

note of experiment in week9

zxp

November 9, 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 batch 128 of im2win's gflops and memory usage, using tensor1dV1 management memory and Lu's kernel. When running conv1-2, latitude can only be divisible by 32 when partitioned. conv7 cannot be calculated directly at once; it needs to be calculated in two parts. See the attachment for the results.

2.1 Analysis

Basically in line with expectations. The memory usage is small, and the initial input and filter tensor are not moved to the GPU