

PR #27055 完整报告

sgl-project/sglang

[diffusion] CI: ad lingbot case

合并时间: 2026-06-02 22:26

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27055>

执行摘要

- 一句话: LingBot 实时测试用例内联至 ONE_GPU_CASES
- 推荐动作: 值得快速合并, 无阻塞点。这是测试基础设施的清理, 提升了 CI 用例注册的可发现性和一致性, 适合在单 GPU 测试普遍采用此模式。

功能与动机

PR 标题和 body 明确指出, 此变更旨在将 LingBot realtime Plastic Beach 测试用例内联至 ONE_GPU_CASES, 并移除独立的工厂注册逻辑, 以便 AST 驱动的 diffusion 分区解析器能直接看到该用例, 从而自动将其分配到正确的 CI 资源分区。

实现拆解

变更涉及两个测试文件, 均属于测试基础设施范畴:

1. 移除工厂函数并内联用例 (`gpu_cases.py`) - 删除从 `test_utils.py` 导入的 `_make_lingbot_realtime_plastic_beach_case` 工厂函数 (该函数封装了 `DiffusionTestCase` 创建)。- 将工厂函数返回的 `DiffusionTestCase` 对象直接以字面量形式插入 ONE_GPU_CASES 列表的末尾 (排在 `ltx_2_3_hq_pipeline` 用例之后), 使列表结构扁平化。- 删除文件末尾的 `ONE_GPU_CASES.append(_make_lingbot_realtime_plastic_beach_case())` 语句, 消除动态追加。
2. 更新一致性 GT 数据版本 (`test_utils.py`) - 将 `SGL_TEST_FILES_CI_DATA_REVISION` 从 `4a62a21f0f8bcc54c3bc6de5dbf25515bcb92b7e` 更新为 `8dbe6bc44cda46065fde2168a3d10fd792e2b289`, 以便 CI 使用与新用例匹配的 ground-truth 数据。这是独立提交, 无代码逻辑变更。

关键文件:

- `python/sglang/multimodal_gen/test/server/gpu_cases.py` (模块 GPU 测试用例; 类别 test; 类型 test-coverage; 符号 `_make_lingbot_realtime_plastic_beach_case`): 主要变更: 移除工厂函数 `_make_lingbot_realtime_plastic_beach_case`, 将用例内联至 ONE_GPU_CASES 列表, 并删除末尾的 `.append` 调用。
- `python/sglang/multimodal_gen/test/test_utils.py` (模块 测试工具; 类别 test; 类型 test-coverage): 更新一致性 GT 数据 revision, 使 CI 使用与新用例匹配的 ground-truth 数据。

关键符号: `_make_lingbot_realtime_plastic_beach_case` (已删除)

关键源码片段

`python/sglang/multimodal_gen/test/server/gpu_cases.py`

主要变更: 移除工厂函数 `_make_lingbot_realtime_plastic_beach_case`, 将用例内联至 `ONE_GPU_CASES` 列表, 并删除末尾的 `.append` 调用。

```
# 变更后: ONE_GPU_CASES 列表末尾直接内联 LingBot 实时测试用例
# 原工厂函数 _make_lingbot_realtime_plastic_beach_case 被移除,
# 其返回的 DiffusionTestCase 对象被直接写入列表, 便于 AST 解析器识别。
ONE_GPU_CASES: list[DiffusionTestCase] = [
    # ... 其他用例 ...
    DiffusionTestCase(
        "lingbot_world_realtime_plastic_beach",
        DiffusionServerArgs(
            model_path="robbyant/lingbot-world-fast-diffusers",
            modality="video",
            num_gpus=1,
            extras=["--pipeline-class-name LingBotWorldCausalDMDPipeline"],
            text_encoder_cpu_offload=True,
        ),
        LINGBOT_WORLD_REALTIME_sampling_params,
        run_component_accuracy_check=False,
        run_models_api_check=False,
        run_t2v_input_reference_check=False,
    ),
]
```

评论区精华

无人工 review 评论。Gemini Code Assist 自动评论仅摘要了变更内容, 未提出反馈。

- 暂无高价值评论线程

风险与影响

- 风险: 风险极低:
 - 仅涉及测试配置, 无任何运行时或核心逻辑变更。
 - AST 解析器依赖的列表结构未改变, 仅将动态追加改为静态列表字面量, 解析行为完全一致。
 - 一致性 GT revision 更新是常规数据同步, 若 revision 无效, CI 将立即失败并易于回滚。
- 影响: 影响范围仅限于 CI 测试:
 - LingBot realtime Plastic Beach 用例现在能被 AST 解析器自动分配到 1-GPU 测试分区, 而非依赖单独的注册逻辑。
 - 对用户无影响, 对系统无影响。

- 团队可受益于更统一的测试用例管理方式：所有单 GPU 用例均通过 ONE_GPU_CASES 列表注册，降低维护负担。
- 风险标记：暂无

关联脉络

- PR #27026 [diffusion] Add realtime WebUI super resolution controls: 同一功能线（LingBot 实时推理），此前已添加实时 WebUI 支持，本 PR 补充了 CI 测试覆盖。
- PR #27023 [diffusion] Optimize LingBot realtime transformer path: 同一功能线，优化 LingBot 实时推理路径，本 PR 确保优化后的代码有 CI 验证。