

PR #32128 完整报告

sgl-project/sglang

[Kernel] Reclassify kernel tests by ops group + move helpers out of the package (RFC #29630)

合并时间: 2026-07-23 12:18

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/32128>

执行摘要

- 一句话: 内核测试按 ops 组重新分类并移出辅助包
- 推荐动作: 该 PR 侧重重构和整理, 不涉及核心逻辑变更。建议阅读 kernels/README 了解最终布局 and 分类原则。对于贡献内核测试的开发者, 应参考新的 test/registered/kernels/ops/<group>/ 目录放置测试。设计决策 (如 tests 与 benchmarks 分离、辅助包单独放置) 值得关注。

功能与动机

RFC #29630 提出统一内核命名空间和测试组织。本 PR 旨在将内核测试镜像 `sglang.kernels.ops.<group>` 结构, 使测试分类与 ops 分组一致, 并将测试文件彻底移出 `sglang` 运行时包, 符合包内不得包含注册测试的约束。

实现拆解

1. 将测试按 ops 组重新分类: 将 test/registered/jit/ (102 个测试) 和 test/registered/kernels/ (21 个) 下的文件移到 test/registered/kernels/ops/<group>/, 根据每个测试导入的 `sglang.kernels.ops.<group>` 进行分类。
2. 分离 benchmarks: 在代码审查讨论中要求将 benchmarks 与 tests 分开, 后续 commit 将 test/registered/kernels/ops/<group>/benchmark/ 迁移到 test/registered/kernels/benchmark/<group>/, 并移动了 nvfp4 bench 数据夹具。
3. 将测试辅助包移出运行时包: 将原本位于 python/sglang/kernels/jit/tests/ 的共享测试辅助代码 (utils.py, kv_canary/, deepseek_v4/) 转移到 python/sglang/test/kernels/, 并更新所有测试文件中的导入路径。
4. 注册未注册的测试: 发现了 6 个测试函数 (如 minimax_m3_mxfp8、sconv_extend_metadata、moe_preprocess 等) 没有 CI 注册, 添加了 register_cuda_ci 或 register_amd_ci 装饰器, 确保它们被 CI 覆盖。
5. 解决命名冲突: 将两个同名的 test_sigmoid_gate_mul.py 之一重命名为 test_sigmoid_gate_mul_broadcast.py。
6. 更新文档与 CI 过滤器: 更新 kernels/README 反映最终布局, 将 CI 修改过滤器从 test/registered/jit/** 改为 test/registered/kernels/**。

关键文件:

- `test/registered/kernels/ops/kv_canary/test_verify_hand.py` (模块 验证 hand; 类别 test; 类型 rename-or-move) : 典型测试搬迁示例, 展示了导入路径从 `sglang.kernels.jit.tests` 到 `sglang.test.kernels` 的变更, 是辅助包移出运行时的直接体现。
- `python/sglang/test/kernels/kv_canary/_canary_helpers.py` (模块 金丝雀辅助; 类别 test; 类型 rename-or-move) : 共享测试辅助包从运行时包移出到 `sglang.test.kernels`, 该文件是辅助包搬迁的代表。
- `test/registered/kernels/ops/layernorm/test_minimax_m3_rmsnorm.py` (模块 层归一化; 类别 test; 类型 test-coverage; 符号 `register_amd_ci`) : 首次注册到 CI 的测试示例, 从包内迁移并添加 `register_amd_ci`, 体现了辅助包搬迁和注册新测试的模式。
- `python/sglang/kernels/README.md` (模块 文档; 类别 docs; 类型 documentation) : 文档更新反映最终布局, 指导贡献者了解当前结构。

关键符号: 未识别

关键源码片段

`test/registered/kernels/ops/kv_canary/test_verify_hand.py`

典型测试搬迁示例, 展示了导入路径从 `sglang.kernels.jit.tests` 到 `sglang.test.kernels` 的变更, 是辅助包移出运行时的直接体现。

```
# test/registered/kernels/ops/kv_canary/test_verify_hand.py (head)
# 搬迁后导入路径全部从 sglang.kernels.jit.tests 变为 sglang.test.kernels

from sglang.kernels.ops.kv_canary import consts
from sglang.kernels.ops.kv_canary.consts import splitmix64, splitmix64_mix3
from sglang.kernels.ops.kv_canary.verify import (
    CanaryLaunchTag,
    RealKvSource,
    VerifyOrWriteContext,
    VerifyPlan,
    launch_canary_verify_kernel,
)
from sglang.kernels.ops.kv_canary.verify_ref import (
    _compute_real_kv_hash_scalar,
    launch_canary_verify_kernel_torch_reference,
)
from sglang.kernels.ops.kv_canary.write_ref import (
    launch_canary_write_kernel_torch_reference,
)
from sglang.test.ci.ci_register import register_amd_ci, register_cuda_ci
from sglang.test.kernels.kv_canary._canary_helpers import (
    FakeViolationLog,
    assert_only_bits_set,
    chain_anchor_signed,
    make_canary_buf,
    make_canary_buf_pair,
    make_log_pair,
```

```

make_real_kv_source,
make_real_kv_sources,
make_verify_plan,
make_verify_plan_pair,
make_write_plan,
read_slot_fields,
stamp_clean_chain,
stamp_pair,
to_signed_int64,
write_slot_fields,
)
from sglang.test.kernels.kv_canary._differential import (
    _run_both_verify,
    run_verify_diff,
)
from sglang.test.kernels.kv_canary._fixtures import clone_real_kv_sources
from sglang.test.kernels.kv_canary._hand_oracle import (
    _hand_fold_all,
    _hand_fold_partial,
)

```

评论区精华

PR 没有公开的 review 评论，但从提交历史可以看出 review 讨论要求将 benchmarks 从 ops/ test 目录中分离出来（commit 8c7935f 标题 'Per review: keep benchmarks out of the ops/ test dirs'）。最终形成了 test/registered/kernels/benchmark// 独立目录。此外，RFC #29630 本身经过多轮讨论，本 PR 是其实施的最后一步。

- Benchmarks 应与 tests 分离 (design): 采纳建议，创建 test/registered/kernels/benchmark// 目录，将 benchmarks 从 ops/ 下移出。

风险与影响

- 风险：主要风险在于大范围的文件移动和导入路径修改可能导致遗漏的引用错误。虽然作者验证了 hooks 和 imports 通过，但 205 个文件的变更量增加了漏改的可能性。6 个测试首次注册可能因环境依赖缺失而失败。CI 修改过滤器的变更也需要关注是否覆盖了所有必要的测试。整体风险较低，但需通过 CI 运行验证。
- 影响：对用户无影响（不涉及用户逻辑）。对开发者：测试目录结构更清晰，便于按 ops 分组查找和添加测试；辅助包从运行时代码中移除，避免了包内测试文件的误引入。对 CI 系统：修改过滤器更新，确保正确触发内核相关测试；新增 6 个测试进入 CI 套件，提升了 coverage。
- 风险标记：大量文件搬迁，导入路径变更，新注册测试可能失败，CI 过滤器更新需验证

关联脉络

- PR #29630 RFC: Unify kernel tests and namespaces: 该 PR 是 RFC #29630 的后续，实现其提出的测试组织方案。

- PR #32072 [Kernel] RFC #29630 finale: retire `sglang.jit_kernel` into `sglang.kernels`:
完成了内核命名空间统一，本 PR 在此基础上进一步整理测试结构。