

# PR #24916 完整报告

sgl-project/sglang

ci: run H20 stage with CUDA 13

合并时间: 2026-05-11 11:18

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24916>

## 执行摘要

移除 H20 CI 阶段强制指定的 `CUDA_VERSION: cu129` 环境变量，使该阶段使用默认的 CUDA 13.0 工具链，修复 DeepEP 扩展因 CUDA 版本不匹配导致的构建失败。

## 功能与动机

H20 CI runner 已升级至 CUDA 13.0，但 `stage-c-test-8-gpu-h20` 仍强制使用 `CUDA_VERSION: cu129`，导致 DeepEP 扩展安装时因 CUDA 版本（13.0）与 PyTorch 编译版本（12.9）不匹配而失败（issue 引用：RuntimeError: The detected CUDA version (13.0) mismatches the version that was used to compile PyTorch (12.9)）。

## 实现拆解

1. 在 `.github/workflows/pr-test.yml` 中找到 H20 CI 阶段的 `env` 配置块。
2. 删除 `CUDA_VERSION: cu129` 行。
3. 变更后，该阶段将使用 `scripts/ci/cuda/ci_install_dependency.sh` 定义的默认 CUDA 版本 `cu130`，与 H20 runner 环境一致。

无可用关键源码片段（仅一行环境变量删除）。

## 评论区精华

无 review 评论，仅 CI 触发和测试通过确认。

## 风险与影响

- 风险极低：仅删除一行环境变量覆盖，默认值 `cu130` 已在其他阶段（如 H100）验证稳定。
- 影响范围：仅限 H20 CI 阶段，不会影响其他平台或生产环境。

## 关联脉络

近期 H20 相关的 CI 问题较多（如 PR#23819、PR#20177），此 PR 属于基础设施适应硬件环境升级的常规维护。