week-2

JX-Ma

2024/8/3

1 Introduce

This weeks'work is as follow:

- I reproduced the environment of Lu, next, I test the libtorch of cublas and cudnn version.
- I attempt different blocking size in im2win.

2 experiment

2.1 experiment1

environment1

- **operational system: centOS7.**
- **CUDA-Version: 11.3**
- **CUDNN-Version: 8.2.1**
- **GCC-Version: 9.5**
- **Openmp-Version: 4.5**
- **pytorch-Version: 2.2.0**

environment2

- **operational system: centOS7.**
- **CUDA-Version: 11.1**
- **CUDNN-Version: 8.0.1**
- **GCC-Version: 9.5**
- **Openmp-Version: 4.5**
- **pytorch-Version: 1.10.0**

The following figure shows the tflops of libtorch for Cublas and Cudnn in Environment 1 and Environment 2: the cublas_lu is environment2, the libtorch_cublas is environment1.

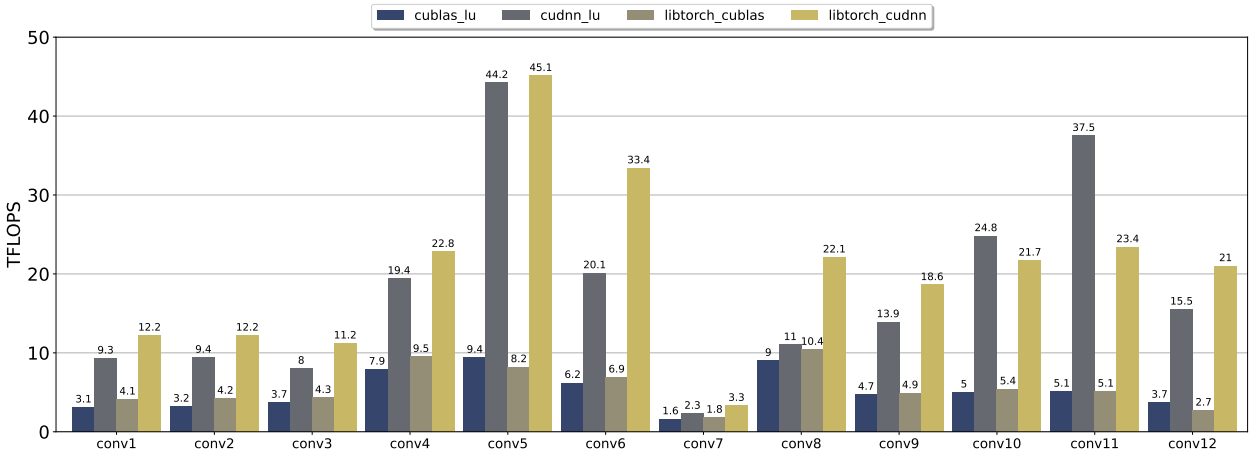

图 1: Tflops of different environment

2.2 experiment2

the experiment adopt the environment1. I attempt different blocking strategy for block on GPU.

The following figure show the blocking is 128x128,the legend is third parameter size.

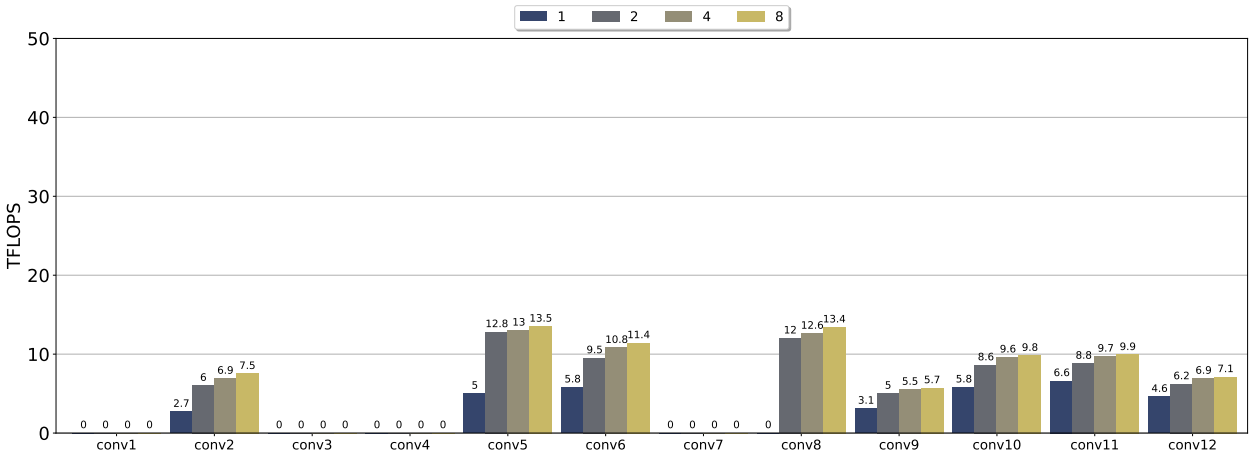

图 2: the first two size is 128x128

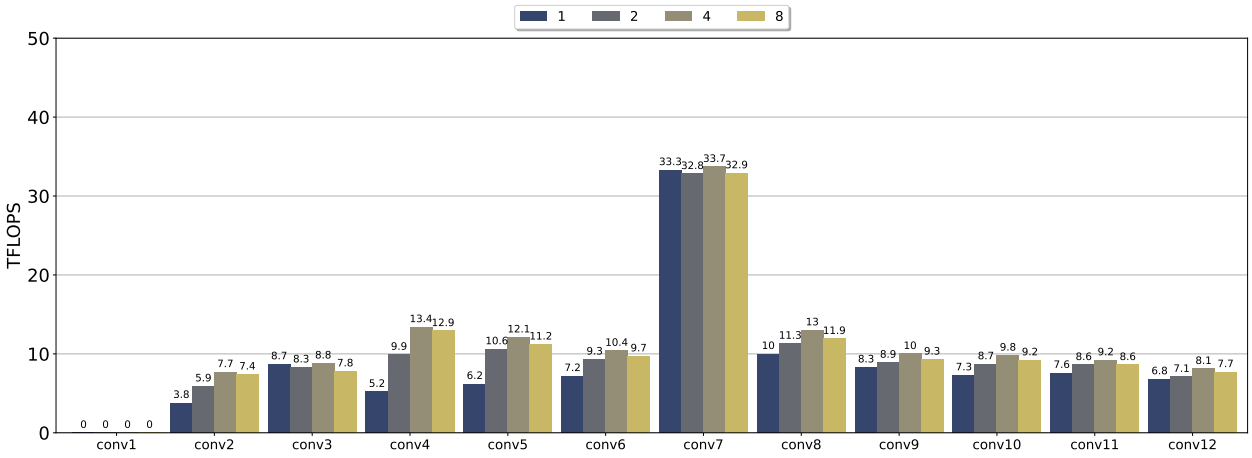

图 3: the first two size is 64x64

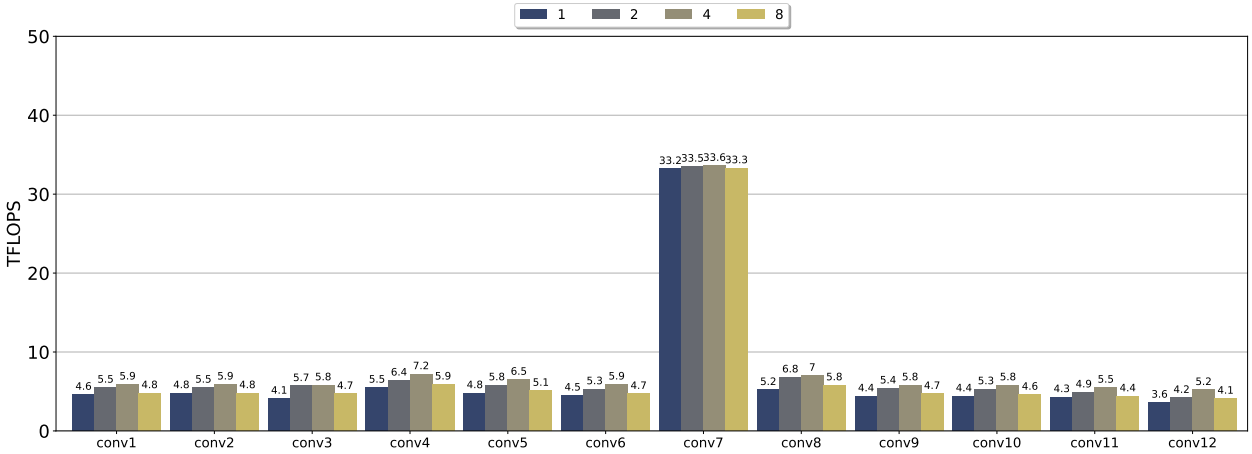

图 4: the first two size is 32x32

3 summary

In the experiment2, we can observe that the conv7 is very high, and in blocking size is 16x16, we can see the Tflops more than 800 in conv4, this is impossible obviously. so i regard there may have two reasons for this. firstly, the error occured in kernel function, the error has been occured but the program not stop, the time record still runing, so the comsumption of time is too short result in the Tflops is too high. the next reason is that timer not suitable for GPU, we can rechoose a timer for GPU.

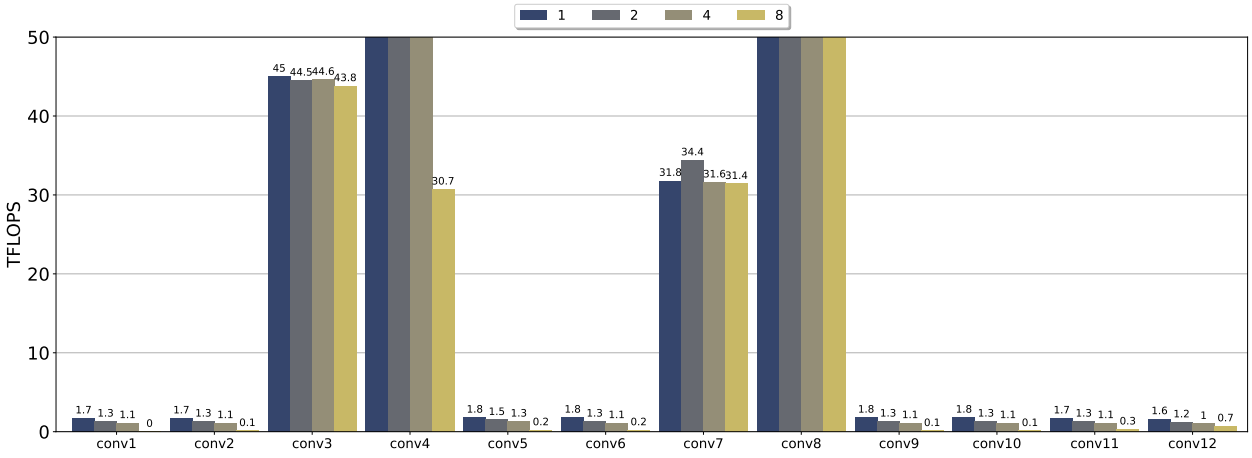

图 5: the first two size is 16x16