

note of experiment in week15

zxp

December 21, 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

I've been trying im2col on GPU for a few weeks. now, there are libtorch im2col methods (which compute one batch at a time), and im2col, which emulates libtorch writing to compute all batches at once, but it doesn't work very well. Please see the attachment for details. There is also a problem with the server, and now it is suspected that CUDA and cublas cannot be used normally

2.1 Analysis

There are a few layers that work well, all at once so batches are a lot faster than batches at a time, but not all of them, some of the layers are just unreadable. I was scared to run the time of only matrix multiplication without im2col transformation, which was in line with expectations. It is faster to calculate a large matrix than to calculate a small matrix separately. I guess my computer performance is not good. Do not know the effect on the service, but this server really do not know how to do, can not find the reason ah, the code running in before can not run now, do not know the reason