### 基本信息

模型名称：*RDN*

原始代码链接：<https://github.com/yjn870/RDN-pytorch>

Gitee链接：*XXX*

承接人信息：*中山大学 计算机（软件）学院 张冬雨老师 学生徐一蘅*

华为初审人：

### 性能数据

训练状态下一个迭代耗时

|  |  |  |
| --- | --- | --- |
| Name | GPU/fps | NPU/fps |
| 1p | 200.667 | 242 |
| 8p | 1715.735 | 1431~1990 |

GPU PROF文件：



NPU PROF文件：

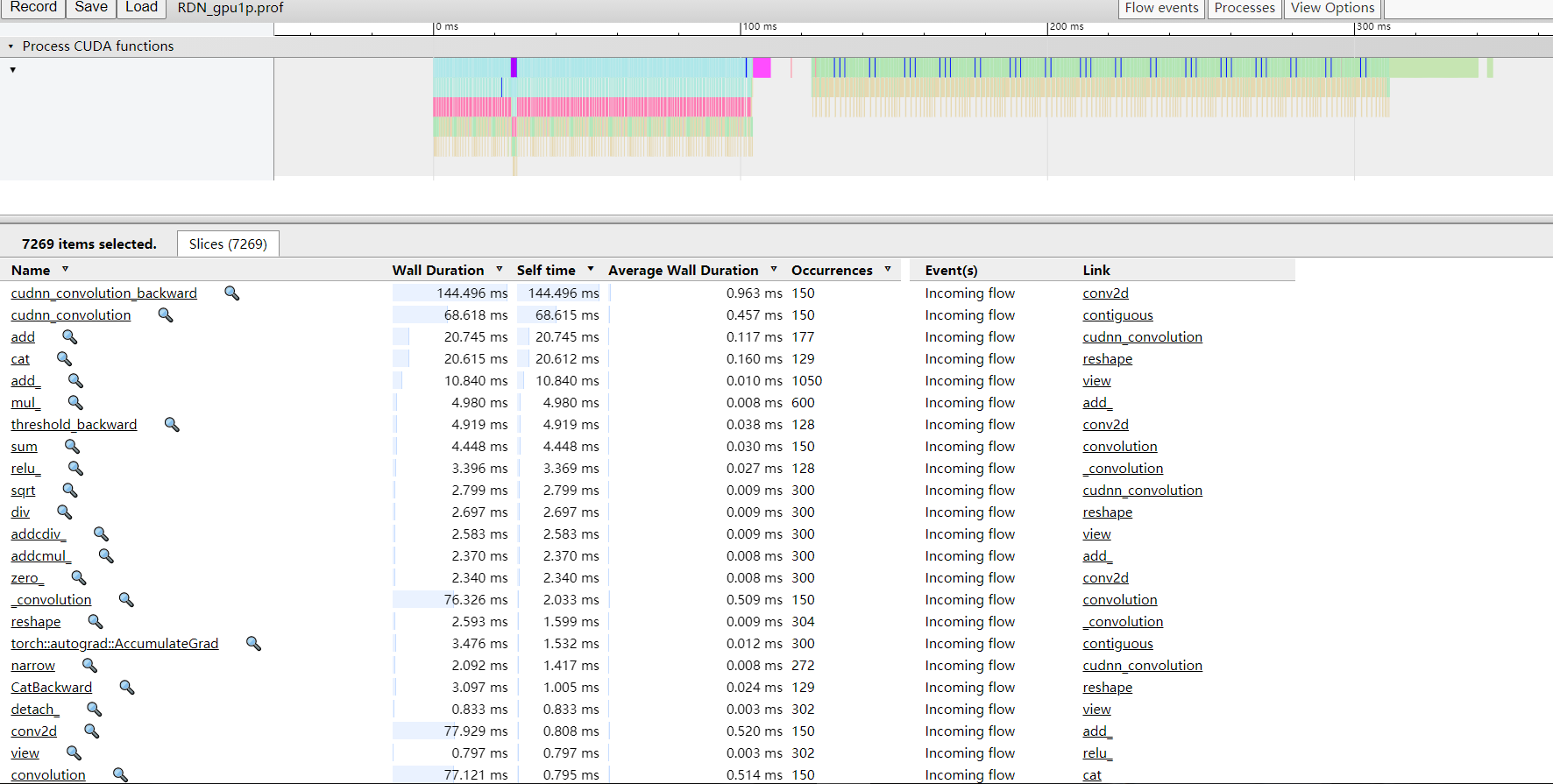


NPU CANN Profiling数据：

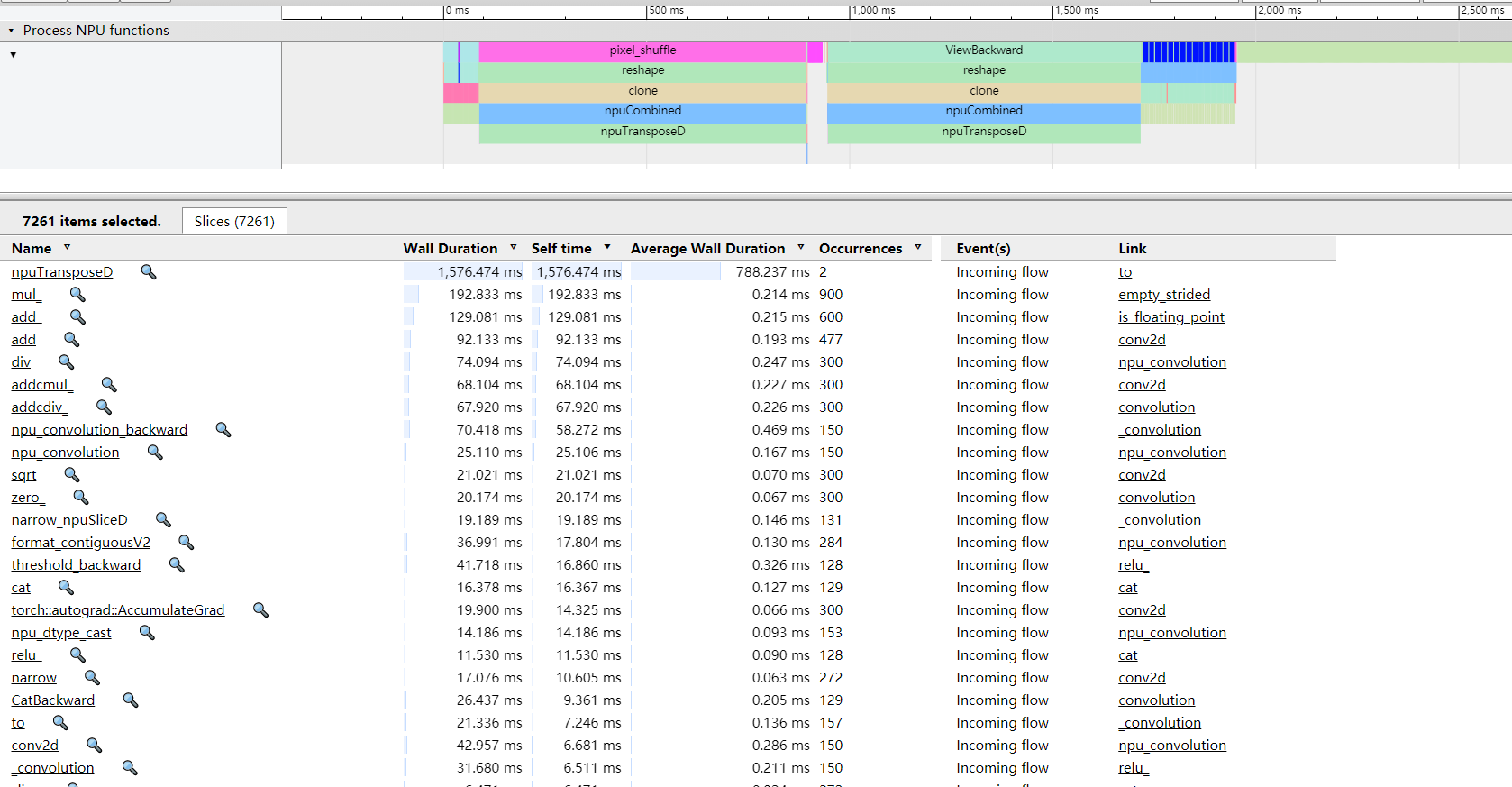


### 性能分析

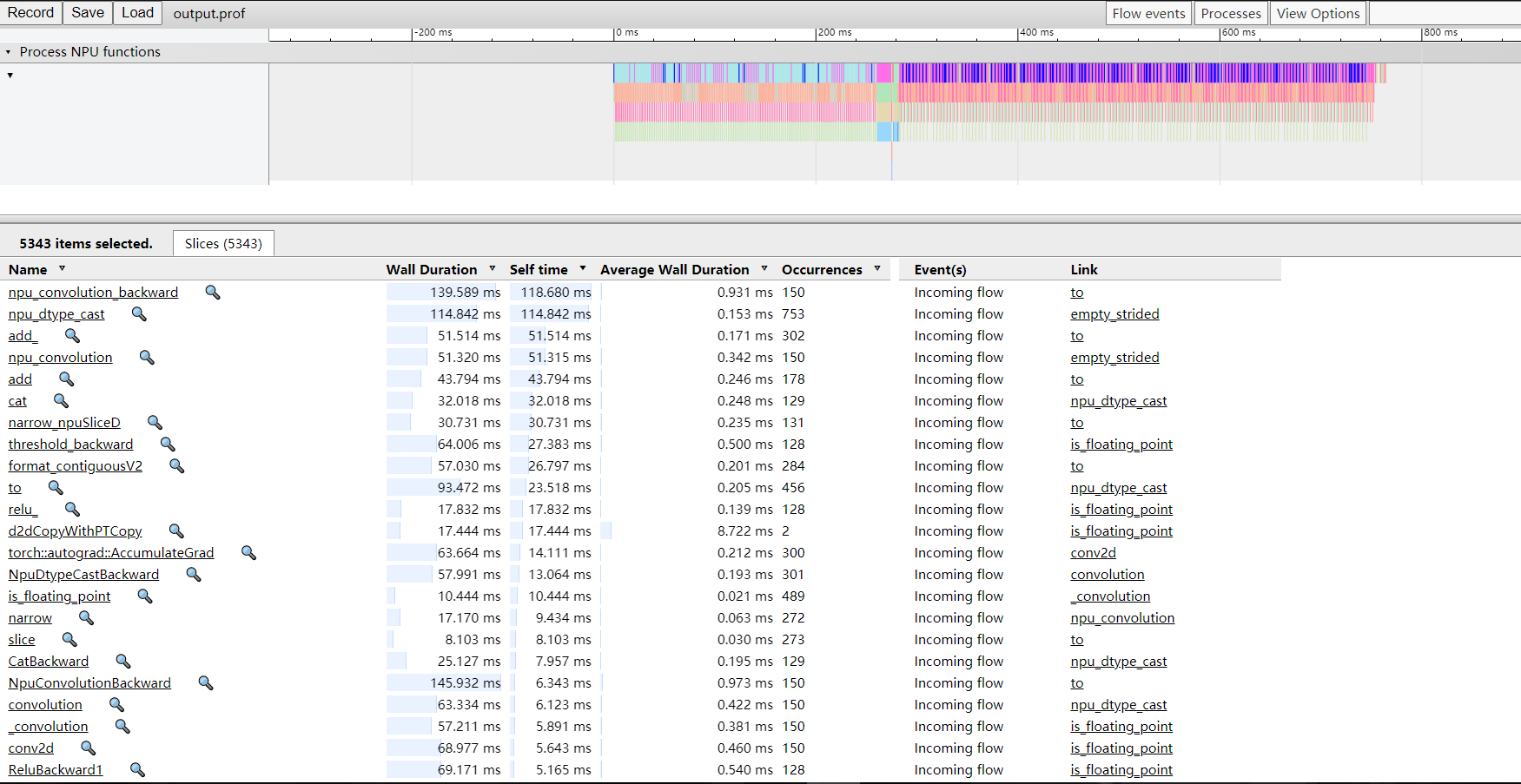
GPU算子执行时间，以下算子通过Wall Duration排序



提供的NPU单卡服务器中算子执行时间：



提供的NPU多卡服务器中算子执行时间：



相比于GPU，NPU上共计多出2300 ms耗时，主要的性能问题主要在于以下几个点：

1. pixel\_shuffle算子在GPU上共计耗时0.13 ms，在NPU上共计耗时806.801 ms，共计额外耗时806.7 ms。目前的解决方法是将模型中pixel\_shuffle转移到cpu中规避。
2. clone算子在GPU上共计耗时0.194 ms，在NPU上共计耗时1577 ms，共计额外耗时1577 ms。
3. NPU与GPU相比增加了npuTransposeD、npuCombined算子。npuTransposeD算子在NPU上共计耗时1576ms，共执行2次。NpuCombined算子在NPU上共计耗时1576ms，共执行2次。Issue已提，链接为：<https://gitee.com/ascend/modelzoo/issues/I44CC8?from=project-issue>。
4. reshape算子在GPU上共计耗时2.6ms，在NPU上共计耗时1577ms，共执行4次。
5. ViewBackward算子在GPU上共计耗时2.1ms，在NPU上共计耗时770ms，共执行一次。

#### 4其他问题：

在提供的单卡npu服务器中出现了pixel\_shuffle计算缓慢的现象，但是在提供的8卡的服务器中却没有出现，因此在8卡的服务器中没有在cpu进行规避。