

# PR #38629 完整报告

vllm-project/vllm

[Fix] handle PaddleOCR-VL image processor max\_pixels across Transformers v4/v5

合并时间: 2026-03-31 23:50

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/38629>

## PR 分析报告: 修复 PaddleOCR-VL 图像处理器 max\_pixels 跨版本处理

### 执行摘要

本 PR 修复了 PaddleOCR-VL 模型在不同 Transformers 版本 (v4/v5) 中 max\_pixels 访问不一致的 bug, 通过基于 trust\_remote\_code 配置的动态逻辑确保兼容性, 影响范围小但提升了模型稳定性。

### 功能与动机

动机源于 Transformers 库的版本差异: 在 v4.x 中, PaddleOCR-VL 需通过 trust\_remote\_code=True 加载, max\_pixels 存储在 image\_processor.max\_pixels; 而在 v5 及以上的原生实现中, 该值位于 image\_processor.size.longest\_edge。若不处理此差异, 可能导致 AttributeError 和模型功能异常。

### 实现拆解

关键改动集中在文件 vllm/model\_executor/models/paddleocr\_vl.py 的 get\_image\_size\_with\_most\_features 方法中。代码逻辑如下: 

```
if self.ctx.model_config.trust_remote_code:
    max_pixels = image_processor.max_pixels
# Transformers v4 路径 else:
    max_pixels = image_processor.size.longest_edge # Transformers v5 路径
max_num_tokens = max_pixels // (factor**2)
```

 通过检查 trust\_remote\_code 配置, 动态选择正确的 max\_pixels 来源, 避免了硬编码版本依赖。

### 评论区精华

review 讨论的核心是属性访问的潜在风险:

- gemini-code-assist[bot] 指出: image\_processor.size.longest\_edge 可能引发 AttributeError, 因为 size 通常为字典, 建议使用字典访问 (如 image\_processor.size['longest\_edge'])。
- Isotr0py 澄清: Transformers v5 引入了 SizedDict, 支持通过属性和键访问, 因此当前实现是正确的。这揭示了外部库的内部演进, 无需修改代码。
- 结论: 基于 Isotr0py 的链接和验证, 团队确认实现无误, 解决了兼容性担忧。

### 风险与影响

- 技术风险：主要涉及兼容性——若 `trust_remote_code` 配置错误或 Transformers 未来版本改变实现，可能引发 `AttributeError`。代码变更虽小，但缺少测试覆盖，有潜在回归风险。
- 影响范围：仅影响使用 PaddleOCR-VL 多模态模型的用户，确保其在 v4 和 v5 版本上正常工作，对系统性能无显著影响。此修复有助于维护跨版本用户体验。

## 关联脉络

从历史 PR 看，本 PR 与近期 bugfix 和模型相关 PR（如 #36178、#36540）有相似模式，但无直接关联。它反映了团队在多模态模型集成中对第三方库版本兼容性的持续关注，可作为后续类似修复的参考。