week-5

JX-Ma

2024/8/24

1 intorduce

This week, the path for CUDNN to call the convolution algorithm was forcibly changed to follow the logic in conv7.cpp.

2 experiment

This week, I was checking on the server what algorithm was called by CUDNN. I output static_cast<int>(fwdAlgPerf. algo) before and after the CUDNnConvolutionFwdAlgPerf_t function. I found that the output in conv1-conv12 is 1, and the corresponding algorithm is FFT. Since I previously consulted the documentation and found out that CUDNN adaptively searches for algorithms during the first convolution, I could see that CUDNN had very low Gflops in the first experiment, only single digits, so I changed the number of runs to 10 later, but the output result was still 1. Afterwards, I forcibly modified the parameters of fwdAlgPerf.algo, but running it this way would result in an error in the first few layers, with the error message terminating called after throwing an instance of 'c10:: Error' what(): Unable to find a valid cuDNN algorithm to run convolution，I only ran the program once here, and theoretically it should output the value of the algorithm twice. However, the first few layers output multiple times before reporting an error, and the output of the later layers is inconsistent. For detailed files, please refer to debug.txt