note of experiment in week1

zxp

September 14, 2024

1 environment

cpu:Intel(R) Xeon(R) Gold 6330 CPU @ 2.00GHz (56 cores were applied)
gpu:rtx3090(a piece was applied)
System:CentOS7
Compiler:9.5

2 code

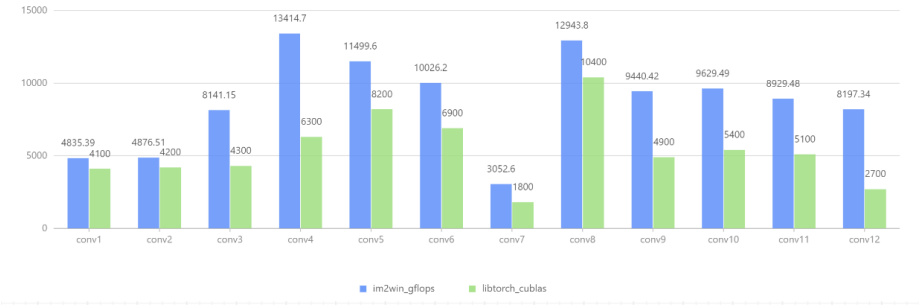

Figure 1: gflops

3 Experiment

Fixed the previous bug, not undercount, overcount, array overreach but CUDA did not report error. Repaired the loop ending logic. However, the conv7 was not able to run with batch set to 128, maybe the size of conv7 was too large, the batch of conv7 was set to 64, and the other was set to 128

3.1 Analysis

except conv1 and conv2 are slower. this speed is still in line with expectations, most are faster than the cubls in libtorch, and are close to the data in the paper.

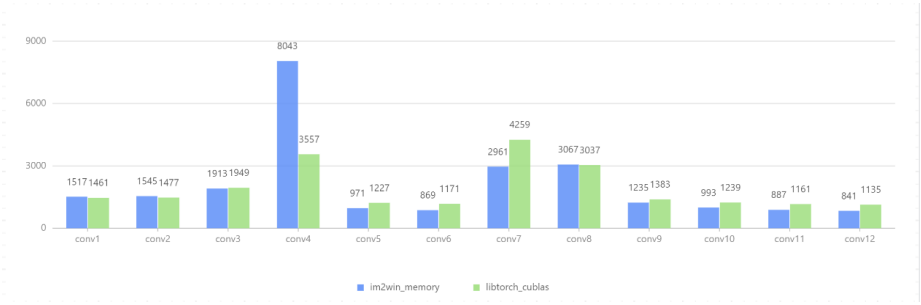

Figure 2: memory

4 Experiment2

This part is to test memory in GPU, open two terminals a run program a test memory used. Test confirmed that the conv can be run on experiment1, and also tested im2win and cublas in libtorch, and some of the im2win I wrote. But the situation is very strange and I can't understand it at all. I suspect there may be something wrong with the method of testing memory.

4.1 Analysis

Oddly enough, the speed of convolution is too fast, and millisecond output may not enough for testing memory