

- 论文名称: VRAG-RL: Empower Vision-Perception-Based RAG for Visually Rich Information Understanding via Iterative Reasoning with Reinforcement Learning
- 第一作者: 阿里
- 论文链接: <https://arxiv.org/abs/2505.22019v2>
- 最新日期: 2025年6月3日
- github: <https://github.com/Alibaba-NLP/VRAG.git>

VRAG-RL: Empower Vision-Perception-Based RAG for Visually Rich Information Understanding via Iterative Reasoning with Reinforcement Learning

Qiuchen Wang, Ruixue Ding, Yu Zeng, Zehui Chen, Lin Chen
Shihang Wang, Pengjun Xie, Fei Huang, Feng Zhao⁺

 Tongyi Lab, Alibaba Group

Part 1 : VRAG- RL 算法原理

一、研究背景

这篇论文的核心内容是提出了一种名为**VRAG-RL**的新型强化学习（RL）框架，旨在解决传统检索增强生成（RAG）方法在处理丰富视觉信息（如包含图表、表格、复杂布局的文档）时的局限性。

1. 问题

论文指出了现有多模态RAG方法面临的三个具体挑战：

- 视觉信息下的推理能力激活不足**：先前的多模态RAG方法通常只是简单地将图像嵌入到上下文中，没有专门的机制来处理视觉特有的感知任务（如关注高信息密度区域），导致VLMs的推理潜力未能被充分利用。
- 低效和脱节的检索过程**：模型在与搜索引擎交互时，由于无法精确表达复杂需求，常常导致检索到的信息不相关，从而进行重复、无效的交互，限制了整体效果。
- VLM多轮推理与训练的不稳定性**：现有的用于语言模型的RL框架在处理复杂的多步推理任务时常出现不稳定，模型难以在多轮交互中保持有效的推理，这一挑战在能力有限的VLM上更为突出。

2. 方法

VRAG-RL框架下的智能体，会自己制定策略：它会先对整个文档进行一个粗略的浏览，然后决定：我是应该用搜索引擎查找更多资料，还是应该聚焦到这张图表的某个特定区域，把它放大看清楚？

其核心创新包括：

- 1 视觉感知动作空间**：定义了一个新的动作空间，允许模型对检索到的图像进行**区域选择、裁剪和缩放**，从而实现从粗到细的信息获取。
- 2 细粒度奖励函数**：设计了一个综合奖励函数，它不仅评估最终答案的正确性（基于模型的奖励），还整合了**检索过程的效率和推理模式的一致性**。
- 3 迭代推理框架**：模型通过“**思考-动作-观察**”的范式与环境交互，并使用**GRPO（Group Relative Policy Optimization）**算法进行训练，优化其多步推理能力。



Question: Which H2 antagonist has a lower bioavailability: the drug that inhibits several CYP-450 isoenzymes and reduces hepatic blood flow or Famotidine?



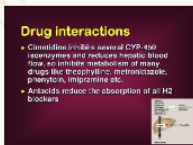
<think> To answer this question, I need to identify the H2 antagonist that inhibits several CYP-450 isoenzymes and reduces hepatic blood flow. Then, I will compare its bioavailability with that of Famotidine.

</think>

<search> H2 antagonist that inhibits several CYP-450 isoenzymes **</search>**



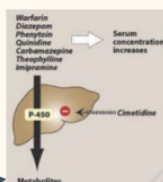
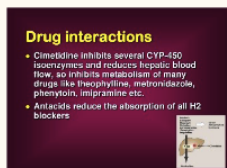
Image Retrieval



<think> The image provides information about drug interactions, specifically mentioning Cimetidine's effect on CYP-450 isoenzymes and hepatic blood flow. This aligns with the search query. I need to look at the lower right corner of the image to confirm this.

</think>

<region> [556, 337, 722, 505] **</region>**



Select & Crop & Zoom in



<think> The diagram confirms that Cimetidine inhibits the P-450 enzymes, leading to increased serum concentrations of drugs metabolized by these enzymes. Now, I need to find information on the bioavailability of Cimetidine and Famotidine to determine which has lower bioavailability.

</think>

<search> bioavailability of Cimetidine and Famotidine **</search>**



Image Retrieval

Comparison of H2 antagonists				
	Cimetidine	Famotidine	Ranitidine	Nizatidine
Bioavailability (%)	80	40	80	80
Relative Potency (nM)	1.5-2.5	1.5-2.5	2.5-4	1.1-1.8
IC50 (nM)	5	5	15	5
Antagonist of	1	1.7	2	9
CYP450				
Clearing (ml/min)	400	100	20	500



<think> Based on the provided table, Cimetidine has a bioavailability of 80%, while Famotidine has a bioavailability of 40%. Therefore, Famotidine has a lower bioavailability compared to Cimetidine. The drug that inhibits several CYP-450 isoenzymes and reduces hepatic blood flow is Cimetidine, as indicated in previous information.

</think>

<answer> Famotidine has a lower bioavailability (40%) compared to Cimetidine (80%). Therefore, the H2 antagonist with lower bioavailability is Famotidine. **</answer>**

Figure 8: Case 2 for VRAG-RL.

Question: 哪种 H2 受体拮抗剂的生物利用度更低：能够抑制多种 CYP-450 同工酶并降低肝血流量的药物，还是法莫替丁？

为回答这个问题，我需要先确定那种能抑制多种 CYP-450 同工酶并降低肝血流量的 H2 受体拮抗剂。然后，我将其与法莫替丁的生物利用度进行比较。

抑制多种 CYP-450 同工酶的 H2 受体拮抗剂

这张图片提供了药物相互作用的信息，特别提到西咪替丁对 CYP-450 同工酶和肝血流量的影响。这与搜索内容一致。我需要查看图像右下角以确认这一点。

[556, 337, 722, 505]

图示证实西咪替丁抑制 P-450 酶，导致这些酶代谢的药物血清浓度升高。接下来，我需要查找西咪替丁和法莫替丁的生物利用度，以确定哪一种更低。

西咪替丁和法莫替丁的生物利用度

根据所给表格，西咪替丁的生物利用度为 80%，而法莫替丁为 40%。因此，法莫替丁的生物利用度低于西咪替丁。此前信息表明，能够抑制多种 CYP-450 同工酶并降低肝血流量的药物是西咪替丁。

法莫替丁的生物利用度更低（40%），西咪替丁为 80%。因此，生物利用度更低的 H2 受体拮抗剂是法莫替丁。

3. 结果

在SlideVQA、ViDoSeek和MMLongBench等多个具有挑战性的基准测试上，VRAG-RL框架取得了显著成果。

- 在使用Qwen2.5-VL-7B模型时，其总体性能（57.1%）远超基线模型Search-R1-VL（37.4%）；
- 在使用Qwen2.5-VL-3B模型时，性能提升更为显著，从21.3%提升至53.5%，展示了该方法的有效性。

Table 1: **Main Results.** The best performance are marked in bold. SlideVQA and ViDoSeek mainly focus on reasoning type, while MMLongBench focuses on the visual type of reference content. OCR-based (🔗) RAG and purely visual (👁️) RAG are evaluated with the same prompt and setting.

Method	SLIDEVQA		ViDoSEEK		MMLongBench					Overall
	Single-hop	Multi-hop	Extraction	Logic	Text	Table	Chart	Figure	Layout	
Qwen2.5-VL-3B-Instruct										
🌀 Vanilla RAG	15.1	12.1	8.8	14.3	3.9	5.1	1.7	3.1	2.5	11.2
🌀 ReAct	11.8	9.9	5.3	7.4	6.5	3.7	3.9	5.2	2.5	8.4
🌀 Search-R1	17.5	13.8	13.3	20.7	3.4	3.2	4.5	4.1	6.8	14.1
👁 Vanilla RAG	19.4	12.2	10.1	17.3	2.2	4.1	5.2	4.7	4.3	13.2
👁 ReAct	15.7	10.9	6.7	14.2	2.7	3.6	3.4	3.1	5.1	10.9
👁 Search-R1-VL	26.3	20.1	20.1	29.8	8.5	7.8	7.9	9.3	7.6	21.3
👁 VRAG-RL	65.3	38.6	63.1	73.8	22.7	16.1	21.9	21.4	19.5	53.5
Qwen2.5-VL-7B-Instruct										
🌀 Vanilla RAG	26.1	10.6	24.7	30.9	8.5	5.4	11.7	4.4	3.3	20.9
🌀 ReAct	21.2	13.3	14.3	21.3	5.9	5.1	7.3	5.5	1.7	15.8
🌀 Search-R1	28.4	19.7	20.8	30.6	9.9	6.0	7.9	10.1	5.9	22.2
👁 Vanilla RAG	29.1	17.4	26.4	41.3	13.1	14.7	15.9	4.3	7.6	24.2
👁 ReAct	34.8	20.4	27.5	42.1	10.1	12.4	10.2	6.2	7.1	26.9
👁 Search-R1-VL	48.3	42.3	40.5	50.3	19.9	13.4	12.9	11.4	10.2	37.4
👁 VRAG-RL	69.3	43.1	60.6	74.8	26.1	26.3	24.8	25.9	21.2	57.1

三个基准测试 (Benchmark)：

- **SlideVQA**：逻辑推理专项考试，考察在结构化文档（如PPT）中的深度理解和推理能力。
- **ViDoSeek**：信息检索与定位专项考试，考察在海量文档中“大海捞针”并找到唯一答案的能力。
- **MMLongBench**：视觉感知与长文本理解综合考试，考察在复杂的长文档中处理多种视觉元素的能力。

Benchmark	核心目标	测试的关键能力	对VRAG-RL的意义
SlideVQA	理解结构化的PPT幻灯片	布局理解、多跳逻辑推理	检验智能体的推理能力
ViDoSeek	在海量文档中检索并回答	高效准确的检索、定位信息	检验智能体的RAG检索效率
MMLongBench	理解长篇幅、多模态文档	长上下文、多样的视觉感知（图表、表格等）	检验其 <region> 视觉感知动作的有效性

二、方法论

作者采用了一套基于强化学习的综合方法来解决上述问题，该方法被称为**VRAG-RL**。

其详细步骤和设计如下：

1. 整体框架：迭代推理与强化学习

论文将VLM与环境的交互建模为一个迭代推理和工具调用的过程，如图1(a)所示。模型（Policy Model）根据历史交互轨迹 H_{t-1} 生成“思考”和“动作” A_t ，与外部环境（包括搜索引擎和视觉感知功能）交互后获得“观察” O_t 。