

# PR #23220 完整报告

sgl-project/sglang

Bugfix: Qwen3-VL-MoE adapt encoder\_only

合并时间: 2026-05-22 11:12

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/23220>

## 执行摘要

- 一句话: 修复 Qwen3-VL-MoE encoder-only 模式启动失败
- 推荐动作: 建议尽快合并并此修复。改动简单且安全, 能够解决特定场景的服务阻塞问题。未来应考虑为 encoder-only 模式添加单元测试, 覆盖模型加载路径。

## 功能与动机

该 bug 由 PR #19135 引入, `self.model` 只在非 encoder-only 模式下创建, 而本次变更未考虑 encoder-only 场景。具体错误信息: `AttributeError: 'Qwen3VLMoeForConditionalGeneration' object has no attribute 'model'`

## 实现拆解

1. 在 `python/sglang/srt/models/qwen3_vl_moe.py` 的 `load_weights` 方法中, 在访问 `self.model.start_layer` 之前增加 `and hasattr(self, 'model')` 条件。
2. 这样, 当 `self.model` 不存在时 (encoder-only 模式), 不会尝试访问 `start_layer`, 从而避免 `AttributeError`。
3. 修改后通过启动测试验证 (参见 PR body 中的成功日志)。
4. 未新增测试文件, 但修改本身是安全的防御性编码。

关键文件:

- `python/sglang/srt/models/qwen3_vl_moe.py` (模块 模型加载; 类别 source; 类型 core-logic; 符号 `load_weights`): 这是唯一的修改文件, 修复了 `load_weights` 方法在 encoder-only 模式下的 `AttributeError`。

关键符号: `load_weights`

## 关键源码片段

`python/sglang/srt/models/qwen3_vl_moe.py`

这是唯一的修改文件, 修复了 `load_weights` 方法在 encoder-only 模式下的 `AttributeError`。

```
# 文件: python/sglang/srt/models/qwen3_vl_moe.py, 方法 load_weights
for name, loaded_weight in weights:
    name = name.replace(r"model.language_model.", r"model.")
    layer_id = get_layer_id(name)
```

```
# 当 `--encoder-only` 启用时，父类不会创建 `self.model`，  
# 所以需要先检查 `hasattr(self, "model")`，避免 AttributeError。  
if (  
    "visual" not in name  
    and layer_id is not None  
    and hasattr(self, "model") # 修复行：在 encoder-only 模式下可能没有 self.model  
    and hasattr(self.model, "start_layer")  
    and (  
        layer_id < self.model.start_layer  
        or layer_id >= self.model.end_layer  
    )  
):  
    continue
```

## 评论区精华

无实质技术讨论。PR 由 `sclang-npu-bot` 和 `ping1jing2` 批准。评论主要是重跑 CI 的指令。

- PR 审核过程 (other): 批准并合并。

## 风险与影响

- 风险：风险极低。改动仅为增加属性存在性检查，不会影响已有逻辑。但若未来重构导致 `self.model` 的属性名称或创建时机变化，本处可能需要同步调整。
- 影响：影响范围仅限 Qwen3-VL-MoE 模型使用 `--encoder-only` 的场景。修复后该配置可正常启动并处理 `/encode` 请求，对相关用户有直接影响。无性能、安全或兼容性影响。
- 风险标记：暂无

## 关联脉络

- 暂无明显关联 PR