**Faster-whisper large v3多进程性能测试**

# 测试前提

* eval同一份数据，common 17 ar test部分，总数10480条。
* 相同的一组checkpoint，总共5个。checkpoint基于openai/whisper-large-v3训练
* eval的基座模型是Systran/faster-whisper-large-v3
* 运行在同一个docker container中
* GPU : 1块nvidia a40

# 测试内容

* 多线程对比单线程是否有速度提升
* 更多的线程对比更少的线程是否有速度提升
* 训练后的模型对比基座模型，wer和cer是否有下降

# 测试结果

## 单进程

总耗时：5.45h

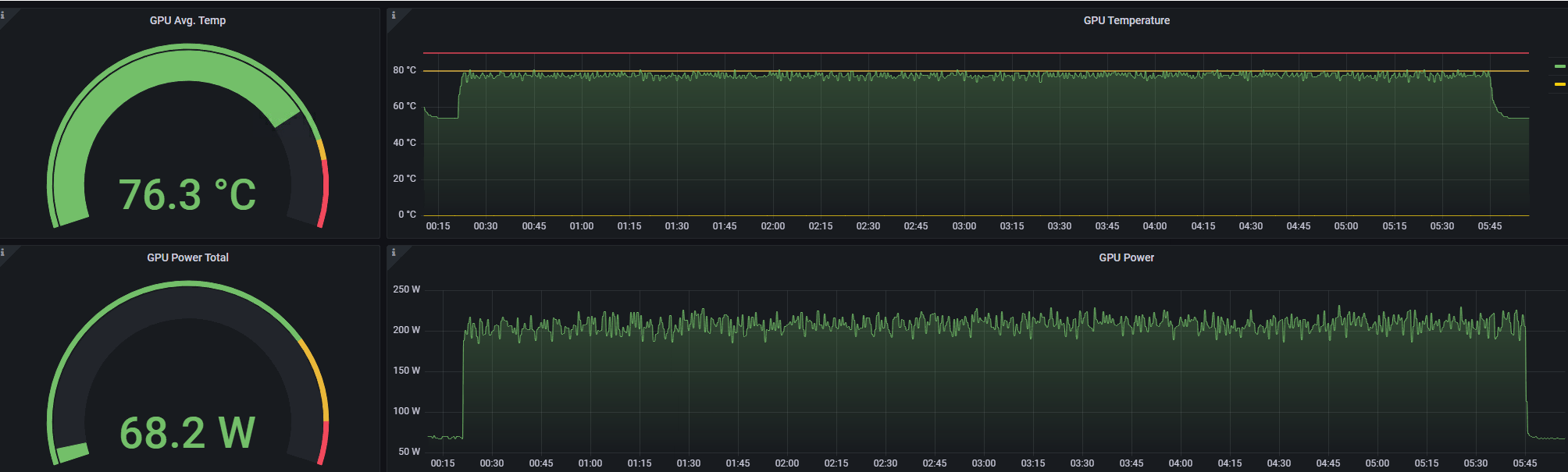
nvidia-smi查看进程情况

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

GPU功率：

温度不超过80度，功率在190w-240w之间波动



## 2个进程

总耗时：4.05h

nvidia-smi查看进程情况

表格

描述已自动生成

GPU功率：

温度不超过80度，功率主要在230w~260w之间波动

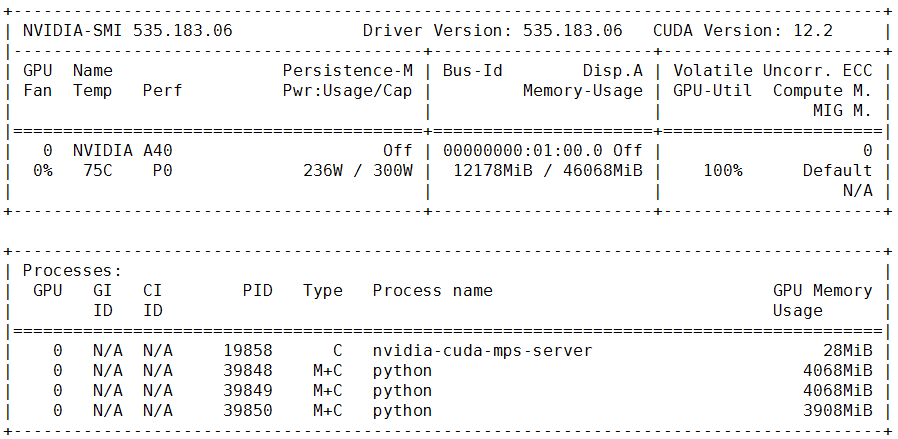
图表

描述已自动生成

## 3个进程

总耗时：3.48h

nvidia-smi查看进程情况



GPU功率：

温度不超过80度，功率主要在245w~275w之间波动

图形用户界面, 图表

描述已自动生成

## 4个进程

总耗时：3.28h

nvidia-smi查看进程情况

图片包含 表格

描述已自动生成

GPU功率：

温度不超过80度，功率主要在275w附近波动

电脑萤幕画面

描述已自动生成

## 5个进程

总耗时：3.01h

nvidia-smi查看进程情况

表格

描述已自动生成

GPU功率：

温度不超过80度，功率主要在270w以上波动，一半时间处于280w~290w之间

电脑游戏画面

描述已自动生成

# 测试总结

## 耗时趋势

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 进程数 | 1 | 2 | 3 | 4 | 5 |
| 总耗时(h) | 5.45 | 4.05 | 3.48 | 3.28 | 3.01 |

## 测试总结

经测试，由于随着进程数的增加，在3进程、4进程、5进程上，耗时减少的幅度明显减小，但功率却在不断的上升，推测算例单元已满，有任务排队，故gpu在压榨功率。

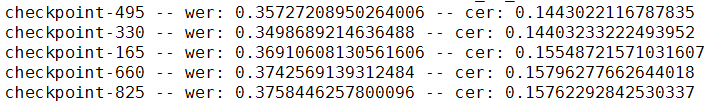
所以推荐执行进程4-5个。

## 错误率CER/WER

基座模型是指：faster-whisper-large-v3

### 4.3.1 微调后的模型

**float16 带标符**



**float16 不带标符**

文本

描述已自动生成

### 4.3.2 基座模型

|  |  |  |
| --- | --- | --- |
|  | wer | cer |
| float16 带标符 | 0.4227 | 0.1654 |
| float16不带标符 | 0.2667 | 0.0751 |

注：基座模型单线程执行1次在1h左右；将数据集分成4份，执行一次0.55h

### 4.3.3 测试总结

微调模型较基座模型错误率有下降，训练有效