

用vllm部署模型后

原视频fps=25 qwen2.5vl instruct 7B模型抽帧率 fps=20

```
===== 统计结果 =====
Prefill 近似耗时 (首 token 延迟) : 6.378 秒
总耗时: 7.782 秒
Decode 近似耗时: 1.404 秒

模型响应内容:
这段视频展示了多个自然景观的航拍镜头:

1. 视频开始时, 展示了一片宁静的山谷, 阳光透过云层洒在草地上, 远处是连绵的山脉和几座小木屋, 环境非常平和。
2. 接着, 镜头切换到一个湖

(qwen2.5vl) l40@k8smaster:~/newdisk1/mf1$ /home/l40/anaconda3/envs/qwen2.5vl/bin/python /home/l40/newdisk1/mf1/qwen2.5vl_agents/test/test_prefill_time/vllm/get_fps_and_frames.py
正在分析视频: /home/l40/newdisk1/mf1/data/data/videos/test.mp4
-----
视频帧率: 25.00 FPS
视频总帧数: 2073 帧
(qwen2.5vl) l40@k8smaster:~/newdisk1/mf1$ /home/l40/anaconda3/envs/qwen2.5vl/bin/python /home/l40/newdisk1/mf1/qwen2.5vl_agents/test/test_prefill_time/vllm/test_videos_prefill_directly.py
使用模型: qwen2.5vl
视频路径: /home/l40/newdisk1/mf1/data/data/videos/test.mp4
对应 URL: file:///home/l40/newdisk1/mf1/data/data/videos/test.mp4
开始发送请求并计时...

===== 统计结果 =====
Prefill 近似耗时 (首 token 延迟) : 0.960 秒
总耗时: 2.361 秒
Decode 近似耗时: 1.402 秒

模型响应内容:
这段视频展示了多个自然景观的航拍镜头:

1. 视频开始时, 展示了一片宁静的山谷, 阳光透过云层洒在草地上, 周围点缀着几座小木屋和松树。整个场景显得非常平和与美丽。
2. 接下来, 镜头切换

(qwen2.5vl) l40@k8smaster:~/newdisk1/mf1$
```

解决了时间上的瓶颈

现在的主要问题是 attention failure:

评估指标:

对象级异常理解指标 (Object-level Anomaly Understanding Score, OAUS)

它由两部分组成:

- 1. 对象级异常检测得分
- 2. 异常推理文本质量

评估维度	指标名称	指标符号	作用说明
对象级异常检测	对象异常识别 F1 值	Obj-F1	衡量模型能否正确识别“某个对象是否发生异常”。
对象级异常检测	对象异常覆盖率	Obj-Cov	在真实发生异常时，模型是否至少一次检测到了该异常对象。
异常推理文本质量	异常要素匹配率	AEMR	判断推理文本中“对象、行为、异常类型”是否与真实情况一致。
异常推理文本质量	文本质量得分 (BLEU/METEOR/ROUGE 综合)	TQS	衡量生成推理文本的流畅性、清晰度和参考文本的一致性。
综合能力指标	综合异常理解得分	OAUS	综合检测能力 (Obj-F1) 与推理能力 (AEMR+TQS) 的统一指标。

接下来要解决的问题：

时间上的attention failure

空间上的attention failure