

PR #32282 完整报告

sgl-project/sglang

[ci][diffusion] Enhance diffusion GT publishing with per-platform directory resolution

合并时间: 2026-07-25 09:54

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/32282>

执行摘要

- 一句话: 扩散 GT 发布支持按平台目录解析
- 推荐动作: 值得合并, 但建议后续添加单元测试覆盖 `_resolve_target_dir` 函数 (例如 `mock get_consistency_platform`) 。

功能与动机

该 PR 解决 diffusion GT 发布时路径与一致性测试读取路径不一致的问题。PR body 没有详细说明, 但从变更可知, 之前的 `check_quality` 和 `publish` 函数允许 `target_dir` 为 `None`, 且没有根据平台子目录进行解析, 导致发布路径与测试端 (已在 `test_utils.py` 中按平台解析) 不同步, 可能引起 GT 不匹配。

实现拆解

1. 修改函数签名: `scripts/ci/utlils/diffusion/publish_diffusion_gt.py` 中 `check_quality` 和 `publish` 函数的 `target_dir` 参数改为必须传入, 去除 `target_dir or DEFAULT_TARGET_DIR` 的默认回退逻辑, 由调用者统一处理。
2. 新增 `_resolve_target_dir` 函数: 该函数接收 `target_dir` 和 `per_platform` 布尔值, 当 `per_platform` 为 `True` 时, 通过导入 `sglang.multimodal_gen.test.test_utils.get_consistency_platform()` 获取当前平台子目录 (如 `h100`), 并将 `target_dir` 拼接为 `<target>/<platform>`。该函数使用与一致性测试相同的解析器, 确保读写路径一致。
3. 修改 `main` 入口: 从环境变量 `SGLANG_DIFFUSION_GT_PER_PLATFORM` 读取是否启用 `per-platform` 功能, 并通过 `_resolve_target_dir` 得到实际 `target_dir`, 再传递给 `check_quality` 或 `publish`。
4. CI workflow 配置: 在 `.github/workflows/diffusion-ci-gt-gen.yml` 的 `env` 中添加 `SGLANG_DIFFUSION_GT_PER_PLATFORM: "1"`, 启用 `per-platform` 目录解析。

关键文件:

- `scripts/ci/utlils/diffusion/publish_diffusion_gt.py` (模块 CI 脚本; 类别 `infra`; 类型 `infrastructure`; 符号 `check_quality`, `publish`, `_resolve_target_dir`, `main`): 核心变更文件: 修改了 `check_quality` 和 `publish` 函数签名, 新增 `_resolve_target_dir` 函数, 重构了 `main` 入口以支持 `per-platform` 目录解析。
- `.github/workflows/diffusion-ci-gt-gen.yml` (模块 CI 配置; 类别 `infra`; 类型 `infrastructure`): 新增环境变量 `SGLANG_DIFFUSION_GT_PER_PLATFORM` 以启用

per-platform 功能。

关键符号: `_resolve_target_dir`, `check_quality`, `publish`

关键源码片段

`scripts/ci/utils/diffusion/publish_diffusion_gt.py`

核心变更文件: 修改了 `check_quality` 和 `publish` 函数签名, 新增 `_resolve_target_dir` 函数, 重构了 `main` 入口以支持 per-platform 目录解析。

```
# scripts/ci/utils/diffusion/publish_diffusion_gt.py

def _resolve_target_dir(target_dir: str | None, per_platform: bool) -> str:
    """Resolve the effective remote target dir.

    With ``per_platform`` set, append the consistency platform subdir
    (``h100``/``b200``/``5090``) so the publish path matches how the consistency
    tests resolve GT: ``<target>/<platform>/<file>`` first, bare ``<target>/<file>``
    only as fallback. Reuses the SAME resolver the tests use so read and write can
    never drift. Lazy import keeps this script light when the flag is off.
    """
    target_dir = target_dir or DEFAULT_TARGET_DIR
    if per_platform:
        from sglang.multimodal_gen.test.test_utils import get_consistency_platform

        target_dir = f"{target_dir}/{get_consistency_platform()}"
    return target_dir

def main():
    # ... argument parsing ...
    per_platform = os.environ.get("SGLANG_DIFFUSION_GT_PER_PLATFORM") == "1"
    target_dir = _resolve_target_dir(args.target_dir, per_platform)

    if args.check_only:
        check_quality(args.source_dir, target_dir)
    else:
        publish(args.source_dir, target_dir)
```

评论区精华

无 review 评论, PR 仅包含一条来自 Gemini Code Assist 的弃用通知, 没有实质性讨论。

- 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 工具脚本, 不作用于运行时路径。但若环境变量 `SGLANG_DIFFUSION_GT_PER_PLATFORM` 未正确设置或 `get_consistency_platform` 逻

辑变更，可能导致发布路径错误。另外，`_resolve_target_dir` 中的惰性导入 (lazy import) 在首次调用时若缺少依赖会抛出 `ImportError`，但只影响 CI 流程。

- 影响：直接影响 diffusion CI 的 GT 发布流程：确保发布到正确平台子目录，与测试端路径匹配。对其他模块无影响。
- 风险标记：缺少测试覆盖

关联脉络

- PR #32297 [ci][diffusion] Read consistency GT from ci-data-diffusion @ per-platform commit: 同属 diffusion CI 改进系列，该 PR 使 GT 读取支持按平台目录，而本 PR 使 GT 发布也支持按平台目录，两者互为读写对称操作。