

week-3

JX-Ma

2024/8/10

## 1 introduce

This weeks'work is as follow:

- Tested the correctness of the function that verifies convs correctness.
- Try using GPU Timer to test CUDNN and CUBLAS

## 2 experiment

### 2.1 experiment1

- CUDA-Version: 11.3
- CUDNN-Version: 8.2.1
- GCC-Version: 9.5
- Openmp-Version: 4.5
- pytorch-Version: 2.2.0

the maximum error:

- conv1 : 0.000106812
- conv2 : 0.000106812
- conv3 : 0
- conv4 : 0.0925999
- conv5 : 1.1837 and The 9948th element 68.3013 : 67.1176
- conv6 : 0.00409698
- conv7 : 0
- conv8 : 0.0442009

- conv9 : 0.00100708
- conv10 : 0.0536995
- conv11 : 0.0716
- conv12 : 0.0951996

I guess the reason for these errors is that when using the CUDNN acceleration library, it will choose the fastest algorithm for convolution calculation on its own, while some algorithms accelerate convolution by cropping some input tensor or filter tensor data, which may lead to a decrease in accuracy.

As introduced in the PyTorch documentation, some of the algorithms in CUDNN may be non deterministic, meaning that running them twice at the same time may result in inconsistent results.

## 2.2 experiment2

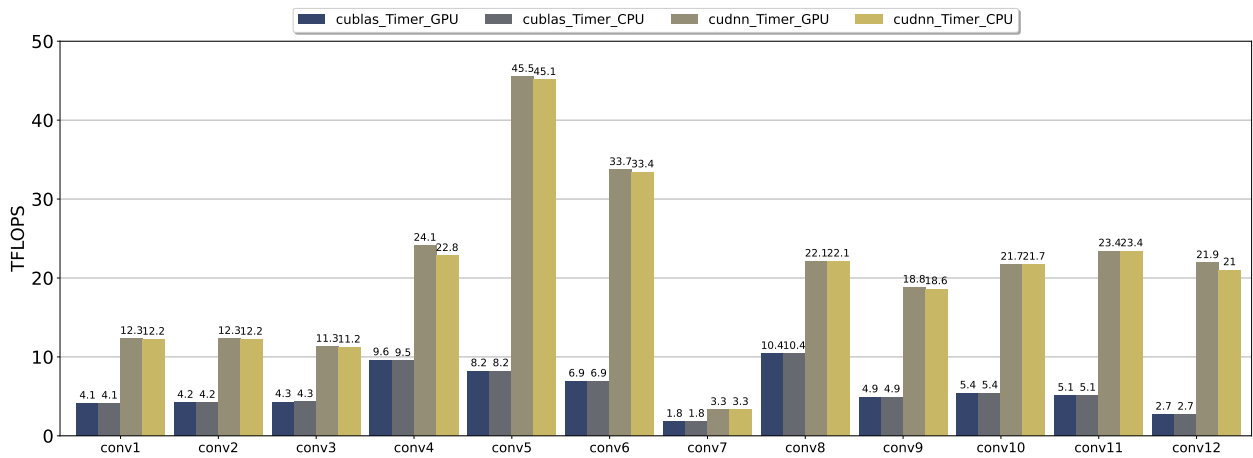


图 1: Timer CPU vs GPU

The unit obtained by GPU timer is ms, while the unit returned by CPU timer is s.

## 3 summary

From the above experiment, our validation function should have no problem. and the Timer have no problem too. so, Regarding the issue of excessive experimental data from last week, I guess the program has issues with running GPU code but does not report any errors, the time will too small to the Tflops too high.