Supervised Fine-Tuning

Image $\mathcal{V}$ **Query** $\mathcal{Q}$

Difference Comparison QA:
How does this X-ray compare to my previous one?

Region aware QA:
Are there any concerning findings in the marked region?

Single Image QA:
Does the right apical zone exhibit any tubes/lines or diseases?

Image Encoder

Projector   Tokenizer

Pre-trained VLM ❄   **LoRA** 🔥

Rationale: [Yes.] [The image demonstrates a right-sided chest tube with its tip ......]

Rationale $\mathcal{T}$ = [Answer] $\mathcal{A}$ + [Explanation] $\mathcal{E}$

Counterfactual Rationale Reinforce

Chest X-ray VLM Response y

Incorrect   Low-confidence correct   High-confidence correct

Incorrect: [No.] [The image does not show any tubes or lines ......]

Correct: [Yes.] [The image shows increased opacities in the ......]

Correct: [Yes.] [The image shows a right-sided chest tube ......]

① **Hard Example Identification**

② **Confidence-Similarity Joint Mining**

y_i y_i $K = 3$

Query $\mathcal{Q}$ →Similar→ $\mathcal{Q}_r$
Rationale $\mathcal{T}$ →Similar→ $\mathcal{T}_r$
Appearance $\mathcal{V}$ →Similar→ $\mathcal{V}_r$

Opposite $\mathcal{A}$, **Similar** $\mathcal{E}$

③ **Counterfactual Rationale Construction**

[Yes.] [The image shows increased opacities in the ......] →Replace→ [No.] [The image shows increased opacities in the ......]

[No.] [The image does not show any increased opacities......] y_l

Semantic Retrieval

Preference Alignment

Preference Alignment Data $\mathcal{D}_{hard}$

Chosen y_w: Ground Truth $\mathcal{T} = [\mathcal{A}; \mathcal{E}]$

Rejected y_l: — Incorrect Response
— Constructed Response

Chosen:[Yes.] [The image demonstrates a right-sided ......]
Image & Query

maximum likelihood

Chosen:[No.] [The image does not show any increased ......]
Image & Query

Preference Fine-Tuning

Stronger CheXVLM