

PR #23749 完整报告

sgl-project/sglang

[CI] release-whl-kernel: strip +cu129 local version before PyPI upload

合并时间: 2026-04-26 13:14

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/23749>

执行摘要

- 一句话: 修复 PyPI 上传因 +cu129 版本号被拒
- 推荐动作: 该 PR 是典型的 CI 故障修复, 值得快速合并。对于团队而言, 建议关注未来新增硬件平台 wheel 版本号时, 是否会产生类似的 PyPI 兼容性问题。

功能与动机

PR #23587 为 cu129 wheel 添加了 +cu129 本地版本标签, 但 PyPI 拒绝任何包含 PEP 440 local version label 的上传 (HTTP 400), 导致 release-whl-kernel.yml 的 twine upload 失败。需要在不影响本地安装和 whl 索引的前提下, 为 PyPI 上传生成干净版本。

实现拆解

1. 在 `.github/workflows/release-whl-kernel.yml` 的 `build-cu129-matrix` job 中新增步骤: 在 Upload to PyPI 之前插入 Strip +cu129 local version for PyPI upload。该步骤使用 `python3 -m wheel unpack` 解压 wheel, 替换 METADATA 中的 Version 字段 (去除 +cu129), 重命名 dist-info 目录, 然后用 `python3 -m wheel pack` 重新打包到 dist-pypi/ 目录。
2. 调整 PyPI 上传路径: 将 twine upload 的目标从 dist/* 改为 dist-pypi/*, 确保上传的是剥离后的 wheel。
3. 保留原 dist/ 目录: dist-pypi/ 是全新生成的, dist/ 保持不变, 从而不影响后续 Upload artifacts → release-cu129 → update_kernel_whl_index.py 流程 (该脚本兼容带 +cu129 和不带 +cu129 的文件名)。
4. 该步骤仅对 cu129 生效, 因为 cu130 矩阵不执行 PyPI 上传 (已有注释说明), 故无需改动。

关键文件:

- `.github/workflows/release-whl-kernel.yml` (模块 CI 脚本; 类别 infra; 类型 infrastructure): 核心变更文件, 新增了剥离 +cu129 版本的步骤, 并修改了 twine upload 路径。

关键符号: 未识别

评论区精华

没有 review 评论或讨论。PR 由作者单人提交并合并，仅有一条 gemini-code-assist 的配额警告，不涉及技术讨论。

- 暂无高价值评论线程

风险与影响

- 风险:

1. 兼容性风险（低）：update_kernel_whl_index.py:check_wheel_cuda_version 已确认兼容带和不带 +cu129 的文件名，但需验证是否还有其他地方依赖 +cu129 标识。
2. 流程断裂风险（低）：剥离过程在 CI 中运行，若 wheel 解压 / 打包失败，整个工作流会失败；但 CI 已有 e2e 验证通过。
3. 无直接业务风险：变更仅涉及 CI 脚本，不修改任何运行时代码。

- 影响:

- 用户 / 外部依赖方：无直接影响，PyPI 上的 sgl-kernel 版本号将不再包含 +cu129，符合 PyPI 规范。
- 系统：修复了 release 流水线的断裂，确保 PyPI 上传继续可用。
- 团队：CI 维护者需要注意未来版本号变更时类似问题的处理。影响范围小，仅限 CU129 矩阵的 PyPI 上传步骤。
- 风险标记：CI 流程变更，依赖上传脚本兼容性

关联脉络

- PR #23587 Add +cu129 local version to cu129 wheel: 本 PR 正是为了解决 #23587 引入的 +cu129 版本号导致 PyPI 上传失败的问题。