

week5实验记录

zxp

March 30, 2024

1 environment

cpu:Inter i5-12400f (2.5 GHz)

GPU:4060ti

System:Ubuntu 22.04.1

Compiler:gcc 12.3

2 code

尝试使用cublas和CUDA

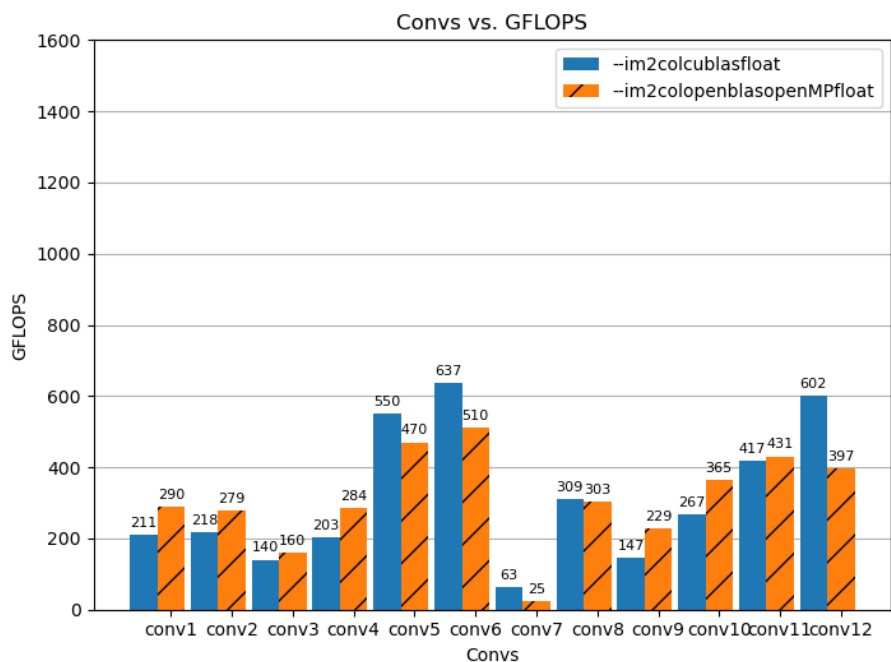


Figure 1: cublals

3 Experiment

尝试使用cublas来实现im2col卷积。

cublas是在GPU上进行运算的，使用cublas需要安装CUDA，但不需要额外安装cublas，安装好的CUDA里面有cublas，引入对应的头文件，编译的时候使用nvcc编译即可使用。cublas和其他的线性代数库一样有各种线性代数的方法，实现的也是level3的矩阵乘法。

但和其他线性代数库不同的是，需要先把数据从内存中搬运到GPU的显存中，CUDA中有配套的函数cudaMalloc来申请显存，和cudaMemcpy来搬运数据。还有一点和其他线性代数库不同的是cublas是列优先，需要注意是否需要将矩阵转置。

3.1 Analysis

在自己电脑上跑的，虽然显卡一般但在GPU上应该特别快才对，应该是数据从内存搬到显存然后把结果搬回来花了很多时间，然后im2col变换和将矩阵乘法的结果转换成数据张量在cpu上完成的，只有矩阵乘法是在GPU上运算的，所以才是这样的结果。

即使这样也可以看出GPU运算矩阵乘法很快

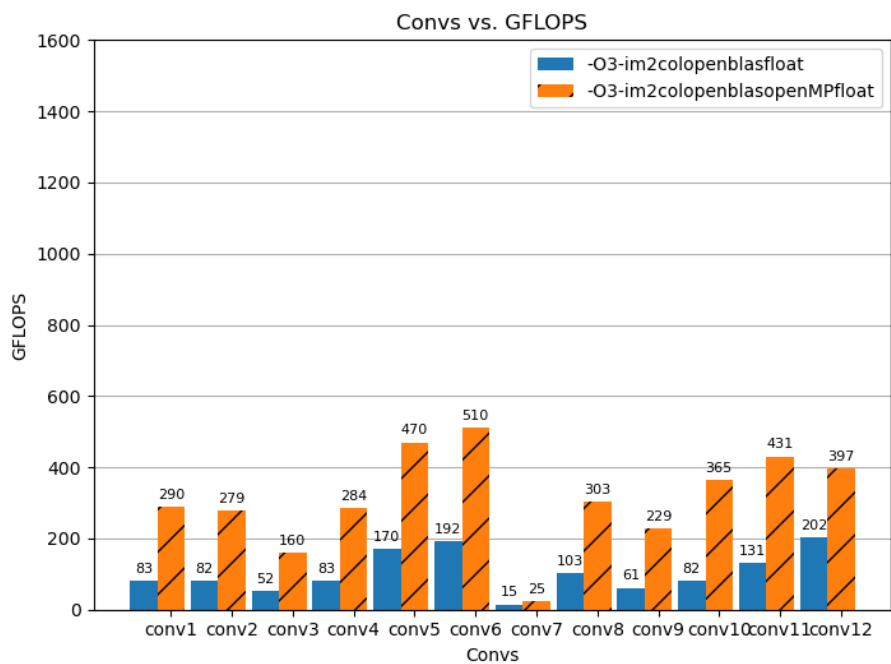


Figure 2: cubals

4 Experiment2

并行版本的openblas之前特别慢是因为openmp和pthread冲突了。

4.1 Analysis

去掉后openblas虽然快了，但im2col变换和将矩阵运算的结果转换成输出张量的并行都是靠openMP实现的，不使用openmp的话，这个版本的im2col卷积肯定不如其他版本快。