

PR #2044 完整报告

THUDM/slime

support rich image config for vlm

合并时间: 2026-06-10 14:06

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2044>

执行摘要

- 一句话: VLM 支持丰富图像配置
- 推荐动作: 值得精读。该 PR 展示了如何以最小侵入方式扩展多模态输入格式, 设计简洁。后续可考虑添加单元测试覆盖新分支。

功能与动机

目前 VLM 仅支持简单的图像路径 /URL/base64 字符串列表格式, 无法满足 Qwen3-VL 等模型对每张图像精细控制 (如像素限制) 的需求。PR body 指出该功能缺失, 并引用 Qwen3-VL README 作为新支持的格式参考。

实现拆解

1. 入口变更: 在 slime/utils/data.py 的 _build_messages 函数中, 处理多模态占位符替换的循环内, 将原先直接构造 dict 的逻辑拆分为先检查 item 数据类型。
2. 核心逻辑: 如果 item 已是 dict (新格式), 则直接将整个 dict 加入 content_list; 否则 (字符串), 依旧按原有方式包装为 {"type": mt.name, mt.name: item}。
3. 测试配套: 无新增测试文件, 但旧测试应仍通过 (因为修改是兼容的)。

关键文件:

- slime/utils/data.py (模块 数据处理; 类别 source; 类型 core-logic): 核心数据预处理文件, 实现图像配置格式扩展, 是本次变更的唯一文件。

关键符号: _build_messages

关键源码片段

slime/utils/data.py

核心数据预处理文件, 实现图像配置格式扩展, 是本次变更的唯一文件。

```
# slime/utils/data.py (line 164-171 附近)
item = content.pop(0)
# Support rich image config from https://github.com/QwenLM/Qwen3-VL/blob/main/README.md
# 如果 item 已经是 dict (新格式), 直接追加到 content_list
if isinstance(item, dict):
    content_list.append(item)
```

```
# 否则（字符串等），按原有方式包装为统一格式
else:
    content_list.append({"type": mt.name, mt.name: item})
```

评论区精华

无 review 评论，PR 由作者直接合并，无讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低。修改仅 9 行，逻辑向后兼容：当 item 不是 dict 时行为完全不变。但缺乏测试覆盖新格式的分支，未来重构可能引入回归。另外如果数据中 dict 的键名与 mt.name 冲突可能导致意外行为，但可能性小。
- 影响：对用户：支持更丰富的图像配置，提升 VLM 使用灵活性。对系统：无性能或兼容性影响。对团队：小范围易维护的功能扩展。
- 风险标记：缺少测试覆盖

关联脉络

- 暂无明显关联 PR