arXiv:2602.01200v1 [cs.CV] 1 Feb 2026

Med3D-R1：激励3D医学视觉语言模型进行异常诊断的临床推理

Haoran Lai[1,2,] Zihang Jiang[2,*] Kun Zhang[2] Qingsong Yao[3] RongshengWang[2]Zhiyang He[4] Xiaodong Tao[4] Wei Wei[5] S.Kevin Zhou[1,2,3,*]1 中国科学技术大学生物医学工程学院、生命科学与医学部，安徽省合肥市，230026，中国2 中国科学技术大学苏州高等研究院，江苏省苏州市，215123，中国3 斯坦福大学，美国加利福尼亚州帕洛阿尔托，94305[4] 科大讯飞股份有限公司医疗业务部，中国合肥，230088[5] 中国科学技术大学附属第一医院、生命科学与医学部，中国科学技术大学，安徽省合肥市，230001，中国{haoranlai, rongsheng wang}@mail.ustc.edu.cn{zyh, xdtao}@iflytek.com, jzh0103@ustc.edu.cn, s.kevin.zhou@gmail.com

摘要

开发具备稳健临床推理能力的三维视觉语言模型仍是一项挑战，原因在于体积医学影像本身复杂，模型容易过拟合报告中的表面模式，且缺乏具备可解释性的奖励设计。本文提出 *Med3D-R1*，这是一种采用两阶段训练流程的强化学习框架：监督微调（*SFT*）和强化学习（*RL*）。在监督微调阶段，我们引入残差对齐机制，以桥接高维三维特征与文本嵌入之间的差距，并采用异常加权策略，突出临床信息性词元，同时减少报告中的结构偏差。在强化学习阶段，我们重新设计一致性奖励，明确鼓励连贯、循序渐进的诊断推理。我们在两个三维诊断基准 *CT-RATE* 和 *RAD-ChestCT* 上，使用医学多项选择视觉问答任务评估了该方法。模型在 *CT-RATE* 和 *RAD-ChestCT* 上分别达到 *41.92 %* 和 *44.99%* 的最先进准确率。这些结果表明，该方法提升了异常诊断和临床推理能力，并在两个基准上均优于先前方法。总体而言，我们的方法有望通过构建更可靠、透明的三维医学视觉语言系统，改进真实世界中的诊断工作流程。

*通讯作者

1. 引言

随着体积医学影像数据日益普及，基于计算机断层扫描（CT）的三维诊断已成为临床决策中的一项关键任务 [4, 8, 21]。近年来，大语言模型（LLM）越来越多地被用于整合视觉与语言模态，以增强诊断推理能力，开创了医学人工智能的新范式 [15, 23]。在这一范式下 [22], ，通常采用主流的两阶段训练策略：首先，监督微调（SFT）用于对齐视觉和文本特征，同时赋予模型完成关键任务所需的专业医学知识；其次，强化学习（RL）进一步优化模型输出，以满足特定临床目标并提升其在现实世界中的适用性。

然而，尽管这一范式展现出良好的潜力，现有方法在应用于三维医学影像时仍存在若干关键局限。首先，与二维图像相比，三维体积数据在空间结构、尺度变化和语义粒度方面复杂得多 [17, 29]。这种复杂性显著增加了提取具有诊断显著性的特征并将其与相应文本表征对齐的难度。临床报告的特点进一步加剧了这一挑战：此类报告主要由高层次的抽象描述构成，且往往缺乏详细的空间标注。文本描述中的这种语义抽象加剧了高维视觉输入与抽象语言输出之间的不对齐，

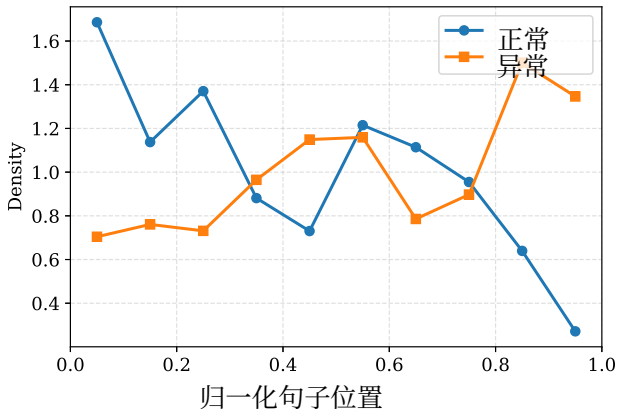


图1. 正常句子和异常句子在报告归一化位置上的密度分布。正常句子主要出现在报告开头，而异常句子则越来越集中于较后的位置，这揭示了放射学报告撰写中存在一致的位置偏差。

最终导致诊断推理中的优化不稳定和性能欠佳。

其次，我们发现，监督微调策略往往会过拟合表层模式，模型高度依赖训练集中频繁重复的公式化语言结构，而未能培养真正的推理能力。虽然正常病例与异常病例的整体类别平衡相对适中，但医学报告的结构组成存在一个更为关键的问题。如图1所示，正常发现主要集中在报告的前面部分，而异常发现则更常出现在后面部分。这种不对称分布揭示了一种固定的“正常优先”叙述模式：模板化的正常描述往往先于更多变的异常观察结果。这类报告惯例可能会隐性地促使模型一开始就默认作出正常预测，从而削弱其识别并推理那些打破既有模板、具有临床意义的异常的能力。

此外，当前的强化学习策略在很大程度上忽视了临床诊断背后的推理过程。现有的奖励函数 [15, 23]主要关注最终输出的正确性，而未对中间决策步骤进行建模，这些步骤本应反映放射科医生如何分析三维体数据并从中进行推断。这种缺乏具备可解释性的奖励设计，导致推理链的学习不够理想，并削弱了整体诊断一致性与可靠性。

为解决这些局限性，我们提出 Med3D-R1，这是一种统一的 3D 医学视觉语言建模框架，旨在显著增强临床推理能力。

在监督微调阶段，我们提出了两个关键模块：残差对齐机制（RAM），用于弥合高维 3D 特征与文本嵌入空间之间的表征差距；以及异常重加权（ARW）策略，通过优先处理具有临床重要性的词元，抵消放射学报告的结构化叙述风格所带来的偏差。为进一步增强模型的推理能力，我们引入了一个强化学习阶段，直接优化诊断正确性和连贯性。不同于以往的方法 [15, 23] ，我们的强化学习设计包含一种新颖的一致性奖励，使推理链与临床逻辑保持一致，从而实现更值得信赖的决策。总之，这些改进构成了一个协调统一的框架，显著提升了 3D 视觉语言模型的可解释性和诊断表现。

为验证我们的方法，我们在医学多选视觉问答（MMVQA）设定下，于CT-RATE [10] 和RAD-ChestCT [7] 两个诊断基准上进行了广泛实验。我们的模型在CT-RATE上达到41.92%的准确率，在RAD-ChestCT上达到44.99%的准确率，在两个基准上均取得最先进水平的结果。这些结果表明模型在异常诊断和临床推理方面具有较强能力，并进一步凸显了该方法支持真实世界诊断工作流程的潜力，即弥合原始体积数据与可解释的医学决策之间的差距。

我们的主要贡献总结如下：

- 据我们所知，我们首次指出，结构化报告中正常陈述和异常陈述的排列顺序会产生位置偏差，从而损害自回归训练，并提出*ARW* 来缓解这一问题。
- 我们首次将强化学习应用于3D 医学视觉语言模型，并引入一致性奖励，以促进基于临床依据且连贯的推理，从而确保推断更加可信。
- 我们的方法在CT-RATE和RAD-ChestCT上分别达到41.92%和44.99%的准确率，在MMVQA设定下创下新的最先进水平。

2. 相关工作

2.1. 医学视觉语言模型

视觉语言模型在多种医学影像任务中展现出巨大潜力，包括报告生成、视觉问答（VQA）和图文检索。早期工作（如 CheXzero [30] 和 CARZero [13] ）采用对比学习，将放射影像与对应报告对齐，主要针对胸部X光片等二维模态。在此基础上，近期研究通过将大语言模型与视觉编码器相结合，推动医学视觉语言模型向更全面的方向发展。代表性

模型，如 Med-Flamingo [20]，LLaVA-Med [16]，Med-PaLM [31]，MedGemma [25]，Lingshu [35]，和华佗GPT-Vision [5] 通过指令微调，将通用架构扩展至医学领域，从而实现更具表现力的理解、细粒度推理和多任务泛化。

尽管这些方法显著提升了医学领域的多模态理解能力，但它们主要聚焦于二维成像，或以表层视觉-语言对齐为核心的任务。与二维图像相比，计算机断层扫描等三维医学成像模态带来了新的挑战，包括高维度、空间冗余以及更高的诊断复杂性。为此，越来越多的研究开始探索如何在视觉-语言框架中有效建模并对齐三维数据。典型例子包括 RadFM [32]，M3D [2]，E3D-GPT [12]，CT-CHAT [10],和 Med3DVLM [33]，，这些研究证明了将视觉-语言建模拓展至三维空间的可行性，并在下游任务中取得了令人鼓舞的表现。

尽管近期取得了进展，当前的三维医学视觉语言模型在临床推理能力方面仍然有限。大多数现有方法仅依赖监督微调，这往往会导致模型对训练数据中的表层模式过拟合。这不仅削弱了底层大语言模型的语言理解能力，也未能培养结构化的诊断推理能力。因此，模型往往更注重记忆模式，而非真正的决策过程。相比之下，我们的工作迈出了重要一步：引入强化学习以明确增强推理能力，从而使模型在三维场景下能够做出更忠实、可解释且立足临床的预测。

2.2. 三维医学视觉-语言对齐

将三维医学图像与相应的文本语义对齐，仍是多模态学习中一个基础却具有挑战性的问题。近期的方法探索了多种策略，将体积特征投影到语言空间，通常依赖于从三维视觉嵌入到文本表示的直接映射。RadFM [32] 采用基于Qformer的架构，在对齐之前聚合三维特征。M3D [2] 引入空间平均池化以压缩体积特征，而E3D-GPT [12] 则通过自适应加权池化对此加以增强，以改进表示。Med3DVLM [33] 进一步应用 MLP-Mixer，以整合各切片间的空间信息与语义信息。除这些直接对齐策略外，一些近期工作还引入辅助机制来减轻学习负担。例如，3DHRG [19] 利用低分辨率体积引导高分辨率特征学习，而Med-2E3 [28]则利用二维图像-文本对齐为三维切片选择提供信息，从而提供一种形式的间接监督。

尽管这些方法在一定程度上有效，但通常会学习从高维图像特征到全文嵌入的直接投影。对于三维医学数据而言，这种映射可能不稳定或效率低下，因为这类数据往往稀疏、结构复杂且语义多样。为解决这一问题，我们提出一种残差对齐机制，引入可学习的文本锚点作为中间语义目标。模型不再直接与完整的文本表示对齐，而是学习将三维特征映射到相对于这些锚点的残差，从而有效地将对齐任务转化为局部细化过程。这种设计降低了优化难度，增强了语义可解释性，并有助于在三维场景中实现更准确的视觉-语言对齐。

2.3. 医学推理与可解释性

推理在医学决策中发挥着核心作用，模型不仅必须给出正确答案，还必须生成符合临床逻辑的解释。尽管监督微调一直是使视觉语言模型适应医学任务的主流方法，但它往往会助长捷径学习，无法捕捉临床工作流程的内在逻辑。这一问题在医学场景中尤为严重，因为可解释性和领域对齐至关重要。

为克服这些局限，近期研究引入了强化学习，以提升二维医学视觉语言模型的推理泛化能力。MedVLM-R1 [23] 和Med-R1 [15]等工作表明，奖励引导学习能够提升模型在不同模态和任务中的可解释性与鲁棒性。通过采用组相对策略优化（GRPO）[26, 27]，，即使没有明确的思维链监督，这些模型也能学会生成符合临床指南的结构化推理依据。GRPO强化学习并非只优化答案准确率，而是鼓励模型探索多样且正确、逻辑连贯的推理路径。

然而，这些进展迄今仍局限于胸部X光片或皮肤科图像等二维领域。据我们所知，我们的工作首次将基于强化学习的推理监督扩展到三维医学视觉语言场景。除了常用的格式奖励和准确率奖励外，我们还提出了一种新颖的一致性奖励，直接监督模型中间推理与参考放射学报告之间的语义对齐。这一设计有助于生成忠实、结构化且基于临床的解释，弥合体积图像解读与可靠诊断决策之间的鸿沟。

3. 方法

我们提出了 Med3D-R1，这是一种三维医学视觉语言模型，旨在通过三个互补模块同时提升诊断准确率和可解释性

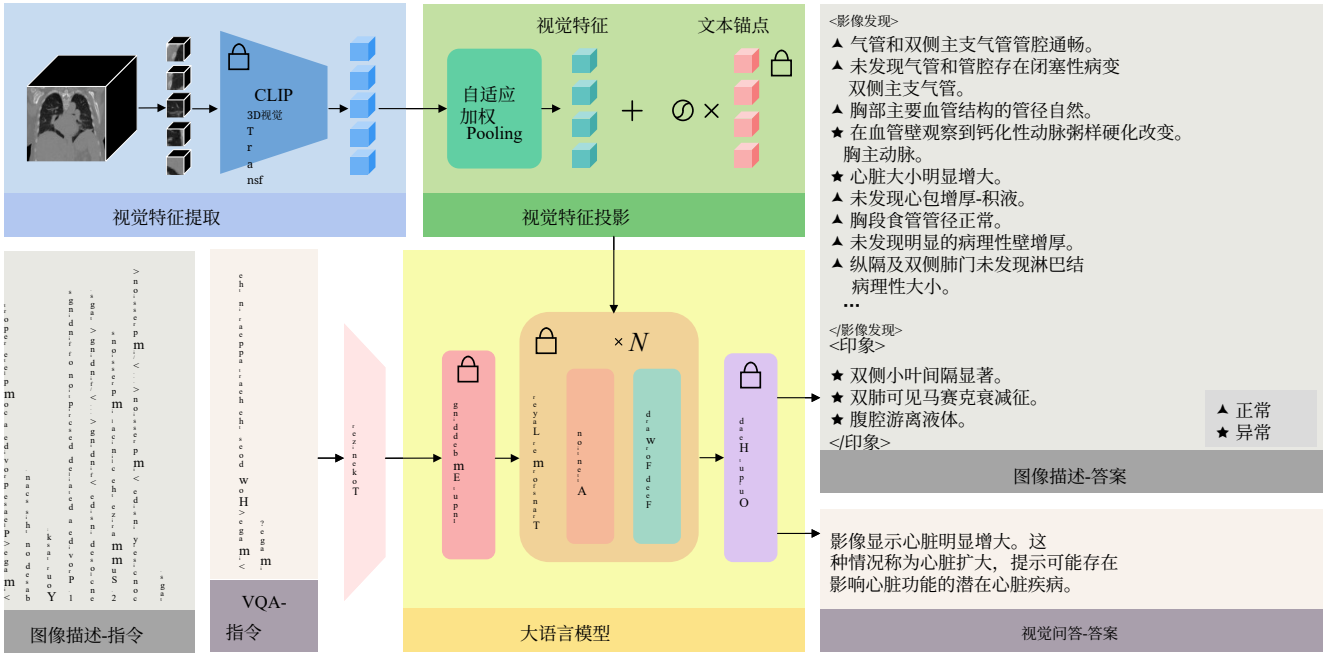


图2。监督微调框架概览。残差对齐机制优化视觉变换器（ViT）输出的三维图像特征与固定文本锚点之间的映射，从而增强跨模态对齐。异常加权策略进一步提高生成损失中具有临床意义的发现的权重。在此阶段，只有自适应加权池化模块可训练，其他所有组件均保持冻结。

首先，残差对齐机制（RAM）通过将图像特征插值至固定的文本锚点，增强跨模态对齐。其次，异常加权（ARW）策略在监督微调（SFT）期间重新平衡训练损失，以强调具有临床意义的异常发现，如图所示2。最后，在强化学习（RL）阶段，我们引入一致性引导奖励，明确鼓励准确且有临床依据的诊断推理。下面将详细介绍各个组成部分。

3.1. 残差对齐机制

三维CT体积包含丰富的解剖和病理信息，因此形成了高维且多样化的视觉特征空间。相比之下，临床报告相对稀疏，且变化较小，这使得直接对齐这两种模态颇具挑战。

为弥合这一差距，我们首先将自适应加权池化模块 [12]应用于三维视觉编码器提取的特征。该模块学习将空间分布的词元聚合为一组更紧凑的表示，从而捕捉最具诊断价值的线索。在此基础上，我们提出RAM，将对齐重新定义为学习图像特征与文本空间中固定语义锚点之间的差异，而不是直接映射。受残差学习启发 [11], , 其中$h(x) = x + f(x)$通常比$h(x) = f(x)$单独使用更容易优化，因此我们专注于残差变换，以降低连接三维图像嵌入与低维文本嵌入的难度。

具体而言，设 $E \in \mathbb{R}^{V\times D}$ 为从预训练语言模型的分词器获得的词嵌入权重矩阵。每一行向量 $E_j \in \mathbb{R}^D$ 是第 j 个词汇表词元的 D 维嵌入， V 是词汇表大小。我们计算一个单一的**文本锚点**

$$t_{\text{anchor}} = \frac{1}{V}\sum_{j=1}^{V} E_j \in \mathbb{R}^D, \tag{1}$$

即所有词汇嵌入的均值向量，并在整个训练集过程中保持固定。

对于每个投影后的图像词元 $x_i \in \mathbb{R}^D$，RAM 通过词元专属门控在 x_i 和 t_{anchor} 之间进行插值:

$$z_i = \sigma_i\, t_{\text{anchor}} + (1-\sigma_i)\, x_i, \tag{2}$$

$$\sigma_i = \text{Sigmoid}\big(\text{MLP}([x_i;\, t_{\text{anchor}}])\big), \tag{3}$$

其中， $\sigma_i \in [0,1]$ 决定了每个词元在多大程度上依赖语义锚点。这种残差融合使模型能够自适应地将视觉线索与领域无关的文本参考相结合，从而在三维视觉特征与临床语言表征之间实现更稳定的跨模态对齐和更强的语义一致性。

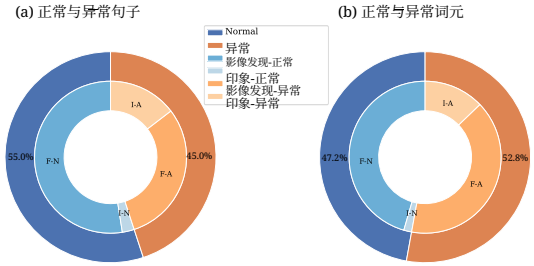


图 3. CT-RATE 报告中句子级和词元级标签的分布。每个环形图的外环展示正常 和异常 句子（左）及词元（右）的总体比例，表明这两类的数量大致相当。内环则进一步将每一类细分为影像发现 和 印象 两部分。

3.2. 异常加权

在 SFT 期间，模型通过预测序列中的下一个词元，学习将体积扫描与临床报告进行对齐。在结构化放射学报告中，无论是否存在影像发现，每个解剖区域都会按照固定顺序进行描述。因此，正常描述通常较早出现，而异常发现虽然至关重要，却往往在文本后部才出现。我们的句子级分析显示，正常陈述略多于异常陈述（如图 3 所示，在 CT-RATE 数据集中二者的比例约为 1:1.22），而自回归语言建模可能会进一步放大这一差异。因此，模型倾向于复现正常模式，却忽视临床上重要的异常。

为缓解这一问题，我们引入一种 ARW 策略。首先，将每份报告拆分为多个句子；随后使用 Deepseek-V3 [9]将这些句子标记为正常 或异常 。在监督微调期间，我们将属于异常句子的词元的损失乘以一个因子 $\lambda > 1$进行调整：

$$\mathcal{L} = -\sum_{t=1}^{T} \alpha_t \log P(\hat{y}_t = y_t), \tag{4}$$

$$\alpha_t = \begin{cases} \lambda, & \text{if } y_t \text{ is from an abnormal sentence,} \\ 1, & \text{otherwise,} \end{cases} \tag{5}$$

其中， α_t 表示词元级权重。该重新加权方案旨在提高模型对临床上有意义的异常的敏感度，从而抵消模型对报告开头常见的重复性正常句子模式过拟合的倾向。

3.3. 奖励推理一致性

尽管大多数面向医学视觉语言模型的基于强化学习的方法 [15,23] 侧重于生成结构化响应并实现较高的最终答案准确率，但它们很少验证模型的推理过程是否真正基于临床证据。为弥合这一差距，我们引入了一致性奖励 ，用于监督模型推理轨迹与相应放射学报告之间的语义对齐。

任务表述。 我们将异常诊断任务表述为医学多选题视觉问答（MMVQA）。给定一组三维CT扫描和一条临床提示，模型必须从四个选项（A/B/C/D）中选出一个正确答案。输出分为两部分：一段用 <think>...</think> 标签括起的逐步推理依据，随后是在 <answer>...</answer> 标签中的最终判断（详情见补充材料）。将推理与最终结论明确分开，有助于评估并提升可解释性。

一致性奖励。 在监督微调之后，我们加入一个强化学习阶段，以同时优化准确率和可解释性。标准监督学习虽然鼓励模型忠实于参考答案，却不会直接促进高质量的诊断推理。为此，我们定义了一种一致性奖励，用于衡量 <think> 推理过程与放射学报告的契合程度。预训练奖励模型（RM-Mistral-7B [34] ）会量化语义重叠；然而，由于该奖励模型是在通用领域数据上训练的，其直接评分可能反映出与医学领域并不完全一致的偏好。为了在利用该模型语言理解能力的同时，减轻此类偏见的影响，我们将其输出离散化为少量等级（例如 0、0.5、0.8、1.0）。这种离散化缓解了领域失配，并确保奖励信号保持稳健且符合临床实际。有关离散化的消融研究见补充材料。

总体目标。 强化学习阶段优化三项奖励之和：(i) 格式奖励 $r_{\text{fmt}} \in \{0, 1\}$，用于检查是否符合标签模板；(ii) 准确率奖励$r_{\text{acc}} \in \{0, 1\}$ ，用于验证所选选项；以及 (iii) 上文定义的一致性奖励 r_{con} 。总奖励为 $r = r_{\text{con}} + r_{\text{fmt}} + r_{\text{acc}}$， 我们使用 GRPO [26]对其进行最大化：采样多个响应 $\{o_k\} \sim \pi_\theta$ ，计算其经组归一化的优势，并通过 KL惩罚使 π_θ 接近冻结的参考策略 π_{ref}。

通过引入一致性奖励，我们的模型不仅能选出正确答案，还能生成有临床依据、逻辑连贯的推理过程，从而得出更值得信赖且更易于理解的诊断。

4. 实验

4.1. 数据集

CT-RATE [10] 公开发布的 CT-RATE 语料库包含来自 21,304 名患者的 50,188 个非增强胸部CT体积（其中 25,692 次扫描采用不同的卷积核重建了两次），每个体积均配有一份放射学报告。该语料库提供了 18 种自动提取的异常标签。我们遵循官方 划分：47,149 个体积用于训练集，3,039 个体积用于内部验证集。在监督微调阶段，我们联合使用报告生成任务和由 GPT-4 [1]生成的 713,150 个视觉问答样本进行训练。在强化学习阶段，诊断被设定为 MMVQA：每个查询包含一个正确的异常类别和三个随机干扰项。我们使用覆盖 5 % 患者（1,065 个病例）的子集进行强化学习，而官方验证集则用于评估报告生成和诊断推理。

RAD-ChestCT [7] RAD-ChestCT包含3,630个公开发布的胸部CT体积，每个均标注了84个异常标签。由于该数据集缺少放射学报告，它仅用于基于视觉输入评估诊断推理。为确保与CT-RATE的标签一致性，按照[14]中提出的映射协议，将RAD-ChestCT中的27种异常映射到CT-RATE中的18种异常。

4.2. 实现细节

我们首先遵循BrgSA[14], 训练一个三维CLIP [24] 模型，BrgSA是CLIP的三维扩展，旨在将CT扫描与其相关的放射学报告对齐。具体而言，我们采用三维ViT-base架构 [6] 作为视觉编码器，并在CT-RATE数据集上对其进行训练。训练完成后，我们仅保留预训练的3D视觉Transformer编码器，并将其用作后续三维医学视觉语言模型中的视觉主干网络。该视觉编码器与Qwen-2.5-3B[36]集成，以支持下游临床推理任务。所有训练与评估均在配备8块NVIDIA GPU的单台H20服务器上进行。

在第一阶段，我们首先对图注和视觉问答数据进行监督微调，仅更新投影层。这种方法在保留大语言模型语义能力的同时，增强了视觉文本对齐。 λ 在 ARW 中经实证设定为 1.10（见补充材料）。在第二阶段，我们在 MMVQA 数据集上引入强化学习，共同最大化三种奖励之和。联合优化这些奖励，可在单次训练过程中提升诊断正确性、模板合规性和具有临床依据的推理能力。

我们使用准确率评估 MMVQA 任务。对于报告生成，我们采用 BLEU、METEOR、ROUGE 和 CIDEr [18]。此外，我们还使用 DeepSeek-V3 [9]计算语义指标 BERT F1分数（BERT F1） [38] 和大语言模型得分（LLM Score）。

4.3. 与最先进方法的比较

根据输入数据，当前的视觉语言模型大致分为三类：（1）二维图像输入；（2）视频输入；（3）三维体积数据输入。本研究根据现有视觉语言模型支持的数据类型，为三维 CT 扫描设计了三种输入处理策略：

首先，对于**三维到二维转换**，我们设计了两种方法：一种是将三维CT扫描数据转换为**多切片拼图**，把多个切片拼接成一张二维图像，同时保留精细的解剖结构；另一种是利用数字重建放射影像（**DRR**），沿三个正交方向投影三维体积数据，模拟来自三个视角的 X 射线图像（详情见补充材料）。对于**视频输入**，我们将深度维度视为时间轴，把沿深度维度排列的 CT 切片组织成视频帧，模拟切片随深度增加而产生的动态变化。最后，**三维体积输入** 保留完整的三维结构信息，包括高度、宽度和深度，确保呈现完整的解剖细节。

如表 1所示，我们在这些不同的输入类型下评估了模型。结果表明，在二维输入场景中，除 MedGemma 外，采用拼图策略的模型表现更好。这主要是因为 DRR 在简化过程中损失了更多信息，而拼图策略保留了更多切片细节。然而，MedGemma 在 DRR 上表现良好，这可能得益于其在大量 X 射线数据上进行训练，因此处理 DRR 输入时具有优势。对于视频输入，尽管这种方法模拟了切片之间的动态关系，但与拼图策略相比，性能并未显著提升。这可能是由于视频输入的帧采样造成信息损失，限制了模型捕捉细粒度细节的能力。这表明，对于三维医学影像，拼图策略比 DRR 和视频输入都更有效，尤其是在捕捉空间细节方面。

此外，我们还对接受视频输入的 Qwen2.5-VL-3B 应用 GRPO 训练，将其命名为 Qwen2.5-VL-3B-RL，并与 Qwen2.5-VL-3B 模型进行比较。结果表明，应用强化学习后性能略有提升：CT-RATE 从 29.30 提高到 31.22，RAD-ChestCT 从 29.34 提高到 30.71。然而，与经过监督微调后已具备扎实医学知识基础的 Med3D-R1 (S1) 模型相比，性能提升幅度要显著得多。Med3D-R1 的性能大幅提高，CT-RATE 从 27.12 提升至 41.92，RAD-ChestCT 从 26.60 提升至 44.99。这一结果进一步验证，具备医学知识基础的模型能够更好地利用强化学习来提升推理能力，从而取得显著的性能增益。

我们的实验还揭示了现有

输入类型	方法	大语言模型骨干网络	CT-RATE		RAD-ChestCT	
			有思考	无思考	有思考	无思考
拼图	Qwen2.5-VL-3B [3]	Qwen-2.5-3B	29.50	27.35	28.48	26.89
	Qwen2.5-VL-7B [3]	Qwen-2.5-7B	29.23	28.98	26.35	26.54
	Qwen3-VL-4B-Instruct [37]	Qwen-3-4B	29.93	33.52	29.42	34.48
	Qwen3-VL-4B-Thinking [37]	Qwen-3-4B	28.00	32.77	30.85	34.88
	MedGemma-4B [25]	Gemma-3-4B	24.00	24.06	27.75	28.61
	MedGemma-1.5-4B [25]	Gemma-3-4B	23.07	21.71	25.60	24.40
	Lingshu-7B [35]	Qwen-2.5-7B	28.54	<u>31.88</u>	29.97	<u>35.79</u>
	Med-R1 [15]	Qwen-2-2B	29.32	27.96	<u>31.51</u>	29.99
	MedVLM-R1 [23]	Qwen-2-2B	26.58	26.13	25.47	26.11
DRR	Qwen2.5-VL-3B [3]	Qwen-2.5-3B	25.70	23.09	28.36	25.17
	Qwen2.5-VL-7B [3]	Qwen-2.5-7B	25.01	25.85	22.28	23.71
	Qwen3-VL-4B-Instruct [37]	Qwen-3-4B	25.85	27.86	23.46	25.16
	Qwen3-VL-4B-Thinking [37]	Qwen-3-4B	23.91	26.54	23.98	25.63
	MedGemma-4B [25]	Gemma-3-4B	26.76	27.10	30.90	28.85
	MedGemma-1.5-4B [25]	Gemma-3-4B	25.53	23.35	24.40	24.47
	Lingshu-7B [35]	Qwen-2.5-7B	26.32	28.14	26.46	28.48
	Med-R1 [15]	Qwen-2-2B	26.86	27.36	29.24	28.17
	MedVLM-R1 [23]	Qwen-2-2B	24.19	22.56	22.21	22.20
视频	Qwen2.5-VL-3B [3]	Qwen-2.5-3B	29.30	27.36	29.34	27.12
	Qwen2.5-VL-7B [3]	Qwen-2.5-7B	26.89	26.63	25.32	24.50
	Qwen3-VL-4B-Instruct [37]	Qwen-3-4B	26.34	28.65	28.74	30.30
	Qwen3-VL-4B-Thinking [37]	Qwen-3-4B	26.45	28.95	27.06	29.07
	Lingshu-7B [35]	Qwen-2.5-7B	26.17	29.14	29.25	35.19
	Qwen2.5-VL-3B-强化学习	Qwen-2.5-3B	<u>31.22</u>	28.73	30.71	28.78
体积	RadFM [32]	Llama-2-13B	–	25.77	–	25.00
	M3D [2]	Llama-2-7B	–	22.68	–	24.28
	E3D-GPT [12]	Vicuna-1.5-7B	–	25.29	–	25.15
	CT-CHAT [10]	Llama-3.1-8B	–	23.14	–	20.52
	Med3DVLM [33]	Qwen-2.5-7B	–	21.59	–	22.17
	Med3D-R1 (S1)	Qwen-2.5-3B	27.12	27.32	26.60	25.32
	Med3D-R1（S2）	Qwen-2.5-3B	**41.92**	**40.30**	**44.99**	**43.75**

表 1. 使用 CT-RATE 和 RAD-ChestCT 验证数据集时，MMVQA任务上的性能（%）对比。“有思考”和“无思考”表示在推理过程中是否应用推理指令。“S1”表示采用 SFT 的第一阶段训练，“S2”表示采用 RL 的第二阶段训练。“–”表示模型在“有思考”设置下无法生成有效的多项选择答案，因为它无法遵循推理格式提示。粗体数值表示最佳性能，下划线数值表示次佳性能。

三维体积上的 3D 医学视觉语言模型存在一个重要问题：过拟合，这会导致指令遵循能力下降。许多现有方法在大语言模型训练期间会出现过拟合，导致模型在处理分布外指令时表现不佳。模型输出无法准确遵循给定的提示，从而导致性能显著下降（详情请参阅补充材料）。为确保在这些条件下进行公平评估，我们表格中报告的结果经过后处理：选择与模型原始输出的余弦相似度最高的选项作为最终答案输出。这一问题凸显了当前 3D 医学视觉语言模型在遵循多模态指令方面的局限性，也彰显了我们的方法在应对这一挑战上的优势。

4.4. 消融研究

表2 对比了完整模型与三个消融变体，以诊断 ARW 和 RAM 的作用。完整系统在每项指标上均取得最高分；移除任一组件后，得分都会显著下降（$p < 0.05$ 或 $p < 0.005$），证实了这两个模块

RAM	ARW	B-1	B-2	B-3	B-4	METEOR	R-1	R-2	R-L	CIDEr	BERT F1	大语言模型得分均值
✓	✓	**40.93**	**30.98**	**25.19**	**21.55**	**36.18**	**56.63**	**34.50**	**39.84**	**23.86**	**88.70**	**80.48**	**43.53**
✓	×	38.51*	28.96*	23.40*	19.92*	35.16*	55.94*	33.59*	39.11*	20.22	88.47*	79.50†	42.07
×	✓	39.83*	29.89†	24.08†	20.42†	35.49*	56.19*	33.59†	38.83†	22.51	88.41†	79.49†	42.61
×	×	37.77†	28.07†	22.46†	18.97†	34.43†	55.19†	32.63†	37.79†	22.07*	88.21†	78.48†	41.46

表2. 报告生成中所提模块的消融研究（指标以 % 表示）。缩写：B-n = BLEU-n；R-n = ROUGE-n。 * 表示 $p < 0.05$，而 † 表示与同时包含两个模块的完整模型相比的 $p < 0.005$，采用配对t检验。

Bridge	LoRA	CT-RATE		RAD-ChestCT	
	LLM	有思考	无思考	有思考	无思考
✓	×	27.12	27.32	26.60	25.32
✓	✓	–	**27.81**	–	**25.46**

表3. 不同训练模块下监督微调在MMVQA任务上的表现（%）。

发挥互补而非冗余的作用。

移除 ARW 并保留 RAM 时，绝对降幅最大的是 CIDEr（–3.64）。由于 CIDEr 会为低频且具有临床特异性的 n 元语法赋予更高权重，这一显著下降表明，许多罕见的异常术语如今被遗漏或生成错误。这一观察结果支持 ARW 的预期作用：通过在训练集中过度加权异常句子，它有助于解码器保留详细的病理描述，而这些描述原本可能得不到充分体现。

相反，移除RAM但保留ARW，会导致基于嵌入的指标BERT F1（–0.29）和大语言模型得分（–0.99）降幅最大。这些指标评估整份报告的语义保真度和事实一致性，因此它们对缺少RAM十分敏感，这表明RAM主要保障全局视觉-语言对齐。禁用这两个模块会使所有得分降至最低，证实稳健的报告生成有赖于ARW对异常的聚焦作用与RAM的整体跨模态基础对齐作用共同发挥。

为验证将监督微调仅限于视觉—语言桥接模块是否足够，我们进行了表3中总结的实验。仅微调桥接模块即可保留模型的推理能力：在 CT-RATE 和 RAD-ChestCT 上，启用推理时的准确率分别为 27.1% 和 26.6%，而直接作答的得分仍大致相当。当我们进一步使用 LoRA 调整大语言模型权重时，推理能力消失，直接作答的准确率最多仅提高 0.5%——这表明，在训练数据有限的情况下，改变大语言模型的参数对指令遵循能力的损害，超过了其带来的性能提升。因此，我们将监督微调限定为仅调整桥接模块，并将完整模型的更新留到后续强化学习阶段。

4.5. 一致性奖励的有效性

表4 总结了两位互补评审者的评估结果：一位拥有十年临床经验、取得委员会认证的放射科医生，以及大语言模型 Deepseek-V3。他们分别独立评估了模型的推理质量。两位评审者均从四个维度对推理进行评分：临床正确性、清晰度、鉴别诊断和简洁性。

引入一致性奖励后，两位评审者在各个维度上的模型表现均有所提升。放射科医生给出的分数提高了 +0.21至 +0.40，其中临床正确性提升最大（+0.40）。Deepseek-V3报告的提升幅度更大，达到 +0.21至 +0.90，其中鉴别诊断 提升最大（+0.90）。

所有提升均具有统计显著性（$p < 0.0001$），表明一致性奖励能够使推理更加准确、清晰，并且更符合临床实际。Deepseek-V3 使用的完整评估提示见补充材料。

评判员	一致性	临床	清晰度	鉴别诊断	简洁性
放射科医师	×	3.13	3.14	3.08	3.26
放射科医师	✓	**3.53***	**3.48***	**3.35***	**3.50***
Deepseek	×	3.53	3.53	2.64	3.69
Deepseek	✓	**4.29***	**4.08***	**3.54***	**3.90***

表4. 不同奖励策略下的评估得分。缩写：临床（临床正确性）、清晰度（推理清晰度）、鉴别诊断（鉴别诊断）、简洁（简洁性）。所有指标的范围均为0至5，并由取得委员会认证的放射科医生和Deepseek-V3独立评估。 * 表示与基线相比的 $p < 0.0001$。

5. 局限性

本研究有两个主要局限性。首先，训练数据在规模、解剖覆盖范围和多样性方面有限：我们主要依赖公开胸部CT–报告对。虽然合成视觉问答增加了文本多样性，但它并未扩大图像池、拓宽器官/区域覆盖范围，也未能捕捉多中心异质性，这可能限制鲁棒性和分布外泛化。在未来工作中，我们

将整理更大规模的多中心CT-报告队列，量化数据规模效应，并通过收集远超胸部范围的大量图像（及报告），系统性地扩展到更多器官和成像模态。

其次，临床推理方面的监督仍然不足。模型的推理轨迹主要体现的是内在语言能力，而非由临床医生撰写、可审查的证据链；奖励模型也主要面向通用领域，因此可能生成流畅却缺乏充分证据支撑的解释。今后，我们将与放射科医生合作，构建高质量、基于评分标准的推理依据，并用其微调奖励模型，同时约束 SFT+强化学习训练，从而提升推理输出的临床忠实度与可解释性。

6. 结论

我们提出了 Med3D-R1，这是首个运用强化学习来增强临床推理能力的 3D视觉语言系统。RAM 和 ARW 在监督微调期间缓解了体积数据偏差，而强化学习中的一致性奖励则进一步确保诊断连贯且符合临床实际。在 CT-RATE 和 RAD-ChestCT 上，Med3D-R1 达到新的最先进水平，并生成逻辑性更强、可解释性更高的报告，展现了在三维医学影像领域迈向透明、可信人工智能的切实路径。

参考文献

[1] Josh Achiam、Steven Adler、Sandhini Agarwal、Lama Ah- mad、Ilge Akkaya、Florencia Leoni Aleman、Diogo Almeida、Janko Altenschmidt、Sam Altman、Shyamal Anadkat 等。GPT-4 *08774*，2023年*23036*技术报告。*arXiv* 预印本 *arXiv:* 。。[2] Fan Bai、Yuxin Du、Tiejun Huang、Max Q-H Meng 和 Bo Zhao。M3D：借助多模态大语言模型推进3D医学图像分析。*arXiv* 预印本 *arXiv:2404.00578*，2024年。3，7[3] Shuai Bai、Keqin Chen、Xuejing Liu、Jialin Wang、Wenbin Ge、Sibo Song、Kai Dang、Peng Wang、Shijie Wang、Jun Tang、Humen Zhong、Yuanzhi Zhu、Mingkun Yang、Zhao- hai 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 和 Jun- yang Lin。Qwen2.5-VL 技术报告。*arXiv* 预印本 *arXiv:2502.13923*，2025年。7[4] Kaustav Bera、Kurt A Schalper、David L Rimm、Vamsidhar Velcheti 和 Anant Madabhushi。数字病理学中的人工智能——用于诊断和精准肿瘤学的新工具。《*Nature Reviews Clinical Oncology*》，16 715，2019年 11 7031( ): – 。[5] Junying Chen、Chi Gui、Ruyi Ouyang、Anningzhe Gao、Shu- nian Chen、Guiming Chen、Xidong Wang、Zhenyang Cai、Ke Ji、Xiang Wan 等。致力于大规模向多模态大语言模型注入医学视觉知识。载于*2024* 年自然语言处理实证方法会议论文集，第7346–7370页，2024年。3

[6] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Syl- vain Gelly 等。《一张图像胜过 16x16 个词：用于大规模图像识别的变换器》。arXiv 预印本 arXiv:2010.11929，2020年。6[7] Rachel Lea Draelos、David Dov、Maciej A Mazurowski、Joseph Y Lo、Ricardo Henao、Geoffrey D Rubin 和 Lawrence Carin。基于机器学习的多种异常预测：大规模胸部计算机断层扫描体积。《医学图像分析》，67:101857，2021 年。2, 6[8] Daniel Thomas Ginat 和 Rajiv Gupta。计算机断层扫描成像技术的进展。《生物医学工程年评》，16(1):431–453，2014年。1[9] Daya Guo、Dejian Yang、Haowei Zhang、Junxiao Song、Ruoyu Zhang、Runxin Xu、Qihao Zhu、Shirong Ma、Peiyi Wang、Xiao Bi 等。Deepseek-r1：通过强化学习激励大语言模型的推理能力。arXiv 预印本 arXiv:2501.12948，2025 年。5, 6[10]Ibrahim Ethem Hamamci、Sezgin Er、Furkan Almas、Ayse Gulnihan Simsek、Sevval Nil Esirgun、I rem Dogan、Muhammed Furkan Dasdelen、Omer Faruk Durugol、Bastian Wittmann、Tamaz Amiranashvili 等。基于多模态数据集开发用于三维计算机断层扫描的通用基础模型。arXiv 预印本 arXiv:2403.17834，2024年。2, 3, 6, 7[11] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。载于《IEEE计算机视觉与模式识别会议论文集》，第770–778页，2016年。4[12] Haoran Lai、Zihang Jiang、Qingsong Yao、Rongsheng Wang、Zhiyang He、Xiaodong Tao、Wei Wei、Weifu Lv 和 S Kevin Zhou。E3d-gpt：面向医学视觉语言模型的增强型三维视觉基础。arXiv 预印本 arXiv:2410 14200，2024 3, 4, 7 . . [13] Haoran Lai、Qingsong Yao、Zihang Jiang、Rongsheng Wang、Zhiyang He、Xiaodong Tao 和 S Kevin Zhou。Carzero：用于放射学零样本分类的交叉注意力对齐。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第 11137 11146页，2024 2 – . [14] Haoran Lai、Zihang Jiang、Qingsong Yao、Rongsheng Wang、Zhiyang He、Xiaodong Tao、Wei Wei、Weifu Lv 和 S Kevin Zhou。用于零样本三维医学图像诊断的桥接语义对齐。arXiv 预印本 arXiv:2501.03565，2025年。6[15] Yuxiang Lai、Jike Zhong、Ming Li、Shitian Zhao 和 Xi- aofeng Yang。Med-r1：用于视觉语言模型中可泛化医学推理的强化学习。arXiv 预印本 arXiv:2503.13939，2025年。1, 2, 3, 5, 7[16] Chunyuan Li、Cliff Wong、Sheng Zhang、Naoto Usuyama、Haotian Liu、Jianwei Yang、Tristan Naumann、Hoifung Poon 和 Jianfeng Gao。Llava-med：一天内训练生物医学领域的大语言模型与视觉助手。《神经信息处理系统进展》，36，2024年。3[17] Jingyang Lin、Yingda Xia、Jianpeng Zhang、Ke Yan、Le Lu、Jiebo Luo 和 Ling Zhang。Ct-glip：利用CT扫描和放射学报告进行三维定位语言—图像预训练，面向

全身场景。arXiv 预印本 arXiv:2404.15272，2024年。1
[18] Tsung-Yi Lin、Michael Maire、Serge Belongie、J ames Hays、Pietro Perona、Deva Ramanan、Piotr Doll´ ar 和 C Lawrence Zitnick。Microsoft coco：情境中的常见物体。载于《计算机视觉—ECCV 2014：第13届欧洲会议论文集》，瑞士苏黎世，2014年9月6日至12日，第V部分第13卷，第740–755页。Springer，2014年。6[19] Che Liu、Zhongwei Wan、Yuqi Wang、Hui Shen、Haozhe Wang、Kangyu Zheng、Mi Zhang 和 Rossella Arcucci。针对三维高分辨率医学图像的放射学报告生成基准评估与性能提升。arXiv 预印本 arXiv:2406.07146，2024年。3[20] Michael Moor、Qian Huang、Shirley Wu、Michihiro Ya- sunaga、Yash Dalmia、Jure Leskovec、Cyril Zakka、Ed- uardo Pontes Reis 和 Pranav Rajpurkar。Med-flamingo：多模态医学少样本学习器。载于《健康机器学习》（ML4H），第353–367页。PMLR，2023年。3[21] Evangelos K Oikonomou、Mohamed Marwan、Milind Y Desai、Jennifer Mancio、Alaa Alashi、Erika Hutt Centeno、Sheena Thomas、Laura Herdman、Christos P Kotanidis、Katharine E Thomas 等。利用计算机断层扫描无创检测冠状动脉炎症并预测残余心血管风险（CRISP CT 研究）：对前瞻性结局数据的事后分析。《柳叶刀》，392(10151):929–939，2018年。1[22] Long Ouyang、Jeffrey Wu、Xu Jiang、Diogo Almeida、Car- roll Wainwright、Pamela Mishkin、Chong Zhang、Sandhini Agarwal、Katarina Slama、Alex Ray 等。训练语言模型遵循指令并利用人类反馈。《神经信息处理系统进展》，35:27730–27744，2022年。1[23] Jiazhen Pan、Che Liu、Junde Wu、Fenglin Liu、Jiayuan Zhu、Hongwei Bran Li、Chen Chen、Cheng Ouyang 和 Daniel Rueckert。Medvlm-r1：通过强化学习激励视觉语言模型（vlms）的医学推理能力。arXiv 预印本 arXiv:2502.19634，2025年。1, 2, 3, 5, 7[24]Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark 等。从自然语言监督中学习可迁移的视觉模型。载于《国际机器学习会议》，第8748–8763页。PMLR，2021年。6[25]Andrew Sellergren、Sahar Kazemzadeh、Tiam Jaroen- sri、Atilla Kiraly、Madeleine Traverse、Timo Kohlberger、Shawn Xu、Fayaz Jamil、C´ıan Hughes、Charles Lau 等。Medgemma 技术报告。arXiv 预印本 arXiv:2507.05201，2025年。3, 7[26] Zhihong Shao、Peiyi Wang、Qihao Zhu、Runxin Xu、Junxiao Song、Xiao Bi、Haowei Zhang、Mingchuan Zhang、YK Li、Y Wu 等。Deepseekmath：突破开放式语言模型数学推理能力的极限。arXiv 预印本 arXiv:2402.03300，2024年。3, 5[27] Haozhan Shen、Zilun Zhang、Kangjia Zhao、Qianqian Zhang、Ruochen Xu 和 Tiancheng Zhao。Vlm-r1：稳定且可泛化的 R1 风格大型视觉语言模型。https:
//github.com/om- ai- lab/VLM- R1，2025年。访问日期：2025年2月15日。3[28] Yiming Shi、Xun Zhu、Ying Hu、Chenyi Guo、Miao Li 和 Ji Wu。Med-2E3：一种二维增强型三维医学多模态大语言模型。arXiv 预印本 arXiv:2411.12783，2024年。3[29] Zhongyi Shui、Jianpeng Zhang、Weiwei Cao、Sinuo Wang、Ruizhe Guo、Le Lu、Lin Yang、Xianghua Ye、Tingbo Liang、Qi Zhang 等。通过大规模、细粒度的视觉语言预训练增强 CT 图像理解。arXiv 预印本 arXiv:2501.14548，2025年。1[30] Ekin Tiu、Ellie Talius、Pujan Patel、Curtis P Langlotz、Andrew Y Ng 和 Pranav Rajpurkar。通过自监督学习，从未标注的胸部X光图像中检测病理，达到专家水平。《自然-生物医学工程》，6(12)：1399–1406，2022年。2[31] Tao Tu、Shekoofeh Azizi、Danny Driess、Mike Schaekermann、Mohamed Amin、Pi-Chuan Chang、Andrew Carroll、Charles Lau、Ryutaro Tanno、Ira Ktena 等。迈向通用生物医学人工智能。NEJM AI，1(3)：AIoa2300138，2024年。3[32] Chaoyi Wu、Xiaoman Zhang、Ya Zhang、Hui Hui、Yanfeng Wang 和 Weidi Xie。利用网络规模的二维和三维医学数据，迈向放射学通用基础模型。《自然·通讯》，16(1)：7866，2025年。3, 7
[33] Yu Xin、Gorkem Can Ates、Kuang Gong 和 Wei Shao。Med3DVLM：一种用于三维医学图像分析的高效视觉语言模型。arXiv 预印本 arXiv:2503.20047，2025年。3, 7[34] Wei Xiong、Hanze Dong、Chenlu Ye、Ziqi Wang、Han Zhong、Heng Ji、Nan Jiang 和 Tong Zhang。基于人类反馈的迭代偏好学习：在 KL 约束下弥合强化学习人类反馈理论与实践之间的鸿沟。国际机器学习会议，2024年。5[35] Weiwen Xu、Hou Pong Chan、Long Li、Mahani Aljunied、Ruifeng Yuan、Jianyu Wang、Chenghao Xiao、Guizhen Chen、Chaoqun Liu、Zhaodonghui Li 等。Lingshu：用于统一多模态医学理解与推理的通用基础模型。arXiv 预印本 arXiv:2506 07044，2025年。3, 7。[36] An Yang、Baosong Yang、Beichen Zhang、Binyuan Hui、Bo Zheng、Bo wen Yu、Chengyuan Li、Dayiheng Liu、Fei Huang、Haoran Wei 等。Qwen2.5 技术报告。arXiv 预印本 arXiv:2412.15115，2024年。6[37]An Yang、Anfeng Li、Baosong Yang、Beichen Zhang、Binyuan Hui、Bo Zheng、Bowen Yu、Chang Gao、Chengen Huang、Chenxu Lv 等。Qwen3 技术报告。arXiv 预印本 arXiv:2505.09388，2025年。7[38] Tianyi Zhang、Varsha Kishore、Felix Wu、Kilian Q Weinberger 和 Yoav Artzi。BERTScore：使用 BERT 评估文本生成。arXiv 预印本 arXiv:1904.09675，2019年。6