

PR #44819 完整报告

vllm-project/vllm

[CI] Consolidate multimodal endpoint tests.

合并时间: 2026-06-08 19:48

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/44819>

执行摘要

- 一句话: 合并多模态入口测试文件并统一资源管理
- 推荐动作: 值得快速阅读。该 PR 展示了测试目录布局规范化的实践: 通过 `conftest.py` 共享常量和按功能域组织测试文件, 对后续多模态测试扩展具有参考价值。但由于无核心逻辑变动, 不强制深度审查。

功能与动机

PR 描述指出合并多模态入口测试的动机: "These tests have more in common. Reduce the number of files in the tests/entrypoints/openai/chat_completion folder. These all make the endpoint tests easy to maintain." 同时, review 评论进一步确认多模态输入必须使用 chat 请求, 因此测试函数名需要修正。

实现拆解

1. 创建共享资源层: 新建 `tests/entrypoints/multimodal/conftest.py`, 定义 `TEST_IMAGE_ASSETS` 常量列表 (包含四种不同格式和色彩的测试图片), 供所有多模态测试文件统一引用, 消除跨文件重复定义。
2. 迁移 LLM 多模态测试: 从 `tests/entrypoints/llm/test_chat.py` 中移除 `vision_llm fixture` 和 `test_chat_multi_image` 函数 (删除 42 行), 并在 `tests/entrypoints/multimodal/llm/test_chat.py` 中重新创建 (新增 50 行), 将 `import` 由旧的 `tests.entrypoints.openai.chat_completion.test_vision` 改为新的 `tests.entrypoints.multimodal.conftest`。
3. 重组 OpenAI Chat Completion 测试: 将原 `tests/entrypoints/openai/chat_completion/test_completion_with_image_embeds.py` 重命名为 `tests/entrypoints/multimodal/openai/chat_completion/test_chat_completion_with_image_embeds.py`, 并将测试函数名从 `test_completions_with_image_embeds` 改为 `test_chat_completions_with_image_embeds`; 将 `test_vision.py` 重命名并移除本地 `TEST_IMAGE_ASSETS` 定义, 改为从 `conftest` 导入。
4. 重组 Responses 测试: 将 `tests/entrypoints/openai/responses/test_image.py` 重命名至 `tests/entrypoints/multimodal/openai/responses/test_image.py`, 移除重复资源定义并改为从 `conftest` 导入。
5. 补充包初始化文件: 在新增目录中添加空的 `__init__.py` 文件。

6. 更新 CI 配置：修改 `.buildkite/test-amd.yaml` (+29 行) 和 `.buildkite/test_areas/entrypoints.yaml` (+13 行)，将旧路径替换为新路径。

关键文件：

- `tests/entrypoints/multimodal/llm/test_chat.py` (模块 LLM 测试；类别 test；类型 test-coverage；符号 `vision_llm`, `test_chat_multi_image`)：新增文件，将多图片聊天测试从纯文本测试文件中分离，并复用共享资源常量，体现测试组织重构的核心目标。
- `tests/entrypoints/llm/test_chat.py` (模块 LLM 测试；类别 test；类型 test-coverage；符号 `vision_llm`, `test_chat_multi_image`)：被修改的文件，删除了多图片测试相关的 fixture 和函数，减少了该文件的责任范围，使其专注于纯文本聊天测试。
- `tests/entrypoints/multimodal/conftest.py` (模块 测试共享层；类别 test；类型 test-coverage；符号 `TEST_IMAGE_ASSETS`)：新增的共享配置文件，集中定义了多模态测试所需的图片资源常量，避免了在多个测试文件中重复定义，降低了维护成本。
- `tests/entrypoints/multimodal/openai/chat_completion/test_chat_completion_with_image_embeds.py` (模块 Chat 测试；类别 test；类型 rename-or-move；符号 `test_completions_with_image_embeds`, `test_chat_completions_with_image_embeds`)：被重命名并修改了测试函数名，以兼容 Chat Completion API 语义，体现了测试正确性改进。
- `tests/entrypoints/multimodal/openai/chat_completion/test_vision.py` (模块 视觉测试；类别 test；类型 rename-or-move)：被重命名并移除了重复的资源定义，改为从共享 `conftest` 导入，是统一资源管理的关键步骤。
- `tests/entrypoints/multimodal/openai/responses/test_image.py` (模块 响应测试；类别 test；类型 rename-or-move)：重命名并移除重复资源定义，与 `test_vision.py` 类似的清理动作，确保所有多模态测试共享同一资源集。
- `.buildkite/test-amd.yaml` (模块 CI 配置；类别 config；类型 configuration)：更新了 CI 测试路径，确保新的 multimodal 测试路径被 AMD 测试流水线覆盖。
- `.buildkite/test_areas/entrypoints.yaml` (模块 CI 配置；类别 config；类型 configuration)：更新了 entrypoints 测试区域的路径，确保通用 CI 流水线能发现迁移后的测试。

关键符号：`vision_llm`, `test_chat_multi_image`, `test_chat_completions_with_image_embeds`

关键源码片段

`tests/entrypoints/multimodal/llm/test_chat.py`

新增文件，将多图片聊天测试从纯文本测试文件中分离，并复用共享资源常量，体现测试组织重构的核心目标。

```
# SPDX-License-Identifier: Apache-2.0
# SPDX-FileCopyrightText: Copyright contributors to the vLLM project
import weakref

import pytest

from tests.entrypoints.multimodal.conftest import TEST_IMAGE_ASSETS
from vllm import LLM
```

```
from vllm.distributed import cleanup_dist_env_and_memory
```

```
@pytest.fixture(scope="function")
```

```
def vision_llm():
```

```
    # pytest caches the fixture so we use weakref.proxy to
```

```
    # enable garbage collection
```

```
    llm = LLM(
```

```
        model="microsoft/Phi-3.5-vision-instruct",
```

```
        max_model_len=4096,
```

```
        max_num_seqs=5,
```

```
        enforce_eager=True,
```

```
        trust_remote_code=True,
```

```
        limit_mm_per_prompt={"image": 2},
```

```
        seed=0,
```

```
    )
```

```
    yield weakref.proxy(llm)
```

```
    del llm
```

```
    cleanup_dist_env_and_memory()
```

```
@pytest.mark.parametrize(
```

```
    "image_urls", [[TEST_IMAGE_ASSETS[0], TEST_IMAGE_ASSETS[1]], indirect=True  
)
```

```
def test_chat_multi_image(vision_llm, image_urls: list[str]):
```

```
    messages = [
```

```
        {
```

```
            "role": "user",
```

```
            "content": [
```

```
                *(
```

```
                    {"type": "image_url", "image_url": {"url": image_url}}
```

```
                    for image_url in image_urls
```

```
                ),
```

```
                {"type": "text", "text": "What's in this image?"},
```

```
            ],
```

```
        }
```

```
    ]
```

```
    outputs = vision_llm.chat(messages)
```

```
    assert len(outputs) >= 0
```

tests/entrypoints/multimodal/conftest.py

新增的共享配置文件，集中定义了多模态测试所需的图片资源常量，避免了在多个测试文件中重复定义，降低了维护成本。

```
# SPDX-License-Identifier: Apache-2.0
```

```
# SPDX-FileCopyrightText: Copyright contributors to the vLLM project
```

```
# 多模态测试公用图片资源常量，覆盖不同图片格式（JPG/PNG）和色彩模式（灰度 /RGB/RGBA）
# 所有多模态测试文件通过以下语句导入：
# from tests.entrypoints.multimodal.conftest import TEST_IMAGE_ASSETS
TEST_IMAGE_ASSETS = [
    "2560px-Gfp-wisconsin-madison-the-nature-boardwalk.jpg", # 典型自然风景 JPG
    "Grayscale_8bits_palette_sample_image.png", # 8 位灰度 PNG
    "1280px-Venn_diagram_rgb.svg.png", # RGB 彩色 PNG
    "RGBA_comp.png", # 带有 Alpha 通道的 PNG
]
```

评论区精华

- 目录布局建议：审核者 DarkLight1337 建议镜像 tests/entrypoints 目录结构（即在 multimodal 下使用 openai/chat_completion 而非 online），作者 noooop 采纳并调整。
- 测试函数名修正：作者 noooop 指出原文件 `test_completion_with_image_embeds.py` 应更名为 `test_chat_completions_with_image_embeds`，因为多模态输入必须使用 Chat Completion API。DarkLight1337 最初因文件截断产生误解，讨论后确认正确并批准。
- 测试目录结构布局建议 (design): 作者 noooop 采纳了建议，调整了目录结构，采用了与 tests/entrypoints 一致的布局。
- 测试函数名更正：Chat API vs Completions API (correctness): 文件被重命名为 `test_chat_completion_with_image_embeds.py`，测试函数名改为 `test_chat_completions_with_image_embeds`。

风险与影响

- 风险：风险极低。主要风险是 CI 配置未同步更新导致测试丢失，但本 PR 已同步修改 `.buildkite/test-amd.yaml` 和 `.buildkite/test_areas/entrypoints.yaml`，确保新路径被覆盖。测试本身无核心逻辑变更，回归风险可忽略。
- 影响：
 - 用户影响：无。纯测试文件和 CI 配置变更，不影响运行时功能。
 - 系统影响：CI 管道将按新路径执行测试，旧路径不再被覆盖，无重复风险。
 - 团队影响：多模态测试集中在 tests/entrypoints/multimodal/ 下，共享资源减少冗余，提升了可维护性和一致性。
 - 风险标记：CI 配置更新，测试路径变更，无核心逻辑变更

关联脉络

- 暂无明显关联 PR