

PR #31899 完整报告

sgl-project/sglang

[AMD] Add msgpack to ROCm diffusion deps (fix multimodal-gen unit test
ModuleNotFoundError)

合并时间: 2026-08-06 16:52

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/31899>

执行摘要

- 一句话: 为 ROCm diffusion 依赖补充 msgpack, 修复 AMD 单测失败
- 推荐动作: PR 只有一行依赖声明, 逻辑直白, 无需精读。值得关注的是 AMD 依赖清单 `pyproject_other.toml` 与主清单 `pyproject.toml` 的同步维护问题, 建议后续考虑自动化校验或合并清单, 避免类似遗漏。

功能与动机

PR body 明确指出 `multimodal-gen-unit-test-amd-rocm720` 任务每晚失败 (CI 窗口 7/8 天), 报错 `ModuleNotFoundError: No module named 'msgpack'`。根因是 `runtime/entrypoints/vla/protocol.py` 使用 `pack_msgpack/unpack_msgpack`, 而 AMD 专用的 `python/pyproject_other.toml` 的 `diffusion_common` 组缺少该依赖; ROCm 7.0 镜像恰好传递提供 msgpack, 而 7.2 镜像没有, 因此必须显式声明。

实现拆解

本变更包含以下步骤:

1. 定位根因: 确认 msgpack 是 multimodal_gen 的运行时依赖, 被 `runtime/entrypoints/vla/protocol.py` 中的 `pack_msgpack/unpack_msgpack` 使用, 对应单元测试 `test/unit/test_pi05_action_api.py` 会导入它。
2. 修改依赖清单: 在 `python/pyproject_other.toml` 的 `diffusion_common` 列表中按字母序 (位于 `moviepy` 与 `opencv` 之间) 添加 "msgpack"。该清单用于 ROCm wheel/ 镜像构建 (通过 `docker/rocm.Dockerfile` 替换主清单后 `pip install python[srt_hip,diffusion_hip]`)。
3. 覆盖与验证: 所有 AMD diffusion 变体 (`diffusion_hip/diffusion_musa/diffusion_mps`) 均继承 `diffusion_common`, 因此一次添加覆盖全平台。维护者触发 ROCm 7.0 与 7.2 的定向 `targeted validation`, 两者均通过。

关键文件:

- `python/pyproject_other.toml` (模块 依赖清单; 类别 `config`; 类型 `configuration`): AMD ROCm 专属依赖清单, 新增 msgpack 是唯一的变更, 修复 nightly 单测的 `ModuleNotFoundError`。

关键符号: 未识别

关键源码片段

python/pyproject_other.toml

AMD ROCm 专属依赖清单，新增 msgpack 是唯一的变更，修复 nightly 单测的 ModuleNotFoundError。

```
# python/pyproject_other.toml —— AMD/ROCm 使用的依赖清单（构建时替换主 pyproject.toml）
# diffusion_common 被 diffusion_hip / diffusion_musa / diffusion_mps 共享
# 新增 msgpack: VLA 协议运行时依赖（runtime/entrypoints/vla/protocol.py 的 pack_msgpack /
unpack_msgpack）
# ROCm 7.2 基础镜像不传递提供该包，需显式声明；主清单 pyproject.toml 已有此项
diffusion_common = [
    "addict",
    "cloudpickle",
    "diffusers==0.37.0",
    "imageio==2.36.0",
    "imageio-ffmpeg==0.5.1",
    "moviepy>=2.0.0",
    "msgpack", # 按字母序插入，位于 moviepy 与 opencv 之间
    "opencv-python-headless==4.10.0.84",
    "PyYAML==6.0.1",
    "remote-pdb",
    "scikit-image==0.25.2",
    "trimesh>=4.0.0",
    "xatlas",
]
```

评论区精华

作者在 PR 评论中说明：[Pi0.5 VLA protocol](#) (#30633) 在运行时使用 [msgpack](#)，ROCm 7.2 镜像不传递提供该依赖，导致 nightly 失败；并因是外部 fork PR，需要维护者添加标签。维护者 [bingxche](#) 随即针对 head SHA 触发了 ROCm 7.0 与 7.2 的定向验证，结果均为 [Passed](#)。Review 过程中无实质代码讨论，[gemini-code-assist\[bot\]](#) 无反馈，[HaiShaw](#) 直接批准。

- msgpack 缺失导致 nightly 失败 (other): 维护者确认根因，补丁添加依赖后触发验证。
- ROCm 7.0/7.2 定向验证 (testing): 依赖补充在 ROCm 7.0 和 7.2 均生效。

风险与影响

- 风险：风险极低：msgpack 是纯 Python 包，无编译依赖，且主清单 python/pyproject.toml 的 diffusion 组已包含它，说明与现有依赖无冲突。唯一潜在风险是 AMD 清单与主清单的依赖漂移，未来若主清单移除 msgpack，此处可能残留声明，但影响极小。安装镜像体积增加可忽略。
- 影响：影响范围限于 AMD ROCm 平台的 diffusion（多模态生成）安装与 CI：修复 multimodal-gen-unit-test-amd-rocm720 任务的 nightly 稳定性，使 ROCm 7.2 镜像上的单元测试不再因缺包失败。对主流程、非 AMD 平台无影响。团队层面可消除一个持续的

CI 噪声，减少夜间告警。

- 风险标记：依赖清单双源维护，AMD 安装依赖变更

关联脉络

- PR #31483 [AMD] ci: run vetted nested multimodal_gen unit tests on AMD: 该 PR 启用 AMD CI 运行嵌套多模态单测，使 test_pi05_action_api.py 在 ROCm 任务中执行，从而暴露了 msgpack 缺失的问题。