

PR #22631 完整报告

sgl-project/sglang

[codex] Add LTX-2.3 benchmark skill recipes

合并时间: 2026-04-13 12:23

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/22631>

执行摘要

此 PR 为 LTX-2.3 扩散模型添加了一阶段和二阶段基准测试技能配方，通过更新脚本和文档，提升了基准设置的明确性。然而，GPU 数量设置保持为 2（而非 review 建议的 1），可能影响基准比较的一致性，建议关注此决策。

功能与动机

PR 旨在解决 LTX-2.3 模型在基准测试中技能文档覆盖不全的问题。由于 LTX-2.3 演化出两个原生路径（一阶段基准和二阶段基准路径），现有文档未清晰说明，导致用户设置基准时命令发现不一致。PR body 引用：“The skill docs did not spell out both LTX-2.3 cases clearly, which made benchmark setup and command discovery inconsistent。”因此，通过添加明确的预设和文档，改善用户体验。

实现拆解

改动涉及四个文件，按模块拆解如下：

- 核心脚本(bench_diffusion_denoise.py)：在 MODELS 字典中添加 ltx23-one-stage 和 ltx23-two-stage 预设，配置参数如分辨率、GPU 数量（保持 2），并更新 required_gpus_for_model 函数返回 2 GPU。python "ltx23-one-stage": { "path": "Lightricks/LTX-2.3", "extra_args": ["--num-gpus=2", ...], },
- 基准文档(benchmark-and-profile.md)：添加两个预设的运行示例和描述，包括手动命令，但 GPU 数量仍为 2。
- 技能文档(两个 SKILL.md 文件)：更新引用和添加性能命令配方，确保跨技能一致性。

评论区精华

review 由 gemini-code-assist[bot] 提出，核心讨论围绕 GPU 数量设置：

“The ltx23 preset is configured to use 2 GPUs, whereas the ltx2 preset uses only 1 GPU... Consider setting this to 1 GPU to match the existing LTX-2 benchmark configuration.”

讨论强调性能一致性，建议改为 1 GPU 以避免多 GPU 开销。但 PR 最终未采纳，代码中保持 2 GPU，这可能反映了作者的不同权衡或硬件假设。

风险与影响

- 风险：主要风险是基准测试不一致性，因为 LTX-2 使用 1 GPU 而 LTX-2.3 使用 2 GPU，可能导致性能比较偏差。无代码回归风险，但配置可能误导用户。
- 影响：用户现在可以更方便地运行 LTX-2.3 基准测试，提升文档清晰度；但对系统无直接影响，团队需注意 GPU 配置的潜在问题。

关联脉络

与历史 PR #22182 (“[diffusion] model: support LTX2.3 two stage”) 直接相关，该 PR 引入了 LTX-2.3 的两阶段支持，本 PR 在此基础上扩展了基准测试文档。这表明仓库在持续完善扩散模型功能，尤其是 LTX 系列的性能评估工具链。