

week-1

JX-Ma

2024/3/1

1 本周工作

对比了一下直接卷积和 im2win 的性能差异，分别计算了它们在输出张量宽分块大小相同时时候的算术强度。

2 算术强度分析

假设卷积核高度 x 宽度为 4x4，步长大小为 1，输出张量的宽进行 1x2 分块，如下图，从图中可以看到，当我们按照 1x2 分块后两者算术强度应该是一样的，但是相比较直接卷积，im2win 的数据连续性更好。

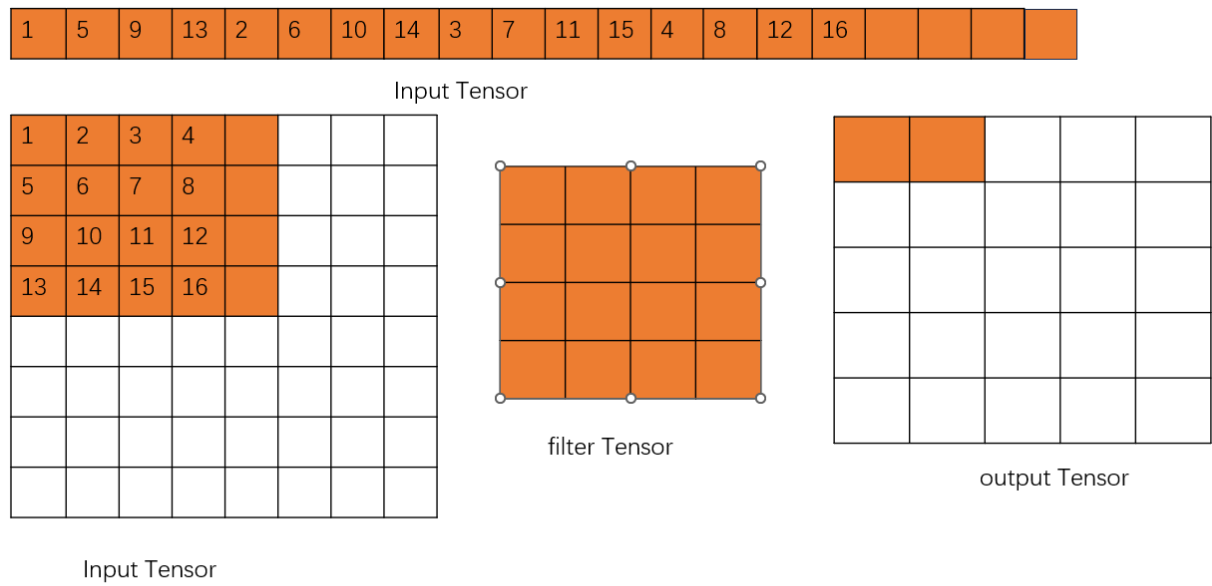


图 1: directConvolutuion

3 实验部分

3.1 实验环境

- 系统: Ubuntu 22.01
- gcc version : 9.5.0
- 优化选项: -O3 -fopenmp -avx2 -fmadd
- cpu:AMD Ryzen 7 6800H 3.20GHz

3.2 实验结果

这周实验把直接卷积上优化结果和 im2win 优化结果进行对比

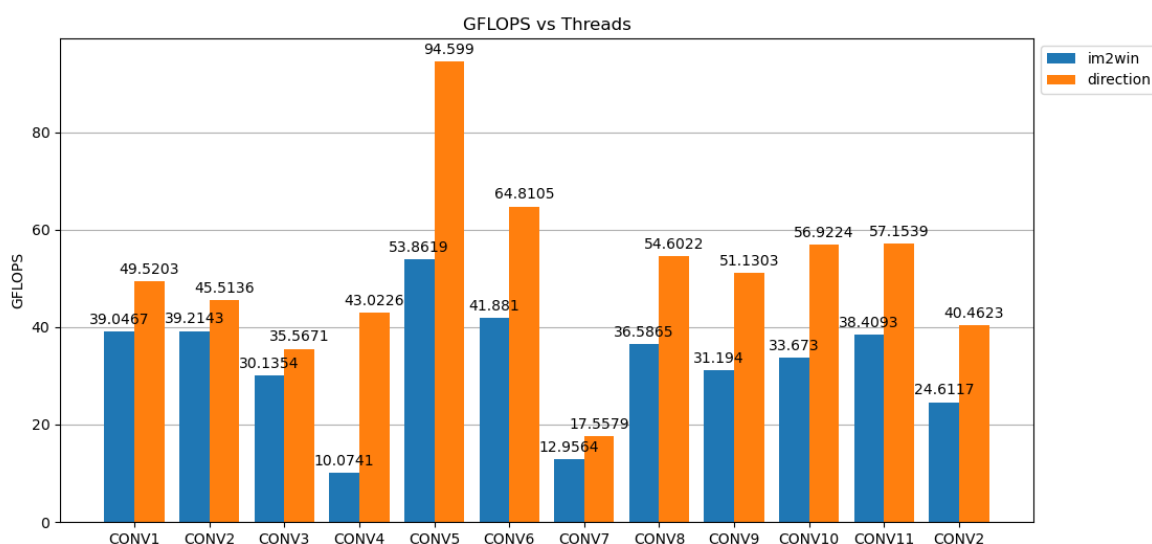


图 2: gflops

3.3 总结

im2win 的数据连续性比直接卷积的好，但是性能却达不到想要的结果，猜测有可能的原因是因为 im2win 的步长变大了，例如 CONV1 的步长为 4，我们使用 im2win 后的步长就会变为 44，当继续在 im2win 把通道展开后步长就会变为 132 此时虽然数据连续性好，但是相邻的输入张量窗口之间的距离变长了。