RAPO: Risk-aware Anatomical Prior Optimization via Reinforcement Learning for CAC Detection in Rheumatoid Arthritis

Jiashu Xu[1], Haipeng Zhou[1], Jianda Ma[2], Yaowei Zou[2], Lie Dai[2], and Lei Zhu[1,3] ✉

[1] The Hong Kong University of Science and Technology (Guangzhou), China
leizhu@hkust-gz.edu.cn
[2] Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, China
[3] The Hong Kong University of Science and Technology, Hong Kong SAR, China

Abstract. Rheumatoid Arthritis (RA) markedly increases the risk of cardiovascular disease (CVD), with coronary artery calcification (CAC) serving as a pivotal biomarker for early risk stratification. RA patients exhibit significantly elevated CAC-related mortality rates, making opportunistic CAC screening on routine chest CT scans highly valuable in clinical practice. However, to the best of our knowledge, automated CAC detection has not been systematically investigated in the RA cohorts, where calcifications often present as minute and low-contrast. To fill this critical gap, we construct the first high-quality CAC dataset for RA patients, comprising 415 cases with professional annotations. We propose Risk-aware Anatomical Prior Optimization (RAPO), a reinforcement learning framework that addresses the challenges of detecting minute and low-contrast lesions. RAPO incorporates a Chan-Vese Anatomical Prior that enforces geometric compactness to improve localization precision, and an Intrinsic Risk Penalty that modulates predictions based on uncertainty to reduce false negatives. Extensive experiments demonstrate that RAPO achieves superior performance, substantially outperforming state-of-the-art methods. The resources will be available at https://github.com/Jiashu-Xu/RAPO.

Keywords: Reinforcement Learning · Lesion Detection · Multimodal Large Language Model.

1 Introduction

Rheumatoid Arthritis (RA) is associated with a substantially increased risk of cardiovascular disease (CVD), which contributes significantly to excess mortality within this patient population [1]. Clinical data indicate that individuals with RA face a 50% to 70% higher risk of developing CVD and an approximate 50% increase in mortality from cardiovascular events compared to the general

✉ Lei Zhu (leizhu@hkust-gz.edu.cn) is the corresponding author.

public [5,21]. In the RA cohort, coronary artery calcification (CAC) serves as a pivotal imaging biomarker for early risk stratification. While the incidence of CAC in RA patients reaches up to 60% [18], detecting its early manifestations on routine non-contrast chest CTs—known as opportunistic screening—remains a formidable challenge. In RA cohorts, calcifications often present as particularly minute and low-contrast lesions in early stages due to chronic systemic inflammation [11]. This, combined with motion artifacts and limited spatial resolution of non-gated acquisitions, makes automated detection exceptionally challenging [6,9].

In recent years, Multimodal Large Language Models (MLLMs) have achieved remarkable success in the medical domain. Generalist models like Qwen-VL [2] and specialized adaptations like LLaVA-Med [13] have demonstrated exceptional proficiency in high-level semantic tasks, such as generating coherent radiology reports [23] and visual question answering [8]. However, directly deploying these models for fine-grained CAC detection reveals critical limitations inherent to their training paradigms. First, a severe Semantic-Spatial Gap persists: general MLLMs rely on Natural Image Priors learned from normal scenarios [7]. When transferred to the low-contrast medical domain, they fail to precisely delineate lesion boundaries, resulting in detection performance degradation. Moreover, current optimization strategies typically rely solely on Intersection over Union (IoU) [20] as the training objective. However, IoU is a purely geometric metric that does not account for prediction uncertainty. In screening scenarios, errors carry asymmetric clinical costs: a false negative can lead to missed diagnosis [28], yet standard models lack mechanisms to suppress such errors in ambiguous regions where uncertainty is high.

Reinforcement Learning (RL) has emerged as a robust alignment technique to bridge these gaps [16,24]. Advanced algorithms like Group Relative Policy Optimization (GRPO) [19] have proven effective in improving reasoning capabilities. Nevertheless, most existing RL approaches in medical imaging focus on report generation logic or utilize sparse rewards, lacking mechanisms to explicitly constrain geometric precision and uncertainty awareness [22,15].

To address these challenges, we propose **RAPO** (Risk-aware Anatomical Prior Optimization), a reinforcement learning framework designed for reliable opportunistic CAC detection. Departing from conventional "SFT-only" paradigms, RAPO introduces a coarse-to-fine curriculum. We initialize the model with Direct Preference Optimization (DPO) [17] to establish foundational discriminative capability, then introduce a novel **Risk-aware GRPO** mechanism that integrates two specialized components: an **Anatomical Prior** based on the Chan-Vese variational model to enforce geometric compactness, and a **Risk-aware Penalty** derived from intrinsic uncertainty to reduce false negatives in ambiguous regions.

Our contributions are summarized as follows:

- We propose **RAPO**, a reinforcement learning framework with a coarse-to-fine training strategy. Crucially, it employs a Risk-aware GRPO mechanism

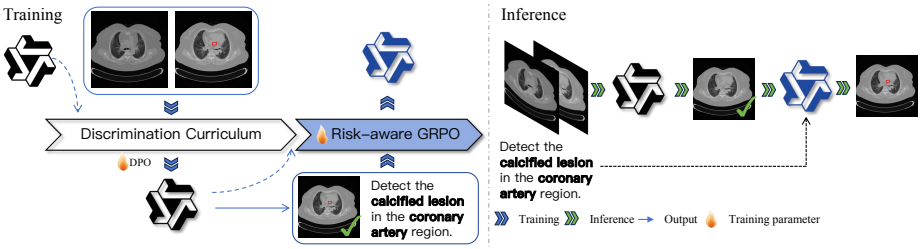

Fig. 1. Overview of the proposed RAPO framework. Left: Two-stage training with DPO-based discrimination and Risk-aware GRPO. Right: Inference pipeline for CAC localization.

with decoupled reward normalization to effectively balance geometric constraints and uncertainty-driven penalties.
– We construct the first high-quality CAC dataset specifically for RA patients, comprising 415 cases with professional annotations on routine chest CTs, establishing a novel benchmark for opportunistic screening in autoimmune populations.
– Extensive experiments demonstrate that RAPO establishes a new state-of-the-art, significantly outperforming general and medical MLLMs in detection performance.

2 Method

2.1 Overview

As shown in Fig 1, we formulate RAPO as a two-stage coarse-to-fine training curriculum for 2D slice-level CAC detection. Given a CT slice, the model predicts CAC presence or absence and generates bounding-box coordinates for positive slices, without introducing an additional multi-stage clinical inference workflow. We initialize the model with Direct Preference Optimization (DPO) to learn CAC presence/absence discrimination, using positive responses as preferred and negative responses as rejected for CAC-positive slices, and the reverse construction for CAC-negative slices. Building on this, we advance to Risk-aware GRPO for fine-grained localization optimization, as shown in Fig 2. In this phase, to resolve the gradient dominance arising from the heterogeneous scales of geometric and safety constraints, we propose a Decoupled Reward Normalization strategy. Without this normalization, the disparate magnitudes of rewards impede effective convergence. For instance, our localization reward R_{loc} is bounded in positive unit interval $[0, 1]$, whereas R_{risk} can take larger-magnitude negative values. Direct summation would cause the high-variance risk gradients to overwhelm the fine-grained geometric signals. Consequently, we compute the total advantage A_{total} by independently normalizing each reward component k:

$$A_{total}(y_i) = \sum_k \lambda_k \cdot \frac{R_k(y_i) - \mu_k}{\sigma_k + \epsilon} \ , \tag{1}$$

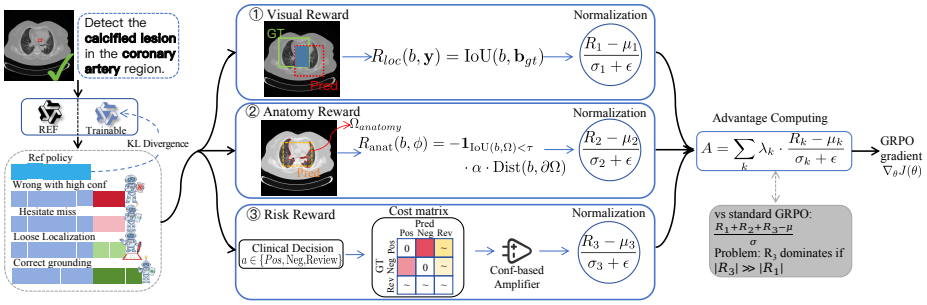

Fig. 2. Overview of the Risk-aware GRPO architecture. Left: Candidate generation from trainable policy. Middle: Three independently normalized rewards. Right: Advantage computing with decoupled normalization to prevent gradient dominance.

where y_i denotes the i-th sampled response from a group of candidates generated for the same input, and μ_k, σ_k are the mean and standard deviation computed across this group. This decoupled estimator ensures balanced optimization of the comprehensive objective $J(\theta)$:

$$J(\theta) = \mathbb{E}_{\mathbf{X}\sim\mathcal{D}, a\sim\pi_\theta}[R_{loc}(b, \mathbf{y}) + R_{risk}(a, c, \mathbf{y}) + R_{anatomy}(b, \phi)] \ , \tag{2}$$

where b, a, c, and ϕ denote the predicted bounding box, clinical action, intrinsic confidence, and variational anatomical prior, respectively.

2.2 IoU-Driven Localization

To ensure the model accurately localizes the calcification region, we employ the Intersection over Union (IoU) as the fundamental localization reward.

The model outputs a structured response containing: (1) a bounding box b, (2) a clinical decision $a \in \{\text{Positive, Negative, Review}\}$, and (3) an intrinsic confidence c derived from the policy logits. Given the ground truth bounding box $\mathbf{b}_{gt}$ annotated by radiologists, the localization reward is defined as:

$$R_{loc}(b, \mathbf{y}) = \text{IoU}(b, \mathbf{b}_{gt}) \ . \tag{3}$$

This term serves as the baseline supervision, encouraging the model to maximize the geometric overlap with the target region.

2.3 Risk-Aware Clinical Decision Mechanism via Intrinsic Uncertainty

While R_{loc} provides essential geometric supervision, relying solely on IoU is insufficient for robust CAC detection. Medical diagnosis differs from general object detection due to asymmetric error costs: in CAC screening, a high-confidence miss leads to irreversible missed diagnosis, whereas a deviated prediction can still trigger human review. IoU fails to distinguish between these clinically distinct failure modes. To address this limitation, we construct a decision mechanism modulated by Intrinsic Uncertainty to explicitly penalize high-risk errors.

Uncertainty-Modulated Asymmetric Risk High-confidence errors, particularly False Negatives in screening contexts, severely compromise clinical safety and trust. To mitigate this, we extract the clinical decision a directly from the model's structured text response and the Intrinsic Confidence $c \in [0, 1]$ from the policy logits. We formulate the risk-aware reward R_{risk} as:

$$R_{risk}(a, \mathbf{y}, c) = -(1 + \gamma \cdot c) \cdot \mathcal{C}_{asym}(a, \mathbf{y}) \ , \tag{4}$$

where $\mathcal{C}_{asym}$ encodes the pathology-specific cost matrix (e.g., penalizing False Negatives more heavily than False Positives). Mathematically, this formulation imposes heavier penalties on high-confidence errors while allowing the model to express uncertainty in ambiguous regions, thereby reducing the risk of committing false negatives.

2.4 Variational Anatomical Guidance via Active Contour Models

Constraining localization to physiologically plausible soft-tissue regions remains challenging in the absence of pixel-level supervision. Conventional gradient-based edge detection methods frequently fail in non-contrast CT due to the low contrast-to-noise ratio and ill-defined anatomical boundaries. To address this ill-posed problem, we formulate the region of interest (ROI) extraction as a variational energy minimization problem using the Chan-Vese Active Contour Model [3]. Unlike edge-based methods, this formulation leverages global image statistics, making it robust to noise and weak boundaries. The energy functional $E(c_1, c_2, \phi)$ is defined as:

$$\begin{aligned} E(c_1, c_2, \phi) = \mu \cdot \mathcal{L}(\phi) + \lambda_1 \int_{\Omega} |I(x) - c_1|^2 H(\phi(x)) dx \\ + \lambda_2 \int_{\Omega} |I(x) - c_2|^2 (1 - H(\phi(x))) dx \ , \end{aligned} \tag{5}$$

where $\mathcal{L}(\phi)$ represents the length regularization term, $H(\phi)$ is the Heaviside function, and c_1, c_2 denote the mean intensities inside and outside the contour, respectively. Minimizing this energy drives the zero-level set of ϕ to converge upon the mediastinum boundaries while maintaining topological continuity. The resultant anatomical mask $\Omega_{anatomy}$ serves as a variational prior. We translate this geometric constraint into a penalty reward $R_{anatomy}$:

$$R_{anatomy}(b, \phi) = -1_{\mathrm{IoU}(b, \Omega_{anatomy}) < \tau} \cdot \alpha \cdot \mathrm{Dist}(b, \partial\Omega_{anatomy}) \ , \tag{6}$$

where $-1_{(\cdot)}$ triggers the penalty if the overlap with $\Omega_{anatomy}$ is insufficient. $\mathrm{Dist}(\cdot)$ measures the Euclidean deviation from the box center to the nearest valid boundary, and α is a scaling factor. Since the Chan-Vese evolution is computationally intensive and acts as a geometric constraint rather than a learnable objective, we implement it as a non-differentiable environmental signal. To minimize computational overhead, the anatomical mask $\Omega_{anatomy}$ is computed using 30 morphological iterations and cached for each image, ensuring zero latency

during the RL training loop. This variational prior is particularly valuable for opportunistic screening, where lesions may be confused with rib calcification or vascular structures.

3 Experiment

3.1 Dataset

We constructed a high-quality proprietary CAC dataset collected from Sun Yat-Sen Memorial Hospital with ethical approval, comprising 415 patients, 5,821 CAC-positive slices, with 13.7 lesions per case on average. Each sample consists of a non-contrast chest CT scan and bounding boxes. The bounding boxes were annotated by one radiologist and verified by a senior radiologist. The data is randomly split into training, validation, and test sets with a ratio of 7:1:2 at the patient level to prevent data leakage.

3.2 Implementation Details

Our framework is built upon the Qwen-2.5-VL-7B [2]. The model was trained using the AdamW optimizer with an initial learning rate of 1e-6, decayed via a cosine schedule. We train DPO for 1000 steps, then GRPO for 2000 steps. We generated $K = 4$ candidates following previous work [24]. For the Chan-Vese functional, we use $\lambda_1 = \lambda_2 = 1.0$ and smoothing $\mu = 0.2$ [3]. The cost matrix is configured with $C_{FN} = 1.5$ and $C_{FP} = 0.7$. The intrinsic confidence c is derived from the average per-token log-probability then normalized. To ensure a rigorous and equitable comparison, all baseline models were subjected to task-specific supervised fine-tuning on our training dataset. All experiments were conducted on 4 × NVIDIA A6000 GPUs (48GB) with a batch size of 4.

3.3 Comparison with State-of-the-Art Methods

We compared our proposed framework with four categories of state-of-the-art (SOTA) methods: 1) **Conventional Detectors**: Pure vision-based object detection models, including YOLO [10], RetinaNet [14] and DINO [26]; 2) **General MLLMs**: Large-scale pre-trained models, including Qwen2.5-VL-7B [2], LLaVA-Onevision-7B [12], and InternVL3-8B [27]; 3) **Medical MLLMs**: Large-scale pre-trained models (including Lingshu-7B [25] and Huatuo-7B [4]) specifically tuned for the clinical domain; 4) **Task-Specific Models**: Models specialized for detection, including MedGround-r1 [24] and VLM-R1 [20]. We adopt four commonly used metrics for quantitative comparisons, and they are mean Intersection over Union (mIoU), Accuracy (Acc), F1-score, and Recall.

Quantitative Comparisons. Tab. 1 summarizes the quantitative results on the RA-CAC dataset. Conventional detectors exhibit a stark trade-off: YOLO and RetinaNet achieve a competitive mIoU but suffer from extremely low Acc

Table 1. Quantitative comparisons between our method and state-of-the-art models on our dataset. We **Bold** the best ones and underline the second best.

Category	**Method**	mIoU (↑)	ACC (↑)	F1 (↑)	Recall (↑)
Conventional	RetinaNet [14]	0.5285	0.1436	0.2511	0.1492
	YOLO [10]	0.5512	0.2431	0.4035	**0.8012**
	DINO [26]	0.4031	0.5832	0.6708	0.6673
General MLLMs	Qwen2.5-VL-7B [2]	0.2315	0.1651	0.1655	0.1659
	LLaVA-Onevision-7B [12]	0.3549	0.2495	0.2310	0.2150
	InternVL3-8B [27]	0.4485	0.4609	0.4266	0.3971
Medical MLLMs	Lingshu-7B [25]	0.5459	0.4646	0.4358	0.4796
	Huatuo-7B [4]	0.4708	0.3498	0.3281	0.3090
Specialized Methods	MedGround-r1 [24]	0.4940	0.4938	0.4950	0.4938
	VLM-R1 [20]	0.5011	0.5136	0.5140	0.5136
Ours	−−	**0.6042**	**0.7660**	**0.7311**	0.6995

and F1, whereas DINO shows the exact opposite trend. Notably, YOLO achieves the highest Recall, but its low Acc and F1 indicate excessive false positives. MLLMs are expected to mitigate these extreme trade-offs and provide a more balanced performance due to their reasoning capabilities, but their overall performance remains suboptimal, lacking the strict geometric constraints needed for minute calcifications. In contrast, our RAPO successfully bridges this gap, achieving the best mIoU of 0.6042, Accuracy of 0.7660, and F1 of 0.7311, while maintaining the second-best Recall of 0.6995. Specifically, RAPO delivers substantial relative improvements of +9.6% in mIoU, +31.4% in Accuracy, and +9.0% in F1 score over the best baselines. This proves its clinical effectiveness for opportunistic screening, validating the integration of variational anatomical priors and intrinsic risk penalties.

Visual comparisons. Fig. 3 presents a qualitative comparison. General MLLMs, for example Qwen2.5-VL-7B in the second row, exhibit severe instability, often hallucinating targets in unrelated regions due to the domain gap. Medical MLLMs, while correctly focusing on the cardiac area, suffer from loose localization and spatial drift, lacking box-level precision. In contrast, RAPO generates compact bounding boxes that tightly adhere to the ground truth. This confirms that our Chan-Vese variational prior effectively refines geometric precision, ensuring the boundaries are anatomically plausible, while the Risk-aware mechanism successfully suppresses the false positives observed in baseline methods.

3.4 Ablation Study

To disentangle the specific contributions of each component, we defined six experimental configurations: 1) Baseline (ID 1) employs Qwen2.5-VL fine-tuned via standard SFT; 2) +DPO (ID 2) applies DPO prior to SFT to discriminate

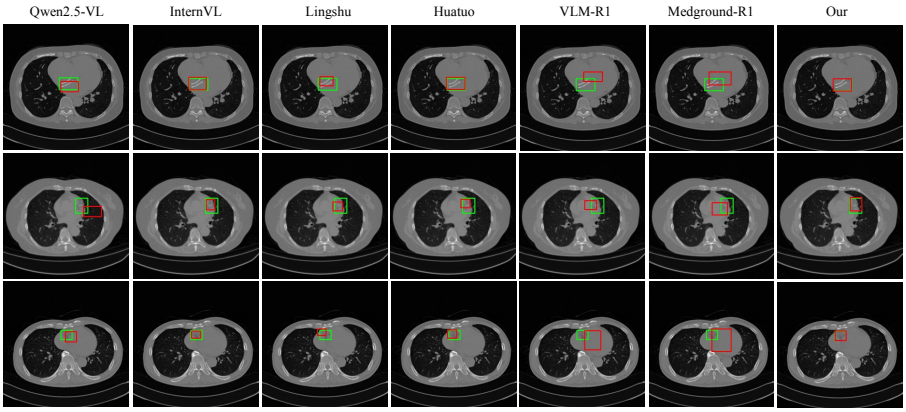

Fig. 3. Visual comparisons between our method and SOTA methods, where the green box denotes the ground truth and red box represents the prediction.

Table 2. Ablation study of the proposed RAPO framework.

ID	Training Stage		Reward		Performance			
	DPO	GRPO	R_{anat}	R_{risk}	mIoU (↑)	ACC (↑)	F1 (↑)	Recall (↑)
1	-	-	-	-	0.2315	0.1651	0.1655	0.1659
2	✓	-	-	-	0.4388	0.2600	0.2497	0.2401
3	✓	✓	-	-	0.5084	0.5745	0.5987	0.6250
4	✓	✓	✓	-	0.5825	0.5350	0.5179	0.5018
5	✓	✓	-	✓	0.5682	0.6954	0.6926	0.7100
6	✓	✓	✓	✓	**0.6042**	**0.7660**	**0.7311**	**0.6995**

CAC presence/absence; 3) +GRPO Base (ID 3) introduces reinforcement learning after DPO using solely the standard IoU reward (R_{loc}); 4) +Anatomy (ID 4) incorporates the Anatomical Prior (R_{anat}) for geometric constraints; 5) +Risk (ID 5) adds Risk Awareness (R_{risk}) for reducing false negatives; and 6) RAPO (ID 6) represents the complete framework.

Tab. 2 validates our design. DPO initialization (ID 2) significantly boosts mIoU over the SFT baseline (ID 1), confirming that establishing discriminative capability is beneficial for effective localization. The subsequent transition to GRPO (ID 3) marks a shift from imitation to reasoning, driving a substantial leap in Accuracy and Recall, which verifies the necessity of RL exploration for handling low-contrast lesions. Regarding reward shaping, the Anatomical Prior (ID 4) enhances geometric compactness but can be conservative for low-contrast or tiny CAC, reducing Recall and Acc when used alone. In contrast, Risk Awareness (ID 5) prioritizes reducing missed detections, achieving the peak Recall by penalizing confident false negatives, although localization quality is lower than the full RAPO. Finally, RAPO (ID 6) achieves the optimal synergy, combining anatomical constraints with uncertainty-driven risk awareness to secure state-of-the-art results across all metrics. Paired bootstrap tests between RAPO and each ablated variant (IDs 1–5) were all significant ($p < 0.001$).

4 Conclusion

We present RAPO, a reinforcement learning framework for reliable opportunistic CAC detection in RA patients. We contribute the first detection dataset for Rheumatoid Arthritis cohorts. To address the semantic-spatial gap and the limitations of IoU-only optimization, we introduce two key innovations: (1) DPO initialization to establish discriminative capability, and (2) a Risk-aware GRPO mechanism leveraging decoupled reward normalization to effectively balance localization fidelity (R_{loc}), anatomical priors (R_{anat}), and intrinsic risk penalties (R_{risk}). Extensive experiments on our RA CAC dataset demonstrate the superiority of our method, achieving state-of-the-art detection performance. Ablation studies further validate the effectiveness of each component.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Project No.82572383), the Guangdong Science and Technology Department (2024ZDZX2004), and the Program of Guangdong Education Department.

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

References

1. Aronov, A.G., Kim, Y.J., Zahid, S., Michos, E.D., Nazir, N.T.: Cardiovascular disease risk factors in newly diagnosed rheumatoid arthritis: A retrospective cohort study. American Journal of Preventive Cardiology p. 101002 (2025)
2. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., et al.: Qwen2.5-vl technical report. arXiv preprint arXiv:2502.13923 (2025)
3. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing **10**(2), 266–277 (2001). https://doi.org/10.1109/83.902291
4. Chen, J., Cai, Z., Ji, K., Wang, X., Liu, W., Wang, R., Hou, J., Wang, B.: Huatuogpt-o1, towards medical complex reasoning with llms (2024), https://arxiv.org/abs/2412.18925
5. Chung, C.P., Oeser, A., Raggi, P., Gebretsadik, T., Shintani, A.K., Sokka, T., Pincus, T., Avalos, I., Stein, C.M.: Increased coronary-artery atherosclerosis in rheumatoid arthritis: relationship to disease duration and cardiovascular risk factors. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology **52**(10), 3045–3053 (2005)
6. Foraker, R., Sperling, L., Bratzke, L., Budoff, M., Leppert, M., Razavi, A.C., Rodriguez, F., Shapiro, M.D., Whelton, S., Wong, N.D., et al.: Opportunistic detection of coronary artery calcium on noncardiac chest computed tomography: an emerging tool for cardiovascular disease prevention: a scientific statement from the american heart association. Circulation **152**(19), e391–e401 (2025)
7. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
8. Hao, P., Wang, H., Li, S., Xing, Z., Yang, G., Wu, K., Zhu, L.: Surgical-mamballm: Mamba2-enhanced multimodal large language model for vqla in robotic surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 573–583. Springer (2025)

9. Hecht, H.S., Cronin, P., Blaha, M.J., Budoff, M.J., Kazerooni, E.A., Narula, J., Yankelevitz, D., Abbara, S.: 2016 scct/str guidelines for coronary artery calcium scoring of noncontrast noncardiac chest ct scans: a report of the society of cardiovascular computed tomography and society of thoracic radiology. Journal of cardiovascular computed tomography **11**(1), 74–84 (2017)
10. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics yolov8 (2023), https://github.com/ultralytics/ultralytics
11. Karpouzas, G.A., Malpeso, J., Choi, T.Y., Li, D., Munoz, S., Budoff, M.J.: Prevalence, extent and composition of coronary plaque in patients with rheumatoid arthritis without symptoms or prior diagnosis of coronary artery disease. Annals of the Rheumatic Diseases **73**(10), 1797–1804 (2014)
12. Li, B., Zhang, Y., Guo, D., Zhang, R., Li, F., Zhang, H., Zhang, K., Zhang, P., Li, Y., Liu, Z., et al.: Llava-onevision: Easy visual task transfer. arXiv preprint arXiv:2408.03326 (2024)
13. 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 Neural Information Processing Systems **36**, 28541–28564 (2023)
14. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (Oct 2017)
15. Liu, Q., Zhang, S., Qin, G., Gu, Y., Jin, Y., Preston, S., Xu, Y., Kiblawi, S., Yim, W.w., Ossowski, T., et al.: Scaling medical imaging report generation with multimodal reinforcement learning. arXiv preprint arXiv:2601.17151 (2026)
16. 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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 337–347. Springer (2025)
17. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S., Finn, C.: Direct preference optimization: Your language model is secretly a reward model. Advances in neural information processing systems **36**, 53728–53741 (2023)
18. Semb, A.G., Ikdahl, E., Wibetoe, G., Crowson, C., Rollefstad, S.: Atherosclerotic cardiovascular disease prevention in rheumatoid arthritis. Nature Reviews Rheumatology **16**(7), 361–379 (2020)
19. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
20. Shen, H., Liu, P., Li, J., Fang, C., Ma, Y., Liao, J., Shen, Q., Zhang, Z., Zhao, K., Zhang, Q., et al.: Vlm-r1: A stable and generalizable r1-style large vision-language model. arXiv preprint arXiv:2504.07615 (2025)
21. Vrints, C., Andreotti, F., Koskinas, K.C., Rossello, X., Adamo, M., Ainslie, J., Banning, A.P., Budaj, A., Buechel, R.R., Chiariello, G.A., et al.: 2024 esc guidelines for the management of chronic coronary syndromes: developed by the task force for the management of chronic coronary syndromes of the european society of cardiology (esc) endorsed by the european association for cardio-thoracic surgery (eacts). European heart journal **45**(36), 3415–3537 (2024)
22. Wang, P., Ye, S., Naseem, U., Kim, J.: Mrg-r1: Reinforcement learning for clinically aligned medical report generation. arXiv preprint arXiv:2512.16145 (2025)

23. Wang, Y., Sun, Y., Tan, T., Hao, C., Cui, Y., Su, X., Xie, W., Shen, L., Yu, Z.: TRRG: Towards Truthful Radiology Report Generation With Cross-modal Disease Clue Enhanced Large Language Models . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15966. Springer Nature Switzerland (September 2025)
24. 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)
25. 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)
26. Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L.M., Shum, H.Y.: Dino: Detr with improved denoising anchor boxes for end-to-end object detection (2022)
27. Zhu, J., Wang, W., Chen, Z., Liu, Z., Ye, S., Gu, L., Tian, H., Duan, Y., Su, W., Shao, J., et al.: Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. arXiv preprint arXiv:2504.10479 (2025)
28. Zhu, Z., Zhang, Y., Zhuang, X., Zhang, F., Wan, Z., Chen, Y., QingqingLong, Q., Zheng, Y., Wu, X.: Can we trust ai doctors? a survey of medical hallucination in large language and large vision-language models. In: Findings of the Association for Computational Linguistics: ACL 2025. pp. 6748–6769 (2025)