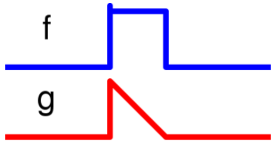
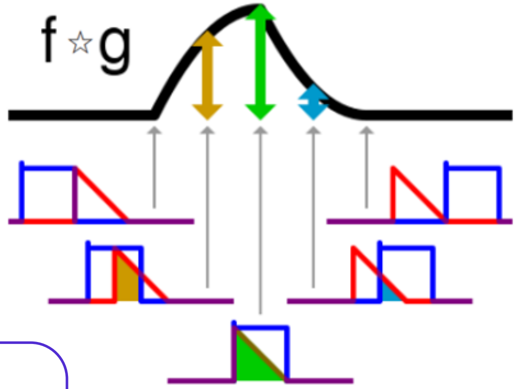
Convolutional Neural Network

Dr. Tushar Sandhan

Correlation & Convolution

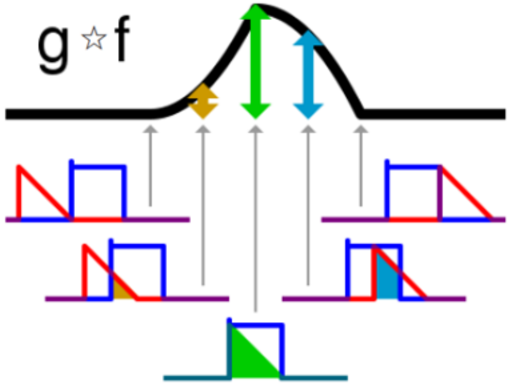
- Correlation



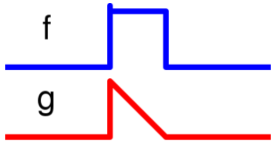


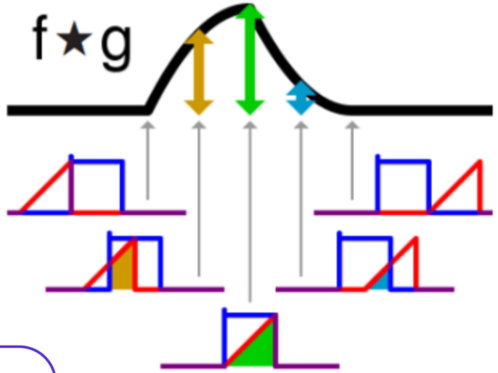
$$R(x) = f(x) \star g(x)$$

$$R(x) = \int_{-\infty}^{\infty} f(z)g(x+z)dz$$



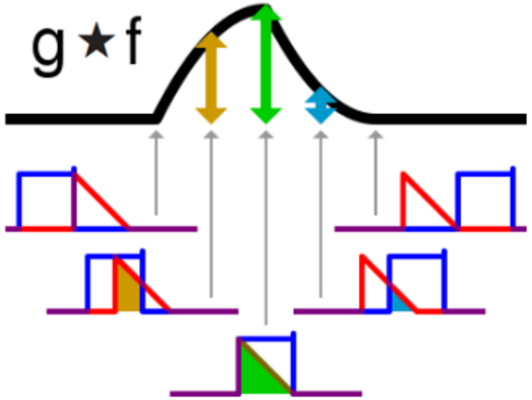
- Convolution





$$G(x) = f(x) \ast g(x)$$

$$G(x) = \int_{-\infty}^{\infty} f(z)g(x-z)dz$$

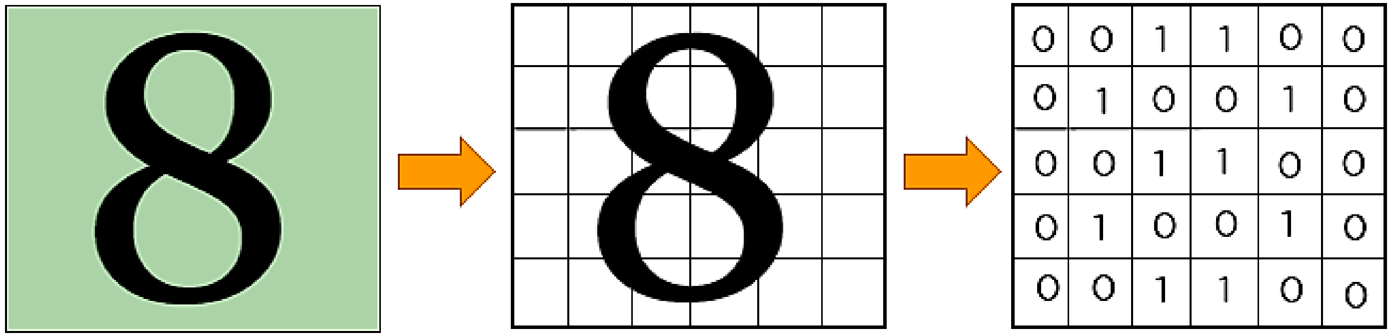


credit: wikipedia

Convolution

- 2D convolution
  - $w \;\; \rightarrow m \times n$
  - $a = \frac{m-1}{2}, b = \frac{n-1}{2}$
    - a, b are assumed to be odd integers
    - note the kernels do not depend on (x, y)

$$(w \star f)(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(x-s, y-t)$$

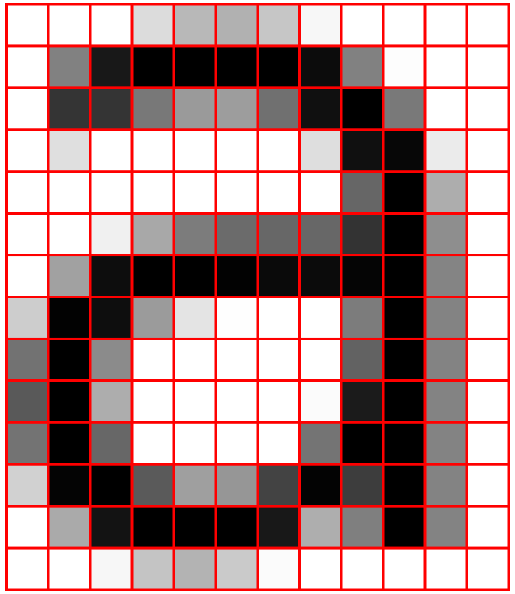


Real Image of the digit 8 | Represented in the form of an array | Digit 8 represented in the form of pixels of 0's and 1's

What is Convolutional Neural Network?

- A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be.

a



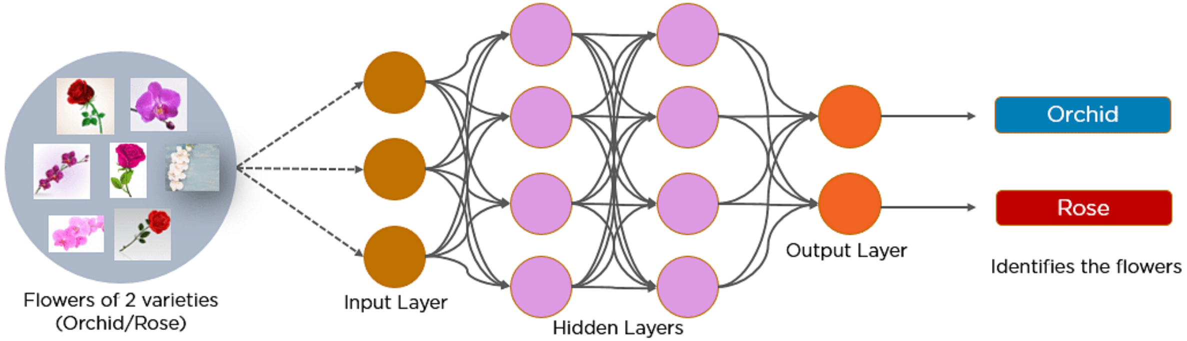
1.0	1.0	1.0	0.9	0.6	0.6	0.6	1.0	1.0	1.0	1.0	1.0
1.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0	1.0	1.0
1.0	0.2	0.2	0.5	0.6	0.6	0.5	0.0	0.0	0.5	1.0	1.0
1.0	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.0	0.0	0.9	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
1.0	1.0	1.0	0.5	0.5	0.5	0.5	0.5	0.4	0.0	0.5	1.0
1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0
0.9	0.0	0.0	0.6	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
0.5	0.0	0.6	1.0	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
0.5	0.0	0.7	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.5	1.0
0.6	0.0	0.6	1.0	1.0	1.0	1.0	0.5	0.0	0.0	0.5	1.0
0.9	0.1	0.0	0.6	0.7	0.7	0.5	0.0	0.5	0.0	0.5	1.0
1.0	0.7	0.1	0.0	0.0	0.0	0.1	0.9	0.8	0.0	0.5	1.0
1.0	1.0	1.0	0.8	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0

sandhan@iitk.ac.in

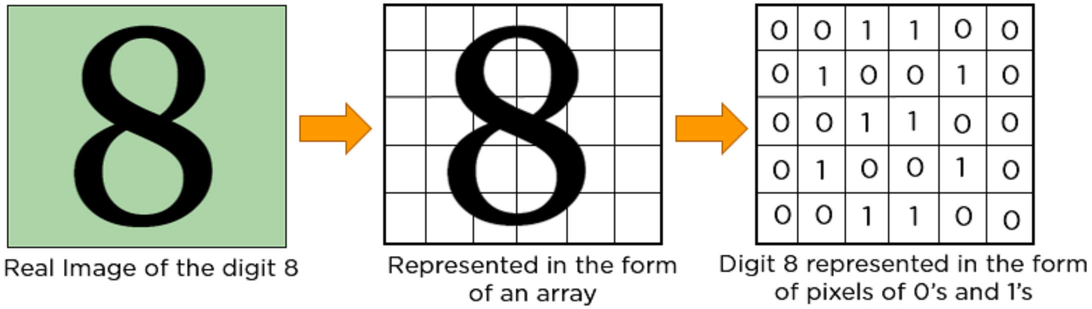
What is Convolutional Neural Network?

- A convolutional neural network is a feed-forward neural network that is generally used to analyze visual images by processing data with grid-like topology. It is also known as a ConvNet. A convolutional neural network is used to detect and classify objects in an image.

- A neural network that identifies two types of flowers: Orchid and Rose.



- In CNN, every image is represented in the form of an array of pixel values.



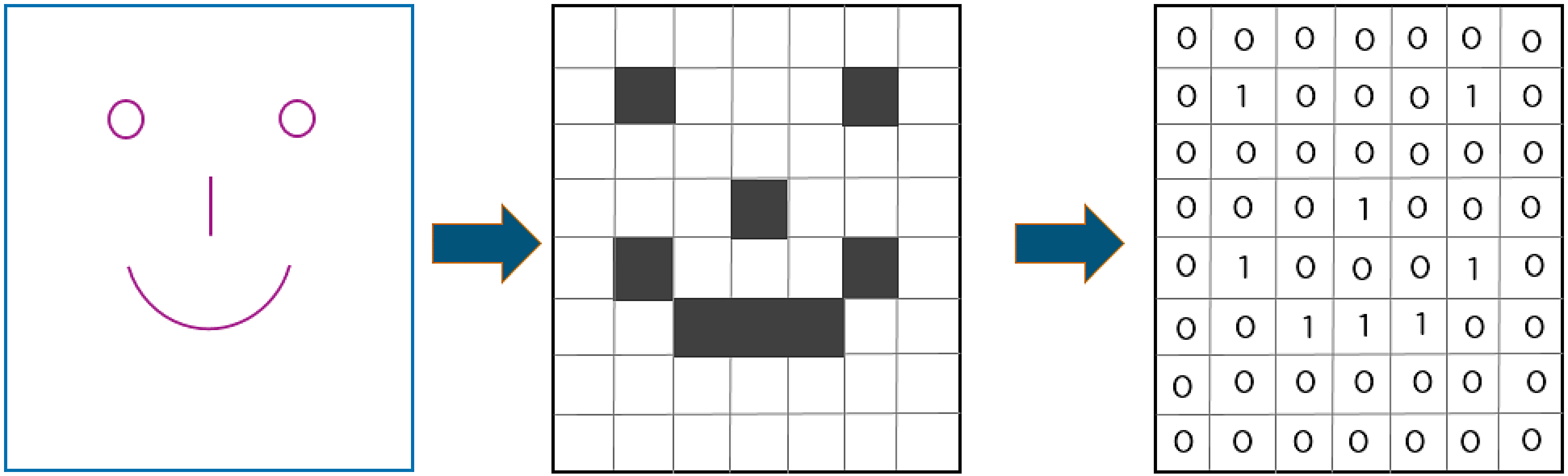
CNNs vs. Neural networks

- TThe biggest problem with regular neural networks (NNs) is a lack of scalability. For smaller images with fewer color channels, a regular NN may produce satisfactory results. But as the size and complexity of an image increases, the need for computational power and resources also increases which requires a larger and more expensive NN.

- Moreover, the problem of overfitting also arises over time, wherein the NN tries to learn too many details in the training data. It may also end up learning the noise in the data, which affects its performance on test data sets. Ultimately, the NN fails to identify the features or patterns in the data set and thus the object itself.

- In contrast, a CNN uses parameter sharing. In each layer of the CNN, each node connects to another. A CNN also has an associated weight; as the layers' filters move across the image, the weights remain fixed -- a condition known as parameter sharing. This makes the whole CNN system less computationally intensive than an NN system.

How Does CNN Recognize Images?



Real Image

Represented in the form of black and white pixels

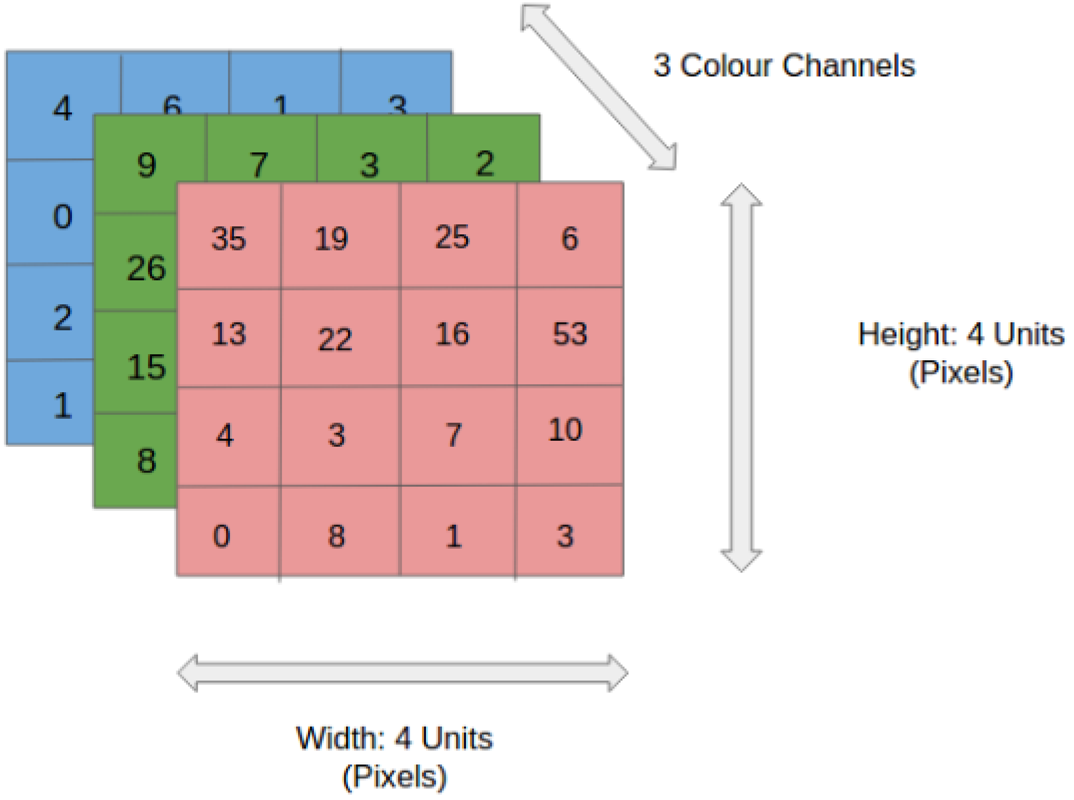
Image represented in the form of a matrix of numbers

- As you can see from the above diagram, only those values are lit with black that have a value of 1.

sandhan@iitk.ac.in

Color image: RGB 3 channels

- TWe have an RGB image that has been separated by its three color planes — Red, Green, and Blue. There are a number of such color spaces in which images exist — Grayscale, RGB, HSV, CMYK, etc.



sandhan@iitk.ac.in

Layers in a Convolutional Neural Network

- TA convolution neural network has multiple hidden layers that help in extracting information from an image.

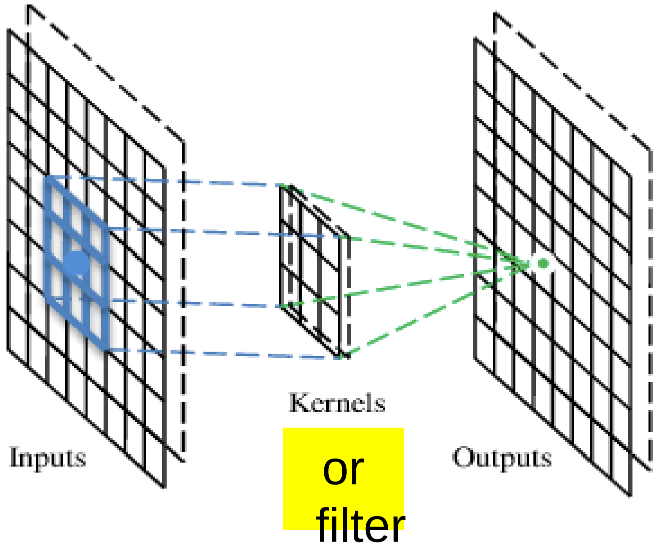
- The four important layers in CNN are:

1. Convolution layer
2. Activation function (ex. ReLU)
3. Pooling layer
4. Fully connected layer

sandhan@iitk.ac.in

Convolution layer:

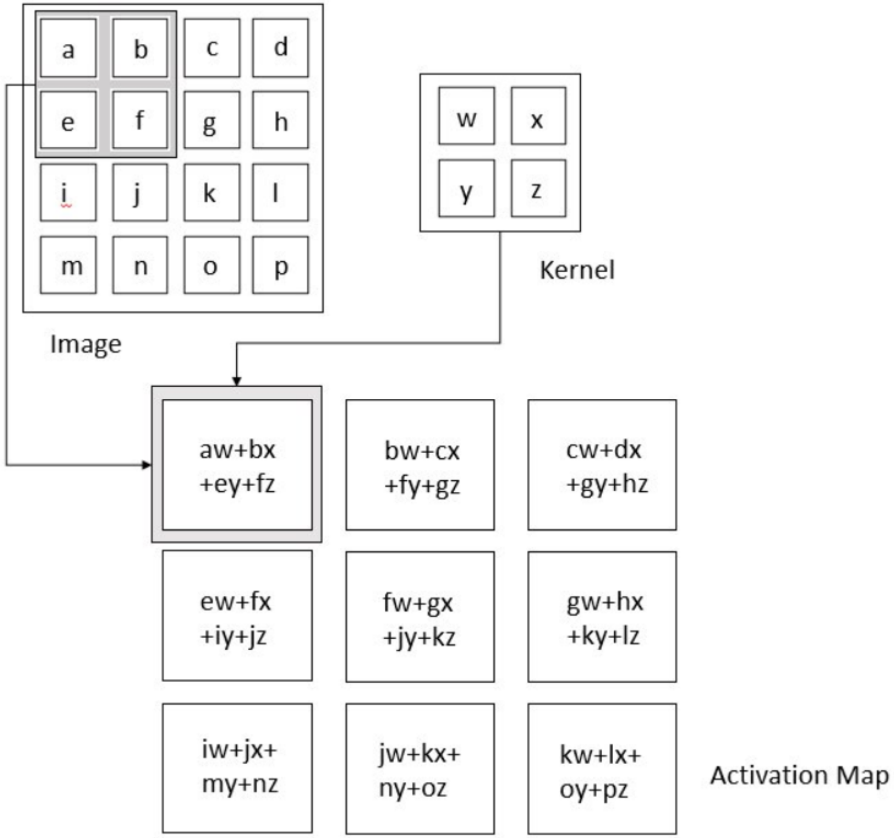
- A CNN is a neural network with some convolutional layers (and some other layers). A convolutional layer has a number of filters that does convolutional operation.



- **What is kernel?**
  - The kernel is nothing but a filter that is used to extract the features from the images. The kernel is a matrix that moves over the input data, performs the dot product with the sub-region of input data, and gets the output as the matrix of dot products.

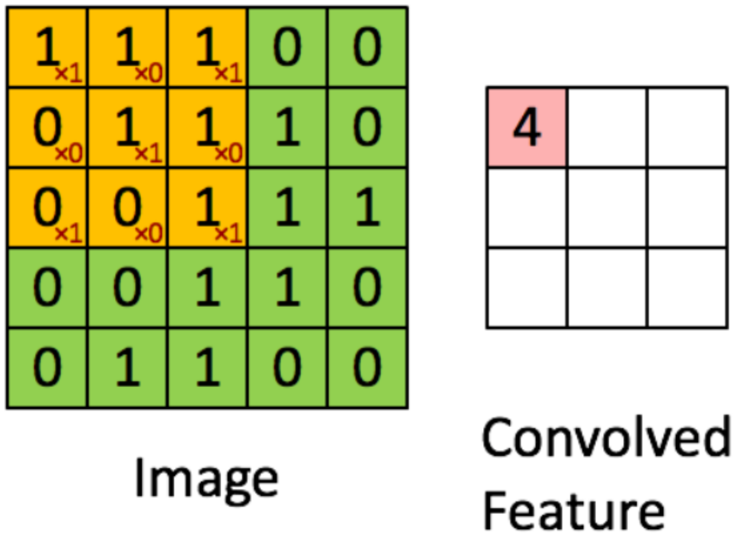
Convolution layer:

- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.



1	0	1
0	1	0
1	0	1

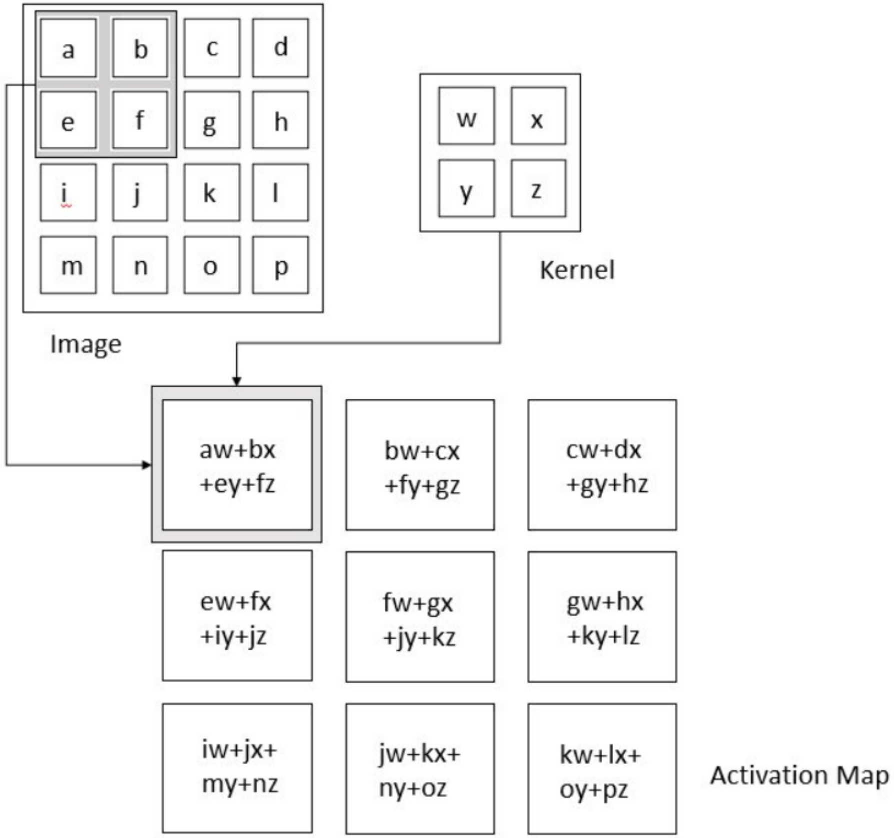
3 x 3
Filter or Kernel

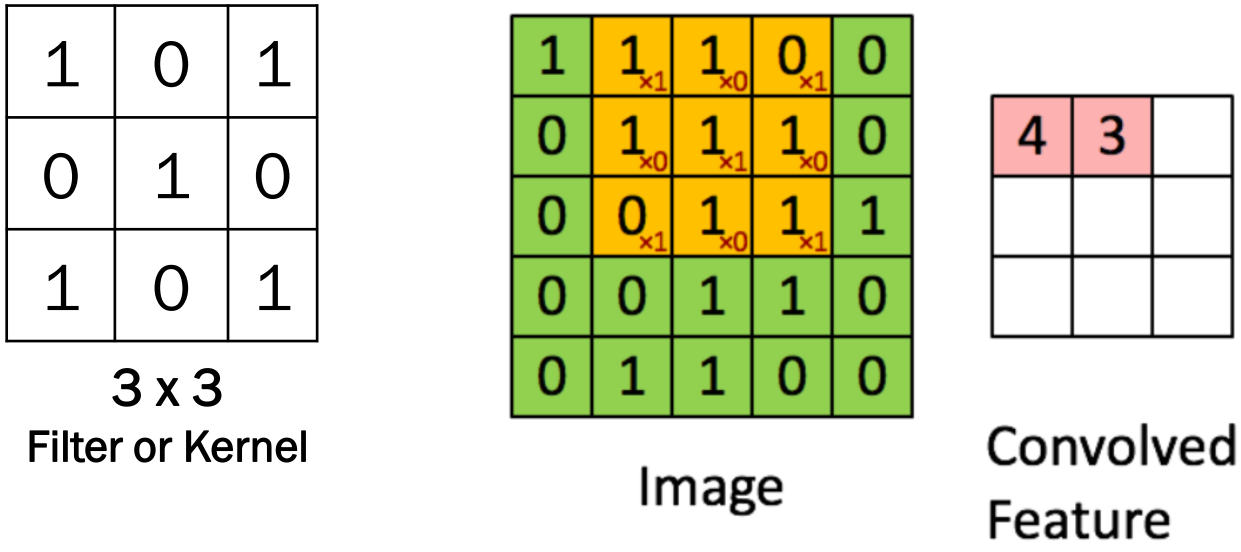


sandhan@iitk.ac.in

Convolution layer:

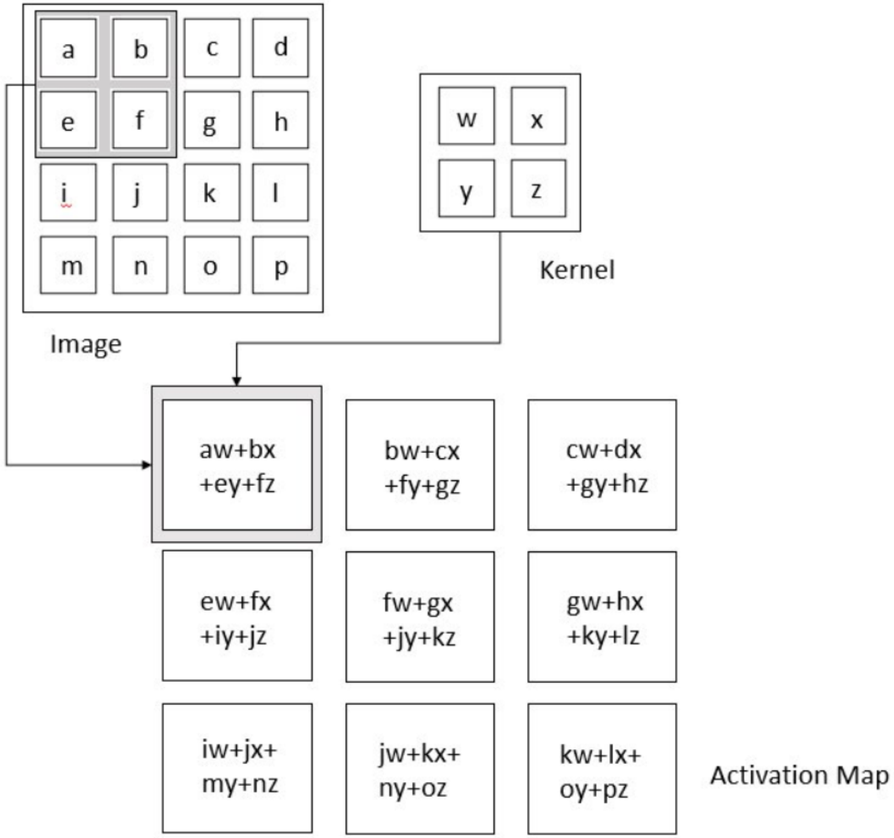
- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.





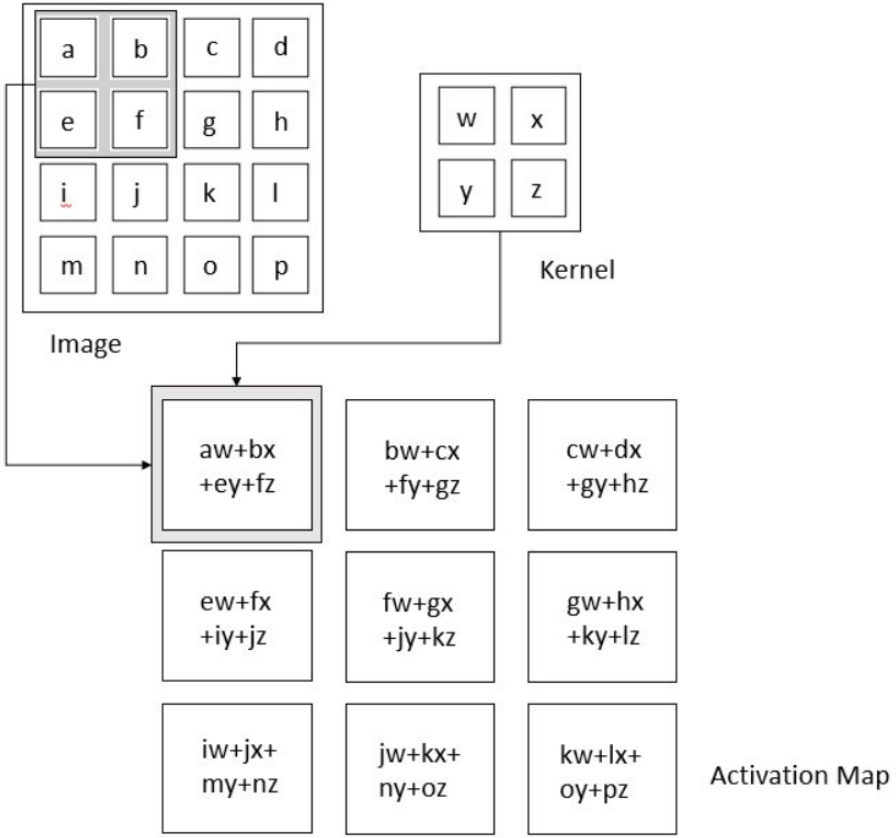
Convolution layer:

- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.



Convolution layer:

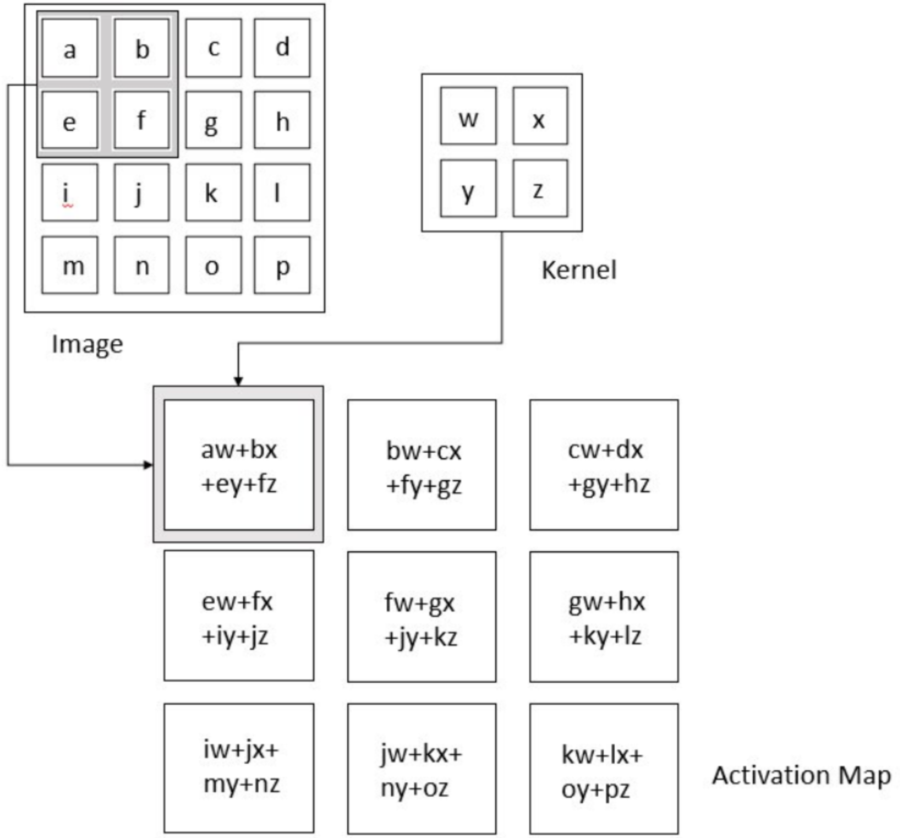
- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.

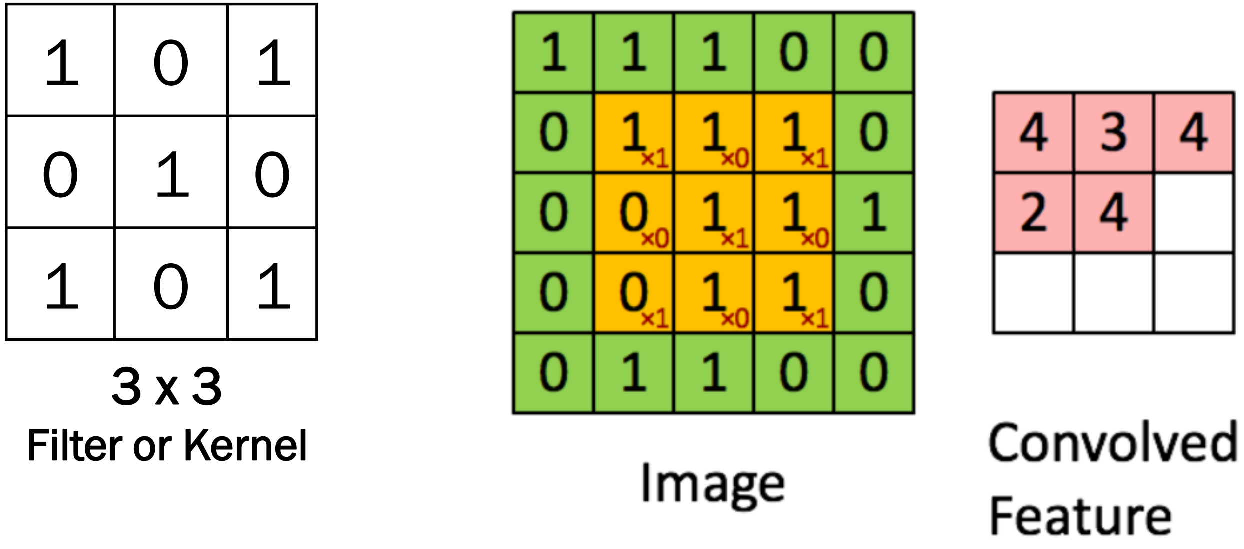


sandhan@iitk.ac.in

Convolution layer:

- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.

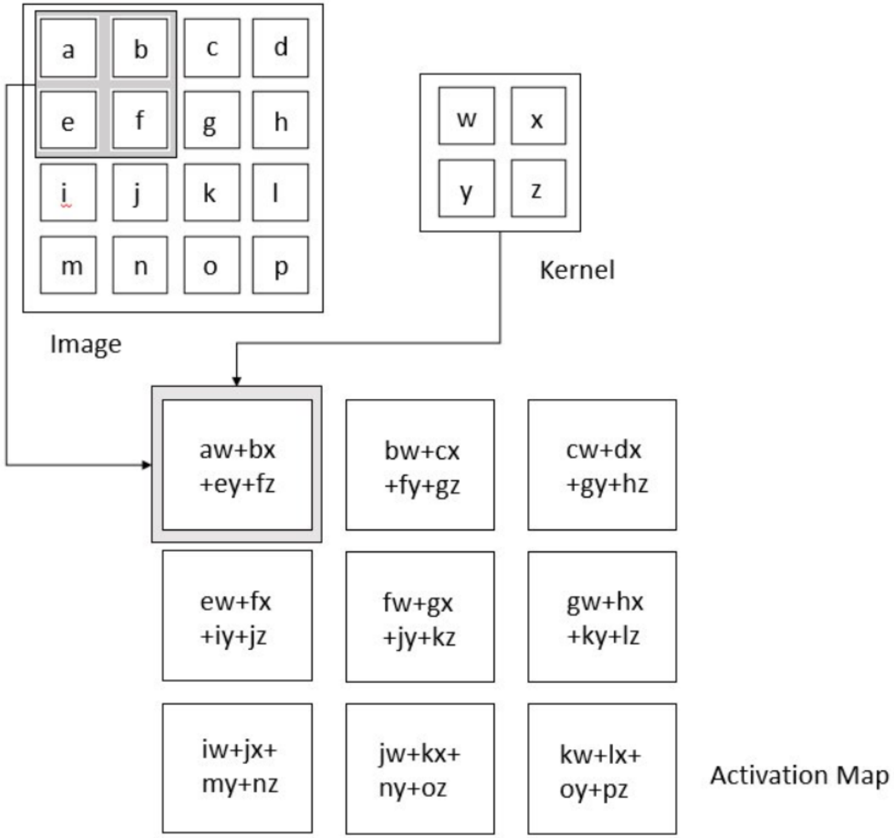


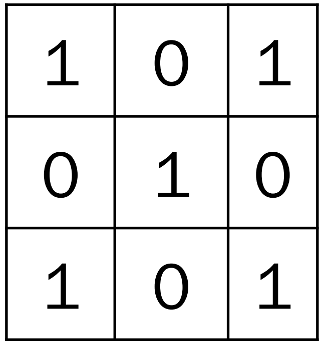


sandhan@iitk.ac.in

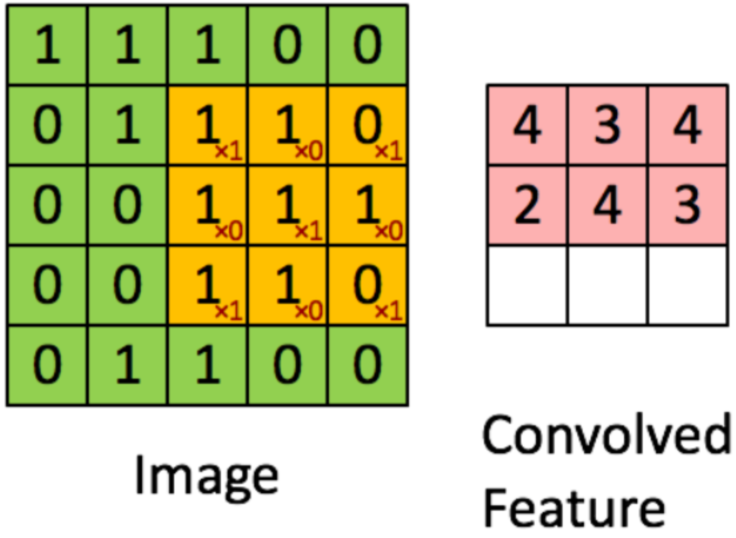
Convolution layer:

- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.





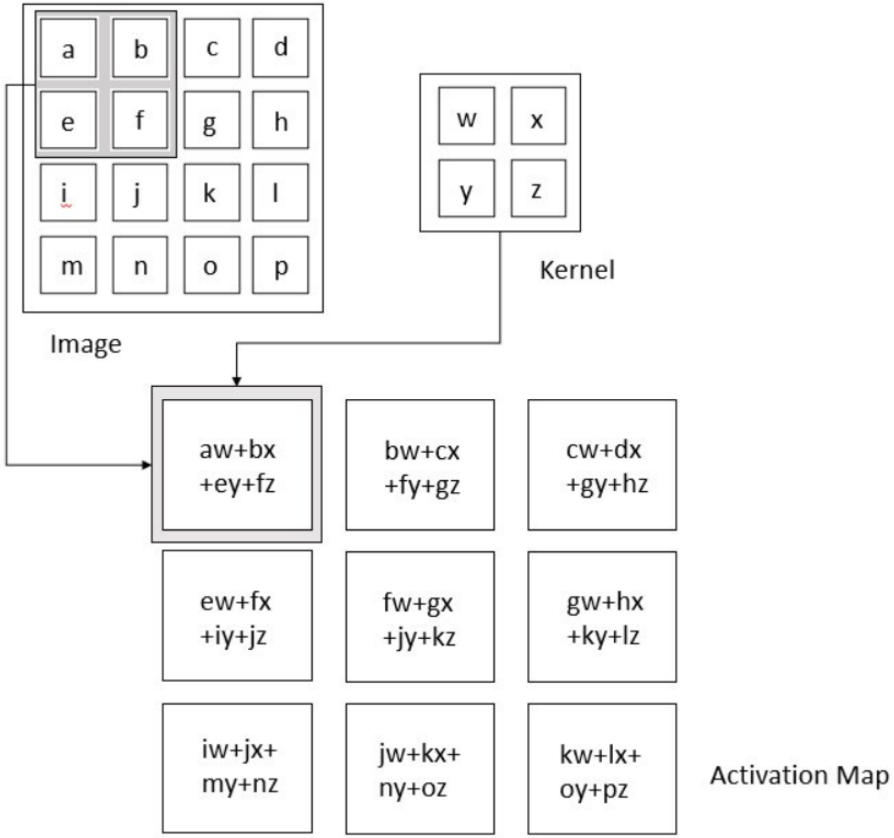
3 x 3
Filter or Kernel



sandhan@iitk.ac.in

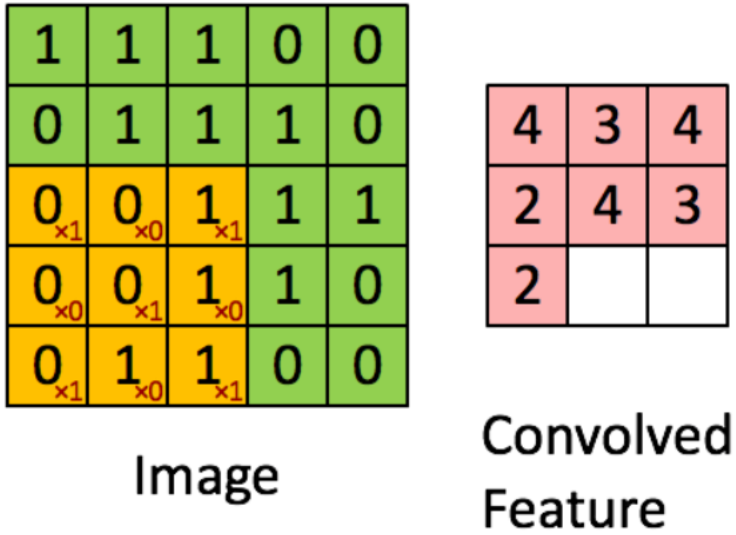
Convolution layer:

- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.



1	0	1
0	1	0
1	0	1

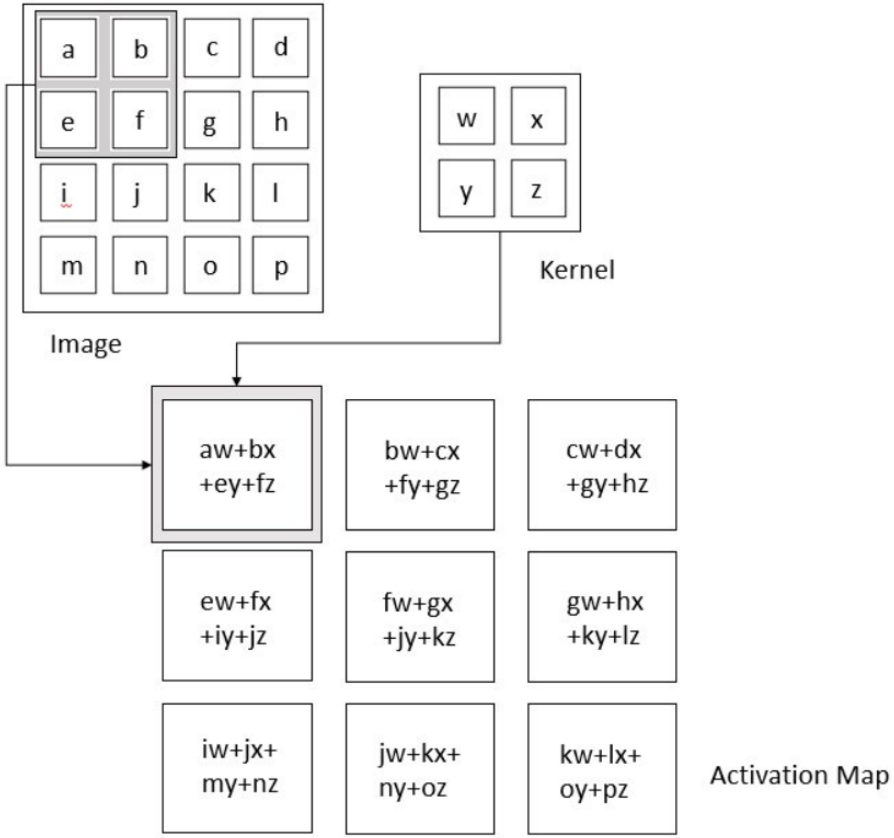
3 x 3
Filter or Kernel

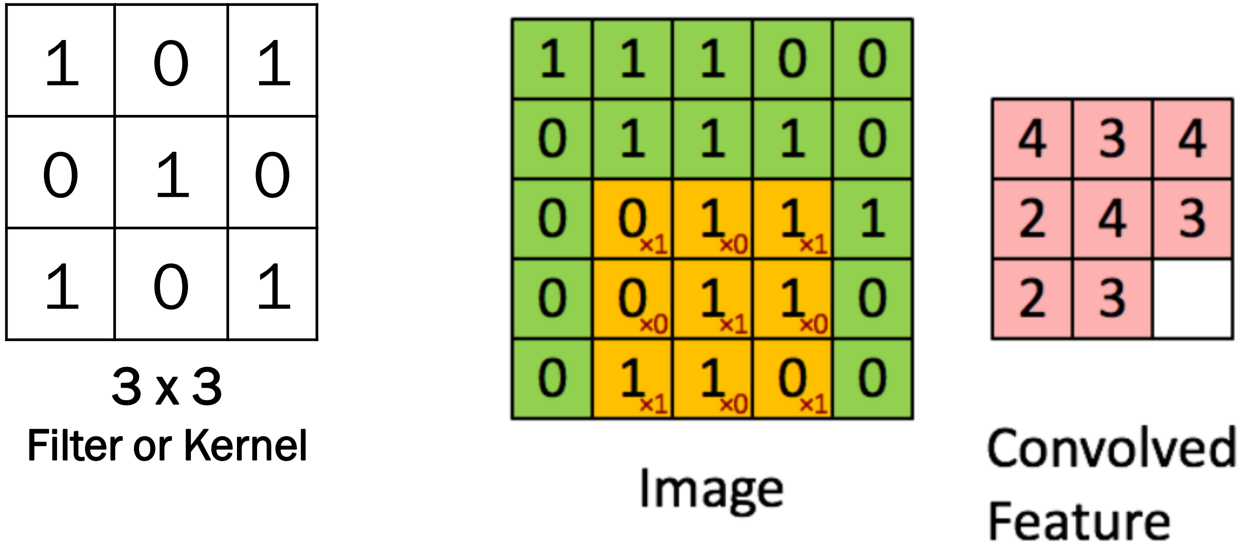


sandhan@iitk.ac.in

Convolution layer:

- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.

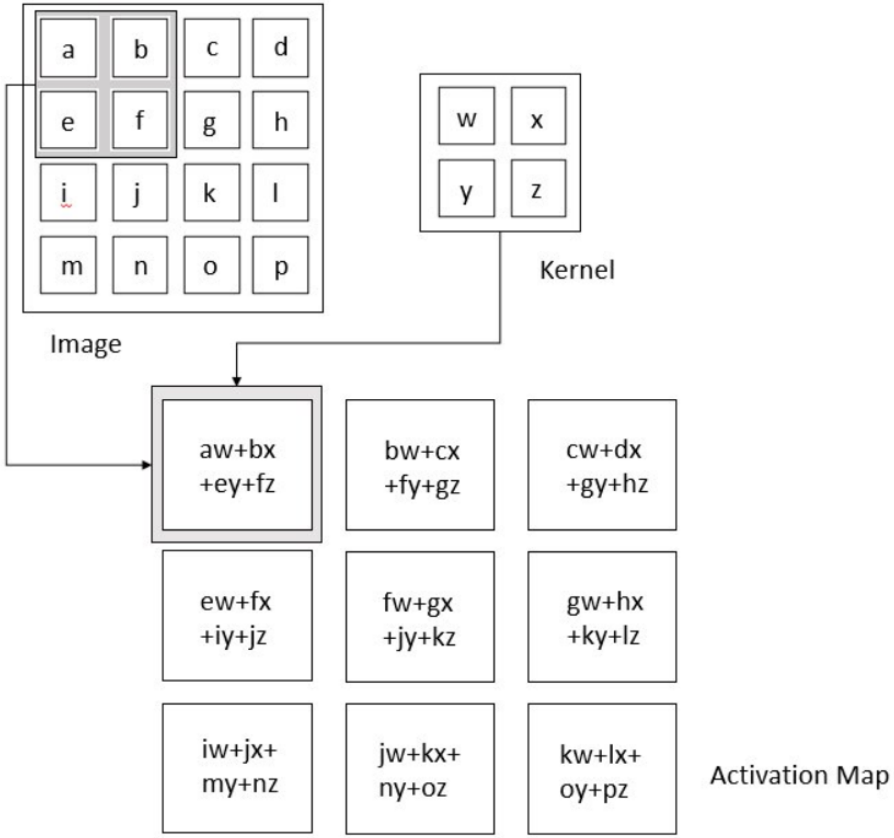


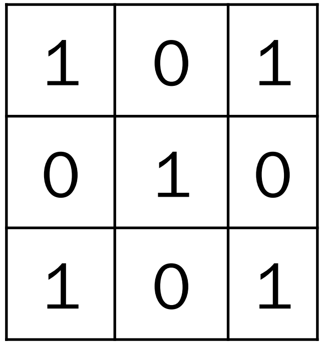


sandhan@iitk.ac.in

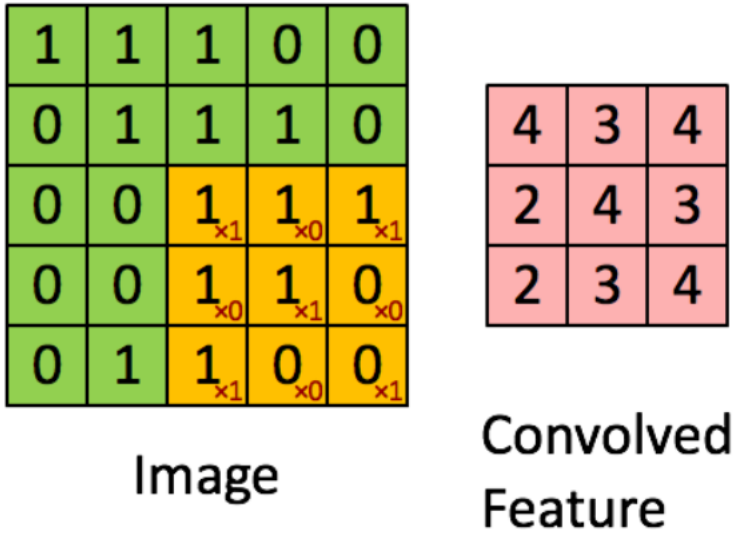
Convolution layer:

- This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.





3 x 3
Filter or Kernel



Convolution layer:

These are the network parameters to be learned.

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

Filter 1

1	-1	-1
-1	1	-1
-1	-1	1

Filter 2

-1	1	-1
-1	1	-1
-1	1	-1

⋮ ⋮

Each filter detects a small pattern (3 x 3).

Convolution layer:

stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

Dot product → 3 -1

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

If stride=2

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

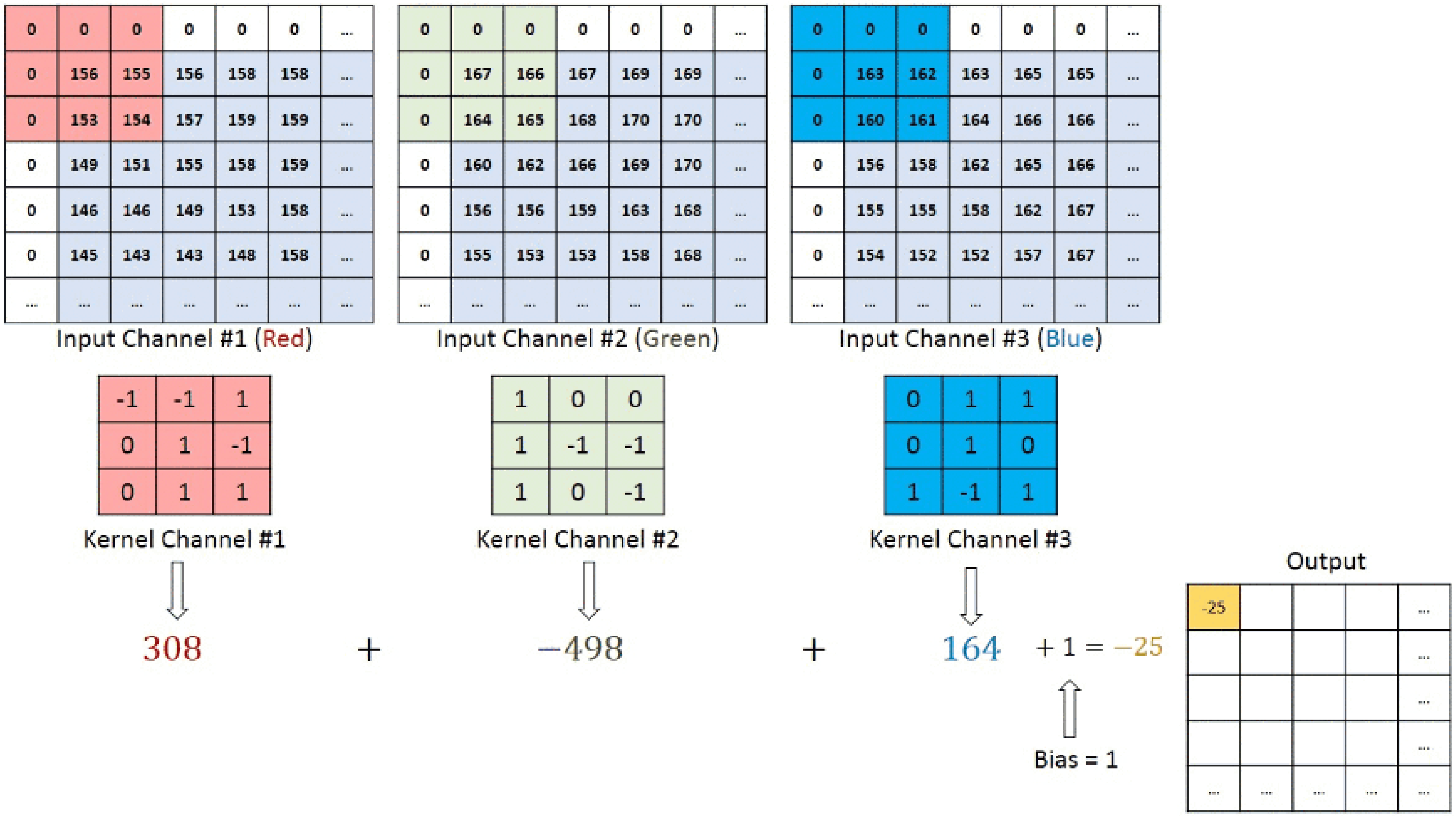
3 -3

6 x 6 image

- **What is stride?**
  - Stride is how far the filter moves in every step along one direction. How does a computer read an image? Basically a computer read an image from left to right and from top to bottom. Therefore it starts from the top-left corner all the way to bottom-right corner.

sandhan@iitk.ac.in

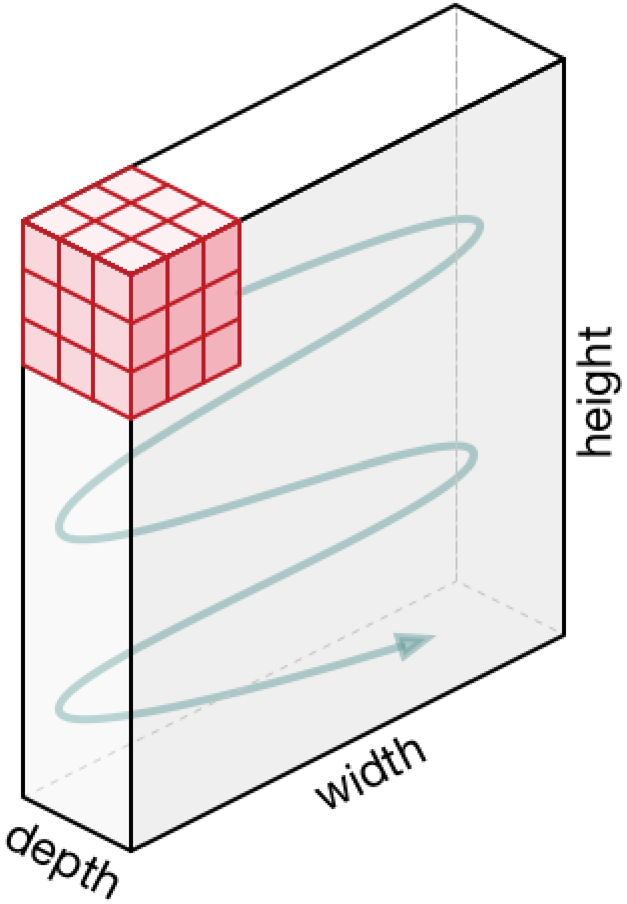
Convolution layer:



Convolution layer:

- TThe filter moves to the right with a certain Stride Value till it parses the complete width.

- Moving on, it hops down to the beginning (left) of the image with the same Stride Value and repeats the process until the entire image is traversed.

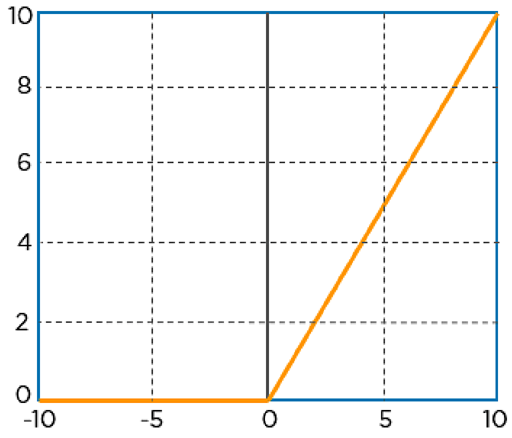


sandhan@iitk.ac.in

Activation function (ex. ReLU)

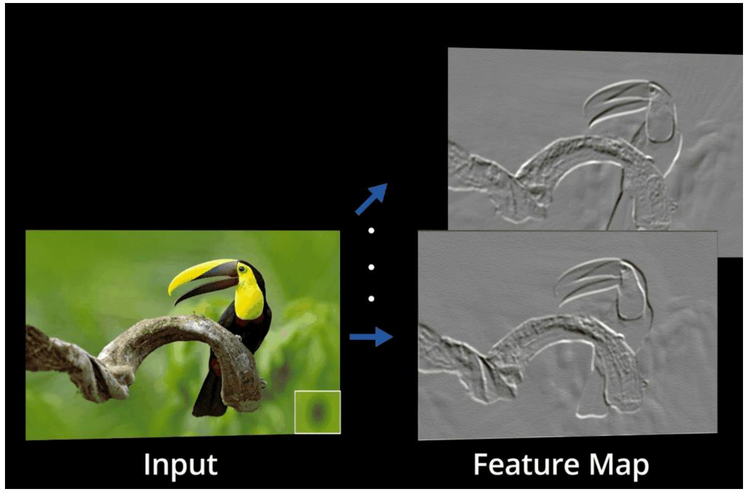
- AReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer.

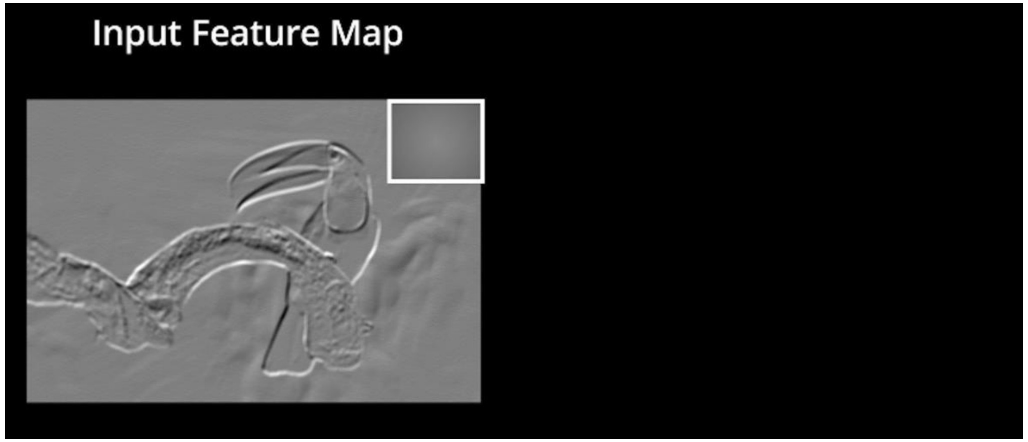
- ReLU performs an element-wise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map. Below is the graph of a ReLU function:



R(z) = max(0, z)

- The original image is scanned with multiple convolutions and ReLU layers for locating the features.

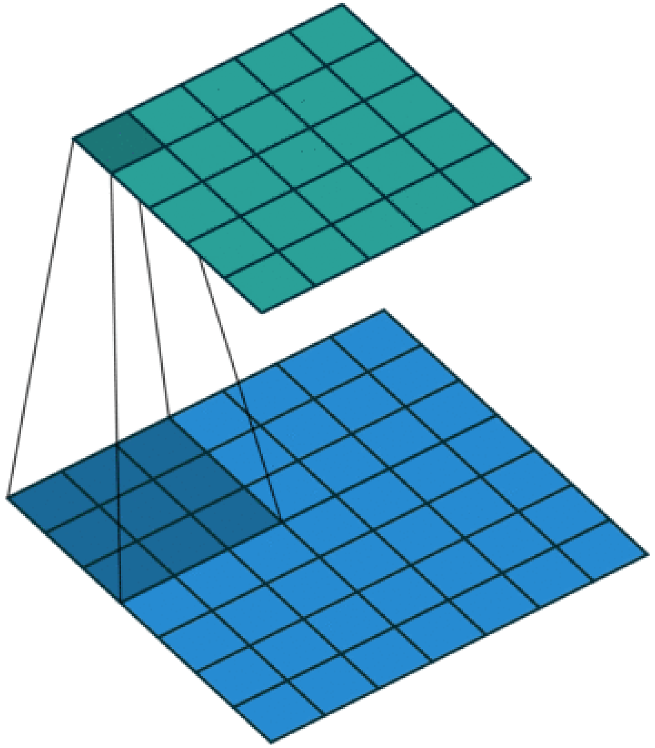




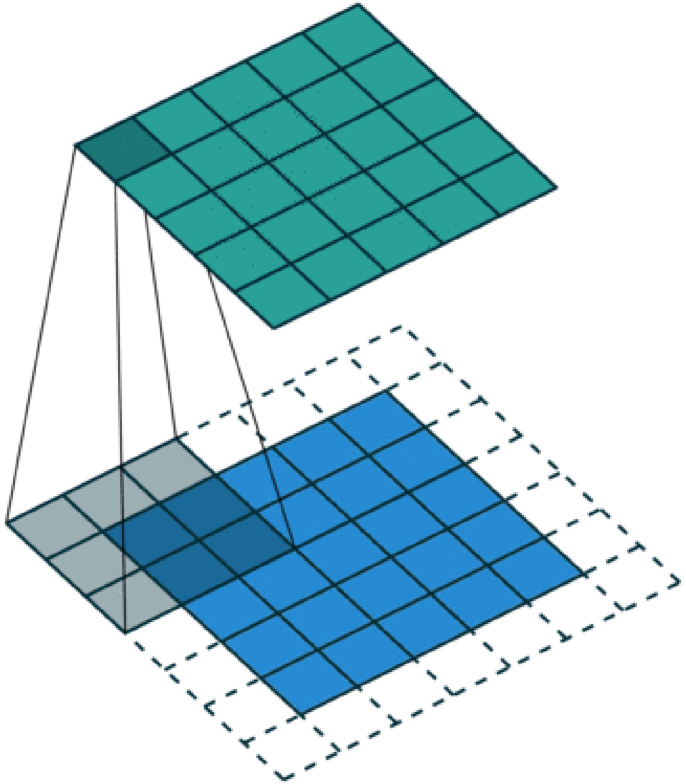
sandhan@iitk.ac.in

Padding

- VALID : Don't apply any padding for image border
- SAME : Apply padding for image border



valid padding

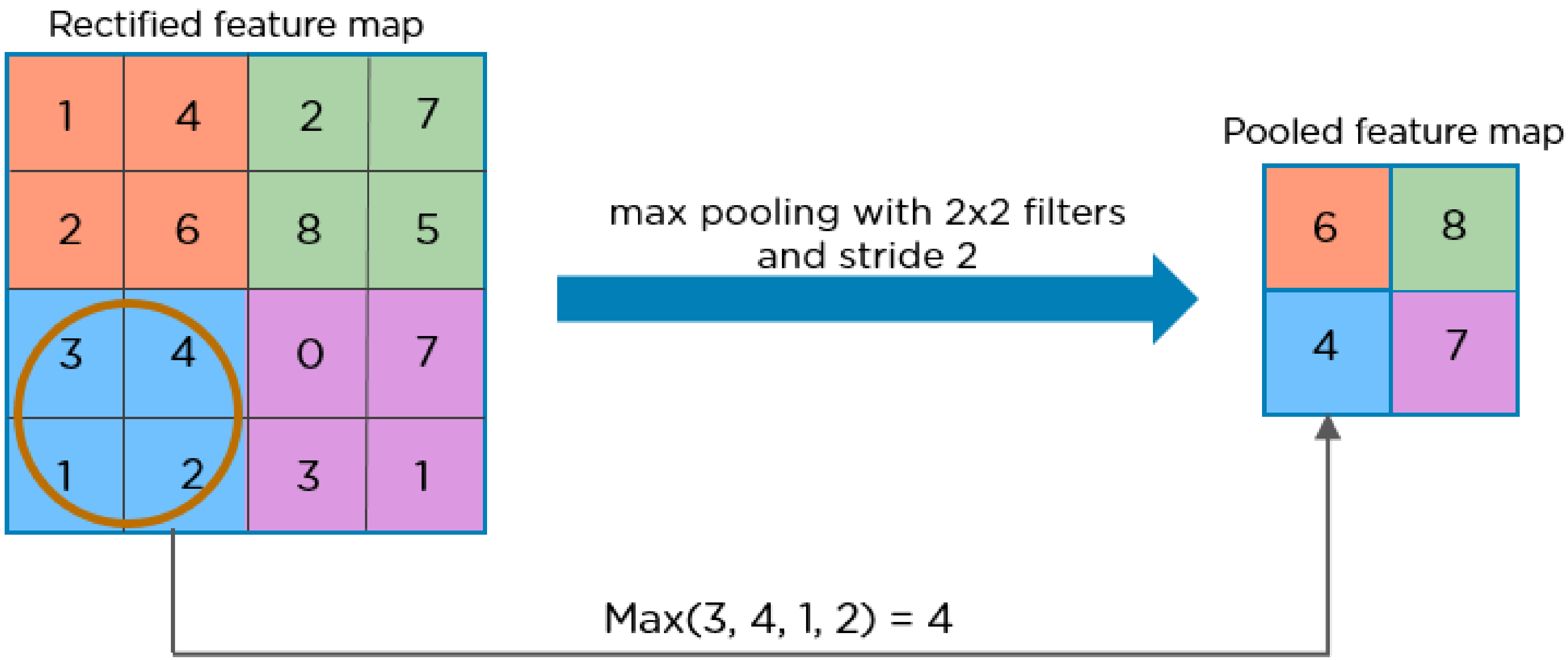


SAME padding

sandhan@iitk.ac.in

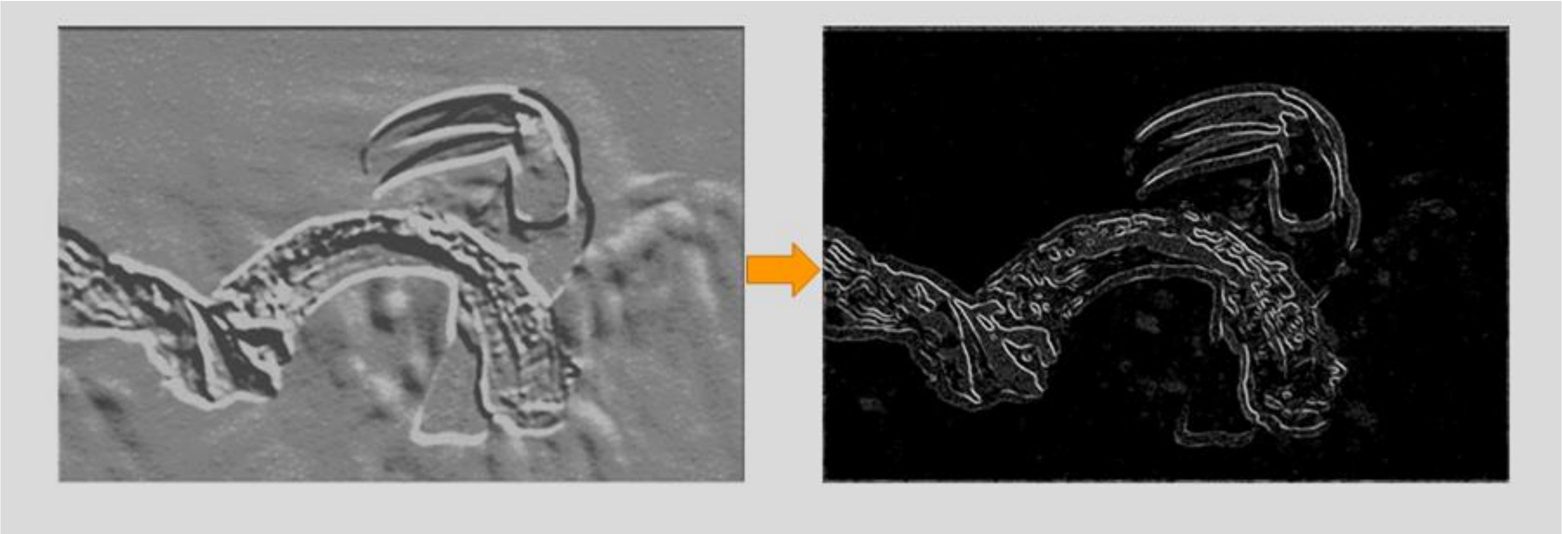
Pooling layer

- Pooling is a down-sampling operation that reduces the dimensionality of the feature map. The rectified feature map now goes through a pooling layer to generate a pooled feature map.

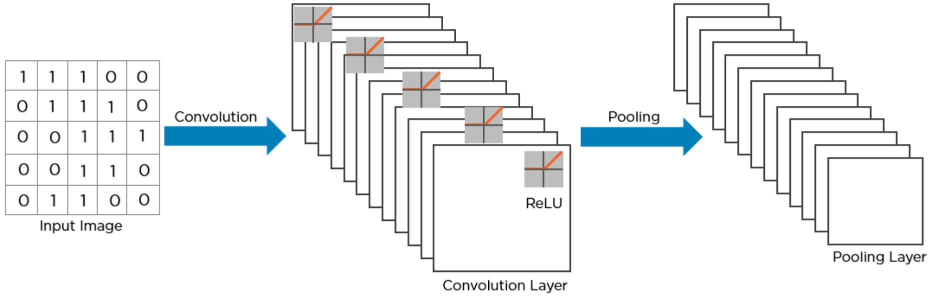


Pooling layer

- The pooling layer uses various filters to identify different parts of the image like edges, corners, body, feathers, eyes, and beak.

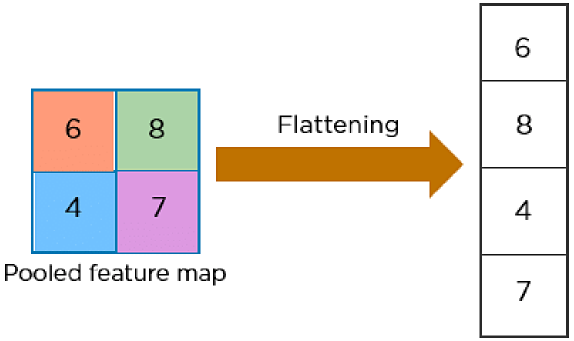


- Here's how the structure of the convolution neural network looks so far:

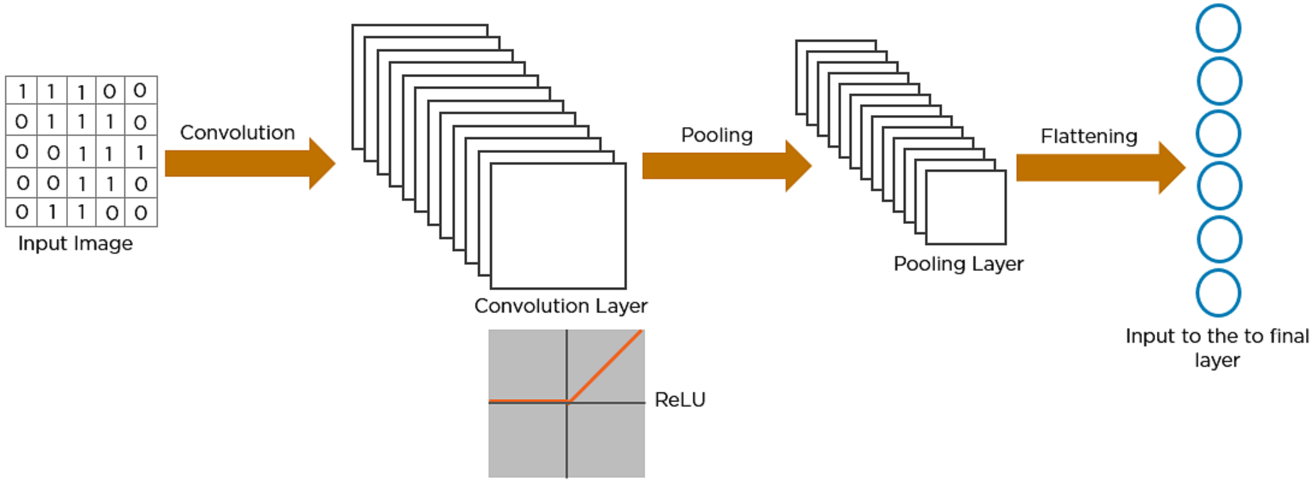


Pooling layer

- The next step in the process is called flattening. Flattening is used to convert all the resultant 2-Dimensional arrays from pooled feature maps into a single long continuous linear vector.



- The flattened matrix is fed as input to the fully connected layer to classify the image.

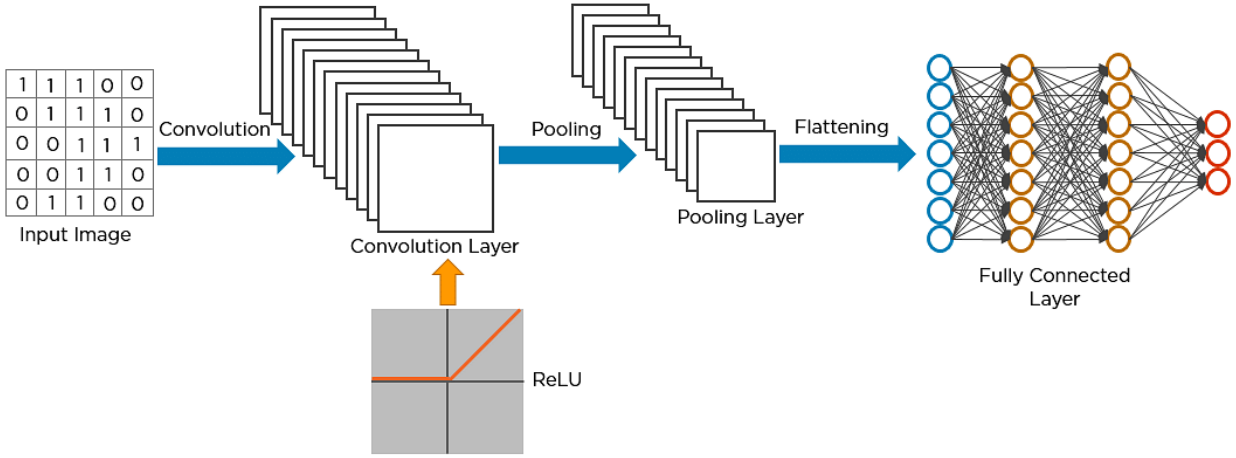


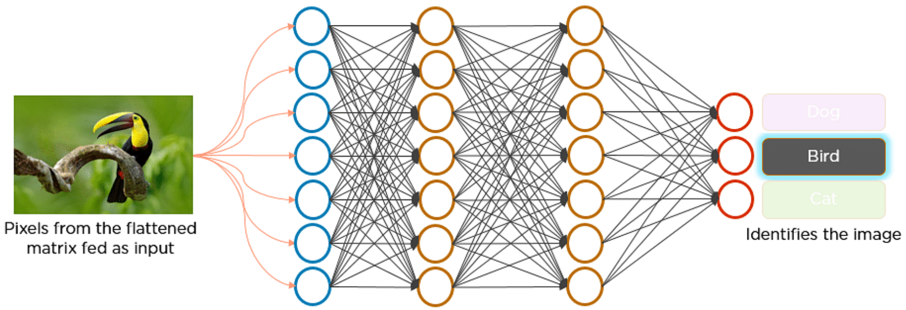
sandhan@iitk.ac.in

Fully connected layer

What is a fully connected layer?

- A fully connected layer refers to a neural network in which each neuron applies a linear transformation to the input vector through a weights matrix. As a result, all possible connections layer-to-layer are present, meaning every input of the input vector influences every output of the output vector.



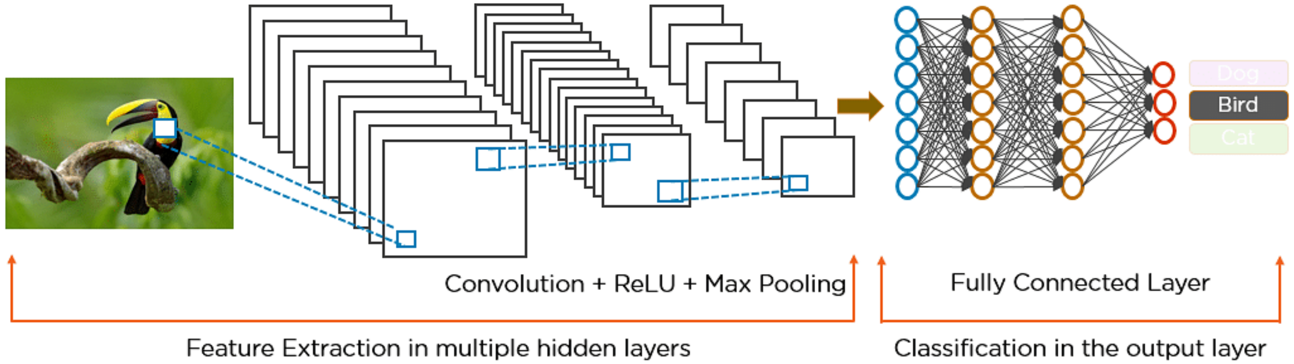


sandhan@iitk.ac.in

Convolutional Neural Network

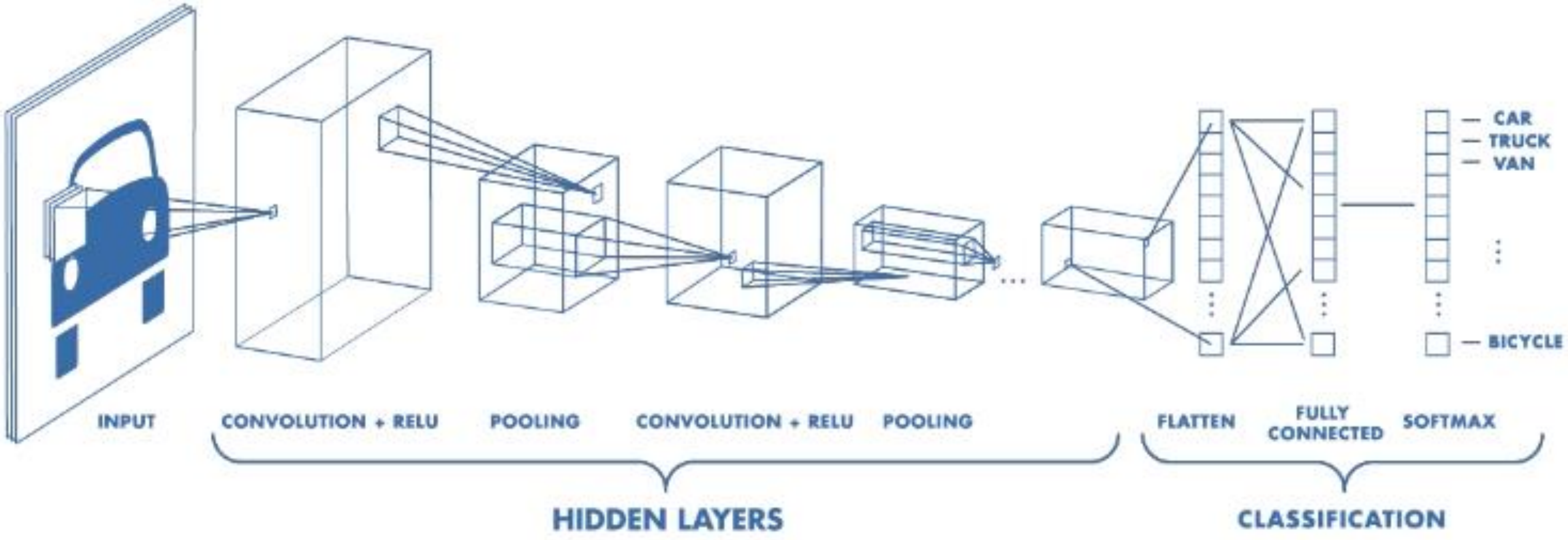
Steps to describe how exactly CNN recognizes a image:

- The pixels from the image are fed to the convolutional layer that performs the convolution operation
- It results in a convolved map
- The convolved map is applied to a ReLU function to generate a rectified feature map
- The image is processed with multiple convolutions and ReLU layers for locating the features
- Different pooling layers with various filters are used to identify specific parts of the image
- The pooled feature map is flattened and fed to a fully connected layer to get the final output



CNN Structure

- These operations are repeated over tens or hundreds of layers, with each layer learning to identify different features.



The most common applications of CNNs are:

The pixels from the image are fed to the convolutional layer that performs the convolution operation

- Healthcare. CNNs can examine thousands of visual reports to detect any anomalous conditions in patients, such as the presence of malignant cancer cells.
- Automotive. CNN technology is powering research into autonomous vehicles and self-driving cars.
- Social media. Social media platforms use CNNs to identify people in a user's photograph and help the user tag their friends.
- Retail. E-commerce platforms that incorporate visual search allow brands to recommend items that are likely to appeal to a shopper.
- Facial recognition for law enforcement. Generative adversarial networks (GANs) are used to produce new images that can then be used to train deep learning models for facial recognition
- Audio processing for virtual assistants. CNNs in virtual assistants learn and detect user-spoken keywords and process the input to guide their actions and respond to the user.

sandhan@iitk.ac.in