zhao, B.: Segvol: Universal and interactive volumetric medical image segmentation. Advances in Neural Information Processing Systems **37**, 110746–110783 (2024)
4. Gong, S., Zhang, L., Zhuge, Y., Jia, X., Zhang, P., Lu, H.: Reinforcing video reasoning segmentation to think before it segments. arXiv preprint arXiv:2508.11538 (2025)
5. Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., et al.: Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025)
6. Han, S., Huang, W., Shi, H., Zhuo, L., Su, X., Zhang, S., Zhou, X., Qi, X., Liao, Y., Liu, S.: Videoespresso: A large-scale chain-of-thought dataset for fine-grained video reasoning via core frame selection. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 26181–26191 (2025)
7. Lai, X., Tian, Z., Chen, Y., Li, Y., Yuan, Y., Liu, S., Jia, J.: Lisa: Reasoning segmentation via large language model. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9579–9589 (2024)
8. Liu, F., Zhou, P., Baccei, S.J., Masciocchi, M.J., Amornsiripanitch, N., Kiefe, C.I., Rosen, M.P.: Qualifying certainty in radiology reports through deep learning–based natural language processing. American Journal of Neuroradiology **42**(10), 1755–1761 (2021)

9. Liu, Y., Peng, B., Zhong, Z., Yue, Z., Lu, F., Yu, B., Jia, J.: Seg-zero: Reasoning-chain guided segmentation via cognitive reinforcement. arXiv preprint arXiv:2503.06520 (2025)
10. Liu, Z., Sun, Z., Zang, Y., Dong, X., Cao, Y., Duan, H., Lin, D., Wang, J.: Visual-rft: Visual reinforcement fine-tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2034–2044 (2025)
11. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024)
12. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: Medsam2: Segment anything in 3d medical images and videos. arXiv preprint arXiv:2504.03600 (2025)
13. Marinov, Z., Jäger, P.F., Egger, J., Kleesiek, J., Stiefelhagen, R.: Deep interactive segmentation of medical images: A systematic review and taxonomy. IEEE transactions on pattern analysis and machine intelligence **46**(12), 10998–11018 (2024)
14. Nath, V., Li, W., Yang, D., Myronenko, A., Zheng, M., Lu, Y., Liu, Z., Yin, H., Law, Y.M., Tang, Y., et al.: Vila-m3: Enhancing vision-language models with medical expert knowledge. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 14788–14798 (2025)
15. Pons, E., Braun, L.M., Hunink, M.M., Kors, J.A.: Natural language processing in radiology: a systematic review. Radiology **279**(2), 329–343 (2016)
16. Shi, H., Han, S., Huang, S., Liao, Y., Li, G., Kong, X., Zhu, H., Wang, X., Liu, S.: Mask-enhanced segment anything model for tumor lesion semantic segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 403–413. Springer (2024)
17. Wei, F., Zhang, X., Zhang, A., Zhang, B., Chu, X.: Lenna: Language enhanced reasoning detection assistant. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2025)
18. Xu, H., Nie, Y., Wang, H., Chen, Y., Li, W., Ning, J., Liu, L., Wang, H., Zhu, L., Liu, J., et al.: Medground-r1: Advancing medical image grounding via spatial-semantic rewarded group relative policy optimization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 391–401. Springer (2025)
19. Yan, Z., Diao, M., Yang, Y., Jing, R., Xu, J., Zhang, K., Yang, L., Liu, Y., Liang, K., Ma, Z.: Medreasoner: Reinforcement learning drives reasoning grounding from clinical thought to pixel-level precision. arXiv preprint arXiv:2508.08177 (2025)
20. Zhang, M., Wu, X., Luo, H., Wang, F., Lv, Y.: Medground: Bridging the evidence gap in medical vision-language models with verified grounding data. arXiv preprint arXiv:2601.06847 (2026)
21. Zhao, T., Kiblawi, S., Usuyama, N., Lee, H.H., Preston, S., Poon, H., Wei, M.: Boltzmann attention sampling for image analysis with small objects. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 25950–25959 (2025)
22. Zhao, Z., Zhang, Y., Wu, C., Zhang, X., Zhou, X., Zhang, Y., Wang, Y., Xie, W.: Large-vocabulary segmentation for medical images with text prompts. NPJ Digital Medicine **8**(1), 566 (2025)