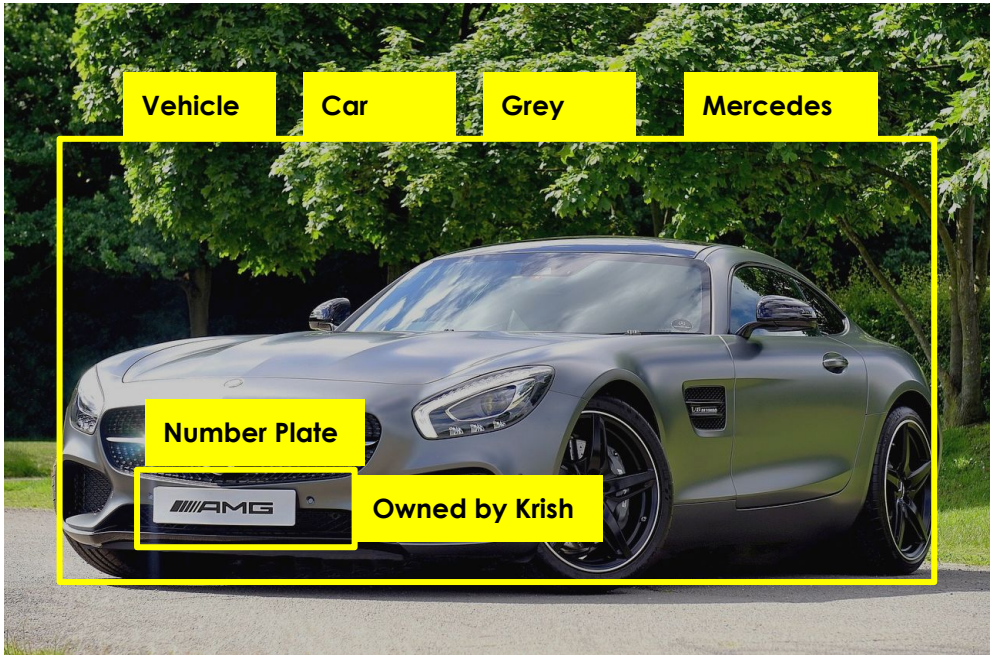
Introduction to Computer Vision

Image as A Data source

"A picture is worth a thousand words" - In a world where data is the new oil, images can be used as a rich source of information to make business decisions.

Images are a great source of information, as they:

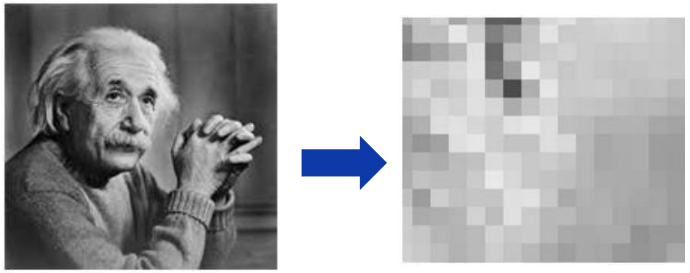
1) Quickly and concisely inform about people, places, objects, and events
2) Provides information that is difficult to convey through written formats (fashion, decor, art, etc.)
3) Sometimes record details of life of people that are not captured in written records
4) Can evoke memories and/or emotions in the viewer

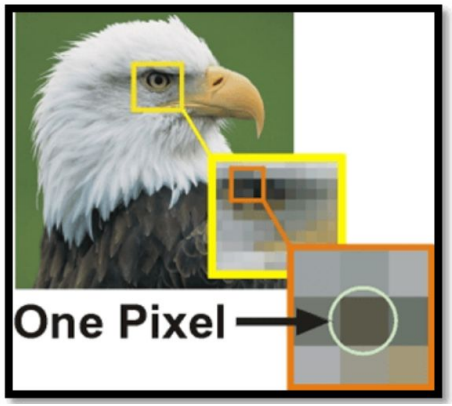


What information can you gather from the image above?

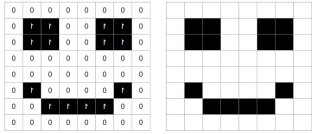
Pixel - the building block of an Image

- A pixel is a smallest unit of a digital image or graphics. Pixel stands for picture element
- A combination of pixels is used to complete images or videos or anything on the display of a digital display device
- **Pixel intensity is the primary value stored in pixels. It generally describes how bright the pixel is , / or what color should it be**
- Every pixel in a grayscale image has an intensity value which ranges from 0 to 255.
- To represent color images, separate red, green and blue components must be specified for each pixel (assuming an RGB colorspace), and so the pixel `value' is actually a vector of three numbers





Types of Image



Binary Images

Binary images are images that have been quantised to two values, usually denoted 0 and 1, but often with pixel values 0 and 255, representing black and white.

8 Bit color format

It has 256 shades of colors in it and is commonly called as the Grayscale images. The format of these images is PGM (portable grey map)

16 Bit color format

It has 65,536 different shades in it. It is also known as High color format. A 16 Bit image is further divided into red, blue and green color space.

How do you distribute 16 bits in 3 color ? - It is distributed 5 bits to red, 6 bits to green and 5 bits

24 Bit color format

They are also known as true color format.A 24 bit image has three different matrices of R, G, B.8 bits for R, 8 bits for G, 8 bits for B.

What are Image channels?

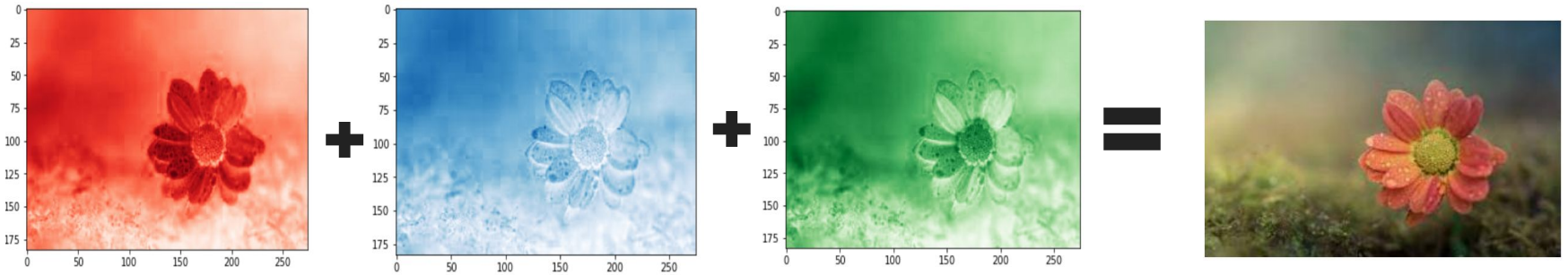
- Pixels are created by a combination of primary colors, this combination defines a color space (RGB, HSV etc), attributes of this color space is defined by channels
- For example an image from a digital camera would have red, green and blue channel whereas a grayscale image would have one channel.

1 channel - Grayscale image

3 channels, RGB

RGB Image

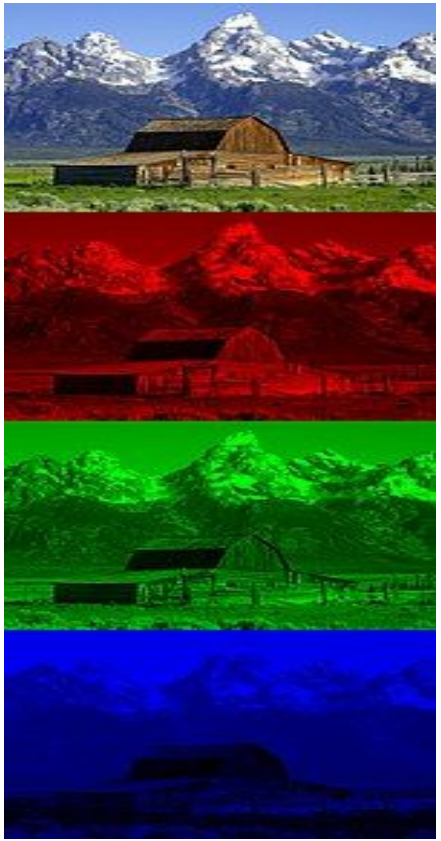
An RGB image has three channels: **red, green, and blue**.



Describe an RGB Image

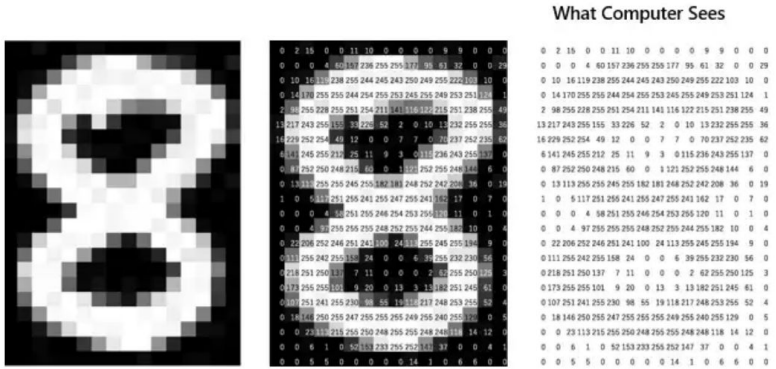
- Shape of the image : (924, 1280, 3)
- Image Height 924
- Image Width 1280
- Dimension of Image 3

- Image size 3548160
- Maximum RGB value in this image 255
- Minimum RGB value in this image 0



How Do computers look at the image?

- For computers images are just numbers
- Suppose we have a picture of handwritten No.8
- It is made of pixels & since it is a **grayscale image, each of these pixels can be represented by a single number.**
- Now **we can represent our image as a 2-dimensional matrix of numbers, one for each pixel in that image**
- **That's how a computer sees that image**.
- It sees a matrix of 2-dimensional numbers.



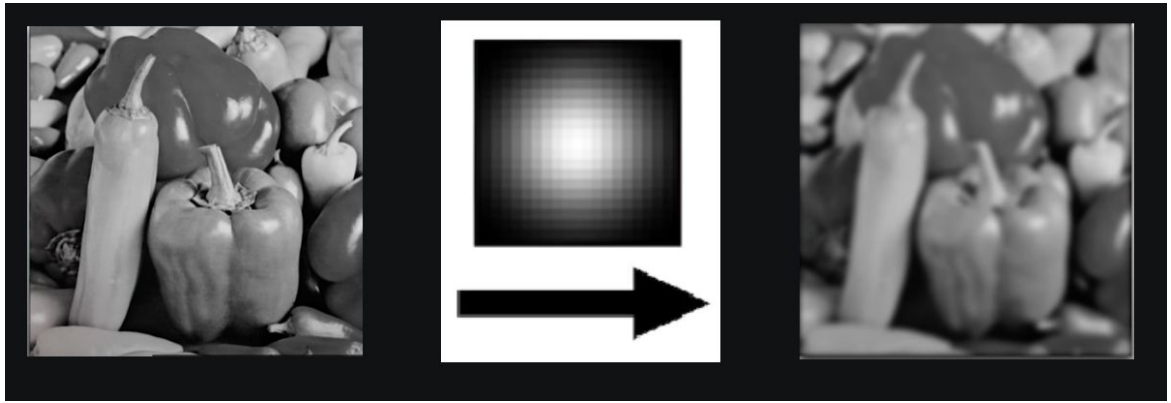
Images are just numbers

Image Processing

Image processing is a method to perform specific operations on an image. This can be done to enhance the image quality or extract some useful information from the image.

In this process the input is an image and the output may be an image or features of an image.

- Gaussian Image processing - Gaussian blur is also know gaussian smoothing, and results in blurring of an image by a gaussian function. It reduces the image noise in the image.



Edge Detection

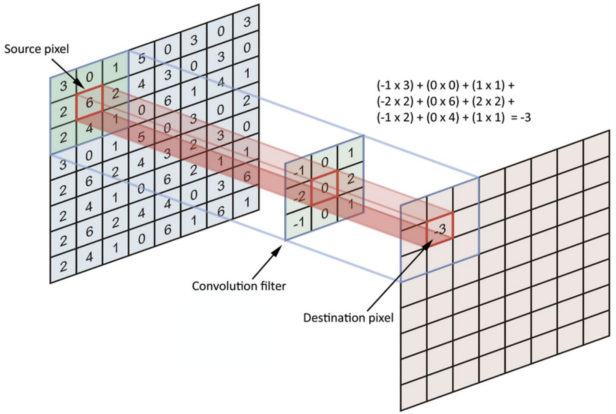
Edge Detection in image processing - Edge detection is an image processing technique for finding the boundaries of objects within images. It works by detecting discontinuities in brightness.

This could be very beneficial in extracting useful information from the image because most of the shape information is enclosed in the edges. **Classic edge detection methods work by detecting discontinuities in the brightness.** Edge detection can be done using various techniques, including the Sobel operator, the Canny edge detector, and the Hough transform.

Edge detection using Sobel operator

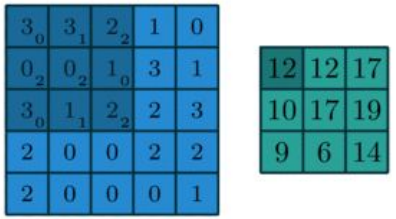
Edge detection using Canny edge

Image Filters



In a convolutional neural network (CNN), filters are used in the convolutional layers to extract features from the input image. **A filter is a small matrix of weights** that is applied to a small region of the input image, called a receptive field, in order to extract specific features from the image.

For example, a filter that is designed to detect edges would have weights that are similar to the **edge detection kernel,** this filter would be applied on the input image, and produce a **new feature map with edges highlighted.**



The Kernel - Feature extraction

A kernel is a small matrix of numbers, typically used to apply a convolution operation to an image. The convolution operation takes the kernel and slides it over the image, element-wise multiplying each pixel in the image with the corresponding pixel in the kernel and summing the results.

A kernel is a specific type of filter that is used for convolution operation,

Edge detection

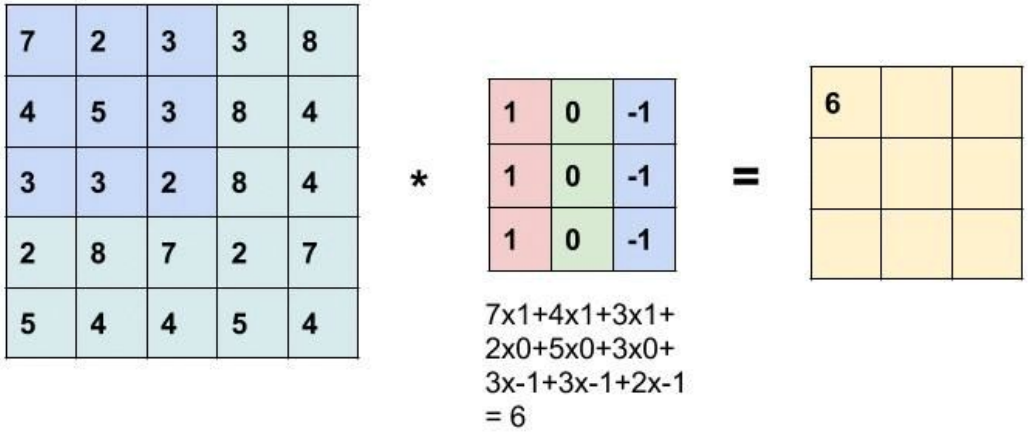
Kernel

$$* \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} =$$

Sharpen

$$* \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} =$$

Feature extraction - Kernel Math

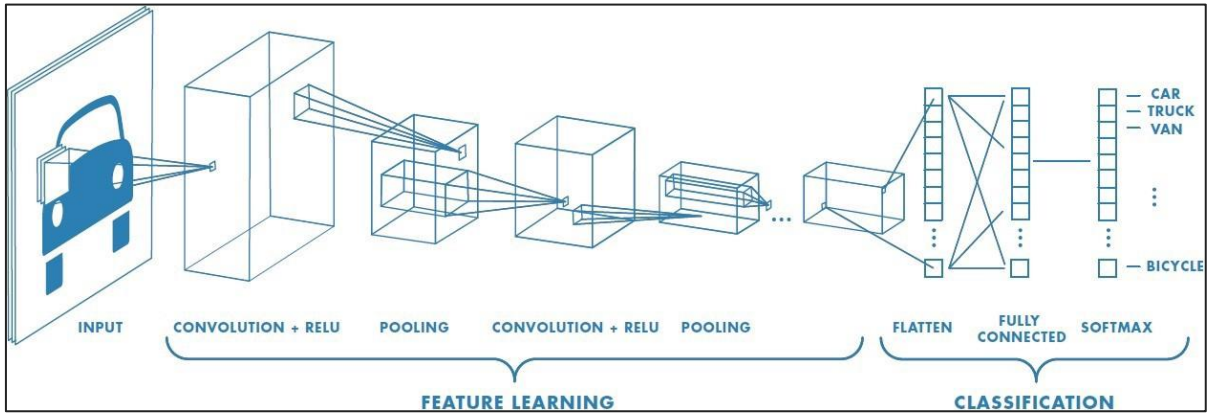


Convolutional Neural Network

A convolutional neural network (CNN) is a type of deep learning neural network that is commonly used for image and video recognition tasks. CNNs are designed to process data with a grid-like topology, such as an image.

A CNN is made up of several layers, including:

- Convolutional layer
- Pooling layer
- Fully connected layer
- Activation functions

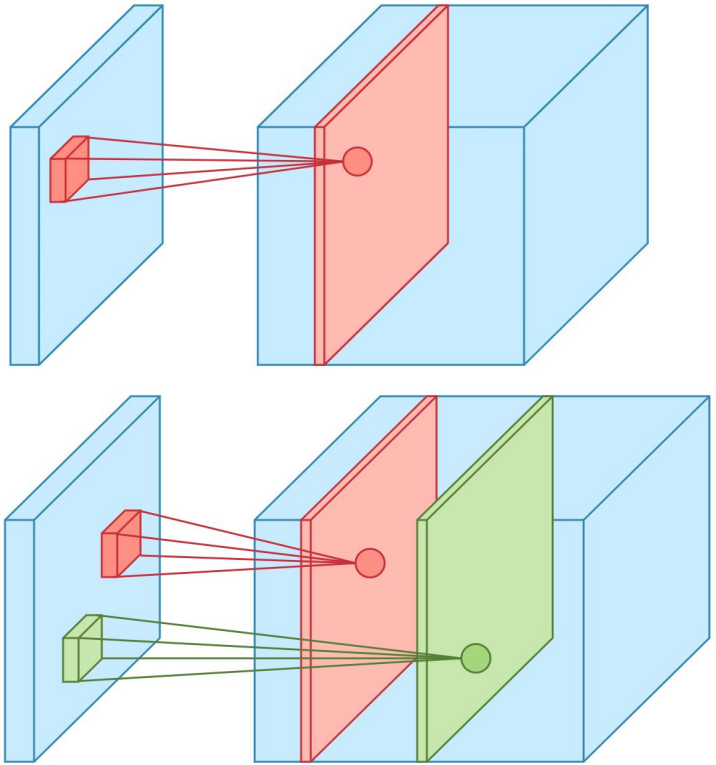


Convolutional Layer

A convolutional layer is a type of layer commonly used in CNNs. These networks are designed to process data that has a grid-like structure, such as an image.

In a convolutional layer, the input data is passed through a **set of filters or kernels** or weights, that are used to extract features from the input data.

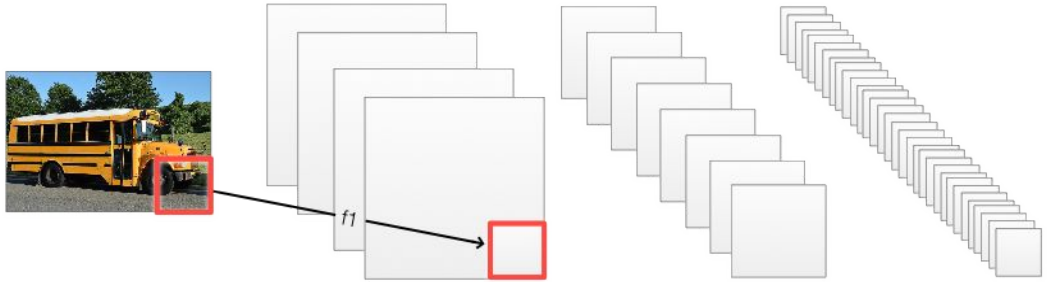
The convolutional layers often also include **pooling layers,** which are used to reduce the spatial dimensions of the data and make the network more robust to small changes in the position of the features.



Multiple Convolutional layers

Multiple convolutional layers are used in a CNN for reasons like:

- To extract more abstract and complex features
- Increase the depth of feature maps
- Introduce non-linearity
- Prevent overfitting



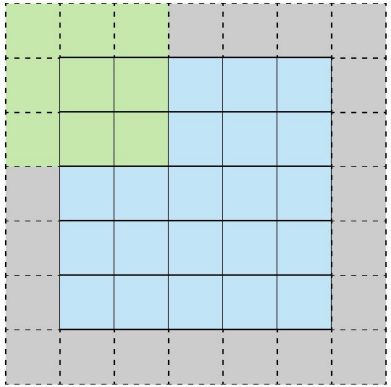
RGB Image	Filter 1 (4 filters)	Convolution layer 1	Filter 2	Convolution layer 2	Filter 3	Convolution layer 3
32x32x3	2x2x4 (s=2)	16x16x4	2x2x8 (s=2)	8x8x8	8x32	2x2x32 (s=2)

$$\text{output width} = \frac{W - F_w + 2P}{S_w} + 1$$

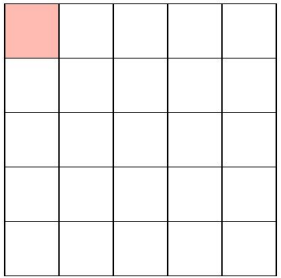
$$\text{output height} = \frac{H - F_h + 2P}{S_h} + 1$$

Padding In CNN

- When a filter moves towards the edges of the input data, it no longer has full overlap with the input data, and this leads to a reduction in the spatial dimensions of the output feature maps. This is commonly referred to as the **"shrinking spatial dimensions" problem.**

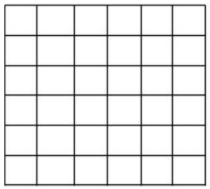


Stride 1 with Padding

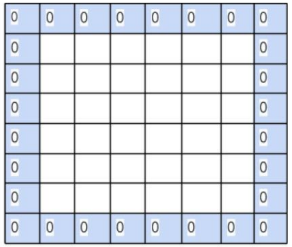


Feature Map

- Padding addresses this problem by adding a certain **number of zero-valued pixels** around the input data, effectively increasing the size of the input data. This allows the convolutional filter to fully overlap with the input data, **even at the edges, and preserves the spatial dimensions** of the output feature maps.



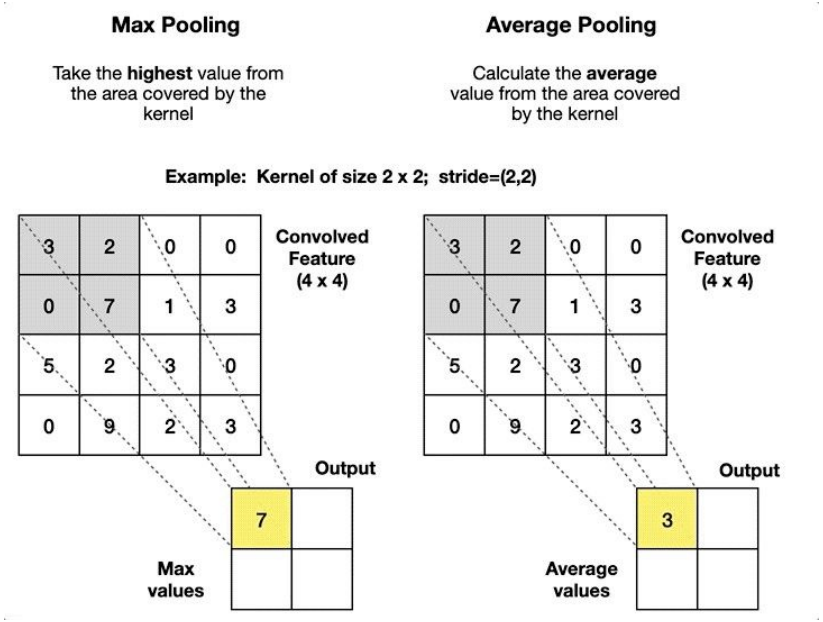
6x6 image



6x6 image with 1 layer of zero padding

Pooling Layers in CNN

- Pooling layers in a convolutional neural network (CNN) are used to reduce the spatial dimensions of the feature maps produced by the convolutional layers.
- The goal of pooling is to retain the most important information in the feature map while reducing its size
- Commonly used pooling layer is maxpooling which takes the maximum value of the window it is applied on. It is used to down-sample the spatial dimensions of the feature maps.



Fully Connected layer

- A fully connected layer in a convolutional neural network (CNN) is a layer that connects every neuron in one layer to every neuron in another layer
- The purpose of fully connected layers in a CNN is to take the high-level features extracted by the convolutional and pooling layers and use them to classify the input image into one of several predefined categories.
- Fully connected layers usually come at the end of the CNN architecture and are connected to the output layer.