RRG-DPO: Direct Preference Optimization for Clinically Accurate Radiology Report Generation

Hong Liu[2,3], Dong Wei[3], Zhe Xu[5], Xian Wu[3(✉)], Yefeng Zheng[3,4], and Liansheng Wang[1,2(✉)]

[1] School of Informatics, Xiamen University, Xiamen, China
lswang@xmu.edu.cn, liuhong@stu.xmu.edu.cn
[2] National Institute for Data Science in Health and Medicine, Xiamen University
[3] Tencent Jarvis Lab, Tencent Healthcare (Shenzhen) Co., Ltd., Shenzhen, China
{donwei,kevinxwu,yefengzheng}@tencent.com
[4] Medical Artificial Intelligence Laboratory, Westlake University, Hangzhou, China
zhengyefeng@westlake.edu.cn
[5] Dept. of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, China
jackxz@link.cuhk.edu.hk

Abstract. Automated radiology report generation (RRG) of routine 2D and 3D radiographs, such as X-ray and computed tomography (CT), has great potential in reducing the workload, variations, and errors of report writing and facilitating patient care. Despite significant advancements in linguistic quality, existing methods may generate reports with hallucinated type I and II (false positive and false negative) errors, which limit clinical efficiency. To mitigate the hallucinations, we propose RRG-DPO, an innovative direct preference optimization procedure with a new loss term, both tailored for effective alignment with the preference for clinically accurate RRG. RRG-DPO retrieves a set of highly relevant reports closest to the preferred response (i.e., the ground truth (GT) report) in a biomedical CLIP embedding space, and selects the one with the most significant abnormality conflicts with the GT as the dispreferred response. Besides being clinically relevant and abnormally aware, this preference data curation process is cost-effective and scalable compared to using large language models for response sampling or evaluation. In addition, we note that except for the abnormality-conflicting sentences, other sentences of the dispreferred report can legibly describe the radiograph of the preferred in a clinically equivalent manner, despite variations in expression. Thus, RRG-DPO creates a sub-preferred report from the dispreferred by deleting the abnormality-conflicting sentences, and promotes its likelihood with a new loss term. RRG-DPO is evaluated on both 2D X-ray and 3D CT data to align a wide range of RRG models. Experiments show that it boosts the clinical efficiency of all assessed models in six metrics: precision, recall, F1 score, RadGraph, RadCliQ, and RaTEScore, effectively reducing hallucinations. Further

H. Liu and D. Wei——Contributed equally; H. Liu contributed to this work during an internship at Tencent.

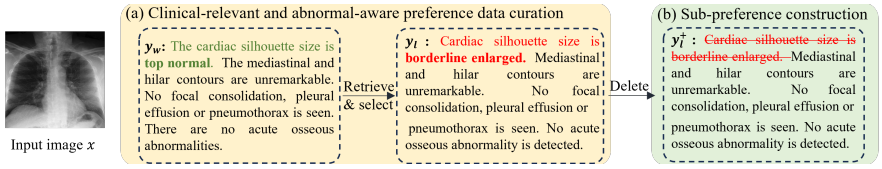

Fig. 1. Our proposed pipeline for preference data curation. (a) Clinical-relevant and abnormal-aware dispreferred response selection. Red highlights abnormality-conflicting sentences in the preferred (y_w, the ground truth report) and dispreferred (y_l, the retrieved and selected report) responses for a chest X-ray image x. (b) Sub-preference response (y_l^+) construction by deleting the conflicting sentence from y_l.

ablation studies show that our method outperforms DPO and DPOP, and its components are helpful.

Keywords: Radiology report generation · Direct preference optimization · Hallucinated type I and II errors.

1 Introduction

Radiology reports of routine radiographs, such as X-ray and computed tomography (CT), serve as essential documentation and facilitate accurate diagnosis and timely treatment in healthcare. However, the interpretation of radiographs and manual creation of the reports demands domain expertise, careful observation, and considerable effort, facing the potential risks of delay or misdiagnosis and inter-observer discrepancy [6].

Various deep-learning methods were proposed for automatic radiology report generation (RRG). They usually follow the classic encoder-decoder paradigm [12, 25, 38]. To improve generated reports, researchers advanced model structures under the paradigm, by introducing cross-modal memory modules [10, 34], multimodal alignment [42], or new attention mechanisms [36]. Alternatively, external domain knowledge in graphs [16, 20, 43], disease tags [18, 39, 44], longitudinal data [26], or retrieved reports [23] was incorporated as guidance. Researchers have recently started leveraging the strong linguistic capability of large language models (LLMs) for RRG [9, 40, 4, 21, 8]. Despite significant progress, these prior approaches treated every word (token) equally for supervised regression of ground truth reports in training. As a result, the trained models may generate reports that are consistent with the training data's writing style and of high-quality linguistics but not well aligned with radiologists' preference for clinical accuracy. Such misalignment causes a high incidence of hallucinated type I (false positive) and II (false negative) errors, which can lead to factuality issues in generated reports.

To mitigate the hallucinations, researchers implemented reinforcement learning (RL) from human (or AI) feedback (RLHF/RLAIF) to align RRG models with preferences for clinically accurate reports [47]. RLHF methods fit a reward

model to a preference dataset and then use RL to optimize a *policy* to produce responses of high rewards [1]. However, the RLHF pipeline is considerably more complex than supervised learning and often unstable. It incurs significant computational costs and major practical challenges. Direct preference optimization (DPO) [33] is a stable, lightweight, single-stage algorithm that implicitly optimizes the same objective as RLHF but without the need to fit a separate reward model. Essentially solving a classification problem, DPO is straightforward to implement and train. Several studies have applied DPO for hallucination suppression and factuality improvement in radiology interpretation/report generation models [2, 15, 37]. However, these works employed the preference data curation processes for natural images, overlooking the clinical relevance of the generated responses. MMedPO [47] proposed incorporating the clinical relevance based on GPT-4o evaluation. Yet, its multi-round policy sampling and frequent GPT-4o evaluation incurred substantial computational and economic costs, limiting scalability. Alternatively, the dispreferred response can be a random report different from the ground truth (GT). However, our experiments indicate that this strategy may even harm the performance of the initial policy.

This paper introduces RRG-DPO, an innovative DPO procedure with a new loss term, both tailored for effective alignment with the preference for clinically accurate RRG, thus reducing hallucinated type I and II errors. Our three primary contributions are the following. (1) **Clinical-relevant and abnormal-aware preference data curation** (Fig. 1(a)). We first retrieve from the training data a set of reports closest to the preferred response y_w (i.e., the GT report) in the BiomedVLP [5] embedding space. Unlike a random report, which can be easily distinguished thus reduces the alignment effect, the retrieved reports are highly clinically relevant to the preferred response. Next, we pick the retrieved report with the most significant discrepancies from the preferred response regarding described abnormalities as the dispreferred response y_l. Despite the overall high similarity, the abnormal-aware selection ensures valid abnormality conflicts between the (dis)preference pair, mimicking hallucinated type I and II errors. In addition, the curation process is computationally and economically efficient compared with employing LLMs for sampling or evaluating dispreferred responses. (2) **Sub-preference optimization** (Fig. 1(b)). We note that, for the first time in the literature on DPO for RRG, except for the abnormality-conflicting sentences, other sentences of y_l can legibly describe the radiograph of y_w in a clinically equivalent manner, despite variations in expression. Therefore, we propose creating a *sub-preferred* report y_l^+ from the dispreferred by deleting the abnormality-conflicting sentences, and promoting its likelihood relative to the reference policy with a new loss term. (3) **Comprehensive study**. We apply our RRG-DPO to both 2D X-ray and 3D CT data to align a wide range of representative RRG models. Experimental results show that our method boosts the clinical efficiency (CE) in precision, recall, and F1 score, and three clinical entity- and relation-based metrics (RadGraph, RadCliQ, and RaTEScore) of all evaluated models, effectively reducing hallucinations. Notably, we achieve a new state of the art for the CE metrics on the MIMIC-CXR dataset [19]. Fur-

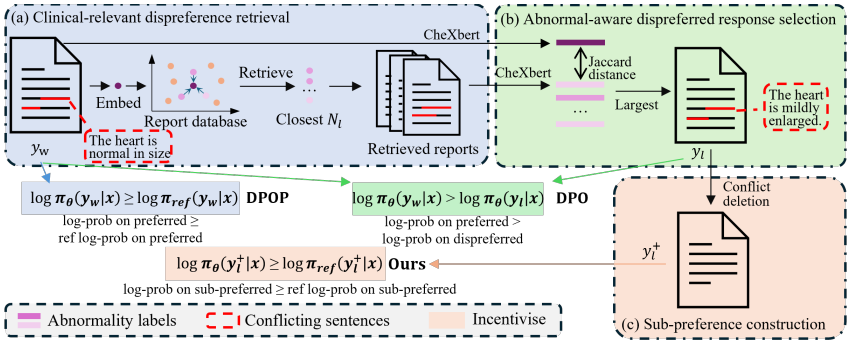

Fig. 2. Our proposed RRG-DPO advances the DPO pipeline for RRG with three novel components: (a) embedding-based clinical-relevant dispreference retrieval, (b) abnormal-aware dispreferred response selection, and (c) sub-preference optimization.

ther ablation study shows that our method outperforms DPO and DPO-positive (DPOP) [29], and its components are helpful.

2 Method

Preliminaries: Direct Preference Optimization. Preference optimization [1] has proven effective in finetuning language models to align model behavior with human preferences. In preference optimization, a relative preference dataset $\mathcal{D} = \{x^{(i)}, y_w^{(i)}, y_l^{(i)}\}_{i=1}^N$ is curated and used for the alignment, where $y_w^{(i)}$ and $y_l^{(i)}$ are a pair of comparatively preferred and dispreferred responses for the i^{th} input $x^{(i)}$. DPO [33] is a theoretically justified algorithm for relative preference optimization without reinforcement learning. It directly optimizes for the policy $\pi_\theta(y \mid x)$ best satisfying the preferences with a simple classification objective:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \beta\log\frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}\right)\right], \quad (1)$$

where σ is the sigmoid function, β is a regularization parameter corresponding to the strength of the Kullback-Leibler (KL)-regularization in RLHF, and π_{ref} is the reference policy. For a native DPO implementation for comparison experiments, this work takes the GT report of a radiograph as the preferred response y_w and another random report from the training set as the dispreferred y_l.

Method Overview. Our RRG-DPO advances the current DPO pipeline for RRG from two aspects. (1) For preference data curation, we propose embedding-based clinical-relevant dispreference retrieval (Fig. 2(a)) followed by abnormal-aware dispreferred response selection (Fig. 2(b)). This approach avoids the cost of fabricating and evaluating dispreferred responses using LLMs while ensuring high clinical relevance and diagnostic discrepancy with the preferred response y_w. (2) For the training objective, we supplement the standard DPO and derived

DPO-positive (DPOP) [29] losses with a new term to maintain the probability of the *sub-preferred* content in the dispreferred response y_l (Fig. 2(c)).

Preference Data Curation. Given a radiology image x, we take its ground-truth report as the preferred response y_w, and curate the clinical-relevant and abnormal-aware dispreferred response y_l from training data in two steps.

Clinical-relevant dispreference retrieval. Studies [28] have shown that vision-language models such as CLIP [32] are an effective source of preference for fixing hallucinations. Being motivated, we retrieve a set of N_l reports $\{y_l^{(i,j)}\}_{j=1}^{N_l}$ whose text embeddings are closest to that of $y_w^{(i)}$ in cosine similarity from the training data, using the text encoder of BiomedVLP [5]. Being authentic clinical reports, the retrieved $\{y_l^{(i,j)}\}$ possess significant clinical relevance compared with the responses generated by processes proposed for natural images [2, 15, 37]. They are also semantically similar to the GT and thus are hard negatives for effective preference optimization. Meanwhile, the retrieval operation is more computationally efficient than LLM sampling [41, 47].

Abnormal-aware dispreferred response selection. To align the policy π_θ with the clinical preference for low type I (false-positive) and II (false-negative) errors, the ideal dispreferred report should present diagnostic hallucinations compared with the GT. However, this requirement cannot be guaranteed by the above-described step alone. Therefore, in the second step, we use the CheXbert [35] model to label each training report for 14 abnormalities (including a "No Finding" category), obtaining a 14-dimension binary multi-hot vector. Then, we compute the Jaccard distance between the label vectors of $\{y_l^{(i,j)}\}$ and that of $y_w^{(i)}$, and pick the $y_l^{(i,j)}$ with the largest distance as the dispreferred response $y_l^{(i)}$. This approach selects the report with the most significant abnormality discrepancies with the GT from a set of highly clinical-relevant candidates, thus maximizing the contrastive value between $y_l^{(i)}$ and $y_w^{(i)}$ for preference optimization.

Sub-Preference Optimization. Studies have shown that the native DPO loss (Eqn. (1)) may lead to a reduction of the model's likelihood of the preferred responses, especially when preferred and dispreferred responses only differ by a few tokens. The analysis of Pal et al. [29] suggested that this happens because DPO increases the probability of the token(s) that differ but decreases that of subsequent tokens. Thus, they proposed the DPO-positive (DPOP) loss, which explicitly penalizes the decrease of the preferred responses' likelihood relative to the reference policy by adding a term $\max\left(0, \log \frac{\pi_{\text{ref}}(y_w|x)}{\pi_\theta(y_w|x)}\right)$ to DPO. DPOP has proven effective on various natural language benchmarks and tasks. However, it only focuses on positive (i.e., preferred) responses but omits the impact of the dispreferred. Take y_w and y_l in Fig. 1 for example. Except for the abnormality-conflicting sentences, other sentences of y_l can also legibly describe the corresponding contents of the image x in a clinically equivalent manner, despite the

differences in expression. As a result, entirely suppressing y_l may undesirably decrease the likelihood of generating its clinically legible content for x.

To tackle this issue, we propose to create a *sub-preferred* report y_l^+ by deleting the abnormality-conflicting sentences from y_l (Fig. 1(b)), and promote y_l^+'s probability with respect to the reference policy. To identify abnormality-conflicting sentences, we first split y_l into sentences separated by periods. Then, we apply the abnormality classifier CheXbert [35] to each sentence and obtain a label vector. A sentence is considered abnormality-conflicting if its label vector differs from y_w. Note that there can be multiple abnormality-conflicting sentences in y_l. Next, we add a penalty term $\max\left(0, \log \frac{\pi_{\text{ref}}(y_l^+|x)}{\pi_\theta(y_l^+|x)}\right)$ to the DPOP loss to maintain a high log-likelihood of y_l^+. This penalty term is 0 when $\frac{\pi_{\text{ref}}(y_l^+|x)}{\pi_\theta(y_l^+|x)} \leq 1$ (i.e., the probability of y_l^+ given the current policy π_θ is not lower than the reference policy π_{ref}), and increases otherwise. Thus, our complete RRG-DPO loss is:

$$\mathcal{L}_{\text{RRG-DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l,y_l^+)\sim\mathcal{D}}\Bigg[\log\sigma\Bigg(\beta\Big(\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\Big) - \lambda\cdot\max\left(0, \log\frac{\pi_{\text{ref}}(y_w|x)}{\pi_\theta(y_w|x)}\right) - \gamma\cdot\max\left(0, \log\frac{\pi_{\text{ref}}(y_l^+|x)}{\pi_\theta(y_l^+|x)}\right)\Bigg)\Bigg], \tag{2}$$

where $\beta, \lambda,$ and γ are weights.

Training Procedure. Our RRG-DPO follows the standard DPO pipeline. First, we train a reference policy π_{ref} via supervised regression on training data. Then, we create a preference dataset from the training data. Lastly, we finetune π_{ref} on the preference dataset with the loss in Eqn. (2), to obtain the optimal policy π_θ aligned with the clinical preference for diagnostically correct reports. The preference data curation and sub-preference computation for all experimental data used in this work are efficiently done offline beforehand in ∼4.6 hours.

3 Experiments

Datasets and Evaluation Metrics. To comprehensively validate our proposed approach to RRG, we conduct experiments on two typical radiographs frequently performed in daily clinics: chest X-ray (CXR) and CT, which represent 2D and 3D radiography, respectively. **MIMIC-CXR** [19] is a large dataset of CXR images and paired reports. We follow the official split and the preprocessing in [11], resulting in a processed dataset of 270,790, 2,130, and 3,858 samples for training, validation, and testing, respectively. **CT-RATE** [13] includes 25,692 non-contrast chest CT volumes with corresponding reports of 21,304 patients. Following [13], we standardize all CT volumes to the voxel spacing of $0.75\times0.75\times1.5$ mm^3. The volumes are center-cropped or padded to a consistent size of $480\times480\times240$ voxels. We use the official training set (24,128 volumes/20,000 patients) for training, and split the official test set into a validation (360 volumes/300 patients) and a testing set (1,204 volumes/1,004 patients).

Table 1. Report generation performance on 2D MIMIC-CXR (top) and 3D CT-RATE (bottom) datasets. We use the official codes for all base models except for [7] (official checkpoint used) and [9] (reimplemented by us). *: $p < 0.05$ for pair-wise comparison between baseline models and our RRG-DPO results by the Wilcoxon signed-rank test.

Methods	CE metrics			NLG metrics			RRG-exclusive metrics		
	Pre. ↑	Rec. ↑	F1 ↑	BL-4 ↑	MTR ↑	RG-L ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
2D MIMIC-CXR									
R2Gen [11]	0.347	0.258	0.276	0.107	0.139	0.277	0.172	2.788	0.483
+ RRG-DPO	**0.405***	**0.364***	**0.360***	**0.117***	**0.154***	**0.282***	**0.197***	**2.714***	**0.505***
EKAGen [7]	0.456	0.367	0.380	**0.119**	0.156	0.286	0.201	2.685	0.487
+ RRG-DPO	**0.462**	**0.487***	**0.443***	0.112	**0.166**	0.286	**0.207***	**2.674***	**0.502***
PromptMRG [18]	0.504	0.507	0.476	0.107	0.154	0.268	0.190	2.733	0.484
+ RRG-DPO	**0.508**	**0.516**	**0.483**	**0.108**	0.154	0.268	**0.191**	**2.731**	**0.488**
R2GenGPT [40]	0.398	0.265	0.318	0.112	0.144	0.262	0.118	2.864	0.430
+ RRG-DPO	**0.409**	**0.296***	**0.343***	**0.114**	**0.147**	**0.266**	**0.159***	**2.772***	**0.454***
3D CT-RATE									
CT2Rep [14]	0.206	0.073	0.099	**0.251**	**0.275**	0.377	0.430	2.001	0.647
+ RRG-DPO	**0.275***	**0.232***	**0.233***	0.248	0.271	**0.378**	**0.442***	**1.978**	**0.652**
3D-CT-GPT [9]	0.335	0.197	0.225	**0.221**	**0.274**	0.333	0.435	2.104	0.619
+ RRG-DPO	**0.342***	**0.263***	**0.252***	0.218	0.272	**0.338**	**0.442**	**2.075***	**0.632***

Following the literature convention [18, 27], we evaluate model performance using both natural language generation (NLG) metrics and example-based CE metrics. The former include BLEU-4 [30], METEOR [3], and ROUGE-L [22]. The latter include precision, recall, and F1 scores, calculated by converting reports into 14 abnormality classification labels using CheXbert [35]. Besides the conventional metrics mentioned above, we also employ three additional metrics that were proposed more recently and exclusively for RRG for a more comprehensive evaluation: RadGraph [17], RadCliQ [45], and RaTEScore [46].

Implementation. The experiments are conducted with PyTorch (2.0.0) [31] on one NVIDIA Tesla V100 GPU with 32GB memory. For the reference policy π_{ref}, we use publicly available codes [11, 14, 18, 40] or checkpoint [7] released by the authors, if available. For preference optimization, the optimizer is AdamW [24] with the learning rate and weight decay set to 10^{-7}. The model is trained for three epochs with our proposed RRG-DPO objective. The batch size is 16 for CXRs and 1 for CT. We only finetune the text decoders (BERT or LLM) for preference optimization. The number of retrieved reports N_l is empirically set to 20. For the weights in Eqn. (2), we set $\beta = 0.3$ following [29] and $\lambda = 1, \gamma = 0.1$ based on preliminary experiments. Our implementation, including visual analysis and the alignment data, is available at: https://github.com/ccarliu/RRG-DPO.

RRG-DPO Performance. We demonstrate the effectiveness of our proposed RRG-DPO by applying it to preference optimization of various RRG models. For 2D CXR images, these include a classical method R2Gen [11], two well-performing up-to-date methods EKAGen [7] and PromptMRG [18], and an LLM-based method R2GenGPT [40]. The results are shown in Table 1 top. Our RRG-DPO improves the CE metrics for all base models, e.g., the recalls of R2Gen and EKAGen increase from 0.258 and 0.367 to 0.364 and 0.487, re-

Table 2. Ablation study results on MIMIC-CXR using the EKAGen model [7]. *: $p < 0.05$ for pair-wise comparison between ablated variants and our full model (f) by the Wilcoxon signed-rank test.

Ablat. config.	Dispreferred response			Training objective			CE metrics			NLG metrics			RRG-exclusive metrics		
	Random	Clinic-relevant	Abnormal-aware	DPO	DPOP	RRG-DPO	Pre.	Rec.	F1	BL-4	MTR	RG-L	RadGraph	RadCliQ↓	RaTEScore
(a)	×	×	×	×	×	×	0.456*	0.367*	0.380*	0.119	0.156*	0.286	0.201*	2.685*	0.487*
(b)	✓	×	×	✓	×	×	0.413*	0.360*	0.359*	0.082*	0.154*	0.256*	0.142*	3.023*	0.474*
(c)	×	✓	×	✓	×	×	0.445*	0.425*	0.407*	0.101*	0.163	0.277*	0.186*	2.750*	0.478*
(d)	×	✓	✓	✓	×	×	0.441*	0.459*	0.420*	0.094*	0.168	0.267*	0.196*	2.713*	0.496
(e)	×	✓	✓	✓	✓	×	0.455	0.467*	0.427*	0.107	**0.172**	0.283	0.200*	2.687	0.496
(f)	×	✓	✓	✓	✓	✓	**0.462**	**0.487**	**0.443**	**0.112**	0.166	**0.286**	**0.207**	**2.674**	**0.502**

spectively. Notably, PromptMRG+RRG-DPO obtains the precision, recall, and F1 score of 0.508, 0.516, and 0.483, respectively, which, so far as we know, establish a new state-of-the-art on the MIMIC-CXR dataset among peer-reviewed literature. Meanwhile, RRG-DPO also improves all the RRG-exclusive metrics while maintaining the NLG performance of the base models. For 3D CT images, the RRG models include CT2Rep (encoder-decoder architecture with relational memory) [14] and 3D-CT-GPT (LLM-based) [9]. As shown in Table 1 bottom, our RRG-DPO significantly improves the CE metrics of both models, e.g., the precision, recall, and F1 score increase by 0.069, 0.159, and 0.134 on CT2Rep. It also improves all RRG-exclusive metrics. Meanwhile, the NLG metrics are comparable after preference optimization. To conclude, the experimental results of our RRG-DPO on both 2D CXR and 3D CT data and various models demonstrate its effectiveness in aligning RRG models with the clinical preference for diagnostically correct reports—without harming linguistic efficiency.

Ablation Study and Comparison with DPO and DPOP. To validate the efficacy of the proposed components in RRG-DPO, we conduct a comprehensive ablation study (Table 2). We compare the performance of several DPO variants using the EKAGen base model [7] (row (a)) on MIMIC-CXR. In row (b), the native DPO uses a random report in the training set as y_l, which degrades all metrics. We conjecture this is because the random report is easily distinguishable from the preferred response, adversely impacting the alignment. Rows (c) and (d) employ our clinical-relevant dispreference retrieval (using a random report from all retrieved ones as y_l) and abnormal-aware dispreferred response selection, both achieving notable improvements in recall and F1 score with minor decreases in precision. Regarding training objectives, row (e) employs the DPOP loss and obtains moderate improvements upon row (d) for most metrics. Finally, row (f), our full RRG-DPO approach, adds the proposed sub-preference optimization, achieving the best performance of all DPO variants (not including the non-DPO baseline) for eight of the nine metrics. Notably, all DPO variants except our full approach decrease the ROUGE-L metric. Our approach maintains the base model's NLG performance while substantially improving CE metrics. These results indicate that our clinical-relevant and abnormal-aware dispreference data curation can effectively improve the clinical efficiency of RRG models via DPO and that the proposed sub-preference optimization helps further improve clinical efficiency while maintaining NLG performance.

4 Conclusion

This work proposed RRG-DPO, a novel direct preference optimization procedure with a new loss term tailored for aligning radiology report generation (RRG) models with the preference for clinically accurate reports. RRG-DPO's effectiveness was validated on 2D and 3D radiography to align various RRG models.

Limitations & Future Work. Due to hardware constraints, this work only evaluated the proposed RRG-DPO on LLMs of moderate sizes (i.e., 7B). It would be helpful to experiment with significantly larger LLMs to study scalability. Based on preliminary experiments, this work employed BiomedVLP [5] for clinical-relevant dispreference retrieval and the classical CheXbert [35] to make abnormal-aware dispreference and sub-preference decisions. There is potential for better performance when using a more advanced retrieval model, and a classifier that can handle more abnormalities with greater accuracy.

Acknowledgments. This work was supported by National Natural Science Foundation of China (Grant No. 62371409) and Fujian Provincial Natural Science Foundation of China (Grant No. 2023J01005).

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

References

1. Bai, Y., et al.: Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862 (2022)
2. Banerjee, O., Zhou, H.Y., Wu, K., Adithan, S., Kwak, S., Rajpurkar, P.: Direct preference optimization for suppressing hallucinated prior exams in radiology report generation. In: Machine Learning for Healthcare Conference. PMLR (2024)
3. Banerjee, S., Lavie, A.: METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. pp. 65–72 (2005)
4. Bannur, S., et al.: Maira-2: Grounded radiology report generation. arXiv preprint arXiv:2406.04449 (2024)
5. Boecking, B., et al.: Making the most of text semantics to improve biomedical vision–language processing. In: ECCV. pp. 1–21. Springer (2022)
6. Brady, A., et al.: Discrepancy and error in radiology: concepts, causes and consequences. The Ulster Medical Journal **81**(1), 3 (2012)
7. Bu, S., et al.: Instance-level expert knowledge and aggregate discriminative attention for radiology report generation. In: CVPR. pp. 14194–14204 (2024)
8. Chaves, J.M.Z., et al.: Towards a clinically accessible radiology foundation model: open-access and lightweight, with automated evaluation. arXiv preprint arXiv:2403.08002 (2024)
9. Chen, H., et al.: 3D-CT-GPT: Generating 3D radiology reports through integration of large vision-language models. arXiv preprint arXiv:2409.19330 (2024)

10. Chen, Z., et al.: Cross-modal memory networks for radiology report generation. In: ACL-IJCNLP (Volume 1: Long Papers). pp. 5904–5914 (2021)
11. Chen, Z., Song, Y., Chang, T.H., Wan, X.: Generating radiology reports via memory-driven transformer. In: EMNLP. pp. 1439–1449 (2020)
12. Donahue, J., et al.: Long-term recurrent convolutional networks for visual recognition and description. In: CVPR. pp. 2625–2634 (2015)
13. Ethem Hamamci, I., et al.: A foundation model utilizing chest ct volumes and radiology reports for supervised-level zero-shot detection of abnormalities. arXiv e-prints pp. arXiv–2403 (2024)
14. Hamamci, I.E., Er, S., Menze, B.: CT2Rep: Automated radiology report generation for 3D medical imaging. In: MICCAI. pp. 476–486. Springer (2024)
15. Hein, D., et al.: Preference fine-tuning for factuality in chest X-ray interpretation models without human feedback. arXiv preprint arXiv:2410.07025 (2024)
16. Huang, Z., Zhang, X., Zhang, S.: KiUT: Knowledge-injected U-Transformer for radiology report generation. In: CVPR. pp. 19809–19818 (2023)
17. Jain, S., , et al.: Radgraph: Extracting clinical entities and relations from radiology reports. arXiv preprint arXiv:2106.14463 (2021)
18. Jin, H., Che, H., Lin, Y., Chen, H.: PromptMRG: Diagnosis-driven prompts for medical report generation. In: AAAI. vol. 38, pp. 2607–2615 (2024)
19. Johnson, A.E., et al.: MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data **6**(1), 317 (2019)
20. Kale, K., et al.: "Knowledge is power": Constructing knowledge graph of abdominal organs and using them for automatic radiology report generation. In: ACL. pp. 11–24 (2023)
21. Lee, S., Youn, J., Kim, H., et al.: CXR-LLAVA: a multimodal large language model for interpreting chest x-ray images. European Radiology pp. 1–13 (2025)
22. Lin, C.Y.: ROUGE: A package for automatic evaluation of summaries. In: Text Summarization Branches Out. pp. 74–81 (2004)
23. Liu, C., Tian, Y., Chen, W., Song, Y., Zhang, Y.: Bootstrapping large language models for radiology report generation. In: AAAI. vol. 38, pp. 18635–18643 (2024)
24. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: ICLR (2019)
25. Lu, J., Xiong, C., Parikh, D., Socher, R.: Knowing when to look: Adaptive attention via a visual sentinel for image captioning. In: CVPR. pp. 375–383 (2017)
26. Nicolson, A., Dowling, J., Anderson, D., Koopman, B.: Longitudinal data and a semantic similarity reward for chest x-ray report generation. Informatics in Medicine Unlocked **50**, 101585 (2024)
27. Nicolson, A., Dowling, J., Koopman, B.: Improving chest X-ray report generation by leveraging warm starting. Artificial Intelligence in Medicine **144**, 102633 (2023)
28. Ouali, Y., Bulat, A., Martinez, B., Tzimiropoulos, G.: CLIP-DPO: Vision-language models as a source of preference for fixing hallucinations in LVLMs. In: ECCV. pp. 395–413. Springer (2025)
29. Pal, A., et al.: Smaug: Fixing failure modes of preference optimisation with DPO-positive. arXiv preprint arXiv:2402.13228 (2024)
30. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: ACL. pp. 311–318 (2002)
31. Paszke, A., et al.: Pytorch: An imperative style, high-performance deep learning library. NeurIPS pp. 8024–8035 (2019)
32. Radford, A., et al.: Learning transferable visual models from natural language supervision. In: ICML. pp. 8748–8763. PmLR (2021)
33. Rafailov, R., et al.: Direct preference optimization: Your language model is secretly a reward model. NeurIPS **36** (2024)

34. Shen, H., Pei, M., Liu, J., Tian, Z.: Automatic radiology reports generation via memory alignment network. In: AAAI. vol. 38, pp. 4776–4783 (2024)
35. Smit, A., et al.: Combining automatic labelers and expert annotations for accurate radiology report labeling using BERT. In: EMNLP. pp. 1500–1519 (2020)
36. Song, X., Zhang, X., Ji, J., Liu, Y., Wei, P.: Cross-modal contrastive attention model for medical report generation. In: COLING. pp. 2388–2397 (2022)
37. Sun, G., et al.: STLLaVA-Med: Self-training large language and vision assistant for medical question-answering. arXiv preprint arXiv:2406.19973 (2024)
38. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: A neural image caption generator. In: CVPR. pp. 3156–3164 (2015)
39. Wang, J., Bhalerao, A., He, Y.: Cross-modal prototype driven network for radiology report generation. In: ECCV. pp. 563–579. Springer (2022)
40. Wang, Z., Liu, L., Wang, L., Zhou, L.: R2GenGPT: Radiology report generation with frozen LLMs. Meta-Radiology **1**(3), 100033 (2023)
41. Xiao, T., Shi, L., Liu, P., Wang, Z., Bai, C.: Radiology report generation via multi-objective preference optimization. arXiv preprint arXiv:2412.08901 (2024)
42. Yang, S., Wu, X., Ge, S., et al.: Radiology report generation with a learned knowledge base and multi-modal alignment. MedIA **86**, 102798 (2023)
43. Yang, S., Wu, X., Ge, S., Zhou, S.K., Xiao, L.: Knowledge matters: Chest radiology report generation with general and specific knowledge. MedIA **80**, 102510 (2022)
44. You, D., Liu, F., Ge, S., Xie, X., Zhang, J., Wu, X.: AlignTransformer: Hierarchical alignment of visual regions and disease tags for medical report generation. In: MICCAI. pp. 72–82. Springer (2021)
45. Yu, F., et al.: Evaluating progress in automatic chest x-ray radiology report generation. Patterns **4**(9), 100802 (2023)
46. Zhao, W., Wu, C., Zhang, X., Zhang, Y., Wang, Y., Xie, W.: RaTEScore: A metric for radiology report generation. arXiv preprint arXiv:2406.16845 (2024)
47. Zhu, K., Xia, P., Li, Y., Zhu, H., Wang, S., Yao, H.: MMedPO: Aligning medical vision-language models with clinical-aware multimodal preference optimization. arXiv preprint arXiv:2412.06141 (2024)