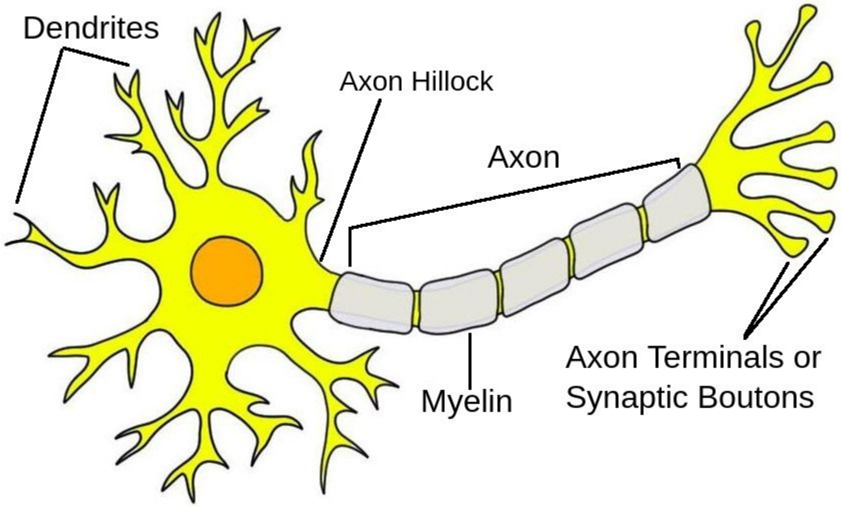
CSE4261: Neural Network and Deep Learning

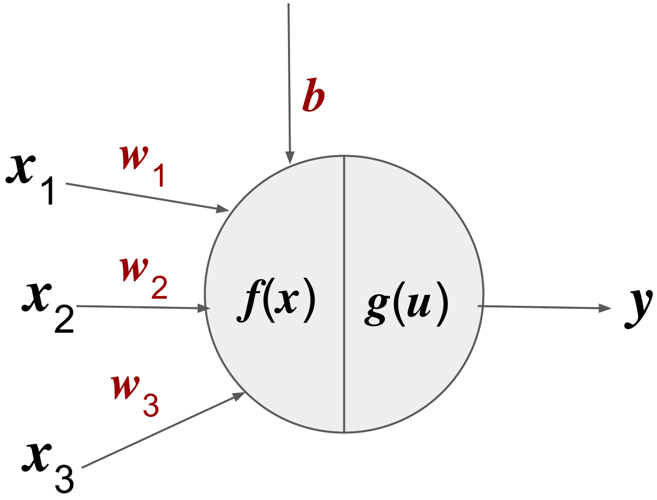
Lecture: 21.05.2025

Sangeeta Biswas, Ph.D.
Associate Professor,
University of Rajshahi, Rajshahi-6205, Bangladesh

Biological Neuron Vs. Artificial Neuron



Biological Neuron



Artificial Neuron

Steps of Traditional Machine Learning

1. Data Preparation:
   a. Data Collection
   b. Data Splitting
2. Data pre-processing
   a. Data Cleaning
   b. Data Transformation
3. Feature Extraction:
   a. Identify Relevant Features
   b. Extract and Transform Features
4. Training Classifier:
   a. Model selection
   b. Searching for optimum parameters
5. Evaluation
   a. Estimate metrics using test set

End-to-End Learning

End-to-end learning or end-to-end training, is a machine learning model that:

- learns directly from raw input data to produce a desired output, without the need for manual feature engineering or pre-processing steps.
- gets popularity in deep learning.
- bypasses few steps of traditional learning such as data pre-processing, feature extraction.

Kernel

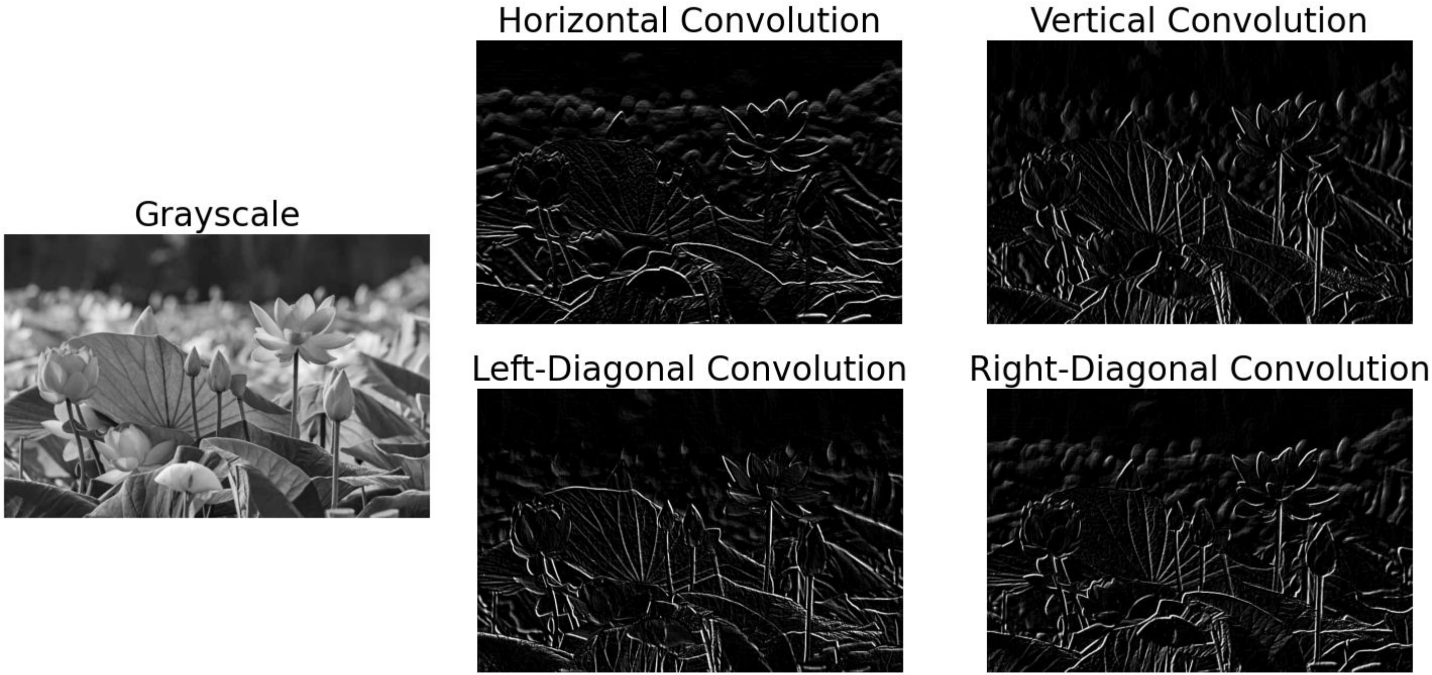
- In computer vision, kernel is a (usually) small matrix of numbers that is used in image convolutions.
- Different sized kernels containing different patterns of numbers produce different results under convolution.
- Generally, small odd numbers such as 3, 5, or 7 are used as the size of a kernel.
- Popular human-decided kernels:
    - Gaussian Blur Kernel
    - Laplacian Kernel
    - Prewitt Kernel
    - Sobel Kernel

Sobel Kernel

$$K_{\text{Vertical}} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad K_{\text{Horizontal}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

$$K_{\text{Left Diagonal}} = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \quad K_{\text{Right Diagonal}} = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Effect of Well-Known Sobel Kernel



Feature Map

A feature map:

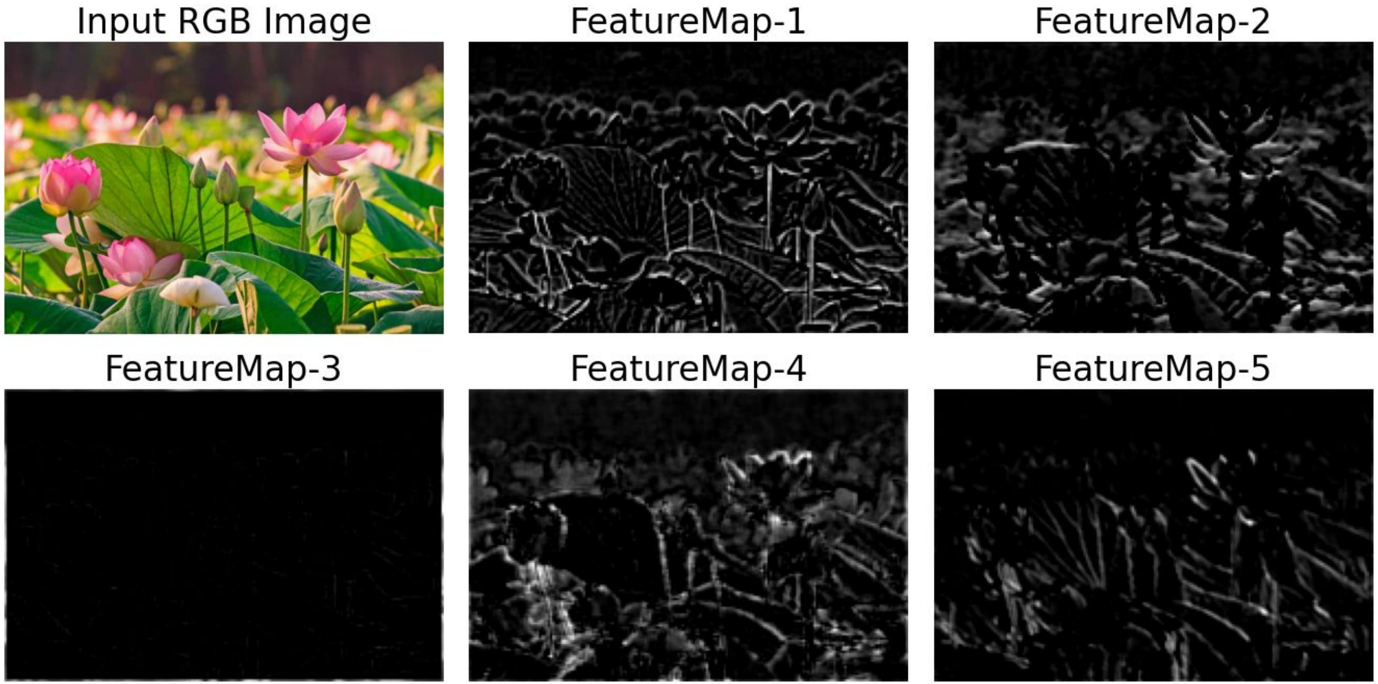
- is a 2D matrix in a hidden layer
- represents the output of a convolutional layer
- captures specific features or patterns in the input data.

In a CNN:

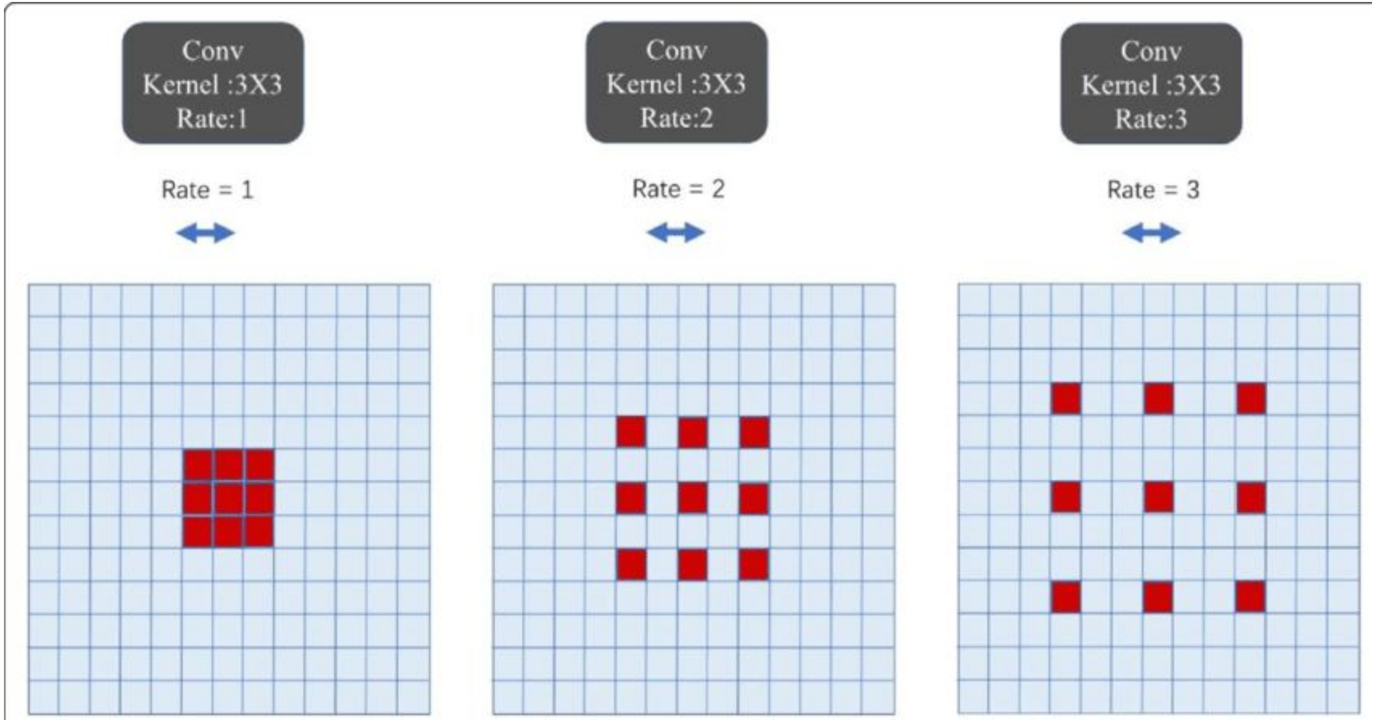
- Instead of human-decided kernels/filters, different automatically learned kernels/filters are used to generate different feature maps

Feature Maps of Pre-trained VGG16

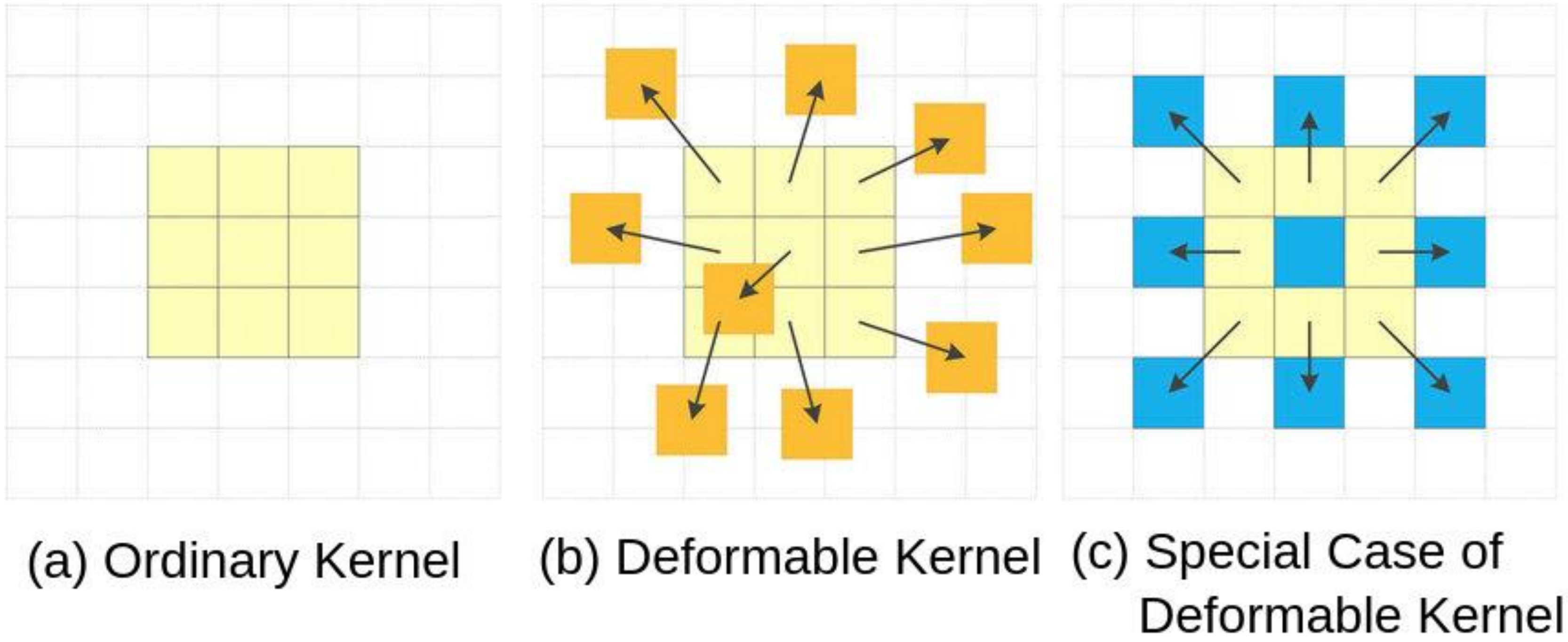
- Layer: block3_conv1



Atrous or Dilated Kernel



Deformable Kernel



Point-Wise convolution

1 x 1 kernel is used for
point-wise convolution.

Advantage: Increase or reduce the depth (or number of channels) of an image.

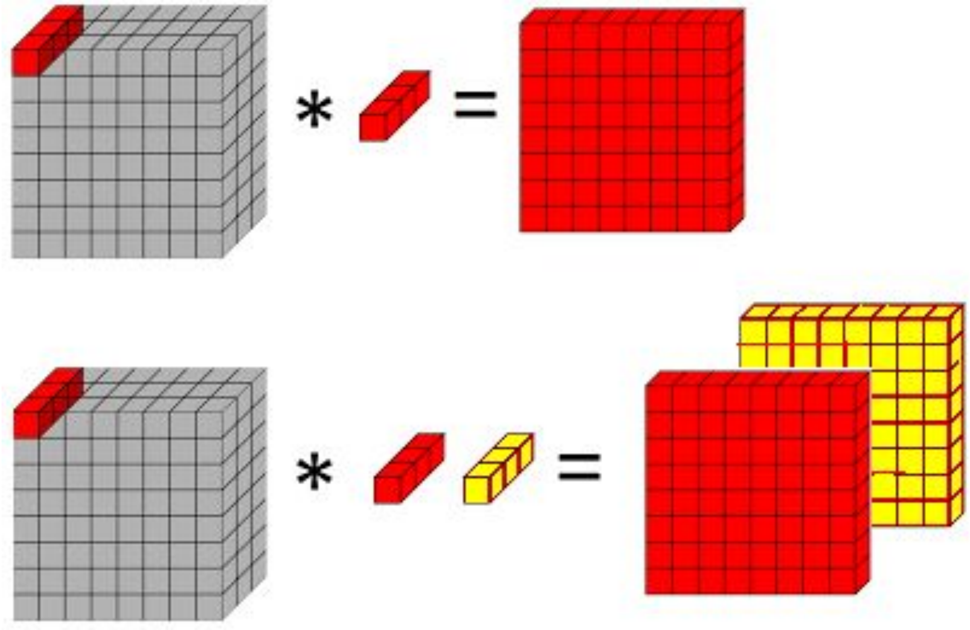


Image Source: Google Search Engine (Modified)

Spatial Separable Convolution

A spatial separable convolution simply divides a kernel into two, smaller kernels.

Advantage: Less parameters, less memory and less computations than regular convolution

Limitations: Not all kernels can be spatially “separated”

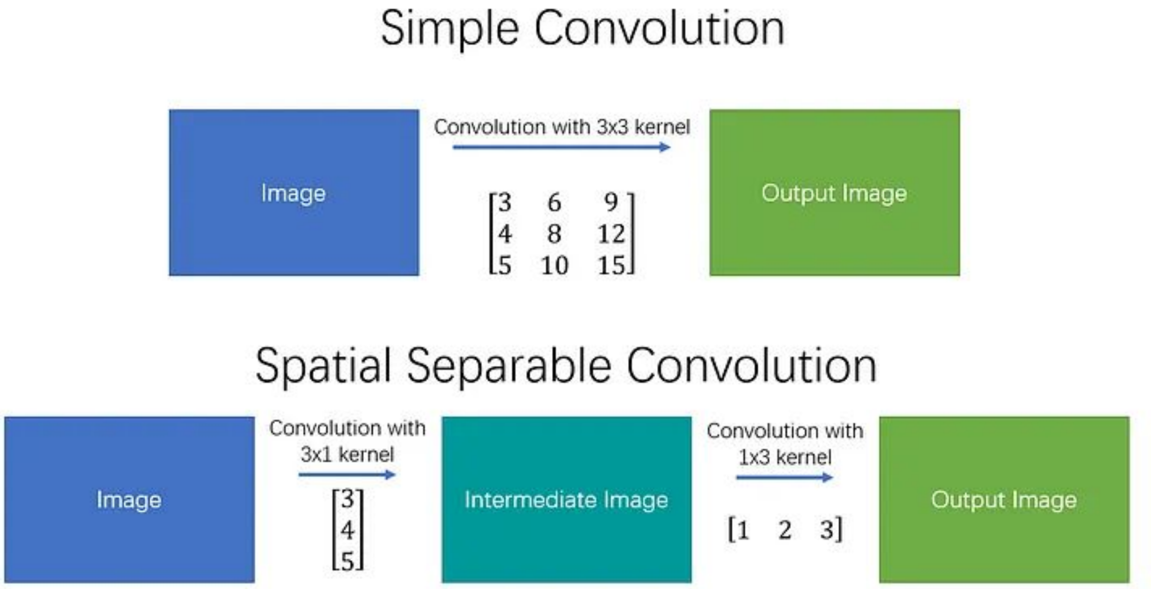


Image Source: Google Search Engine

Depthwise Convolution

Each input channel is convolved with a different kernel (called a depthwise kernel).

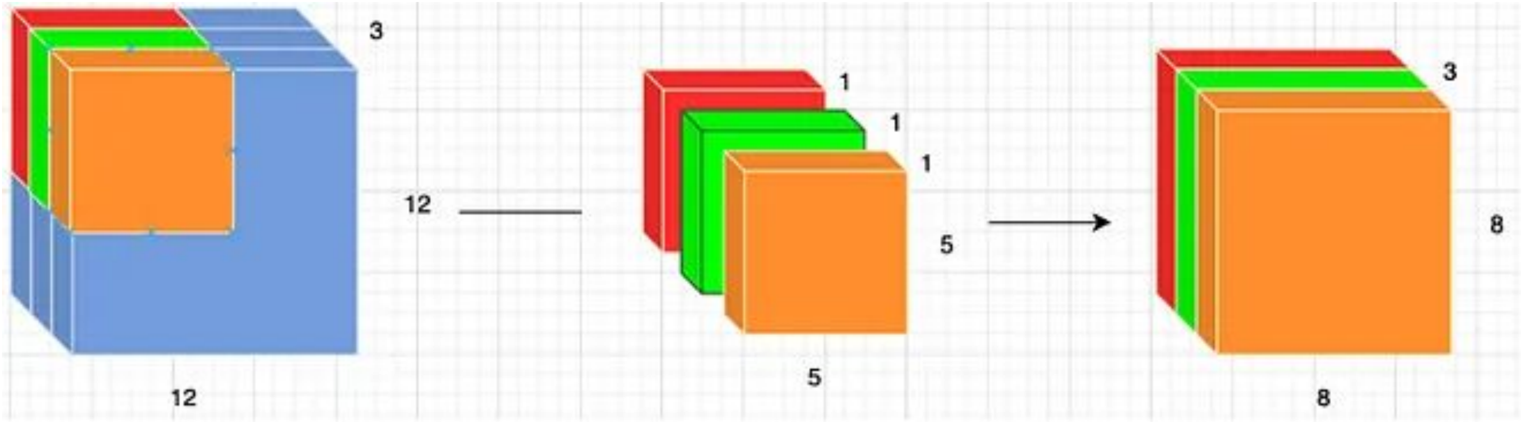


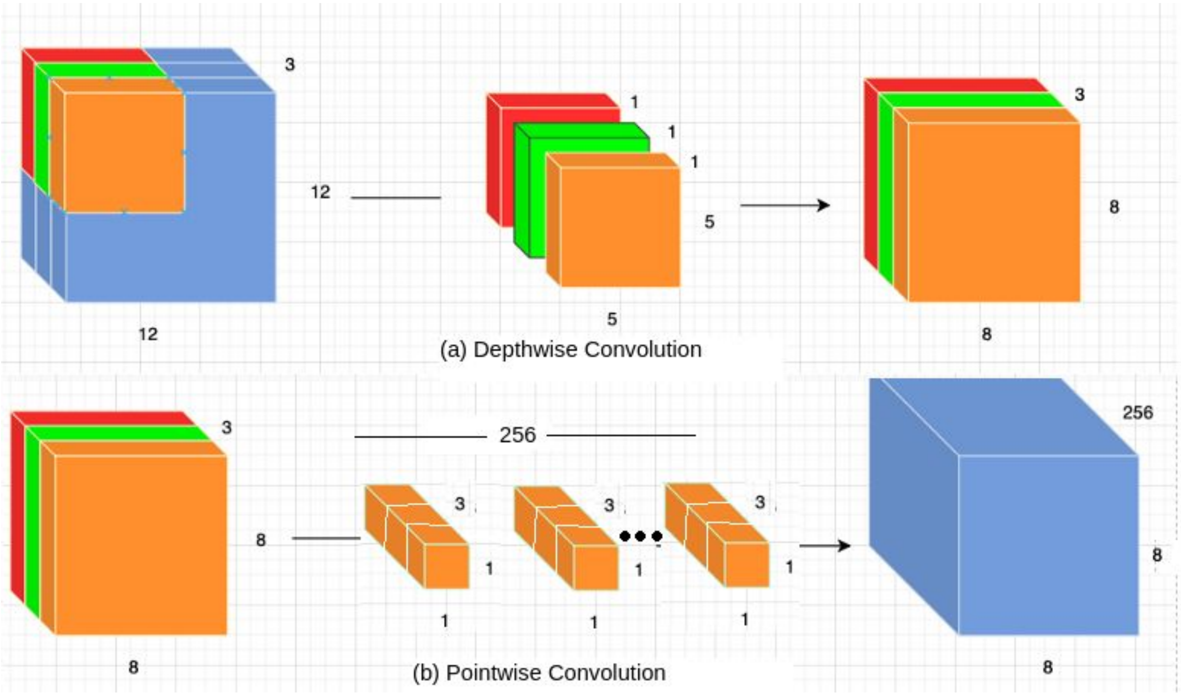
Image Source: Google Search Engine

Depthwise Separable Convolution

A Depthwise Convolution follows Pointwise Convolution.

For standard Convolution, number of multiplications: 8x8x5x5x3x256 = 1228800

For Depthwise Separable Convolution, number of multiplications: 8x8x5x5x3 + 256x1x1x3x8x8 = 53952



(a) Depthwise Convolution

(b) Pointwise Convolution

Image Source: Google Search Engine

Modified Depthwise Separable Convolution

In a Modified Depthwise Separable Convolution is the **pointwise convolution followed by a depthwise convolution**.

Xception Network uses Modified Depthwise Separable Convolution.

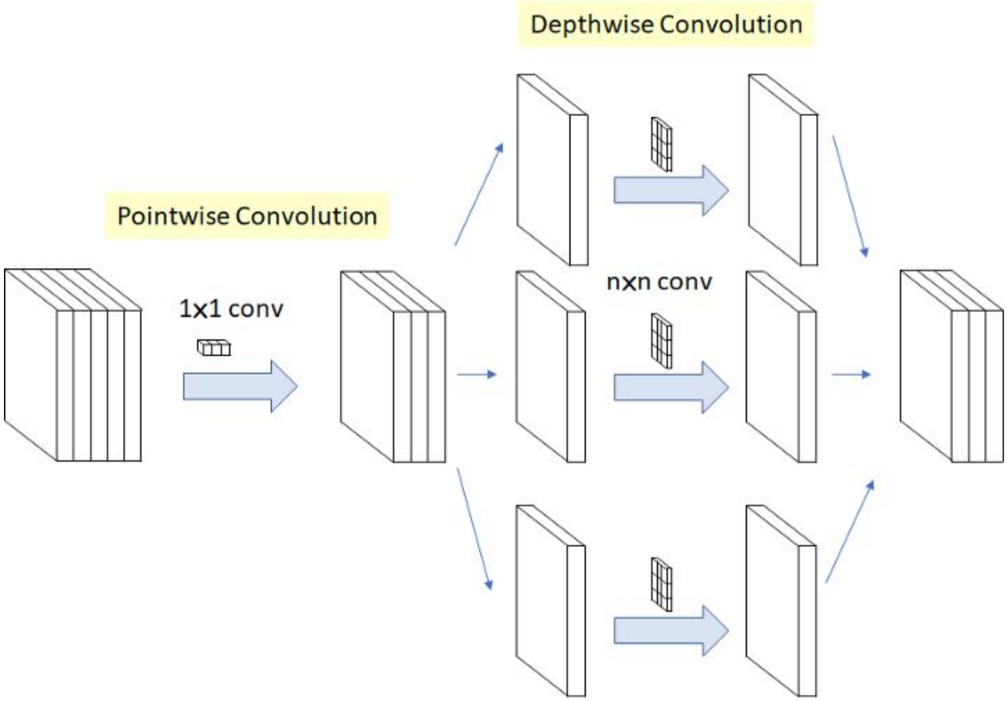


Image Source: Google Search Engine

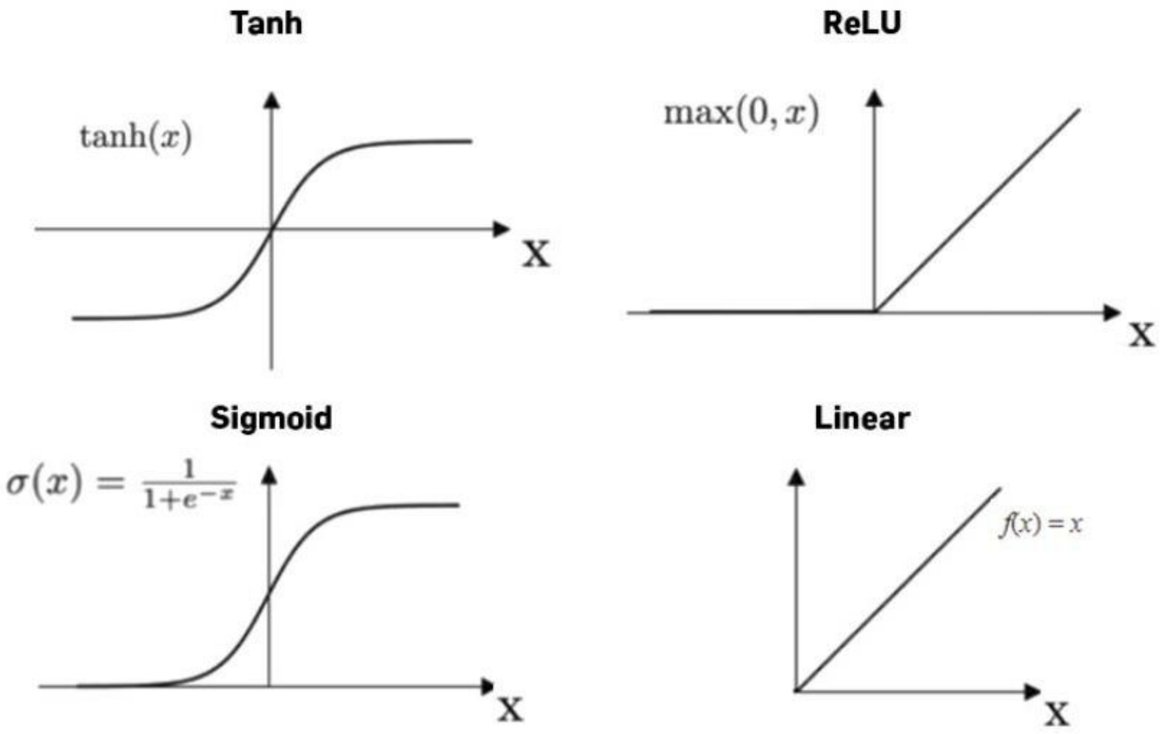
Activation Function

- An activation function is a mathematical equation that determines how much data should be passed from a neuron to the next neuron.
- It is the function that is applied on the weighted sum of the input of the neuron, i.e., it is $g(u)$.
- Generally, it is a nonlinear function for a neuron in the hidden layer and linear/non-linear function for neurons in the output layer.
- A linear activation function, also known as "no activation" or "identity function," is directly proportional to the input.
- Popular non-linear activation functions are:
  - Sigmoid, Tanh, ReLU, ELU, Softmax

Activation Functions

- Linear activation function in the output layer generally used for regression problem.

- Depending on the range of output values, we need to choose activation function.



Some Popular Activation Functions

- **Sigmoid**
  a. Also known as the logistic activation function
  b. Often used for models that predict probability as an output.
  c. Its curve looks like an S-shape and exists between 0 and 1.
  d. Suffers from saturating gradients problem.
- **Hyperbolic tangent (Tanh)**
  a. Has stronger gradients than the sigmoid function, and its output ranges from -1 to +1.
  b. Helps the learning algorithm converge faster.
- **Rectified linear unit (ReLU)**
  a. A non-saturating function, meaning it doesn't become flat at the extremes of the input range.
  b. It is faster than sigmoid and tanh.
- **Softmax**
  a. Converts vectors of real numbers into a probability distribution.
  b. Each output value represents the probability that the input belongs to a specific class.

Activation Functions for Classification

In output layer, generally:

- Sigmoid is used for binary classification
- Softmax is used for multi-class classification

Both generate values in the range of 0-1.

Summation of softmax values is 1.

Summation of sigmoid values in a classifier does not need to be 1.

$$sigmoid,\ y_i = \frac{e^{x_i}}{1 + e^{x_i}}$$

$$softmax,\ y_i = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}}$$