week-7

JX-Ma

2024/9/7

1 introduce

This week, i separated the implementation of im2win data conversion and convolution calculation, and measured the gflops of each benchmark after adding data conversion.

2 experiment

According to the experimental results, it was found that the time of im2win transformation accounted for 1/4 of the total time in data layouts NCHW, NHWC, and CHWN8, which led to a significant decrease in gflops. In the im2win transform, I made optimizations such as openmp and index hashing, but other optimizations are difficult to implement. In the im2win transform, the data we read is not continuous. When the original layout is transformed to im2win, the data we read is elements at a distance of filter. height, so it is not easy to use simd