week-16

JX-Ma

2024/11/16

1 INTORDUCE

This week's work is as follows: We integrated the data and drew memory maps for batches 128, 256, and 512 im2col bubbles, as well as a total memory map for batch 128. Afterwards, we also drew tflops maps for batches 128, 256, and 512 respectively. Read the cublas document, learned several methods of matrix multiplication in cublas, debugged the logic of calling cublas in libtorch, and found the path of pytorch calling cublas. Added references to relevant literature in the paper and included several works related to Winograd and FFT
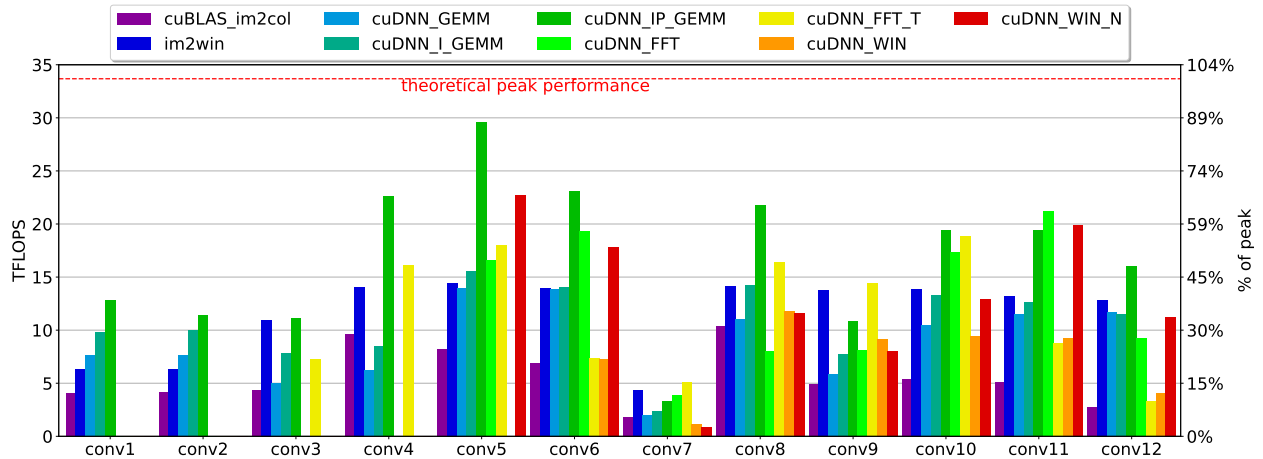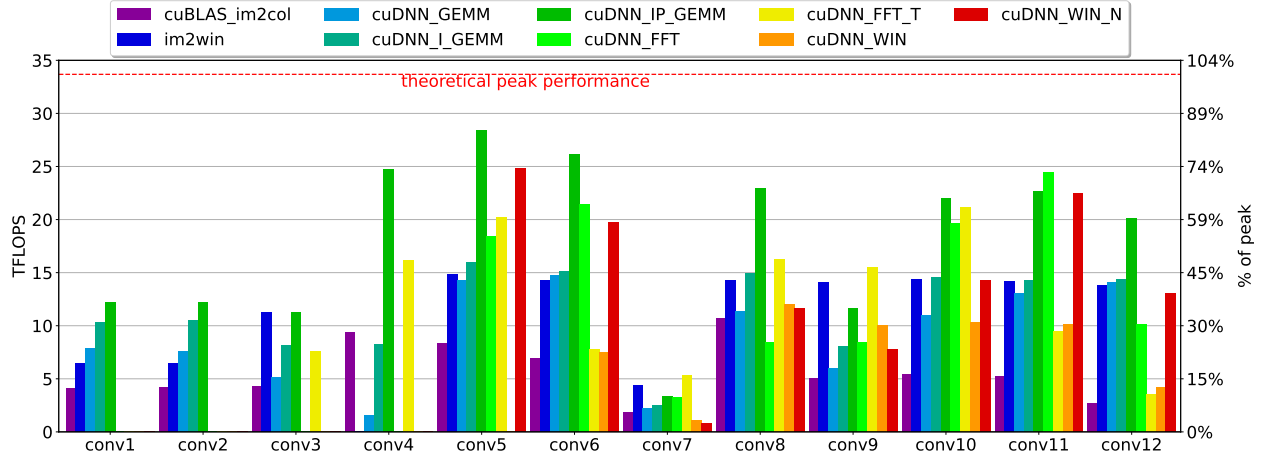
2 experiment

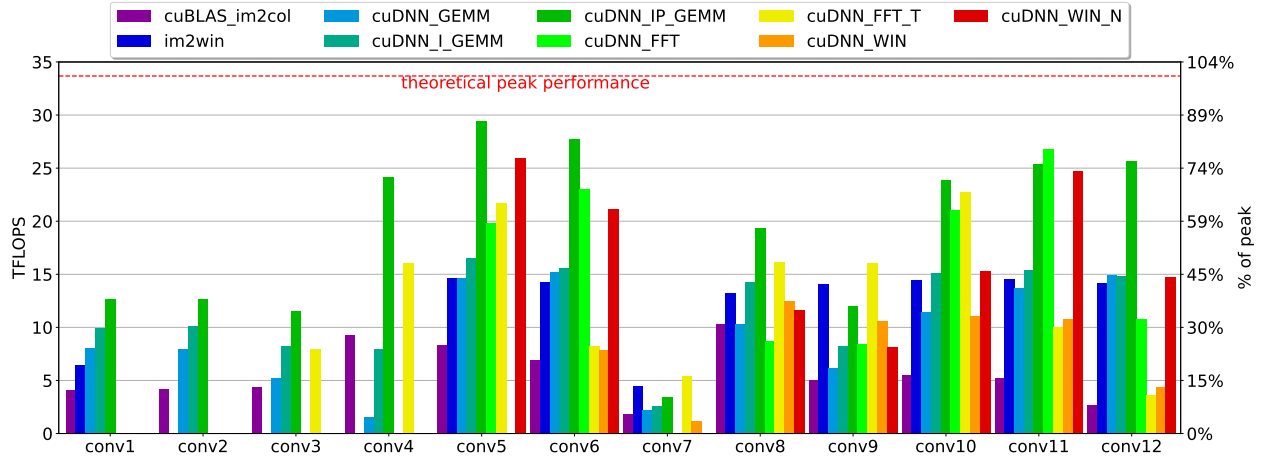
图 1: TFLops of 128 batch

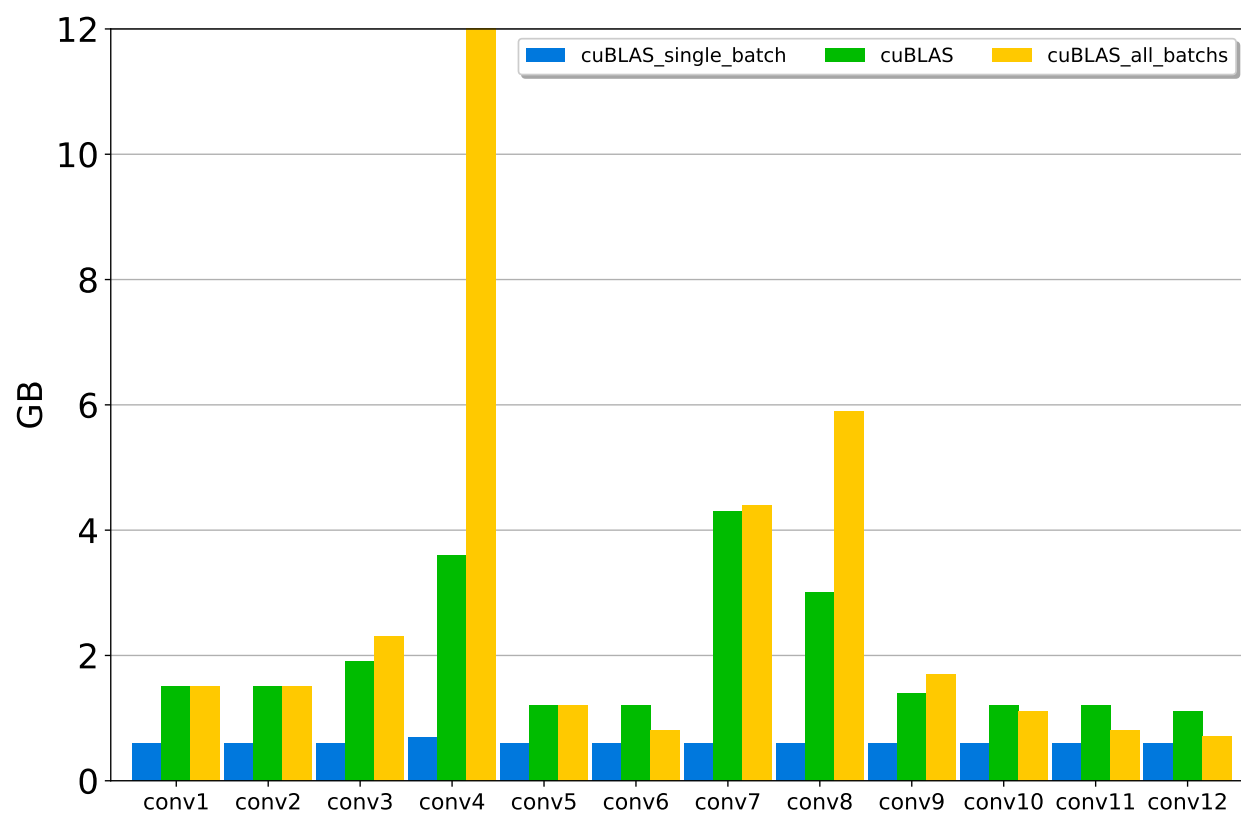图 2: TFLops of 256 batch

图 3: TFLops of 512 batch

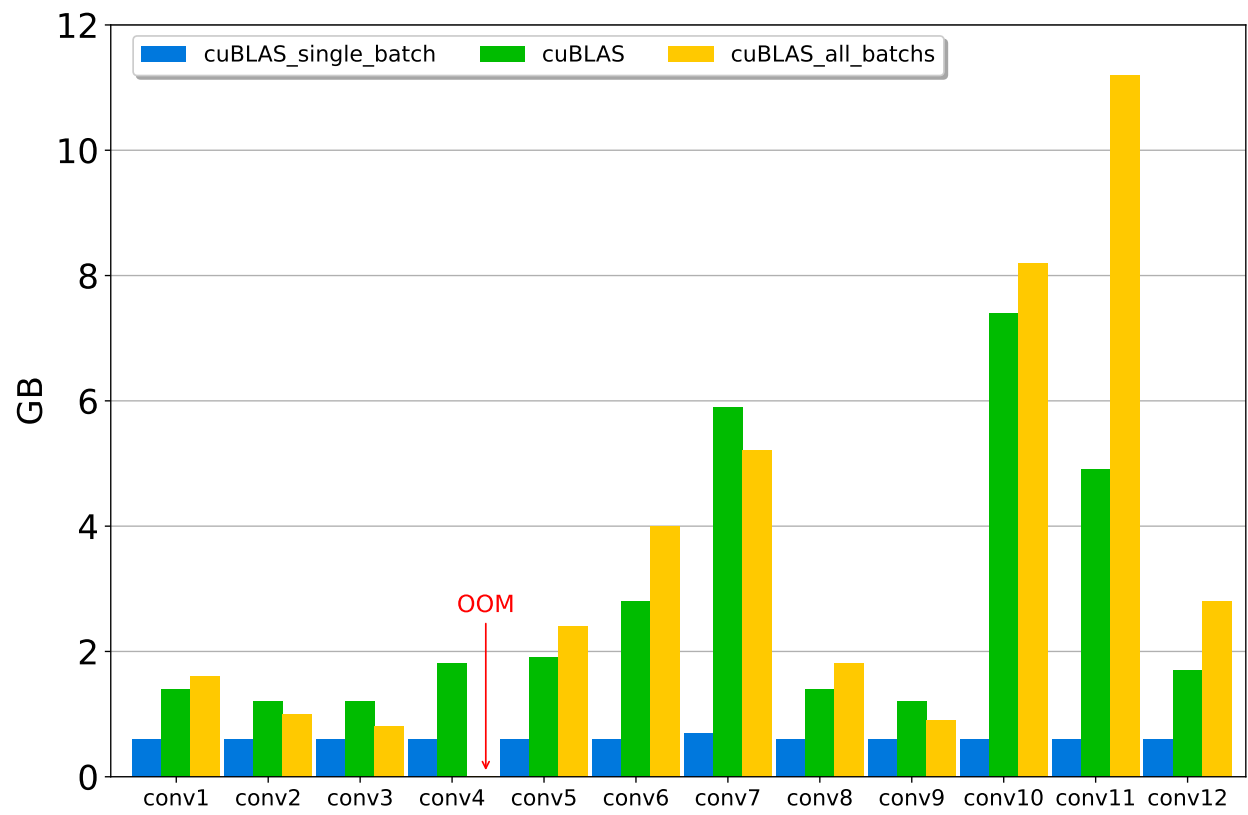图 4: memory of 128 batch only im2col
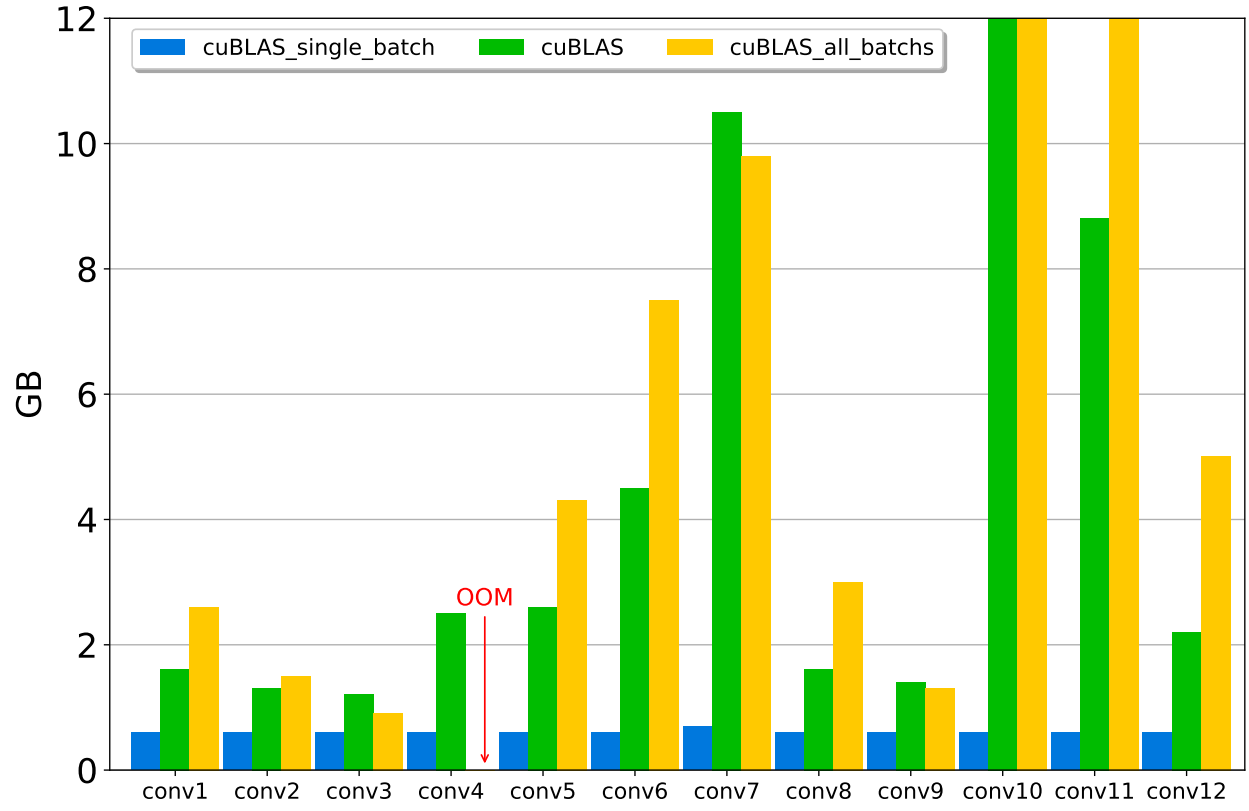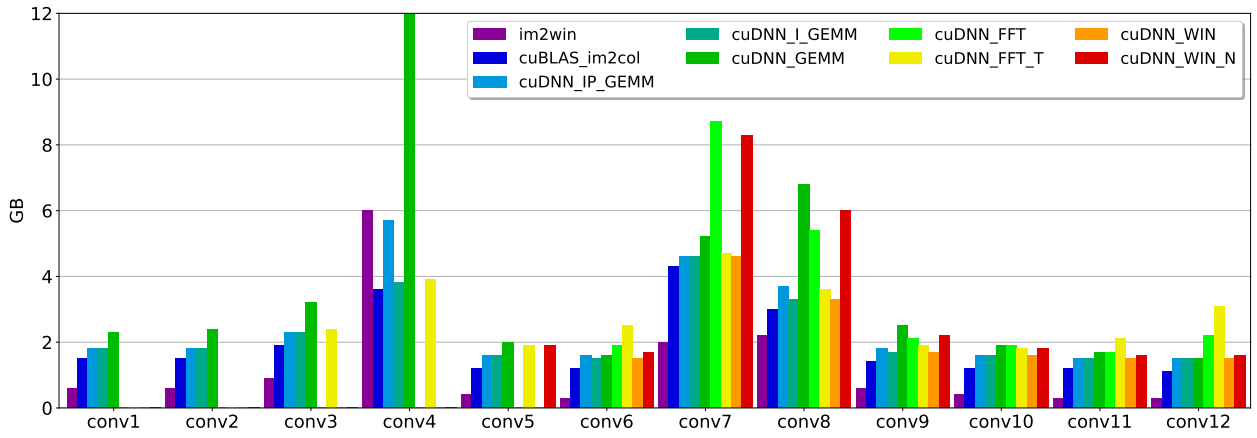
图 5: memory of 256 batch only im2col

图 6: memory of 512 batch only im2col

图 7: all memory of 128 batch