

PR #45918 完整报告

vllm-project/vllm

[Bugfix] MiniCPM-V 4.6: fix grid rows/cols swap in placeholder generation

合并时间: 2026-06-30 23:12

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/45918>

执行摘要

- 一句话: 修复 MiniCPM-V 4.6 网格行列颠倒
- 推荐动作: 值得精读以了解 grid 语义对齐的重要性。对于多模态模型开发人员, 这是一个典型的跨框架数据契约不一致案例。

功能与动机

MiniCPM-V 4.6 是首个手动解包 grids[0]/grids[1] 来构建占位符的模型, 但 grids[0] 被错误地当作列数、grids[1] 当作行数, 而所有 transformers 消费者 (包括 HF 自己的处理器) 使用相反的约定。这导致非方形网格产生错误的占位符格式, 进而影响模型性能。PR body 指出约 42 点 OCRBench 退化。

实现拆解

1. 视频占位符修复: 在 vllm/model_executor/models/minicpmv4_6.py 的 get_video_prompt_texts 方法中, 将 grids[0] 和 grids[1] 的用法互换, 使行数对应 grids[1]、列数对应 grids[0]。
2. 图像占位符修复: 在同一文件的 get_slice_image_placeholder 方法中, 将 num_cols, num_rows = grids[0], grids[1] 改为 num_rows, num_cols = grids[0], grids[1], 并相应调整循环变量。
3. 仅涉及 2 行代码的修改 (+2/-2), 属于纯数据契约修复。

关键文件:

- vllm/model_executor/models/minicpmv4_6.py (模块 模型实现; 类别 source; 类型 data-contract; 符号 get_video_prompt_texts, get_slice_image_placeholder): 核心修复文件, 涉及视频和图像占位符生成中 grids 行列索引的交换。

关键符号: get_video_prompt_texts, get_slice_image_placeholder

关键源码片段

`vllm/model_executor/models/minicpmv4_6.py`

核心修复文件, 涉及视频和图像占位符生成中 grids 行列索引的交换。

```
# vllm/model_executor/models/minicpmv4_6.py (head 版本)
```

```

# 视频占位符修复 (get_video_prompt_texts 方法)
# 原代码: rows = [slice_ph * grids[0] for _ in range(grids[1])]
# grids[0] 被当作列数, grids[1] 被当作行数, 与 HF 约定相反。
# 修正后: grids[1] 作为列数 (每行 slice 数), grids[0] 作为行数
if grids[0] > 0 and grids[1] > 0 and patch_tokens > 0:
    slice_ph = slice_start + video_token * patch_tokens + slice_end
    rows = [slice_ph * grids[1] for _ in range(grids[0])] # 修复点
    per_frame += "\n".join(rows)

# ...

# 图像占位符修复 (get_slice_image_placeholder 方法)
# 原代码: num_cols, num_rows = grids[0], grids[1]
# 修正后: num_rows 对应 grids[0], num_cols 对应 grids[1]
# 注意: 虽然变量名交换了, 但后续使用 num_cols 作为每行 slice 数,
# num_rows 作为行数, 因此需要将 grids[0] 赋给 num_rows。
num_rows, num_cols = grids[0], grids[1] # 修复点
if num_cols > 0 and num_rows > 0 and patch_tokens > 0:
    slice_ph = slice_start + image_token * patch_tokens + slice_end
    slices = [slice_ph * num_cols for _ in range(num_rows)]
    placeholder += "\n".join(slices)

```

评论区精华

PR 获得两位 reviewer (Kavehs1、DarkLight1337) 的批准, 没有 review 评论。提交者请求 DarkLight1337 审查, 对方回复 "Sorry I missed this!" 后批准。没有设计争议。

- 暂无高价值评论线程

风险与影响

- 风险: 风险极低: 修改只涉及两行赋值和索引, 逻辑清晰明确。对于非方形 grid 的修复经过了 OCRBench 退化数据的验证。方形 grid 不受影响。没有测试文件变更, 但修复足够直观。
- 影响: 直接影响使用 MiniCPM-V 4.6 模型且输入为非方形网格图像或视频的用户, 修复了约 27.9% 样本的 OCRBench 性能退化。不影响其他模型或其他 MiniCPM 版本。
- 风险标记: 缺少测试覆盖, 变更极低风险

关联脉络

- PR #46301 [Spec Decode] Fix hidden-state extraction block size for hybrid verifiers: 同为多模态相关 bugfix, 涉及 token 对齐与数据契约一致性。
- PR #47010 fix(security): prevent image decompression bomb OOM denial of service: 同为图像处理相关的多模态安全修复, 关注输入数据格式。