

# note of experiment in week10

zxp

November 23, 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



Figure 1: glops

## 2 Experiment

This week i fixed the im2col convolutions that compute a single batch at one time, now malloc and moving all batches of input tensor memory at once, and all batches of output tensor memory at once, but im2col matrices only malloc one batch size . not tested yet.

Then, on this basis, openmp is used to calculate multiple batches at one time on CPU (malloc the memory size of the im2col matrix of the number of thread batches), and the gflops of conv1-4 under different number of threads are tested

### 2.1 Analysis

glops with different number of threads does not differ much, which is not good. Maybe GPU parallelism still depends on GPU, CPU parallelism has little effect on it

## 3 Experiment

Fixed last week when conv4 for im2win with batch equal to 256 wouldn't work. because cudaMalloc couldn't malloc that much memory at once, but there was enough memory to use. in an ugly way, to malloc it twice and move the im2win tensor to the GPU in two separate malloc

### 3.1 Analysis

It's worth noting here that cudaMalloc can't malloc a particularly large amount of memory at once, even if it's enough