note of experiment in week10

zxp

November 16, 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 Experiment

This week I tested the gflops and memory usage of im2win with batch equal to 256 and 512. The code is the same as last week. Please note that the latitude Settings are different from those of batch equal to 128. With batch equal to 256, conv7 should be set to split into 4 blocks. conv4 doesn't work for unknown reasons. When batch is 512, conv5 and conv8 have to be split into 4 blocks, and conv2 and 3 and 4 don't work for unknown reasons. See the attachment for the results.

2.1 Analysis

When batch is equal to 256, im2win performs well, up to 10% faster than 128, but the biggest problem is that some conv can not run, memory usage is in line with expectations, batch is doubled, memory usage is doubled (some are small, not doubled). Even conv4 with the largest memory usage only uses 6GB of memory with batch equal to 128, so it's not out of memory with high probability, but if you try to split the calculation into chunks, it doesn't work, and not knowing the reason for the error is tricky