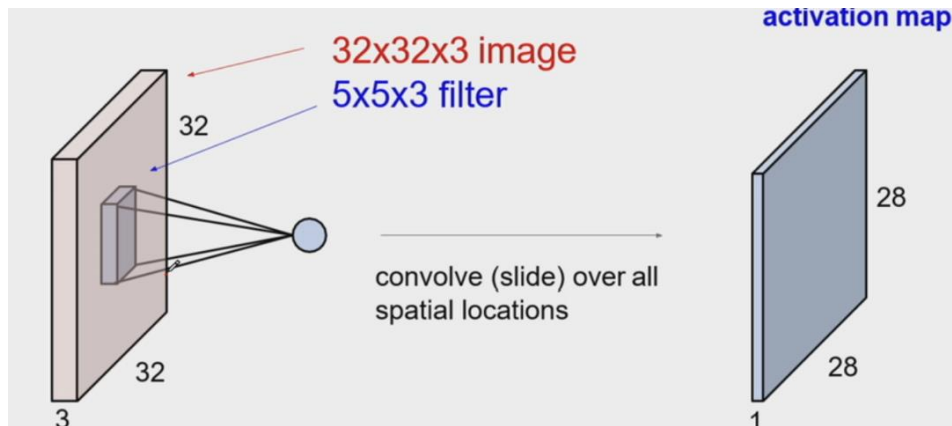# ENPM703- Assignment-2

Part3: Convolution NN

Convolution NN Architecture

Ref: Zaki, George. module06-convolution-layers. University of Maryland.

A Convolutional Neural Network (CNN) is a deep learning model that use convolutional layers to automatically detect and learn spatial feature hierarchies from an input image.

Convolution Layer: The convolution layer uses filters (kernels) that move over the input image, computing dot products between the filter and local slices of the image. This operation captures important spatial features like edges, textures, and shapes.

Activation Function: Following convolution, an activation function (mostly ReLU) is applied to introduce non-linearity.

The implementation of convolution layer is not efficient since the computation is done by 4 nested for loops, but is compared with an optimized implementation which is approximately 500x faster.

Maxpooling

Max Pooling is a form of down-sampling used in CNNs to reduce the spatial dimensions of feature maps. This operation is applied independently across each channel in the feature map. Max pooling divides the input into non-overlapping regions (2x2) and selects the maximum value from each region. For example, a 2x2 pooling operation would reduce a 4x4 input to a 2x2 output.

Max pooling reduces the number of parameters and computational load, while preserving the most essential features. By keeping only the maximum value, max pooling increases the network's robustness to small translations or distortions in the input.

Spatial Normalization

Spatial Normalization (Batch Normalization for spatial data) normalizes activations across channels within a mini-batch, enhancing both training stability and speed in CNNs. For each channel, the mean and variance are calculated across the spatial dimensions and normalize the data. Here the vanilla version of batch normalization function is used by reshaping the inputs accordingly.

Why CNN is better?

CNN are great for image processing tasks due to their ability to capture spatial feature using convolutional and pooling layers.

- Through pooling layers, CNN gain robustness to shifts in feature positions allowing them to recognize elements like edges or objects regardless of their location within an image. This property is essential for consistent and reliable image processing.
- CNN build feature representations at varying abstraction levels within their layered framework. Initial layers detect basic features such as edges and colors, while deeper layers identify more complex structures like shapes, textures, and complete objects.
- CNN are particularly suited for processing high dimensional inputs, like large images by analyzing spatial regions instead of treating each pixel individually. This design enables CNN to handle images more efficiently, as they don't need to connect every pixel to each node in subsequent layers, unlike traditional neural networks, thus making them highly scalable.