**Performance Modelling for Convolution Neural Networks**

* **Convolution**

Q1) How many Flops needs to be done to compute a convolution of dimension k on a image of size

n \*m?

* Our Image size is n\*m and our kernel size which would be used for convolution is k\*k
* In convolution, for a single pixel, we will do 2 times k\*k operations ie with our kernel we will do multiplication first and then addition.
* So for an image with a size of m\*n, we will do a total of 3 \*(m\*n\*k\*k\*2) flops.

Q2) How much memory needs to be moved to compute a convolution of dimension k on an image of

size n\*m?

* We will bring the image of size m\*n from memory also the kernel of size k\*k.
* So we will need to fetch 3\* (((m\*n)+ (k\*k))\*4) bytes from memory.
* Also, we will write back (m\*n)\*12 bytes to memory.

Q3) Calculate Time required for Pixel Processing

* Considering m\*n= 1024 \* 768 and k\*k=3\*3
* Flops done would be = 3 \*(1024 \*768\*9\*2)= 42,467,328 flops
* The number of flops done per second are 1.6 \*10^12 flops
* So flop time is – 26.54 usec
* The amount of bytes that would be required to shift in would be 18874476.
* Consider the Bandwidth to be 100 Gb/sec
* So the time required for mem transfer would be 188.7 usecs.

Q4)Performance Model

* So, performance modeling can be done by
* (m\*n/ floptime)= 88 Gpixelspersecond
* (m\*n/ memtime)= 12 Gpixelspersecond

2