

# PR #24291 完整报告

sgl-project/sglang

[CI] Keep custom sgl-kernel wheel in CUDA CI

合并时间: 2026-05-03 12:53

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24291>

## 执行摘要

- 一句话: 修复 CI 中自定义 sgl-kernel 被覆盖的问题
- 推荐动作: 应合并此修复, 它解决了一个关键的 CI 回归, 使内核相关 PR 的测试更加可靠。建议精读该变更, 以理解 CI 依赖安装的控制流。

## 功能与动机

PR body 指出: 来自 #21247 的回归导致 CUDA CI 总是强制从公共 wheel 索引重新安装 `sglang-kernel`, 这会覆盖当 `CUSTOM_BUILD_SGL_KERNEL=true` 时刚构建的自定义 wheel, 导致 CI 无法测试 PR 的内核变更。同时引用了失败的 CI 运行日志。

## 实现拆解

1. 修改文件: `scripts/ci/cuda/ci_install_dependency.sh` 中的 `install_sglang_kernel` 函数。
2. 添加条件判断: 在原先无条件执行 `pip install sglang-kernel` 之前, 增加 `if [ "${CUSTOM_BUILD_SGL_KERNEL:-}" != "true" ]` 检查。
3. 分支处理:
  - 当 `CUSTOM_BUILD_SGL_KERNEL` 不为 `true` (包括未设置) 时, 保持原有行为: 从与 CUDA 版本匹配的 wheel 索引强制重新安装 `sglang-kernel`。
  - 当 `CUSTOM_BUILD_SGL_KERNEL=true` 时, 跳过安装, 打印日志 `"CUSTOM_BUILD_SGL_KERNEL=true: keeping freshly built sgl-kernel wheel."`, 保留之前构建的自定义 wheel。
4. 影响: 仅影响 CI 安装依赖的流程, 不影响任何运行时逻辑或用户代码。

关键文件:

- `scripts/ci/cuda/ci_install_dependency.sh` (模块 CI 脚本; 类别 `infra`; 类型 `infrastructure`): 这是唯一变更的文件, 修复了 CI 中自定义 sgl-kernel 被公共 wheel 覆盖的问题。

关键符号: 未识别

## 关键源码片段

`scripts/ci/cuda/ci_install_dependency.sh`

这是唯一变更的文件，修复了 CI 中自定义 sgl-kernel 被公共 wheel 覆盖的问题。

```
# 在 install_sglang_kernel 函数中，原无条件安装改为条件判断
install_sglang_kernel() {
    # ... 前面的 torch 安装等逻辑 ...

    # 检查是否设置了 CUSTOM_BUILD_SGL_KERNEL 环境变量
    # 如果为 true，说明 PR 已构建了自定义 sgl-kernel wheel，跳过从公共索引安装
    if [ "${CUSTOM_BUILD_SGL_KERNEL:-}" != "true" ]; then
        # 原逻辑：从与 CU_VERSION 匹配的 wheel 索引强制安装 sglang-kernel
        # 确保 runners 使用与本地 libnvidia链接的正确 wheel（例如 cu129）
        $PIP_CMD install "sglang-kernel==${SGL_KERNEL_VERSION_FROM_SRT}" \
            --index-url "https://docs.sglang.ai/whl/${CU_VERSION}/" \
            --force-reinstall --no-deps $PIP_INSTALL_SUFFIX
    else
        # 保留自定义构建的 wheel，并打印日志
        echo "CUSTOM_BUILD_SGL_KERNEL=true: keeping freshly built sgl-kernel wheel."
    fi
    mark_step_done "${FUNCNAME[0]}"
}
```

## 评论区精华

仅有机器人评论和仓库维护者 Kangyan-Zhou 的感谢回复，无深度讨论。

- 暂无高价值评论线程

## 风险与影响

- 风险：低风险。变更仅涉及 CI 脚本的条件分支，对生产代码无影响。如果 CUSTOM\_BUILD\_SGL\_KERNEL=true 但自定义构建失败，自定义 wheel 可能无效，但 CI 会在构建步骤中报错，因此不会引入静默失败。
- 影响：影响范围：仅影响 CUDA CI 中 CUSTOM\_BUILD\_SGL\_KERNEL=true 的作业，确保它们使用 PR 自定义的内核 wheel 进行测试。其他 CI 作业不受影响。影响程度：中高，因此修复确保内核变更被正确测试，避免误通过或不充分的测试。
- 风险标记：CI 脚本变更，回归修复

## 关联脉络

- PR #21247 [Dependency] Upgrade to Torch 2.11.0: 本 PR 修复了 #21247 引入的回归，该 PR 修改了依赖安装流程，可能导致自定义 kernel 被覆盖。