note of experiment in week25

zxp

August 24, 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

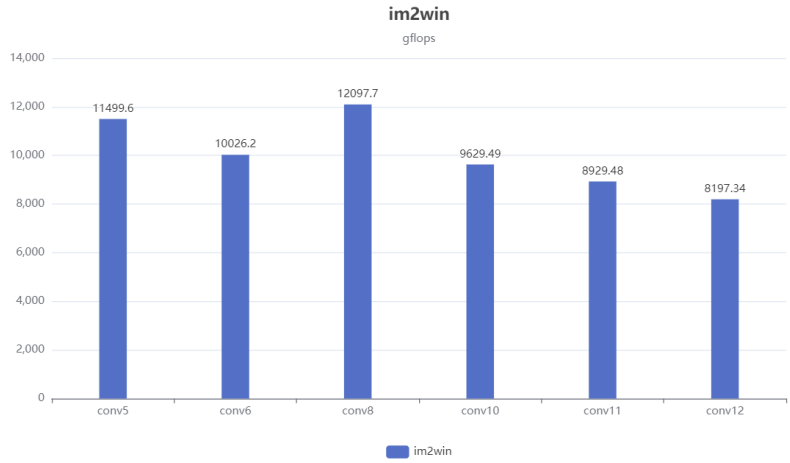

Figure 1: gflops

3 Experiment

Added hoist in the code, commented out the printf during previous debugging, and the speed has significantly improved. But attempts to fix the code for the layers where the latitude cannot be evenly divided have no effect. I can't understand it; I feel like it should be right, but it's not working.

3.1 Analysis

The effect is quite significant, 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.

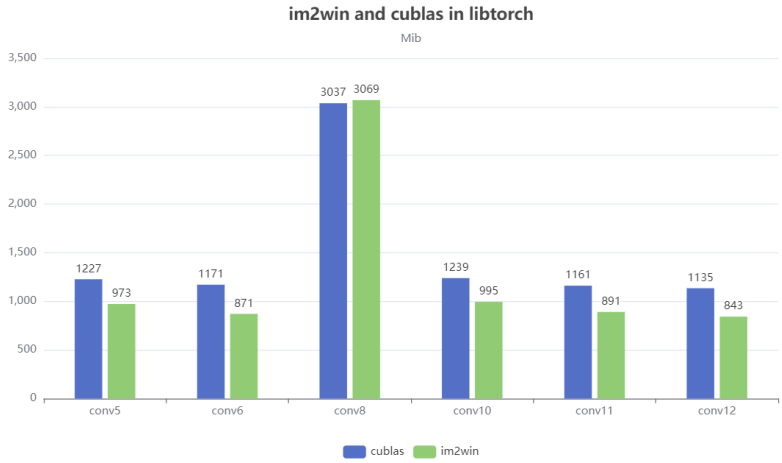

Figure 2: memory

4 Experiment2

This part is to test the memory. Open two terminals, run a program in one, and test the memory usage in the other. Test confirmed that the conv can be run on experiment1, and also tested im2win and cublas in libtorch.

4.1 Analysis

Except for conv8, everything else is as expected. Not sure what's going on with conv8. The connection was disconnected when testing conv8. Will retest this part next week.