

PR #27080 完整报告

sgl-project/sglang

[diffusion] Fix LingBot realtime consistency GT pin

合并时间: 2026-06-03 10:29

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27080>

执行摘要

- 一句话: 更新 LingBot 实时一致性 GT 数据及测试用例修复
- 推荐动作: 此 PR 属于常规的测试数据同步与清理, 技术价值有限。建议相关测试维护者关注 ci-data 仓库的关联 PR#13, 确保 GT 数据版本一致。

功能与动机

根据 PR 描述“This updates the diffusion consistency GT pin to the refreshed LingBot realtime Plastic Beach frames and fixes the realtime unit test to look up the case from `ONE_GPU_CASES` directly”, 是为了同步更新 ci-data 仓库中的 GT 数据, 并清理测试代码中对已废弃辅助函数的依赖, 确保测试的稳定性和可维护性。

实现拆解

1. 更新 GT 数据引用: 在 `python/sglang/multimodal_gen/test/test_utils.py` 中, 将 `SGL_TEST_FILES_CI_DATA_REVISION` 从 `8dbe6bc...` 更新为 `ddaad3f...`, 从而指向 ci-data 仓库中最新版本的 GT 帧数据。
2. 重构采样参数配置: 在 `python/sglang/multimodal_gen/test/server/testcase_configs.py` 中, 将 `LINGBOT_WORLD_REALTIME_sampling_params` 中的 `realtime_events` 列表迁移至 `extras` 字典内的 `condition_inputs.camera_actions` 字段, 使事件定义更加结构化并与引擎要求的输入格式对齐。
3. 修复测试用例注册逻辑: 在 `python/sglang/multimodal_gen/test/unit/realtime/test_realtime_consistency_harness.py` 中, `test_lingbot_realtime_case_is_registered_by_default` 函数不再导入已删除的 `_make_lingbot_realtime_plastic_beach_case`, 而是直接通过 `next()` 从 `ONE_GPU_CASES` 中按 id 定位用例, 并移除了冗余的 `any()` 断言。同时调整了 `test_lingbot_realtime_plastic_beach_params_are_lossless_gt_ready` 中的断言, 以匹配新的配置结构。所有变更均为测试配套调整, 未涉及推理或服务端核心逻辑。

关键文件:

- `python/sglang/multimodal_gen/test/unit/realtime/test_realtime_consistency_harness.py` (模块 测试框架; 类别 test; 类型 test-coverage): 核心测试文件, 修复了用例注册逻辑并调整了配置断言, 移除对已废弃函数的依赖。
- `python/sglang/multimodal_gen/test/server/testcase_configs.py` (模块 测试配置; 类别 test; 类型 test-coverage): 定义了测试用采样参数, 将 `realtime_events` 迁移至

extras.condition_inputs 以匹配新的配置格式。

- python/sclang/multimodal_gen/test/test_utils.py (模块 测试工具; 类别 test; 类型 test-coverage) : 更新 SGL_TEST_FILES_CI_DATA_REVISION, 指向新的 GT 数据版本, 是最基础的配置变更。

关键符号: 未识别

关键源码片段

python/sclang/multimodal_gen/test/server/testcase_configs.py

定义了测试用采样参数, 将 realtime_events 迁移至 extras.condition_inputs 以匹配新的配置格式。

```
# python/sclang/multimodal_gen/test/server/testcase_configs.py
# 变更前: realtime_events 字段定义相机动作事件
# 变更后: 改用 extras.condition_inputs.camera_actions 传递相同信息
LINGBOT_WORLD_REALTIME_sampling_params = DiffusionSamplingParams(
    prompt="A slow aerial orbit around a pastel floating island hotel ...",
    image_path="https://is1-ssl.mzstatic.com/.../825646291038.jpg/600x600bb.jpg",
    output_size="832x480",
    num_frames=9,
    fps=16,
    realtime_num_chunks=4,
    # 旧版本: realtime_events 直接作为顶级字段
    # realtime_events=[
    # {"after_chunk": 0, "kind": "camera_actions", ...},
    # {"after_chunk": 2, "kind": "camera_actions", ...},
    # ],
    realtime_perf_thresholds={
        "p95_chunk_total_ms": 5000.0,
        "p95_scheduler_forward_ms": 4500.0,
        "p95_ws_payload_mb": 16.0,
    },
    realtime_perf_ignore_initial_chunks=2,
    extras={
        "seed": 42,
        "num_inference_steps": 4,
        "guidance_scale": 1.0,
        "realtime_causal_sink_size": 9,
        "realtime_causal_kv_cache_num_frames": 18,
        # 新版本: camera_actions 数据移至 extras.condition_inputs
        "condition_inputs": {
            "camera_actions": [
                ["w"], ["w"], ["w"], ["w"], ["w"], ["w"],
                [], [], [], [], [], []
            ]
        },
    },
)
```

评论区精华

PR 无 review 评论，变更由作者 mickqian 自行合并，未产生公开讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。变更仅影响测试代码和配置定义，不涉及任何核心推理路径或服务端逻辑。更新 GT commit hash 可能导致测试与旧版 ci-data 不兼容，但 PR 已同步关联的 ci-data 仓库 (PR#13)，属于预期同步操作。
- 影响：影响范围仅限于扩散模型的一致性测试套件。影响程度小：测试人员需确保 ci-data 仓库已合并对应的更改，否则测试会因 GT 数据不匹配而失败。对用户无影响。
- 风险标记：暂无

关联脉络

- PR #13 ci-data: Update LingBot realtime consistency GT: 本 PR 引用了 ci-data 仓库的 PR#13 以获取最新 GT 数据，两者需同步合并。