CNN tidbits

# Convolutions in a picture

Below diagram shows the input matrix in blue color of size 5x5

The multiplication of the subset of blue matrix and green kernel to produce one cell of the output is convolution.

A picture containing diagram

Description automatically generated

## Padding

The dotted new cells that are seen are a result of padding of size 1

## Kernel Size

The smaller green matrix is the kernel of size 3x3 or kernel size =3

## Stride

Notice how the green matrix slides by striding 2 cells between each instance.

That is nothing but stride. In this case stride =2

## General formula for determining the size of the output matrix after a convolution

Graphical user interface, application

Description automatically generated

* n is the size of the input matrix and in this case its 5
* pad is size of the padding and in this case its 1
* ks is the size of kernel and here it is 3
* stride is 2

*5+2\*1-3 = 4*

*4/2 = 2*

*2+1 = 3 🡪 The output matrix will be 3 x 3*

Convolution Arithmetic

* Suppose we are passing black and white images (Only one channel. No RGB channels like color images) of size 28 x 28 pixels as input to a CNN.
* Let’s say the batch size is 64
* The input shape will be 64 x 1 x 28 x 28
* This is in the format of NCHW. Pytorch/FastAI use this format. (Tensorflow uses NHWC btw)
  + N Batch Size
  + C Channels
  + H Height
  + W Width
* As we go deep into the layers of CNN the we decrease H and W decrease, but we also increase C
* C is Channels or the features that the CNN is finding out like Eye, Fur etc..
* Below is summary of a CNN model. We can see how the grid size keeps reducing to half and the number of features(channels) keeps doubling.
* In the final two layers, we want just a binary output of True/False or whether it is a 3 or 7 in this case. Note: This was using MNIST grayscale images and hence input channel was 1.

**Table

Description automatically generated**

CNNs for Color Images

* The Kernel Size will be channel\_in x n x n
* ch\_in will be the number of input channels and generally its 3 for RGB or HSV (Hue, Saturation, Variance)
* Everything else we saw before is relevant for Color images too.