**Homework 3 – Performance study of (a small part of ) Neural Network**

1. **Pen and paper method to estimate the complexity**

conv2: We have and , so the complexity is

And we have default batch size of , so the complexity of forward propagation and backward propagation is

Since we have epochs in total, the total complexity is

1. **Run and measure the complexity**

All the data are measured on the Solurm Partition: intereactive with 2 GPUs

* 1. **Time**
     1. Time Measurement - Linux:  
        real 3m31.133s

user 3m54.816s

sys 0m15.881s

* + 1. Time measurement - Python  
       time: 188.59780880300002
  1. **FLOPs**

Total FLOPs: 23984896.0

Total Parameters: 1199882.0

* 1. **Memory**
     1. Model Summary

----------------------------------------------------------------

Layer (type) Output Shape Param #

================================================================

Conv2d-1 [-1, 32, 26, 26] 320

Conv2d-2 [-1, 64, 24, 24] 18,496

Dropout-3 [-1, 64, 12, 12] 0

Linear-4 [-1, 128] 1,179,776

Dropout-5 [-1, 128] 0

Linear-6 [-1, 10] 1,290

================================================================

Total params: 1,199,882

Trainable params: 1,199,882

Non-trainable params: 0

----------------------------------------------------------------

Input size (MB): 0.00

Forward/backward pass size (MB): 0.52

Params size (MB): 4.58

Estimated Total Size (MB): 5.10

* + 1. Nvidia GPU memory utilization measurement

Allocated: 52331346 Bytes

Reserved: 0 Bytes

1. **Try diff batch sizes, draw a chart (x axis: batch size, y axis: flops and/or mem)**
2. **Try diff batch sizes, draw a chart (x axis: batch size, y axis: flops and/or mem)**
   1. Computational Complexity

The result of theoretical complexity is significantly lower than the measured FLOPs and execution time. This may be because pen and paper method provides an upper bound, often underestimated due to ignoring constants

* 1. Memory Usage

Measured memory usage is higher, especially with smaller batch sizes. This may because in our code, the memory are not only dedicated to be used in the cov2d-2 layer. Other code in our program may also use the memory during the process of our code execution.