Image Classification

DigitalSkola
Uncover The World Of Digital Skills With Us

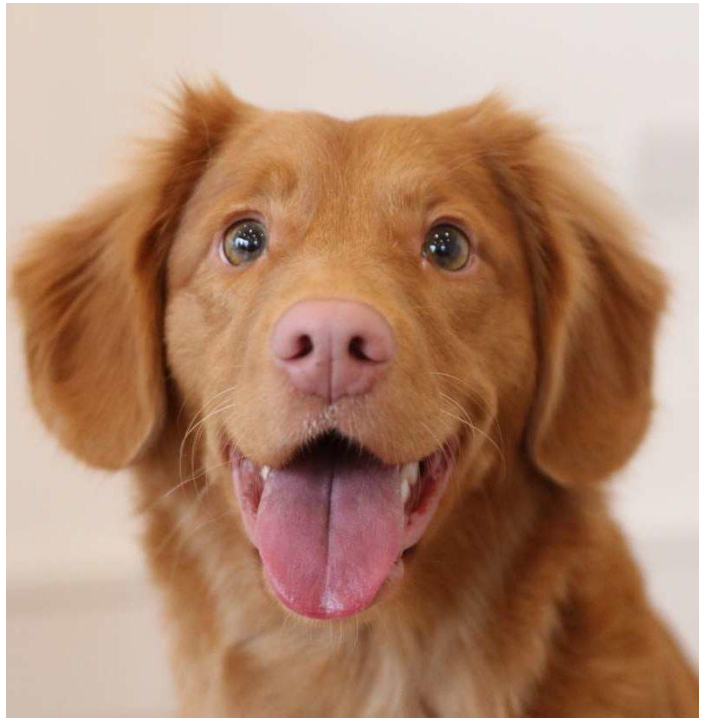
Table of Content

What will We Learn Today?

1. What is image classification?
2. What is Convolutional Neural Network?
3. Who is the inventor of CNN?
4. How CNN works?
5. Hands on CNN

DigitalSkola
Uncover The World Of Digital Skills With Us

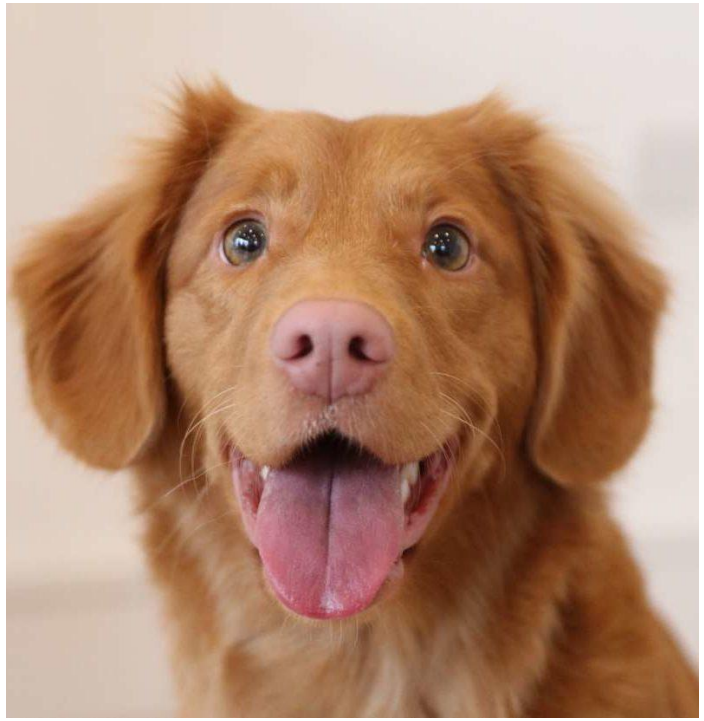
What image is this?



How about this?

DigitalSkola
Uncover The World Of Digital Skills With Us

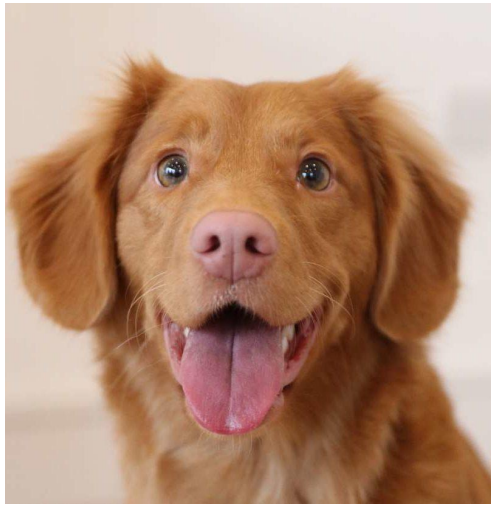
How do you know cats and dogs are different?



What makes them different?

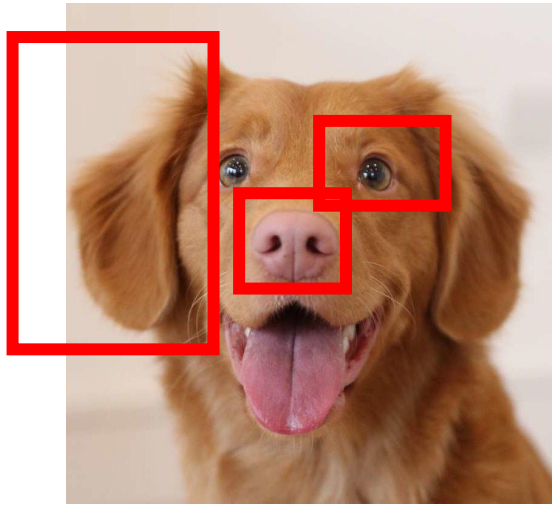
Image classification

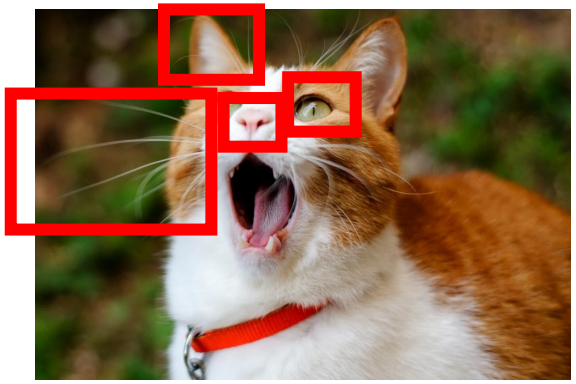
Image classification is the process of categorizing and labeling groups of pixels or vectors within an image based on specific rules. The categorization law can be devised using one or more spectral or textural characteristics.



To put simply...

Image classification is grouping / labeling image from images' features or pattern.





* Red rectanges are potential features that differ dog from cat

DigitalSkola
Uncover The World Of Digital Skills With Us

How computer extract image's feature?

DigitalSkola
Uncover The World Of Digital Skills With Us

Feature of images

image

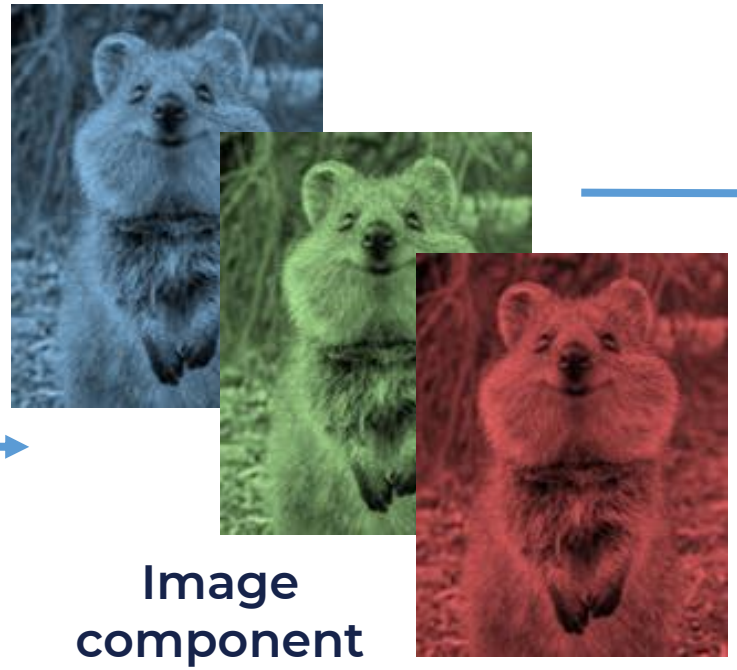
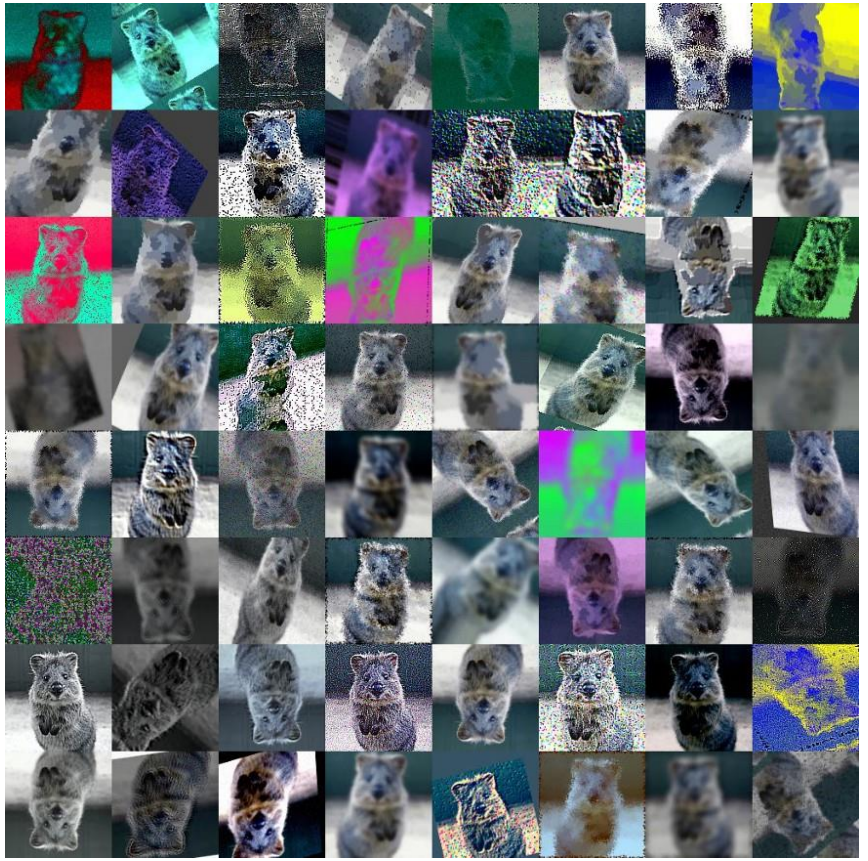


Image component (RGB)



features

Resize, filter, transform are form of feature image methods

DigitalSkola
Uncover The World Of Digital Skills With Us

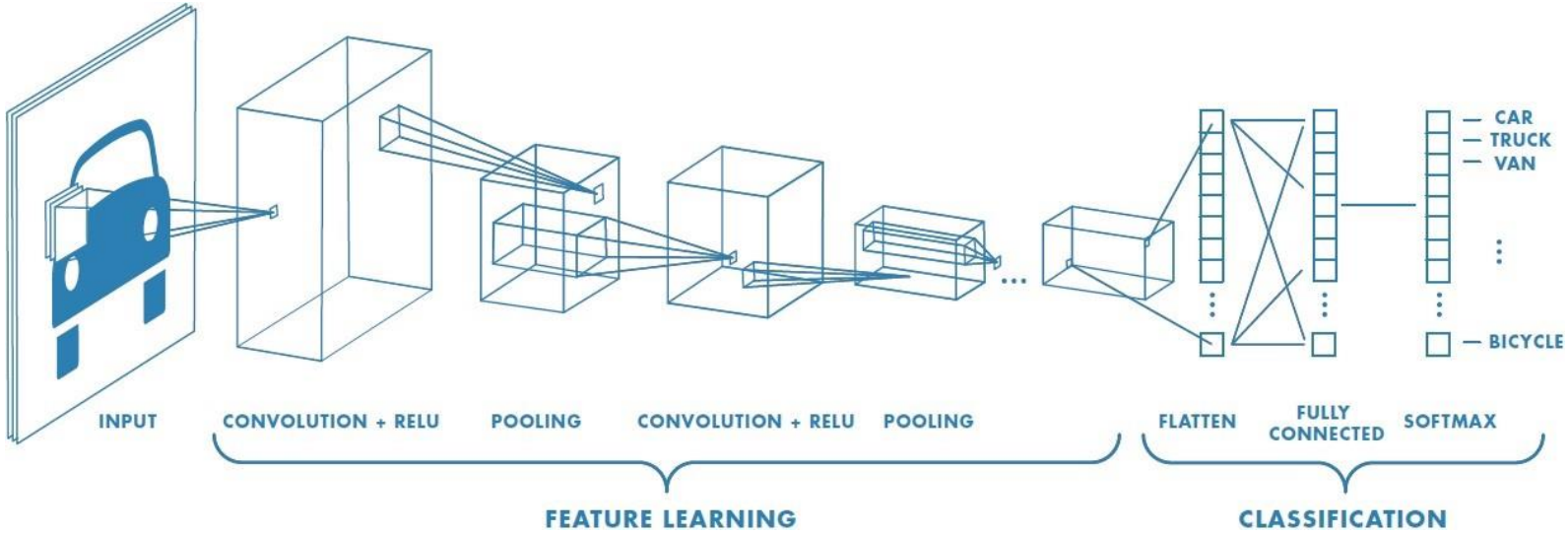
How to extract while classify image?

DigitalSkola
Uncover The World Of Digital Skills With Us

Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data.

CNNs are powerful image processing, artificial intelligence (AI) that use deep learning to perform both generative and descriptive tasks



Inventor of CNN

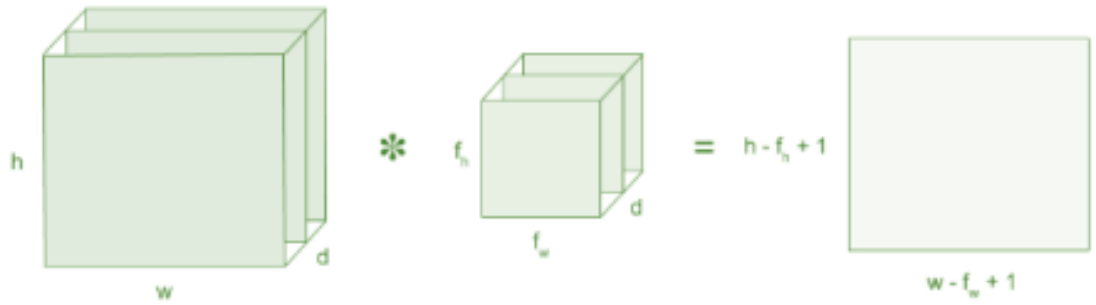
Yann LeCun

Yann André LeCun is a French computer scientist working primarily in the fields of machine learning, computer vision, mobile robotics, and computational neuroscience. He is one of Geoffrey Hinton's student (Geoffrey is godfather of Deep learning).

Convolution Layer

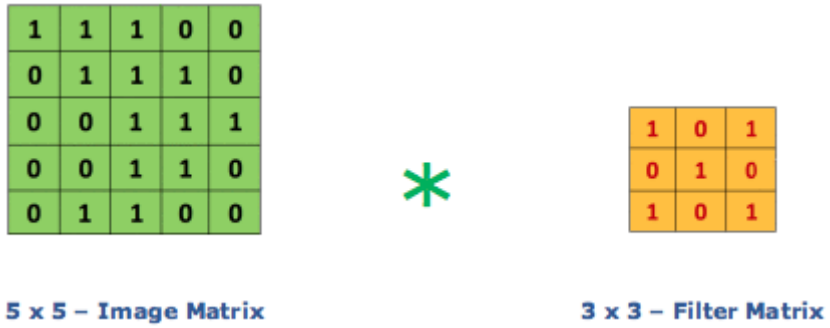
- Convolution is the first layer to extract features from an input image.
- Convolution preserves the relationship between pixels by learning image features using small squares of input data.
- It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel.

  - An image matrix (volume) of dimension **(h x w x d)**
  - A filter **(f_h x f_w x d)**
  - Outputs a volume dimension **(h - f_h + 1) x (w - f_w + 1) x 1**

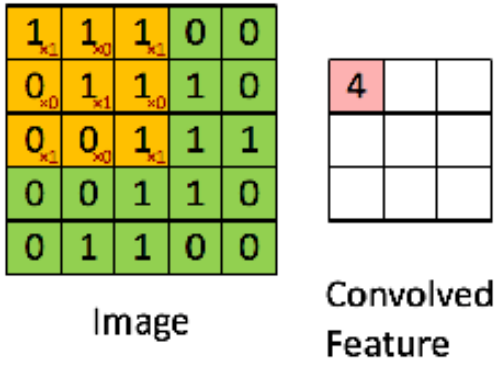


Convolution Layer (2)

- Consider a 5 x 5 whose image pixel values are 0, 1 and filter matrix 3 x 3 as shown in below



- Consider a 5 x 5 whose image pixel values are 0, 1 and filter matrix 3 x 3 as shown in below



Convolution Layer (3)

- Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters.

- Example beside shows various convolution image after applying different types of filters (Kernels).

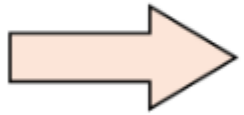
Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Strides

Stride is the number of pixels shifts over the input matrix. When the stride is 1 then we move the filters to 1 pixel at a time. When the stride is 2 then we move the filters to 2 pixels at a time and so on. The below figure shows convolution would work with a stride of 2.

1	2	3	4	5	6	7
11	12	13	14	15	16	17
21	22	23	24	25	26	27
31	32	33	34	35	36	37
41	42	43	44	45	46	47
51	52	53	54	55	56	57
61	62	63	64	65	66	67
71	72	73	74	75	76	77

Convolve with 3x3 filters filled with ones

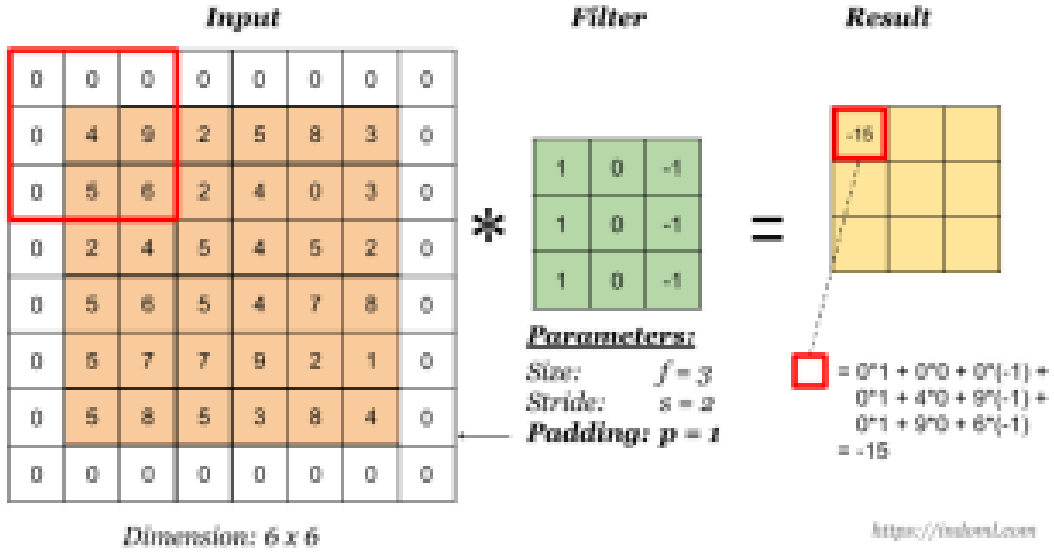


108	126	
288	306	

Padding

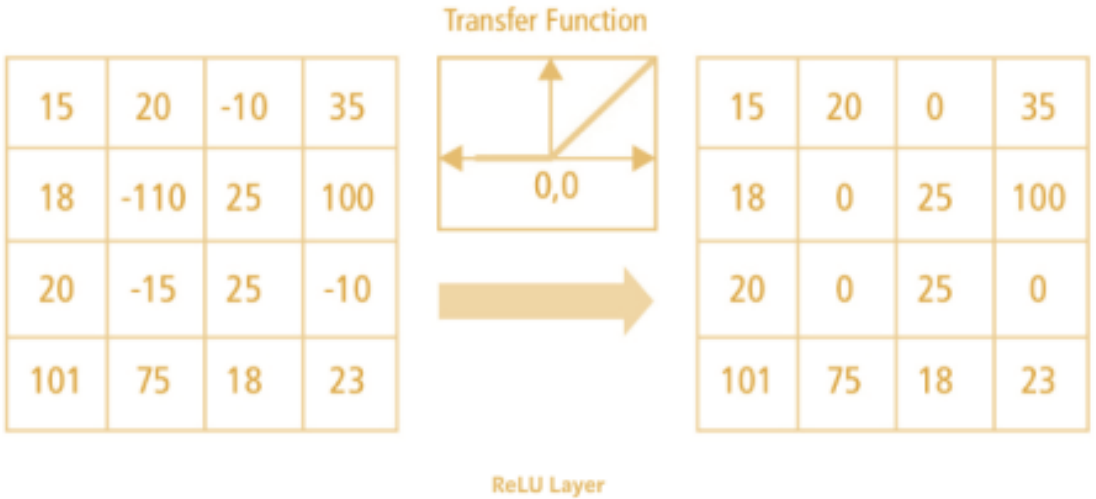
Sometimes filter does not fit perfectly fit the input image. We have two options:

- Pad the picture with zeros (zero-padding) so that it fits
- Drop the part of the image where the filter did not fit. This is called valid padding which keeps only valid part of the image.



Non Linearity (ReLU)

- ReLU stands for Rectified Linear Unit for a non-linear operation.
- The output is $f(x) = max(0,x)$.
- Why ReLU is important : ReLU's purpose is to introduce non-linearity in our ConvNet. Since, the real world data would want our ConvNet to learn would be non-negative linear values.

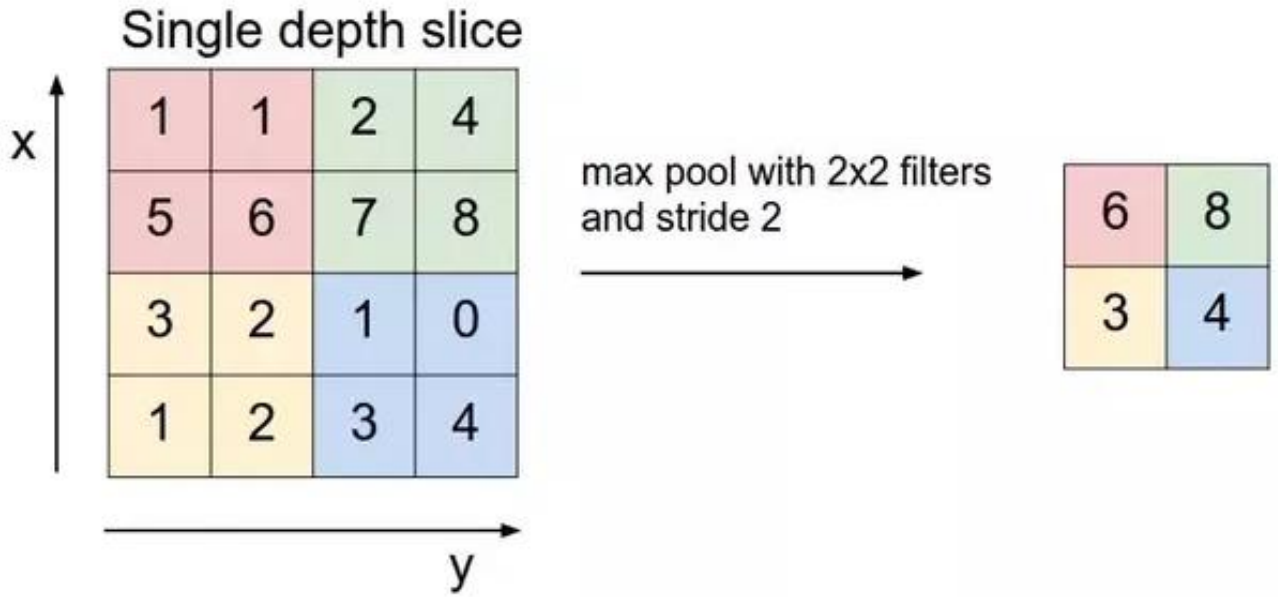


There are other non linear functions such as tanh or sigmoid that can also be used instead of ReLU. Most of the data scientists use ReLU since performance wise ReLU is better than the other two.

Pooling Layer

Pooling layers section would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or downsampling which reduces the dimensionality of each map but retains important information. Spatial pooling can be of different types:

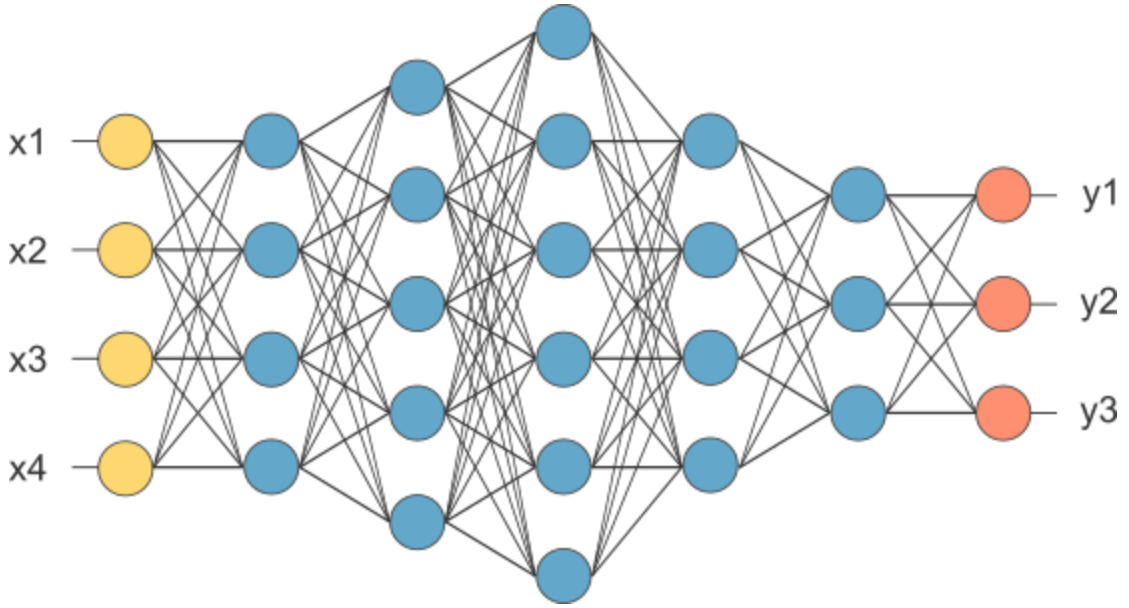
- Max Pooling
- Average Pooling
- Sum Pooling



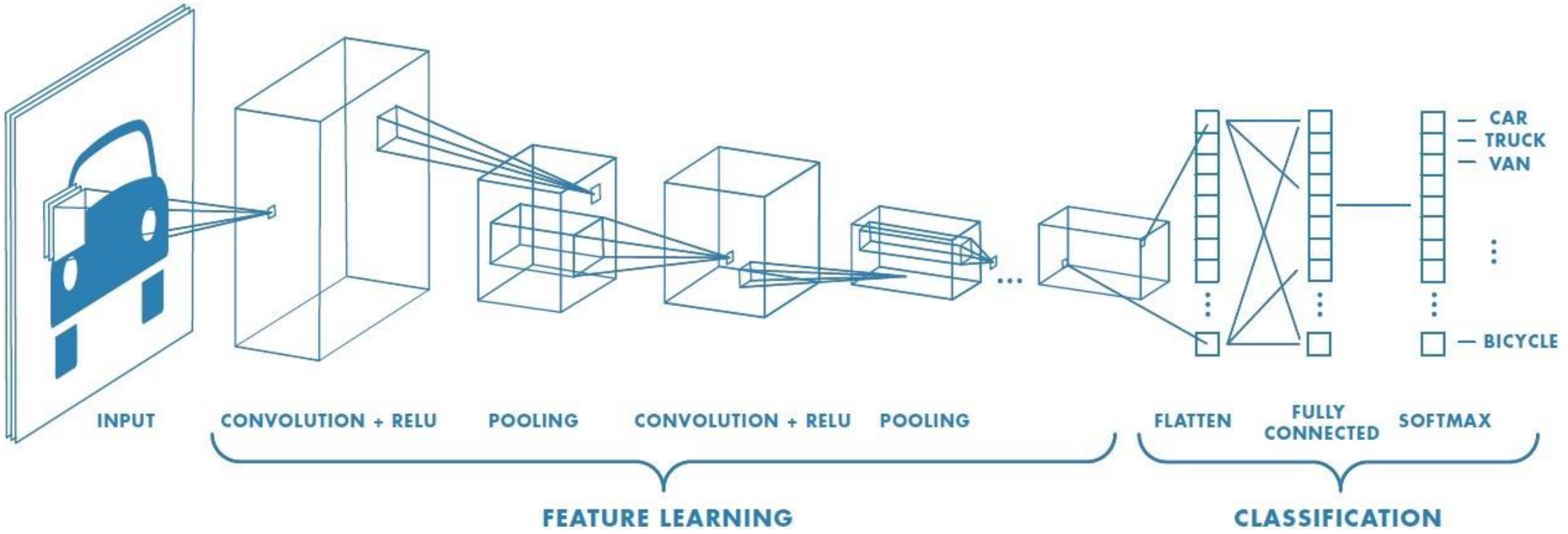
Fully Connected Layer (NN)

The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like a neural network.

With the fully connected layers, we combined these features together to create a model. Finally, we have an activation function such as softmax or sigmoid to classify the outputs as cat, dog, car, truck etc.



Final Architecture



DigitalSkola
Uncover The World Of Digital Skills With Us

Let’s have a hands-on

DigitalSkola

Uncover The World Of Digital Skills With Us

Thank YOU