

PR #34937 完整报告

sgl-project/sglang

Fix Python packaging shadowing in DeepEP wheel builds

合并时间: 2026-08-15 18:27

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/34937>

执行摘要

- 一句话: DeepEP wheel 构建修复 /packaging 遮蔽问题
- 推荐动作: 建议快速阅读, 作为容器挂载点遮蔽 Python 第三方包的典型修复案例。值得关注的决策: 用独立路径规避标准库查找路径的冲突, 比在容器内删除或重命名 overlay 更轻量。可留意 DeepEP 发布链路是否值得补充一个检查 packaging.version 可导入的稳态回归测试。

功能与动机

PR body 说明根因: 打包 overlay 包含 init.py 且挂载在 /packaging, wheel 容器从根目录运行时该目录遮蔽了已安装的 Python packaging 发行版, CUDA 13 构建在 python -m wheel pack 阶段以 ModuleNotFoundError: No module named 'packaging.version' 失败。

实现拆解

按以下步骤拆解:

1. 定位根因: 容器从根目录启动时 /packaging 属于 Python packaging 模块的标准查找路径, 挂载的 overlay 内含 init.py, 导致优先命中 overlay 而遮蔽真实发行版。
2. 修改挂载路径: 在 scripts/build_sgl_deepep.sh 的 docker run 命令中, 将 volume 挂载 \${PACKAGING_OVERLAY}:/packaging:ro 改为 \${PACKAGING_OVERLAY}:/sgl-deep-ep-packaging:ro, 消除名称冲突。
3. 同步调用参数: 容器内执行 bash /packaging/build_sgl_deep_ep.sh 的地方一并改为 bash /sgl-deep-ep-packaging/build_sgl_deep_ep.sh, 并将传给脚本的第二、第三个位置参数由 /packaging 同步为新路径, 保证脚本能定位自身所在目录。
4. 验证配套: bash -n 语法检查与 pre-commit 均通过; CUDA 13 build-only workflow (Run 31878507917) 覆盖 Python 3.10-3.13 与 x86_64/aarch64 共 8 个组合, 全部完成扩展编译、auditwheel repair 与 wheel 重打包。
5. 测试配套: 首次提交曾加入回归测试, 第二个提交 (chore: remove DeepEP build regression test) 将其移除, 最终未保留自动化回归测试。

关键文件:

- scripts/build_sgl_deepep.sh (模块 构建脚本; 类别 infra; 类型 configuration): 唯一变更文件。脚本负责挂载 DeepEP 打包 overlay 并在容器内调用构建命令, 本 PR 的核心修复

点就在这两处路径参数上：将挂载点与调用路径从 /packaging 迁移到 /sgl-deep-ep-packaging，解决 Python packaging 模块被遮蔽的问题。

关键符号：未识别

关键源码片段

scripts/build_sgl_deep.sh

唯一变更文件。脚本负责挂载 DeepEP 打包 overlay 并在容器内调用构建命令，本 PR 的核心修复点就在这两处路径参数上：将挂载点与调用路径从 /packaging 迁移到 /sgl-deep-ep-packaging，解决 Python packaging 模块被遮蔽的问题。

```
# 关键改动：将 DeepEP 打包 overlay 从 /packaging 移到 /sgl-deep-ep-packaging。
# 根因：overlay 内含 __init__.py，挂载在 /packaging 后，容器从根目录运行时
# 会遮蔽已安装的 Python packaging 发行版，导致 python -m wheel pack 阶段
# 报错 ModuleNotFoundError: No module named 'packaging.version'。
docker run --rm \
  --env CUDA_VERSION="${CUDA_VERSION}" \
  --env MAX_JOBS="${MAX_JOBS:-8}" \
  --volume "${DEEPEP_SOURCE}:/deep:ro" \
  --volume "${PACKAGING_OVERLAY}:/sgl-deep-ep-packaging:ro" \
  --volume "${DIST_DIR}:/output/dist" \
  --volume "${PYPI_DIST_DIR}:/output/dist-pypi" \
  "${IMAGE_TAG}" \
  bash /sgl-deep-ep-packaging/build_sgl_deep.sh \
    /deep /sgl-deep-ep-packaging "${raw_dir}" "${CUDA_VERSION}" "${ARCHITECTURE}"
```

评论区精华

PR 没有 review 评论。commit 历史显示第一次提交添加修复逻辑后，第二次提交移除了 DeepEP 构建回归测试（chore: remove DeepEP build regression test），该决策未在 PR 中说明，属于可留意的未决点。

- 暂无高价值评论线程

风险与影响

- 风险：整体风险低：变更仅调整容器内挂载路径与调用参数，不改变构建产物格式或外部接口。两个注意点：一是 /sgl-deep-ep-packaging 成为新的隐式路径约定，若未来其他脚本仍硬编码 /packaging 会重复踩坑；二是回归测试被移除，路径约定一旦变化缺少自动化防线。已通过 8 组合 CI 验证编译、auditwheel repair 与 wheel 重打包均成功。
- 影响：影响范围限定在 DeepEP wheel 发布构建链路：CUDA 13 镜像、Python 3.10-3.13、x86_64 与 aarch64 共 8 个组合。修复前这些组合的发布构建会在 wheel pack 阶段失败，修复后恢复产出；对 SGLang 运行时和用户请求路径无任何影响，构建产物内容也不变。
- 风险标记：缺少回归测试，挂载路径隐式约定

关联脉络

- PR #34914 Update sgl-deep-ep release workflow for DeepEP v2: 同一 DeepEP wheel 发布链路的配套改动，涉及 docker/sgl-deep-ep.Dockerfile 与 release-whl-deepep.yml；本 PR 修复该链路打包阶段的路径遮蔽问题，两者共同保障 CUDA 13 下 DeepEP wheel 的自动化产出。