

OralGPT-Plus: Learning to Use Visual Tools via Reinforcement Learning for Panoramic X-ray Analysis

Yuxuan Fan^{1*} Jing Hao^{2*♣} Hong Chen¹ Jiahao Bao⁴ Yihua Shao⁵

Yuci Liang⁶ Kuo Feng Hung² Hao Tang^{3†}

¹The Hong Kong University of Science and Technology (GZ)

²Faculty of Dentistry, The University of Hong Kong

³School of Computer Science, Peking University

⁴Shanghai Jiao Tong University

⁵Institute of Automation, Chinese Academy of Sciences

⁶College of Computer Science and Software Engineering, Shenzhen University

Abstract

Panoramic dental radiographs require fine-grained spatial reasoning, bilateral symmetry understanding, and multi-step diagnostic verification, yet existing vision-language models operate under a static single-pass paradigm that limits their clinical reliability. In this paper, we introduce OralGPT-Plus, an agentic vision-language model designed to perform iterative and symmetry-aware diagnostic reasoning for panoramic dental radiograph analysis. To support this paradigm, we construct DentalProbe, a five-thousand-image dataset with expert-curated diagnostic trajectories that provide structured supervision for localized inspection and contralateral comparison. We further develop a Reinspection-driven reinforcement learning framework that encourages clinically meaningful re-examination and stabilizes long-horizon reasoning with rubric-based reward and conditioned diagnostic-driven reward. In parallel, we present MMOral-X, the first benchmark for holistic panoramic diagnosis, containing 300 open-ended questions and region-level annotations across multiple difficulty levels. OralGPT-Plus demonstrates consistent and reliable improvements over strong baselines on MMOral-X and established panoramic benchmarks, indicating the effectiveness of interactive and symmetry-informed reasoning. Our work highlights the value of agentic modeling for dental imaging and provides a foundation for future research in clinically aligned panoramic radiograph analysis. Codes will be published in <https://github.com/isbrycee/OralGPT>.

*Equal contribution.

†Corresponding authors: haotang@pku.edu.cn

♣ Project Leader.

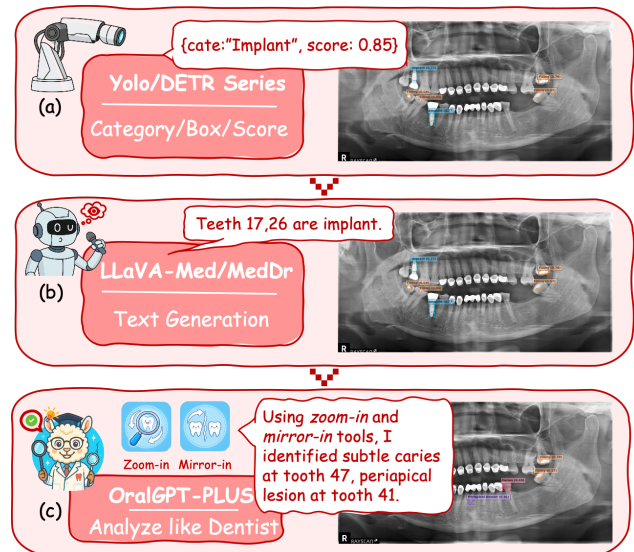


Figure 1. **Paradigm shift in panoramic dental radiograph analysis.** Traditional detectors provide only category-level boxes, while VLMs generate coarse descriptions without structured reasoning. In contrast, OralGPT-Plus performs dentist-like diagnostic reasoning by invoking zoom-in and mirror-in tools to identify subtle findings and produce clinically coherent interpretations.

1. Introduction

Panoramic radiograph is universally recognized as a fundamental tool for oral disease diagnosis and treatment planning, as it provides a comprehensive view of the teeth, alveolar bone, and surrounding anatomical structures within a single image [10, 11, 17, 37]. An exhaustive analysis of panoramic radiographs is indispensable for both screening and therapeutic decision-making [25]. However, interpreting panoramic radiographs is extremely time-consuming and often relies heavily on the experience of

radiologists [10, 11, 17]. The computer vision techniques have promoted the advancement of intelligent dentistry [13, 14]. Figures 1(a) and Figures 1(b) show two widely used anomaly detection paradigms. Traditional object detectors [9, 16] focus on localizing suspected abnormalities and return only categories, bounding boxes, and confidence scores, limiting their ability to support clinically meaningful explanations. Vision-language models (VLMs) [3, 12, 14, 15, 18, 22, 33] have improved semantic expressiveness; yet, the single forward prediction paradigm restricts their ability to revisit ambiguous regions and capture subtle disease patterns. Unlike other dental imaging modalities, panoramic radiographs have a higher image resolution, making dentists frequently magnify areas suspected of containing abnormalities for closer examination. Additionally, they often refer to analogous teeth in other quadrants due to dental symmetry. By comparing corresponding teeth in symmetrical quadrants, they can perceive and extract subtle features, thus identifying minute disease patterns such as carious lesions, apical periodontitis, and others [6]. Inspired by this clinical workflow, we propose OralGPT-Plus, an agentic model that performs diagnostic reasoning through an iterative cycle of thought, action, and observation. The model is equipped with a focused inspection operator that produces localized high-resolution views of suspicious areas, as well as a symmetry-aware operator that retrieves the corresponding anatomical region on the opposite side. These abilities enable the model to refine ambiguous regions, perform bilateral comparisons when needed, and progressively form coherent diagnostic conclusions. This iterative reasoning mirrors core elements of clinical workflows rather than producing a static summary.

Developing such a model requires supervision that reflects real diagnostic behavior. For this purpose, we curate DentalProbe, a dataset that includes more than eight thousand rounds of expert-guided diagnostic trajectories. The trajectories capture the sequence of global inspection, proposal identification, localized examination, and symmetry-based verification that characterizes panoramic radiograph interpretation. Figure 2 illustrates the curation pipeline, which integrates multiple public datasets with expert annotations and multi-agent trajectory refinement. This corpus enables instruction tuning that teaches the model to follow clinically consistent reasoning patterns and to use the diagnostic tools appropriately.

Although instruction tuning provides a foundation for structured behavior, effective multi-turn analysis benefits from further optimization. We therefore introduce a reinforcement learning framework that encourages the model to re-examine regions when necessary and limit unproductive exploratory actions. The reward design offers continuous feedback that reflects diagnostic completeness and accuracy, which alleviates the sparsity that commonly arises

in multi-lesion scenarios. It also regulates the initiation of additional inspections so that the model focuses on clinically meaningful reasoning. These mechanisms contribute to stable long-horizon optimization and support balanced and reliable decision-making.

To evaluate the diagnostic performance in a comprehensive manner, we construct MMOral-X, a benchmark that contains 300 open-ended question-answer pairs covering clinically relevant diseases and findings. The benchmark includes three levels of difficulty that reflect the number, variety, and visual ambiguity of potential abnormalities. It emphasizes holistic analysis from a single panoramic radiograph and enables systematic assessment of multi-turn reasoning, fine-grained lesion interpretation, and symmetry-aware analysis.

Across all components of the framework, OralGPT-Plus pursues interpretability, stability, and alignment with clinical expectations. The experimental results show consistent improvements over strong baselines on MMOral-X, as well as on one established public benchmark. The gains are steady rather than abrupt and reflect the value of interactive and symmetry-aware reasoning strategies in panoramic dental radiograph analysis.

Our contributions are summarized as follows:

- We introduce OralGPT-Plus, an agentic and symmetry-aware model designed to perform iterative diagnostic reasoning on panoramic dental radiographs.
- We curate DentalProbe, a five-thousand image dataset with expert diagnostic trajectories that support clinically aligned instruction tuning and structured tool usage.
- We present MMOral-X, a benchmark for holistic panoramic radiograph interpretation that evaluates holistic understanding across varying levels of complexity.
- Extensive evaluations show that OralGPT-Plus achieves strong and consistent performance, indicating that interactive and symmetry-informed reasoning substantially improves diagnostic quality.

2. Preliminary

Fig. 1 shows the progression of panoramic dental radiograph analysis from object detectors to single-pass VLMs and finally to agentic VLMs. Our work is the first to adopt agentic VLMs for dental panoramic analysis by introducing diagnostic tool reasoning and interactive decision-making.

Paradigm 1: Traditional Object Detection. Early systems relied on detectors such as YOLO [30] and DETR [4] to identify lesions through bounding-box predictions:

$$\mathcal{D}_{\text{obj}}(I_0) = \{(b_i, c_i, s_i)\}_{i=1}^N, \quad b_i \in [0, 1]^4. \quad (1)$$

Although these models can localize lesions accurately, their outputs are limited to box coordinates and confidence scores, leaving the diagnostic rationale unexplained and restricting the system to perception-driven inference.

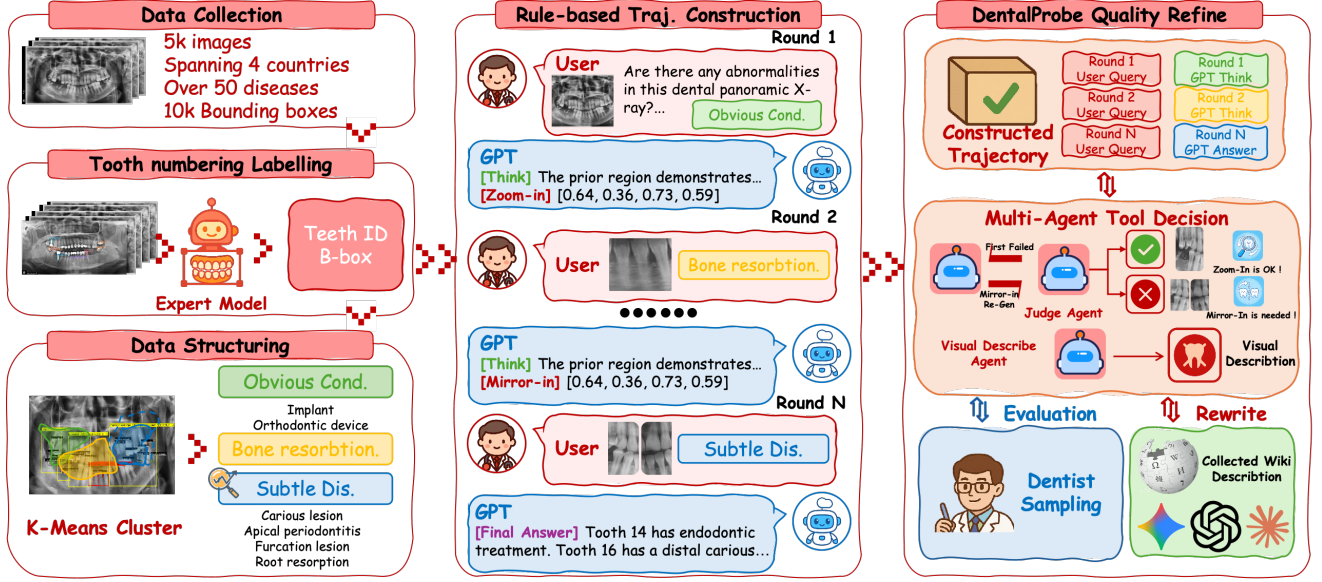


Figure 2. Curation process of DentalProbe dataset with expert trajectory for dentist-like instruction tuning.

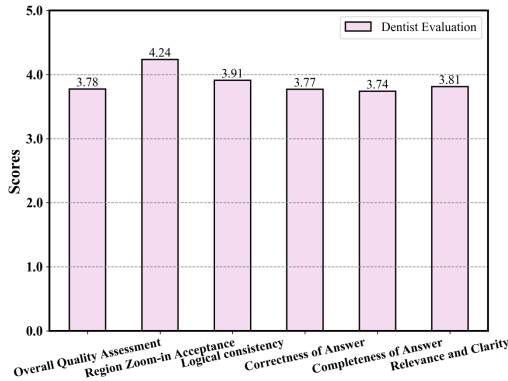


Figure 3. Average score under dentist sampling evaluation on curated DentalProbe dataset.

Paradigm 2: Single Forward VLMs. Single-pass VLMs, represented by LLaVA [24], generate responses directly from an image I_0 and a query q :

$$y = \mathcal{G}_{\text{vlm}}(I_0, q) = \text{Dec}(f_{\theta}(I_0), q). \quad (2)$$

Here, $f_{\theta}(I_0)$ extracts visual features, and $\text{Dec}(\cdot)$ produces the diagnostic text. This paradigm provides semantic interpretation but remains static, as the model responds in one pass without revisiting ambiguous regions.

Paradigm 3: Agentic VLMs. Our OralGPT-Plus introduces an iterative thought-action-observation loop that emulates clinical diagnostic procedures. At each step i , the policy produces a thought T_i , selects an action $\mathcal{A}_i$, and obtains a corresponding observation O_i :

$$\begin{aligned} T_i &= \pi_{\text{thought}}(q, I_0, \mathcal{H}_{i-1}), \\ \mathcal{A}_i &= \pi_{\text{action}}(T_i), \\ O_i &= \text{Env}(\mathcal{A}_i; I_0), \\ \mathcal{H}_i &= \mathcal{H}_{i-1} \cup \{T_i, \mathcal{A}_i, O_i\}. \end{aligned} \quad (3)$$

The environment $\text{Env}(\cdot)$ executes image actions such as “Zoom-In” for close examination. The process ends when the policy issues a Finalize action or reaches K steps, producing the final report:

$$y^* = \pi_{\text{answer}}(q, I_0, \mathcal{H}_K). \quad (4)$$

Through iterative observation and tool-guided inspection, the model reassesses uncertainty, performs bilateral comparisons, and derives clinically interpretable conclusions.

The transition from detectors to agentic VLMs reflects a shift from static perception to interactive reasoning. Traditional detectors localize abnormalities, single-pass VLMs describe them, and agentic VLMs analyze and verify through tool-guided, multi-step exploration. This progression enhances interpretability and strengthens clinical alignment in panoramic dental radiograph analysis.

3. OralGPT-Plus

In this section, we introduce OralGPT-Plus, which emulates dentists’ diagnostic reasoning on panoramic X-rays via tool-augmented learning. We first present Dentist-like Instruction Tuning 3.1, where the model learns core diagnostic behaviors from curated expert trajectories in the DentalProbe dataset. We then describe Reinspection-Driven Reinforcement Learning 3.2, which incorporates rubric-based continuous rewards and conditioned diagnostic incentives to promote clinically meaningful reinspection. These components are unified by a Hybrid Reward System that stabilizes long-horizon optimization. Together, they enable OralGPT-Plus to follow a dentist-like workflow and progressively enhance diagnostic competence.

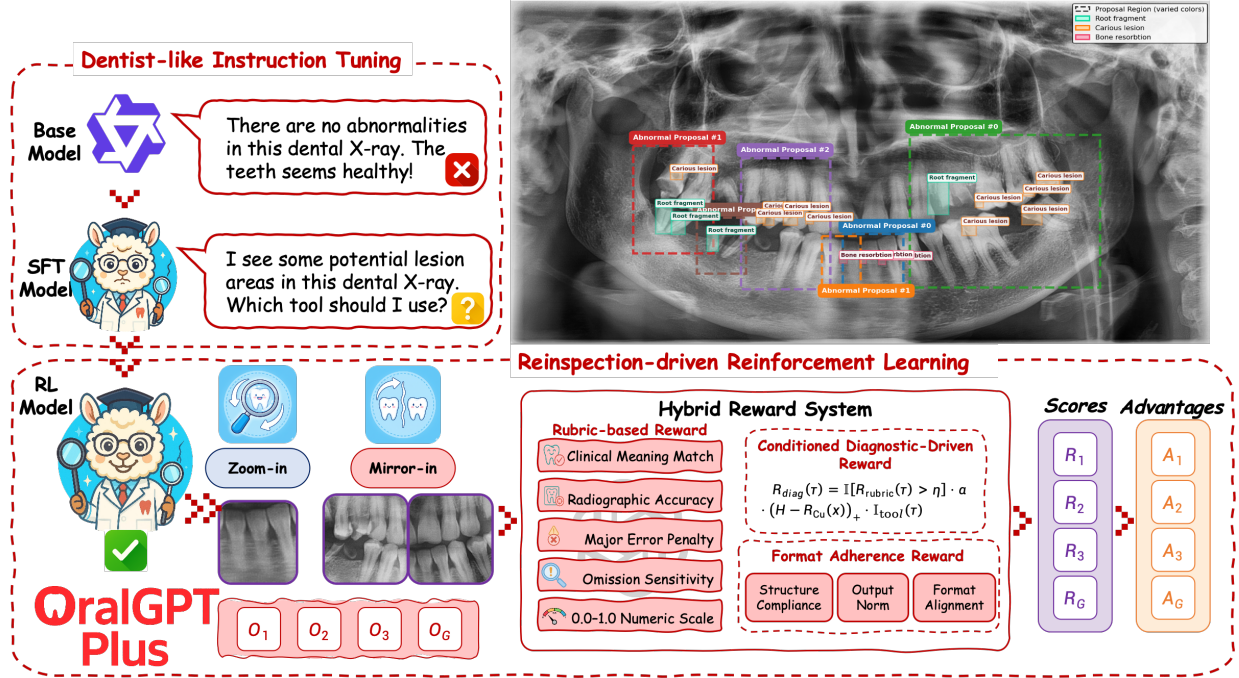


Figure 4. Our training pipeline from dentist-like instruction tuning to re-inspection reinforcement learning.

3.1. Dentist-like Instruction Tuning

Our goal is to enable VLMs to emulate dentists’ diagnostic reasoning on dental X-ray images, especially in subtle lesion regions. The model uses diagnostic tools to locate, zoom in, and compare suspicious areas, improving panoramic X-ray analysis. However, as shown in Fig. 6, models such as Qwen2.5-VL fail to develop tool-based reasoning through prompt engineering alone, consistent with prior observations [36]. This limitation arises from the lack of tool-interaction data during pretraining. To address this, Sec. 3.1.1 introduces a precise diagnostic data construction process that produces high-quality expert trajectories (Sec. 3.1.2, Sec. 3.1.3).

3.1.1. Data Statistics

We integrate four public datasets [8, 26, 29, 44] for abnormality detection in panoramic radiographs, each offering region-level annotations of dental abnormalities. These datasets cover over 50 disease categories across geographic regions, including the United States, China, Switzerland, and Romania, which enhances representativeness and mitigates sampling bias. In addition, we select 2,500 high-quality radiographs from the MMOral-OPG [14] dataset and combine them with 2,562 collected samples to construct the 5k-image DentalProbe dataset.

3.1.2. Dental-Aware Tool Design

Previous agentic VLMs rely primarily on the “Zoom-In” operation to magnify suspicious regions for detailed inspection. However, panoramic dental X-rays exhibit strong

anatomical symmetry, and a purely local enhancement strategy often overlooks this intrinsic characteristic. In clinical practice, dentists routinely perform contralateral comparisons by examining the corresponding tooth or region on the opposite side, which allows them to determine whether subtle shadows or irregularities indicate true lesions.

Motivated by this diagnostic behavior, we introduce the “Mirror-In” tool, which converts oral symmetry from an implicit prior into an explicit model action. After the model identifies a potential lesion, the tool retrieves its horizontally mirrored counterpart across the midline, thereby forming a dual-view pair for comparative reasoning. Formally, given an image $I(x, y)$ of width W and a selected region $[x_1, x_2] \times [y_1, y_2]$, the symmetric view is defined as

$$I_{\text{mirror}}(x, y) = I(W - x, y), \quad (x, y) \in [x_1, x_2] \times [y_1, y_2]. \quad (5)$$

In implementation, we use $1.5 \times$ padded original/mirrored crops to absorb minor misalignment and asymmetry. This explicit symmetry-based inspection enhances the model’s spatial understanding and facilitates the verification of subtle or low-contrast abnormalities through contralateral reference. By embedding clinical comparison habits into the reasoning loop, the agent produces more stable and interpretable diagnostic judgments.

3.1.3. DentalProbe Trajectory Construction

To construct dentist-like diagnostic trajectories, we design a multi-stage pipeline that emulates clinical reasoning on panoramic X-rays as shown in Fig. 2. For each OPG, a

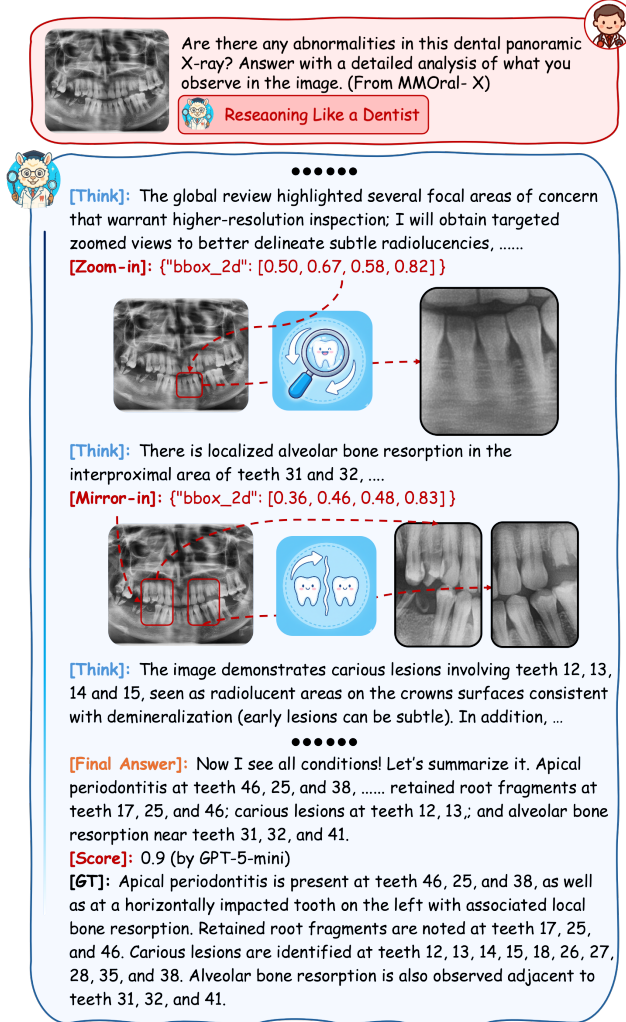


Figure 5. **Diagnostic trajectory generated by OralGPT-Plus on a sample of MMOral-X.**

tooth-number detection model localizes tooth boxes and aligns them with lesion labels. Lesions are categorized into obvious, subtle, and bone-based findings. For subtle and bone-based cases, we apply k -means clustering to tooth-level boxes to generate region proposals that highlight diagnostically challenging areas requiring focused inspection.

We construct multi-turn diagnostic trajectories using explicit rules derived from the proposals. The initial query performs global inspection, followed by turns that progressively incorporate zoomed-in regions and contextual cues to mirror dentists’ stepwise reasoning. After generating these trajectories, we apply a dedicated quality refinement stage. A multi-agent tool decision module evaluates each step and verifies whether the invoked tool is appropriate. If the judge agent deems the current view sufficient, it validates “Zoom-In”; otherwise, it triggers “Mirror-In” for symmetric comparison to support more reliable diagnostics. Next, a visual description agent generates region-level summaries for

all tool-invoked images, which are combined with curated clinical definitions to rewrite each reasoning step. To enhance linguistic diversity and reduce stylistic bias, we use three models [1, 27, 35] for independent rewrites and select the best version. This process produces trajectories that are clinically aligned, stylistically diverse, and instruction-ready. Finally, expert dentists conduct sampling-based evaluation on the DentalProbe dataset, with results shown in Fig. 3. Their evaluations and iterative refinement further improve data quality to a high standard.

3.1.4. Training Strategy

We apply full-parameter SFT to the language module while freezing the vision encoder and projector, ensuring stable visual features and focusing learning on grounded reasoning and correct tool use. Given multimodal input (x, I) and the curated trajectory y^* , the training objective is:

$$\mathcal{L}_{\text{SFT}} = - \sum_{t=1}^T \log \pi_{\theta}(y_t^* | x, I, y_{<t}^*). \quad (6)$$

3.2. Reinspection-Driven Reinforcement Learning

After SFT, the model can invoke basic tools, but naive GRPO [32, 34] remains unreliable for clinical diagnosis. Panoramic radiographs contain multiple spatially distributed lesions, while binary 0,1 rewards offer all-or-nothing feedback, causing sparse signals, weak advantages, and unstable learning. Uncontrolled exploration further drives the agent toward clinically irrelevant regions.

To mitigate reward sparsity and align exploration with diagnostic reasoning, we introduce three complementary components. The rubrics-based reward provides continuous, case-aware supervision; the Conditioned Diagnostic Driven Reward activates exploration only under sufficient rubric confidence; and a unified hybrid reward integrates these signals to ensure stable and interpretable RL.

3.2.1. Rubrics-based Reward

To address reward sparsity in multi-lesion scenarios, we introduce a rubric-based dental reward that provides continuous and informative supervision. Instead of using a binary $\{0, 1\}$ score, we employ a few-shot rubric evaluator aligned with MMOral-X criteria to assess clinical meaning, radiographic accuracy, major error severity, and omission sensitivity. This approach ensures that partially correct predictions, such as identifying some but not all lesions, receive proportional rewards rather than collapsing to zero, thereby enabling stable and effective gradient updates.

Formally, the rubric evaluator (GPT-5-mini) assigns each trajectory τ a continuous score $R_{\text{rubrics}}(\tau) \in [0, 1]$. The scoring follows explicit clinical rules: synonymous diagnostic phrasing is accepted, clinically consistent tooth-indexing

Table 1. Model performance on **MMOral-X** and open-ended VQA for **MMOral-OPG**. Results on MMOral-X are averaged over five runs, whereas MMOral-OPG is evaluated once. In each column, the best score is **bold** and the second best is underlined.

Model	Params	MMOral-X			MMOral-OPG Open-ended VQA						
		Simple	Moderate	Complex	Teeth	Patho	His	Jaw	Summ	Report	Overall
Proprietary VLMs (API)											
GPT-5 [27]	N/A	32.90	8.30	9.30	<u>39.84</u>	29.42	44.08	77.38	<u>40.06</u>	28.16	<u>42.34</u>
Gemini-2.5-Flash [35]	N/A	25.56	9.70	8.70	28.21	24.64	31.95	47.74	13.05	16.73	27.89
Doubao-seed-1-6 ^a [7]	N/A	20.10	5.80	5.20	27.19	18.18	41.72	59.44	22.38	<u>40.20</u>	34.20
Claude-sonnet-4.5 ^b [1]	N/A	27.72	8.14	9.78	28.29	20.45	48.83	71.56	26.67	37.70	37.68
Qwen3-VL-235b-a22b-thinking [39]	235B	10.24	3.88	3.42	27.42	26.36	39.34	58.50	40.48	37.70	35.78
Grok-4 [38]	N/A	15.18	7.96	6.12	22.48	19.08	32.82	54.81	25.36	38.30	30.97
GLM-4.5v [19]	106B	15.68	5.46	4.04	27.53	22.20	<u>46.20</u>	59.56	22.14	42.70	35.95
Medical Specific VLMs											
LLaVA-Med [22]	7B	0.20	0.16	0.10	0.92	1.52	0.00	0.00	0.00	24.50	4.76
MedDr [18]	40B	6.14	4.26	3.08	22.99	<u>32.58</u>	29.57	52.44	20.95	8.70	26.20
HuatuoGPT-V [5]	34B	6.40	2.62	4.02	35.18	24.92	36.32	65.69	25.48	24.80	36.02
HealthGPT [23]	32B	6.34	0.34	0.26	30.64	25.83	27.98	51.12	17.02	8.00	27.80
MedVLM-R1 [28]	2B	16.80	8.70	3.92	22.42	13.71	24.42	43.88	13.57	25.80	24.70
MedGemma-4B [31]	4B	8.48	1.06	3.72	29.69	11.26	28.12	41.88	22.62	39.40	30.45
Open-source VLMs											
LLaVA-OneVision [21]	8B	8.66	3.62	6.08	28.79	13.64	30.12	55.12	20.36	23.20	29.43
Qwen2.5-VL-3B-Instruct [2]	3B	1.82	4.96	3.40	32.01	20.91	27.55	54.50	30.12	5.70	28.79
Qwen2.5-VL-7B-Instruct [2]	7B	7.22	4.58	3.70	17.01	16.10	11.18	29.41	9.07	8.20	15.92
Agentic VLMs											
Mini-o3 [20]	7B	3.14	2.18	2.04	8.72	1.97	15.15	46.62	13.93	9.80	14.72
Our Models											
OralGPT-Plus-3B-SFT	3B	16.45	8.32	14.18	22.17	13.71	16.44	35.25	9.88	19.80	21.02
OralGPT-Plus-7B-SFT	7B	23.06	<u>11.36</u>	<u>14.40</u>	35.02	19.32	29.02	62.37	20.12	11.90	31.44
OralGPT-Plus-3B	3B	<u>33.12</u>	11.24	10.70	21.58	17.35	21.29	35.69	16.07	31.70	24.48
OralGPT-Plus-7B	7B	43.16	20.60	24.96	45.34	33.56	45.58	<u>75.38</u>	34.64	32.70	45.35

^a Evaluated API snapshot dated 2025-10-15.

^b Evaluated API snapshot dated 2025-09-29.

systems are treated as equivalent, minor coordinate omissions incur small penalties, and disease mismatches or false assertions are penalized as major errors. These structured criteria bridge the gap between binary correctness and continuous clinical judgment, offering dense supervision that rewards clinically meaningful outputs while discouraging hallucinations and omissions of key findings. This mechanism improves training efficiency and supports the conditioned diagnostic-driven reward described next.

3.2.2. Conditioned Diagnostic-Driven Reward

Inspired by the clinical workflow that dentists reinspect only when the initial diagnosis is reliable yet incomplete, we design a conditioned exploration reward that prevents unnecessary visual operations at low diagnostic certainty. We adopt the curiosity bonus [36], enabling it only when the rubric score indicates sufficient diagnostic reliability.

For a generated trajectory τ , let $R_{\text{rubrics}}(\tau) \in [0, 1]$ denote the diagnostic confidence evaluated by the rubric-based assessor, and let $\text{Cu}(x)$ represent the current exploration saturation level for query x . We further define $\mathbb{I}_{\text{tool}}(\tau) \in \{0, 1\}$ as an indicator of whether tool is used. The conditioned diagnostic-driven reward is then formulated as:

$$R_{\text{diag}}(\tau) = \mathbb{I}(R_{\text{rubrics}}(\tau) > \eta) \cdot \alpha \cdot (H - \text{Cu}(x))_+ \cdot \mathbb{I}_{\text{tool}}(\tau), \quad (7)$$

where $(\cdot)_+ = \max(\cdot, 0)$ ensures that the curiosity magnitude decreases as exploration becomes saturated, and α controls the strength of the intrinsic reward.

These ensure that the reward encourages clinically motivated reinspection rather than indiscriminate tool use, allowing visual operations only when they meaningfully refine an already credible diagnostic hypothesis.

3.2.3. Hybrid Reward System

Building on the above reward components, we combine the rubric-based clinical reward, conditioned diagnostic incentive, and format signals into a unified Hybrid Reward System:

$$R(\tau) = R_{\text{rubrics}}(\tau) + R_{\text{format}}(\tau) + R_{\text{diag}}(\tau). \quad (8)$$

This composite reward balances diagnostic accuracy, structural correctness, and exploration efficiency. It guides the agent from reliable local assessments to clinically meaningful reinspections while preserving adherence to the diagnostic schema, thereby stabilizing training and preventing tool-use collapse during long-horizon reasoning.

3.2.4. Optimization Objective

We train the model under the GRPO objective [32] with clipped ratios, using standardized advantages denoted as $\mathbb{A}_i$

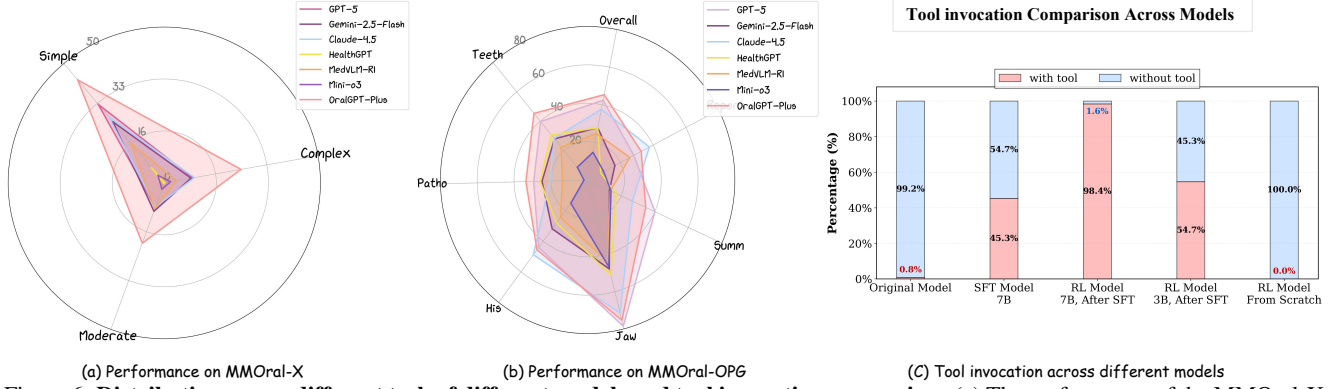


Figure 6. **Distribution across different task of different models and tool invocation comparison.**(a) The performance of the MMOral-X benchmark. (b) Performance comparison on MMOral-Omni benchmark. (c) Performance comparison on MMOral-OPG benchmark.

to avoid notational conflict with earlier action symbols:

$$\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{q \sim \mathcal{D}, o_i \sim \pi_{\theta_{\text{old}}}} \left[\frac{1}{G} \sum_{i=1}^G \min(r_i \mathbb{A}_i, \hat{r}_i \mathbb{A}_i) \right]. \quad (9)$$

Here, G is the number of sampled rollouts per query. The hybrid reward $R(\tau)$ defined above populates the return used for computing advantages.

4. MMOral-X Benchmark

4.1. Benchmark Composition

We introduce MMOral-X, the first benchmark that enables comprehensive diagnosis from a single panoramic radiograph. It contains 300 open-ended Q&A pairs and 686 bounding boxes covering diverse pathological findings, including disease classification, grounding, tooth-level assessments, regional evaluations, and abnormality flags. It also includes clinically relevant non-pathological findings—such as orthodontic appliances, restorations, and surgical devices—that are essential for a complete diagnosis but are largely overlooked in prior datasets. Unlike MMOral-OPG [14], which targets localized, lesion-specific queries, MMOral-X supports holistic diagnostic evaluation. A sample is provided in Fig.5. To assess varying difficulty, MMOral-X is divided into three subsets—Simple, Moderate, and Complex—each containing 100 images. Simple cases require one to two reasoning steps and contain a single proposal; Moderate cases involve two or more proposals; Complex cases include multiple proposal types with substantial visual ambiguity. This stratified organization provides a controlled gradient of diagnostic complexity while maintaining a unified task formulation.

4.2. Evaluation Metrics

Following established evaluation practices and prompts design [14, 41–43], we construct a carefully designed few-shot prompt and employ GPT-5-mini [27] as the judge for

Table 2. Ablation studies.

Model	MMOral-X		
	Simple	Moderate	Complex
wo Instruction tuning	8.02	4.64	4.98
wo R_{rubrics} and R_{cond}	20.82	9.24	9.82
wo R_{rubrics}	21.48	12.08	13.42
wo R_{cond}	24.62	14.14	<u>15.96</u>
wo “Mirror-In” tool	<u>34.68</u>	<u>14.26</u>	14.30
OralGPT-Plus	43.16	20.60	24.96

open-ended assessment. The prompt contains few-shot exemplars covering fully/partially correct and incorrect responses, enabling the judge to score each prediction on a continuous scale from 0.0 to 1.0 with a resolution of 0.1. Beyond this setup, we systematically analyze the choice of the judge, its potential scoring bias, and the variance observed across multiple test runs. We also perform consistency checks to verify that the evaluation procedure remains stable under repeated testing. These analyzes demonstrate that our scoring framework is reliable and robust.

5. Experiments

5.1. Experimental Setups

Settings. All experiments are conducted on four NVIDIA A800 GPUs. For dentist-like instruction tuning, we employ Qwen2.5-VL-3B/7B-Instruct [2]. We perform SFT on 4k DentalProbe samples for three epochs and reserve an additional 1k samples for RL to avoid data leakage. Both the rollout number and global batch size are set to 8, and all other optimization configurations follow [40]. At evaluation, we report Avg@5 for MMOral-X and pass@1 for MMOral-OPG following prior work [14]. For judge, we chose GPT-5-mini because it offers consistent scoring, strong rubric-following, and minimal hallucination.

Baselines. To enable a comprehensive comparison, we evaluate a diverse set of strong baselines, including frontier foundation models, state-of-the-art medical multimodal models, commonly used open-source vision-language

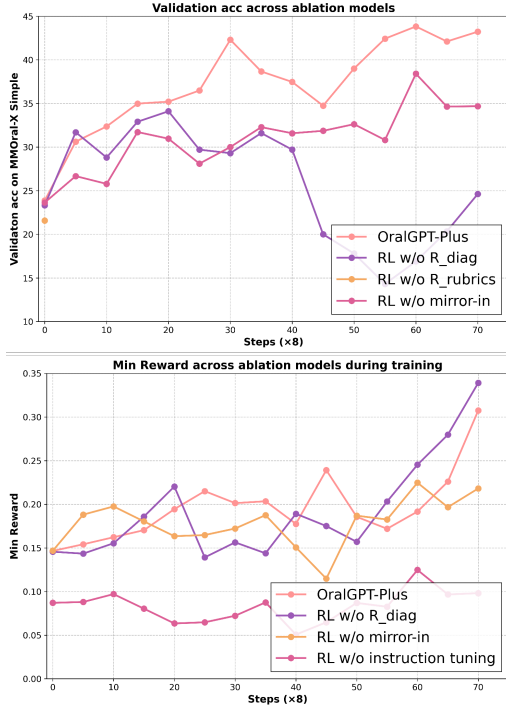


Figure 7. Training dynamics across ablation models.

models, and tool-augmented VLMs.

5.2. Main Results

Table 1 presents the overall results of OralGPT-Plus across MMOral-X and MMOral-OPG benchmarks. On MMOral-X, OralGPT-Plus achieves the highest accuracy under all difficulty levels, surpassing powerful proprietary models such as GPT-5, as well as domain-specific models like MedDr and HuatuoGPT-V. This shows that our two-stage training, including dentist-like instruction tuning and reinspection-driven reinforcement learning, effectively enhances the model’s reasoning ability on complex dental structures. On MMOral-OPG, OralGPT-Plus also obtains the best overall score and leads in key diagnostic aspects such as Teeth, Patho, and Jaw. The results confirm that integrating structured tool interaction enables the model to perform a stepwise and clinically coherent analysis similar to that of human dentists. The distribution across multiple models on different subset are shown in Fig. 6.

We also note that failure cases primarily occur in complex radiographs. Additional analyses are provided in Appendix Sec. ?? (tool-usage dynamics), Sec. ?? (failure analysis), Fig. ??, Fig. ??, and Table ??.

5.3. Findings

In this section, we present four key findings derived from the results in Table 1 and the ablation studies in Table 2.

Dentist-like Instruction Tuning is Essential. As shown in Table 2 and Fig. 6, models without Dentist-like Instruction Tuning exhibit clear performance drops and fail to ac-

tivate tool-usage behaviors, even after reinforcement learning. This demonstrates that tool-based reasoning does not emerge from reward optimization alone but requires clinically structured demonstrations that specify when and how tools should be used. In contrast, instruction-tuned models acquire a stable tool schema that enables reinforcement learning to further refine diagnostic behavior.

Model Capacity Determines the Effectiveness of Reinforcement Learning. With the same instruction-tuning and reward setup, OralGPT-Plus-7B yields markedly larger RL gains than the 3B model. As shown in Fig. 6, the 7B model consistently develops stable tool-based reinspections, while the 3B model fails to do so. As a result, the 7B variant improves across all MMOral-X levels, indicating that sufficient model capacity is crucial for mastering tool-driven diagnostic workflows and fully benefiting from RL.

Conditioned Diagnostic-Driven Reward Prevents Reward Hacking. Ablating the conditioned diagnostic-driven reward induces clear reward hacking. As shown in Fig. 7 and Tab. 2, the model without this component displays a collapse-rebound pattern in which tool usage first drops and then spikes while validation accuracy deteriorates. This failure arises because the agent exploits the reward by invoking a single tool and answering without meaningful inspection. Incorporating the conditioned mechanism suppresses this shortcut, stabilizes training, and triggers reinspections only when diagnostic confidence is sufficient, thereby aligning exploration with clinical reasoning.

Mirror-In Tool Significantly Enhances Diagnostic Reliability. Panoramic radiographs are highly symmetric, and the “Mirror-In” tool effectively exploits this structure. Removing it leads to clear drops across all difficulty levels shown in Tab. 2: accuracy on Simple decreases from 43.22 to 34.68, on Moderate from 20.60 to 14.26, and on Complex from 24.96 to 14.30. These declines indicate the model struggles with low-contrast ambiguities and exhibits unstable reward trajectories. Integrating “Mirror-In” introduces symmetry-aware verification, stabilizes training, and significantly improves diagnostic robustness.

6. Conclusion

This work advances panoramic radiograph analysis from static perception to interactive diagnostic reasoning. We introduce OralGPT-Plus, an agentic model that inspects, compares, and verifies findings through clinically grounded tool use. With dentist-like instruction tuning and reinspection-driven reinforcement learning, OralGPT-Plus achieves strong performance on MMOral-X and MMOral-OPG, surpassing proprietary and medical-domain baselines. The results show that diagnostic capability emerges not only from model capacity but also from the ability to act, reflect, and re-examine. This paradigm provides a promising direction for future diagnostic VLMs.

References

- [1] Anthropic. Introducing claude sonnet 4.5. <https://www.anthropic.com/news/claude-sonnet-4-5>, 2025. 5, 6
- [2] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibao Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-vl technical report. *ArXiv*, abs/2502.13923, 2025. 6, 7
- [3] Zhenyang Cai, Jiaming Zhang, Junjie Zhao, Ziyi Zeng, Yanchao Li, Jingyi Liang, Junying Chen, Yunjin Yang, Jiajun You, Shuzhi Deng, et al. Dentalgpt: Incentivizing multimodal complex reasoning in dentistry. *arXiv preprint arXiv:2512.11558*, 2025. 2
- [4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 2
- [5] Junying Chen, Chi Gui, Ruyi Ouyang, Anningzhe Gao, Shunian Chen, Guiming Hardy Chen, Xidong Wang, Ruifei Zhang, Zhenyang Cai, Ke Ji, et al. Huatuogpt-vision, towards injecting medical visual knowledge into multimodal llms at scale. *arXiv preprint arXiv:2406.19280*, 2024. 6
- [6] Chenlin Du, Xiaoxuan Chen, Jingyi Wang, Junjie Wang, Zhongsen Li, Zongjiu Zhang, and Qicheng Lao. Prompting vision-language models for dental notation aware abnormality detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 687–697. Springer, 2024. 2
- [7] Dong Guo, Faming Wu, Feida Zhu, Fuxing Leng, Guang Shi, Haobin Chen, Haoqi Fan, Jian Wang, Jianyu Jiang, Jiawei Wang, et al. Seed1. 5-vl technical report. *arXiv preprint arXiv:2505.07062*, 2025. 6
- [8] Ibrahim Ethem Hamamci, Sezgin Er, Enis Simsar, Atif Emre Yuksel, Sadullah Gultekin, Serife Damla Ozdemir, Kaiyuan Yang, Hongwei Bran Li, Sarthak Pati, Bernd Stadlinger, et al. Dentex: An abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic x-rays. *arXiv preprint arXiv:2305.19112*, 2023. 4
- [9] Jing Hao and Song Chen. Language-aware multiple datasets detection pretraining for detrs. *Neural Networks*, 179: 106506, 2024. 2
- [10] Jing Hao, Moyun Liu, Lei He, Lei Yao, James Kit Hon Tsoi, and Kuo Feng Hung. Semit-sam: Building a visual foundation model for tooth instance segmentation on panoramic radiographs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 110–121. Springer, 2024. 1, 2
- [11] Jing Hao, Lun M Wong, Zhiyi Shan, Qi Yong H Ai, Xieqi Shi, James Kit Hon Tsoi, and Kuo Feng Hung. A semi-supervised transformer-based deep learning framework for automated tooth segmentation and identification on panoramic radiographs. *Diagnostics*, 14(17):1948, 2024. 1, 2
- [12] Jing Hao, Yuxiang Zhao, Song Chen, Yanpeng Sun, Qiang Chen, Gang Zhang, Kun Yao, Errui Ding, and Jingdong Wang. Fullanno: A data engine for enhancing image comprehension of mllms. *arXiv preprint arXiv:2409.13540*, 2024. 2
- [13] Jing Hao, Yonghui Zhu, Lei He, Moyun Liu, James Kit Hon Tsoi, and Kuo Feng Hung. T-mamba: a unified framework with long-range dependency in dual-domain for 2d & 3d tooth segmentation. *arXiv preprint arXiv:2404.01065*, 2024. 2
- [14] Jing Hao, Yuxuan Fan, Yanpeng Sun, Kaixin Guo, Lizhuo Lin, Jinrong Yang, Qi Yong H Ai, Lun M Wong, Hao Tang, and Kuo Feng Hung. Towards better dental ai: A multimodal benchmark and instruction dataset for panoramic x-ray analysis. *The Thirty-Ninth Annual Conference on Neural Information Processing Systems*, 2025. 2, 4, 7
- [15] Jing Hao, Yuci Liang, Lizhuo Lin, Yuxuan Fan, Wenkai Zhou, Kaixin Guo, Zanting Ye, Yanpeng Sun, Xinyu Zhang, Yanqi Yang, et al. Oralgpt-omni: A versatile dental multimodal large language model. *arXiv preprint arXiv:2511.22055*, 2025. 2
- [16] Jing Hao, Moyun Liu, Jinrong Yang, and Kuo Feng Hung. Gem: Boost simple network for glass surface segmentation via vision foundation models. *IEEE Transactions on Multimedia*, 2025. 2
- [17] Jing Hao, Andrew Nalley, Andy Wai Kan Yeung, Ray Tanaka, Qi Yong H Ai, Walter Yu Hang Lam, Zhiyi Shan, Yiu Yan Leung, Abeer AlHadidi, Michael M Bornstein, et al. Characteristics, licensing, and ethical considerations of openly accessible oral-maxillofacial imaging datasets: a systematic review. *npj Digital Medicine*, 8(1):412, 2025. 1, 2
- [18] Sunan He, Yuxiang Nie, Zhixuan Chen, Zhiyuan Cai, Hongmei Wang, Shu Yang, and Hao Chen. Meddr: Diagnosis-guided bootstrapping for large-scale medical vision-language learning. *CoRR*, 2024. 2, 6
- [19] GLM-V Team Wenyi Hong, Wenmeng Yu, Xiaotao Gu, Guo Wang, Guobing Gan, Haomiao Tang, Jiale Cheng, Ji Qi, Junhui Ji, Lihang Pan, Shuaiqi Duan, Weihang Wang, Yan Wang, Yean Cheng, Zehai He, Zhe Su, Zhen Yang, Ziyang Pan, Aohan Zeng, Baoxu Wang, Boyan Shi, Changyu Pang, Chenhui Zhang, Da Yin, Fan Yang, Guoqing Chen, Jiazhen Xu, Jiali Chen, Jing Chen, Jinhao Chen, Jinghao Lin, Jinjiang Wang, Junjie Chen, Leqi Lei, Letian Gong, Leyi Pan, Mingzhi Zhang, Qinkai Zheng, Shengchao Yang, Shilong Zhong, Shiyu Huang, Shuyuan Zhao, Siyan Xue, Shangqing Tu, Shengbiao Meng, Tianshu Zhang, Tian-Yuan Luo, Tianxiang Hao, Wenkai Li, Wei Jia, Xinpeng Lyu, Xuancheng Huang, Yanling Wang, Ya-Qi Xue, Yanfeng Wang, Yifan An, Yifan Du, Yi Shi, Yiheng Huang, Yilin Niu, Yuan Wang, Yuanchang Yue, Yuchen Li, Yutao Zhang, Yuxuan Zhang, Zhanxiao Du, Zhenyu Hou, Zhao Xue, Zhengxiao Du, Zihan Wang, Peng Zhang, De-Huan Liu, Bin Xu, Juanzi Li, Minlie Huang, Yuxiao Dong, and Jie Tang. Glm-4.5v and glm-4.1v-thinking: Towards versatile multimodal reasoning with scalable reinforcement learning. 2025. 6
- [20] Xin Lai, Junyi Li, Wei Li, Tao Liu, Tianjian Li, and Hengshuang Zhao. Mini-o3: Scaling up reasoning patterns and

- interaction turns for visual search. *ArXiv*, abs/2509.07969, 2025. 6
- [21] Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Yanwei Li, Ziwei Liu, and Chunyuan Li. Llava-onevision: Easy visual task transfer. *ArXiv*, abs/2408.03326, 2024. 6
- [22] Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *ArXiv*, abs/2306.00890, 2023. 2, 6
- [23] Tianwei Lin, Wenqiao Zhang, Sijing Li, Yuqian Yuan, Binhe Yu, Haoyuan Li, Wangui He, Hao Jiang, Mengze Li, Xiaohui Song, et al. Healthgpt: A medical large vision-language model for unifying comprehension and generation via heterogeneous knowledge adaptation. *arXiv preprint arXiv:2502.09838*, 2025. 6
- [24] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023. 3
- [25] Zekai Liu, Andrew Nalley, Jing Hao, Qi Yong H Ai, Andy Wai Kan Yeung, Ray Tanaka, and Kuo Feng Hung. The performance of large language models in dentomaxillofacial radiology: a systematic review. *Dentomaxillofacial Radiology*, page twaf060, 2025. 1
- [26] Sorana Mureşanu, Mihaela Hedeşiu, Liviu Iacob, Radu Eftimie, Eliza Maria Olariu, Cristian Dinu, Reinhilde Jacobs, and On Behalf Of Team Project Group. Automating dental condition detection on panoramic radiographs: Challenges, pitfalls, and opportunities. *Diagnostics*, 14, 2024. 4
- [27] OpenAI. Gpt-5. <https://openai.com/zh-Hans-CN/index/introducing-gpt-5>, 2025. 5, 6, 7
- [28] Jiazhen Pan, Che Liu, Junde Wu, Fenglin Liu, Jiayuan Zhu, Hongwei Bran Li, Chen Chen, Cheng Ouyang, and Daniel Rueckert. 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*, pages 337–347. Springer, 2025. 6
- [29] Karen Panetta, Rahul Rajendran, Aruna Ramesh, Shishir Paramathma Rao, and Sos Agaian. Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. *IEEE journal of biomedical and health informatics*, 26(4):1650–1659, 2021. 4
- [30] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 2
- [31] Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla P. Kiraly, Madeleine Traverse, Timo Kohlberger, Shawn Xu, Fayaz Jamil, Cian Hughes, Charles Lau, Justin Chen, Fereshteh Mahvar, Liron Yatziv, Tiffany Chen, Bram Sterling, Stefanie Anna Baby, Susanna Maria Baby, Jeremy Lai, Samuel Schmidgall, Lu Yang, Kejia Chen, Per Bjornsson, Shashir Reddy, Ryan Brush, Kenneth Philbrick, Mercy Nyamewaa Asiedu, Ines Mezerreg, Howard Hu, Howard Yang, Richa Tiwari, Sunny Jansen, Preeti Singh, Yun Liu, Shekoofeh Azizi, Aishwarya B Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ram’e, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean-Bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Elena Buchatskaya, Jean-Baptiste Alayrac, Dmitry Lepikhin, Vladimir Feinberg, Sebastian Borgeaud, Alek Andreev, Cassidy Hardin, Robert Dadashi, Léonard Hussenot, Armand Joulin, Olivier Bachem, Yossi Matias, Katherine Chou, Avinatan Hassidim, Kavi Goel, Clément Farabet, Joelle K. Barral, Tris Warkentin, Jonathon Shlens, David Fleet, Victor Cotruta, Omar Sanseviero, Gus Martins, Phoebe Kirk, Anand Rao, Shravya Shetty, David Steiner, Can Kirmizibayrak, Rory Pilgrim, Daniel Golden, and Lin Yang. Medgemma technical report. *ArXiv*, abs/2507.05201, 2025. 6
- [32] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. 5, 6
- [33] Yanpeng Sun, Jing Hao, Ke Zhu, Jiang-Jiang Liu, Yuxiang Zhao, Xiaofan Li, Gang Zhang, Zechao Li, and Jingdong Wang. Descriptive caption enhancement with visual specialists for multimodal perception. *arXiv preprint arXiv:2412.14233*, 2024. 2
- [34] Sicheng Tao, Jungang Li, Yibo Yan, Junyan Zhang, Yubo Gao, Hanqian Li, ShuHang Xun, Yuxuan Fan, Hong Chen, Jianxiang He, et al. Moss-chatv: Reinforcement learning with process reasoning reward for video temporal reasoning. *arXiv preprint arXiv:2509.21113*, 2025. 5
- [35] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023. 5, 6
- [36] Haozhe Wang, Alex Su, , Weiming Ren, Fangzhen Lin, and Wenhui Chen. Pixel reasoner: Incentivizing pixel-space reasoning with curiosity-driven reinforcement learning. *arXiv preprint arXiv:2505.15966*, 2025. 4, 6
- [37] Yaqi Wang, Zhi Li, Chengyu Wu, Jun Liu, Yifan Zhang, Jiaxue Ni, Qian Luo, Jialuo Chen, Hongyuan Zhang, Jin Liu, et al. Miccai sts 2024 challenge: Semi-supervised instance-level tooth segmentation in panoramic x-ray and cbct images. *Medical Image Analysis*, page 103986, 2026. 1
- [38] xAI. Grok 4. <https://x.ai/news/grok-4>, 2025. 6
- [39] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxin Yang, Jingren Zhou, Jingren Zhou, Junyan Lin, Kai Dang, Keqin Bao, Ke-Pei Yang, Le Yu, Li-Chun Deng, Mei Li, Min Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shi-Qiang Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng

- Ren, Yang Fan, Yang Su, Yi-Chao Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *ArXiv*, abs/2505.09388, 2025. [6](#)
- [40] Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gao-hong Liu, Lingjun Liu, et al. Dapo: An open-source llm reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*, 2025. [7](#)
- [41] Weihao Yu, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Kevin Lin, Zicheng Liu, Xinchao Wang, and Lijuan Wang. Mm-vet: Evaluating large multimodal models for integrated capabilities. *arXiv preprint arXiv:2308.02490*, 2023. [7](#)
- [42] Weihao Yu, Zhengyuan Yang, Lingfeng Ren, Linjie Li, Jianfeng Wang, Kevin Lin, Chung-Ching Lin, Zicheng Liu, Lijuan Wang, and Xinchao Wang. Mm-vet v2: A challenging benchmark to evaluate large multimodal models for integrated capabilities. *arXiv preprint arXiv:2408.00765*, 2024.
- [43] Jirong Zha, Yuxuan Fan, Tianyu Zhang, Geng Chen, Yingfeng Chen, Chen Gao, and Xinlei Chen. Aircopbench: A benchmark for multi-drone collaborative embodied perception and reasoning. *arXiv preprint arXiv:2511.11025*, 2025. [7](#)
- [44] Yifan Zhang, Fan Ye, Lingxiao Chen, Feng Xu, Xiaodiao Chen, Hongkun Wu, Mingguo Cao, Yunxiang Li, Yaqi Wang, and Xingru Huang. Children’s dental panoramic radiographs dataset for caries segmentation and dental disease detection. *Scientific Data*, 10(1):380, 2023. [4](#)