

PR #48467 完整报告

vllm-project/vllm

remove force channels_last in Idefics3MultiModalProcessor

合并时间: 2026-07-13 22:18

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/48467>

执行摘要

- 一句话: 修复 Idefics3 图片处理 channels_last 强制错误
- 推荐动作: 建议合并。该 PR 是一个精准的 bugfix, 修复了 transformers 升级带来的平台差异问题, 实现简洁、讨论充分、风险低。值得关注的是其对 backend 属性的处理模式, 可作为多模态处理器中类似条件依赖的标准方案。

功能与动机

Transformers 5.13.1 升级后, `Idefics3ImageProcessor` 在 `torchvision>=0.27` 下使用快速后端 (非 pil), 此时输入 tensor 已是 CHW 顺序, 强制 `channels_last` 会导致 `permute` 错误, 进而引发 `RuntimeError: stack expects each tensor to be equal size`。该问题仅出现在 XPU CI (torch 2.12.0) 上。

实现拆解

1. 定位问题: 在 `Idefics3MultiModalProcessor._call_hf_processor` 中, 原先无条件设置 `mm_kwargs = {"input_data_format": "channels_last", **mm_kwargs}`, 导致所有后端都强制 `channels_last`。
2. 修改逻辑: 新增对 `image_processor.backend` 的判断: 仅当后端为 "pil" 时才注入 `channels_last`。通过 `getattr(image_processor, "backend", "pil")` 安全获取默认值。
3. 对应修改: 仅修改 `vllm/model_executor/models/idefics3.py` 文件, 变更集中在第 313-315 行, 增加 1 行条件判断、1 行缩进调整。
4. 讨论扩展: review 中建议对 `_LANCZOS_IMAGE_PROCESSORS` 中其他模型 (如 Chameleon、ColModernVBert) 应用相同修复, 但经确认这些模型原本就无需 `channels_last`, 因此未做改动。

关键文件:

- `vllm/model_executor/models/idefics3.py` (模块 模型执行器; 类别 source; 类型 data-contract): 在 `_call_hf_processor` 中增加条件判断, 仅当 `image_processor` 后端为 pil 时设置 `input_data_format=channels_last`

关键符号: `Idefics3MultiModalProcessor._call_hf_processor`

关键源码片段

vllm/model_executor/models/idefics3.py

在 `_call_hf_processor` 中增加条件判断，仅当 `image_processor` 后端为 `pil` 时设置 `input_data_format=channels_last`

```
# vllm/model_executor/models/idefics3.py
# Idefics3MultiModalProcessor 的图片处理入口

def _call_hf_processor(
    self,
    prompt: str,
    mm_data: Mapping[str, object],
    mm_kwargs: Mapping[str, object],
    tok_kwargs: Mapping[str, object],
) -> BatchFeature:
    # 文本输入直接编码，无需图片处理
    if not (images := mm_data.get("images", [])):
        prompt_ids = self.info.get_tokenizer().encode(prompt)
        prompt_ids = self._apply_hf_processor_tokens_only(prompt_ids)
        return BatchFeature(dict(input_ids=[prompt_ids]), tensor_type="pt")

    # 关键修复：仅当后端为 PIL 缓慢处理器时才强制 channels_last
    # 快速处理器（torchvision>=0.27）使用 CHW 格式，无需 permute
    image_processor = self.info.get_hf_processor().image_processor
    if getattr(image_processor, "backend", "pil") == "pil":
        mm_kwargs = {"input_data_format": "channels_last", **mm_kwargs}

    processed_outputs = super()._call_hf_processor(
        prompt,
        mm_data,
        mm_kwargs,
        tok_kwargs,
    )
    # 后续处理：计算 patch 数、移除 batch 维度等
    ...
    return processed_outputs
```

评论区精华

review 主要围绕三点：

1. 根本原因分析：hmellor 指出，`transformers` 中 `_LANCZOS_IMAGE_PROCESSORS` 仅在 `torchvision<0.27` 时使用慢速 PIL 处理器（需 `channels_last`），而 `torchvision>=0.27` 的快速处理器不需要。CUDA/ROCm CI 仍使用 `torchvision 0.26`，XPU 使用 `0.27`，因此仅 XPU 暴露问题。
2. 修复范围建议：hmellor 最初建议对 `_LANCZOS_IMAGE_PROCESSORS` 中所有 vLLM 支持的模型统一添加条件判断，但后来纠正——Chameleon 和 ColModernVBert 从未需要 `channels_last`，因此仅修复 `idefics3.py` 即可。

3. 版本差异: jikunshang 补充说明 torchvision 版本继承自 torch, XPU 使用 torch 2.12.0, 其他平台仍用 2.11.0。
- 为什么只在 XPU 上出现错误? (question): 版本差异导致, 修复后条件判断兼容两种场景。
 - 修复范围是否应扩展到其他模型? (design): 仅修改 idefics3.py, 其他模型保持原样。

风险与影响

- 风险: 风险较低:
 - 修改仅影响 Idefics3 模型图片处理路径, 添加了 backend 属性检查, 有默认值 "pil", 兼容未定义 backend 属性的旧版本处理器。
 - 在 pil 后端下行为与之前完全一致 (channels_last 被设置)。
 - 无测试文件修改, 但 reviewer 均已 approve。
- 影响:
 - 用户影响: 修复 Idefics3 模型在 torchvision>=0.27 环境 (如 XPU) 下的图片处理错误, 其他平台无影响。
 - 系统影响: 单一文件 5 行修改, 无配置、依赖或性能影响。
 - 团队影响: 明确区分了不同 torchvision 版本下图片处理器的行为差异, 为后续类似问题提供了参考模式。
- 风险标记: 平台差异依赖, 缺少测试覆盖

关联脉络

- PR #47867 Bump transformers to 5.13.1 (or 5.11.0): 本次 bug 的直接触发原因, transformers 升级后引入了 backend 属性变化。