



Prabhu

Mar 4, 2018 . 5 min read . [Listen](#)

Get unlimited access



# Understanding of Convolutional Neural Network (CNN) — Deep Learning

In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used.

CNN image classifications takes an input image, process it and classify it under certain categories (Eg., Dog, Cat, Tiger, Lion). Computers sees an input image as array of pixels and it depends on the image resolution. Based on the image resolution, it will see  $h \times w \times d$  ( $h$  = Height,  $w$  = Width,  $d$  = Dimension). Eg., An image of  $6 \times 6 \times 3$  array of matrix of RGB (3 refers to RGB values) and an image of  $4 \times 4 \times 1$  array of matrix of grayscale image.

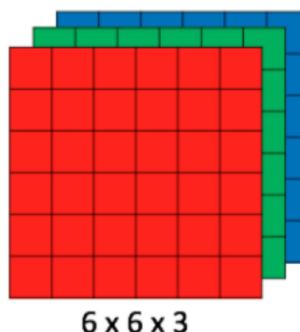


Figure 1: Array of RGB Matrix

Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernels), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. The below figure is a complete flow of CNN to process an input image and classifies the objects based on values.

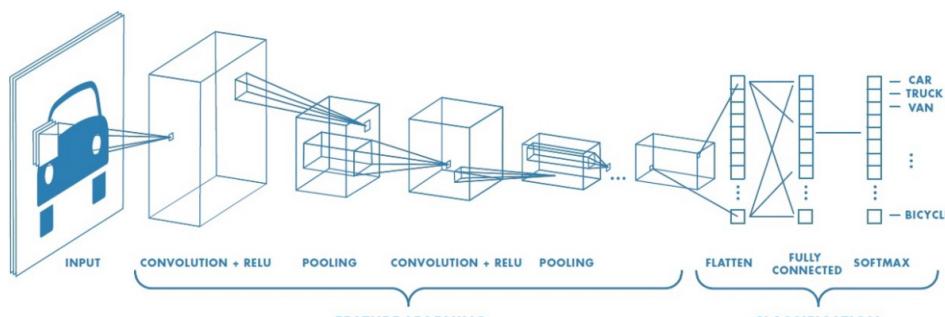


Figure 2: Neural network with many convolutional layers

 Search

Prabhu

1.5K Followers

[deeplore.io](#) — CTO & Data Scientist, Architect, Full Stack Developer — Tech Enthusiast, Learner

[Follow](#)

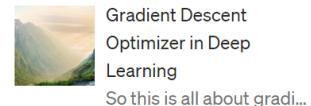
## Related



Gradient Descent:

Going downhill Pt 2

Flavors of Gradient D...

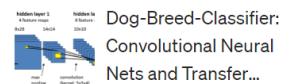


Gradient Descent

Optimizer in Deep

Learning

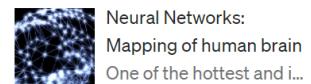
So this is all about gradi...



Dog-Breed-Classifier:

Convolutional Neural

Nets and Transfer...



Neural Networks:

Mapping of human brain

One of the hottest and i...

[Help](#) [Status](#) [Writers](#) [Blog](#) [Careers](#)  
[Privacy](#) [Terms](#) [About](#) [Knowable](#)

## 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  $(h \times w \times d)$
- A filter ( $f_h \times f_w \times d$ )
- Outputs a volume dimension  $(h - f_h + 1) \times (w - f_w + 1) \times 1$

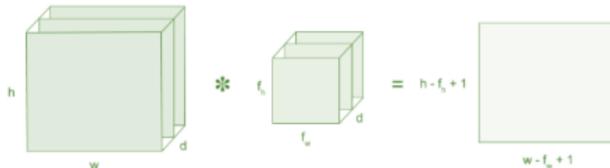


Figure 3: Image matrix multiplies kernel or filter matrix

Consider a  $5 \times 5$  whose image pixel values are 0, 1 and filter matrix  $3 \times 3$  as shown in below

Figure 4 shows a  $5 \times 5$  image matrix and a  $3 \times 3$  filter matrix. The image matrix has values: 1, 1, 1, 0, 0; 0, 1, 1, 1, 0; 0, 0, 1, 1, 1; 0, 0, 1, 1, 0; 0, 1, 1, 0, 0. The filter matrix has values: 1, 0, 1; 0, 1, 0; 1, 0, 1. A green asterisk (\*) is placed between the two matrices, indicating the operation.

Figure 4: Image matrix multiplies kernel or filter matrix

Then the convolution of  $5 \times 5$  image matrix multiplies with  $3 \times 3$  filter matrix which is called “**Feature Map**” as output shown in below

Figure 5 shows an image matrix and a convolved feature map. The image matrix has values: 1, 1, 1, 0, 0; 0, 1, 1, 1, 0; 0, 0, 1, 1, 1; 0, 0, 1, 1, 0; 0, 1, 1, 0, 0. The convolved feature map has a value '4' highlighted in red in the top-left corner of a 3x3 grid, with the rest of the grid empty.

Figure 5:  $3 \times 3$  Output matrix

Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters. The below example 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)	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Figure 7 : Some common filters

## 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.

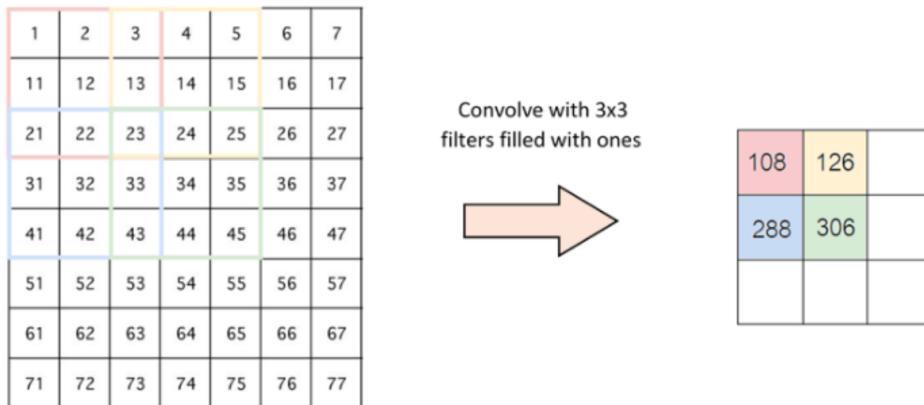


Figure 6 : Stride of 2 pixels

## 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.

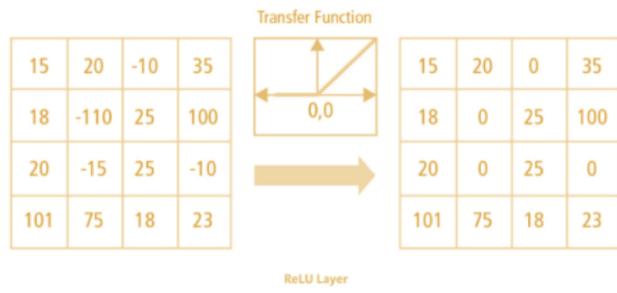


Figure 7: ReLU operation

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

Max pooling takes the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.

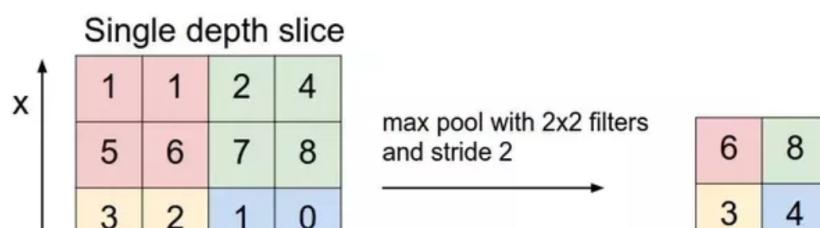




Figure 8 : Max Pooling

### Fully Connected Layer

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.

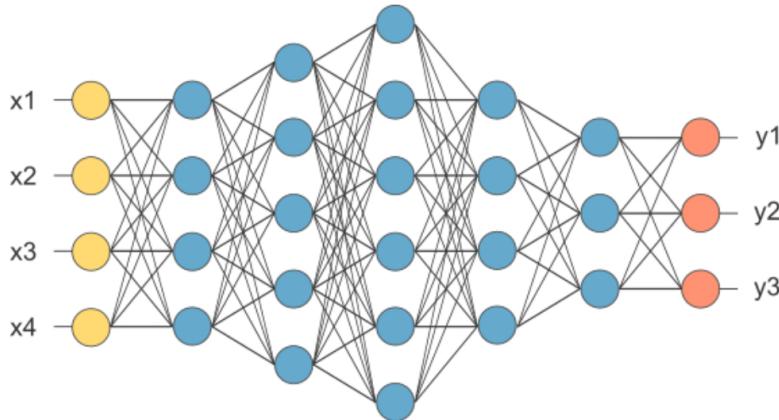


Figure 9 : After pooling layer, flattened as FC layer

In the above diagram, the feature map matrix will be converted as vector ( $x_1, x_2, x_3, \dots$ ). 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.,

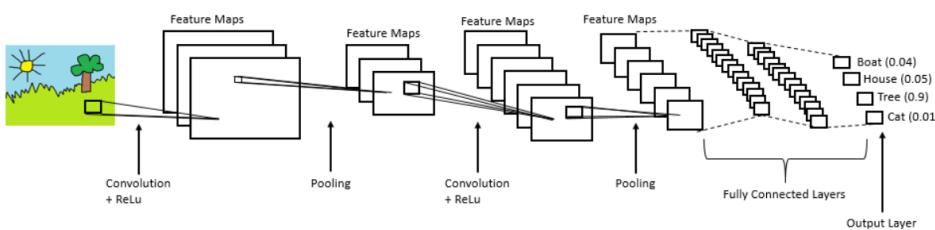


Figure 10 : Complete CNN architecture

### Summary

- Provide input image into convolution layer
- Choose parameters, apply filters with strides, padding if requires. Perform convolution on the image and apply ReLU activation to the matrix.
- Perform pooling to reduce dimensionality size
- Add as many convolutional layers until satisfied
- Flatten the output and feed into a fully connected layer (FC Layer)

- Output the class using an activation function (Logistic Regression with cost functions) and classifies images.

In the next post, I would like to talk about some popular CNN architectures such as AlexNet, VGGNet, GoogLeNet, and ResNet.

## References :

- <https://www.mathworks.com/discovery/convolutional-neural-network.html>
- <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/>
- <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
- <https://blog.datawow.io/interns-explain-cnn-8a669d053f8b>.

7.6K 39



## More from Prabhu

Follow

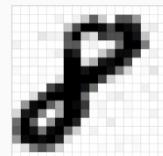


deeplore.io — CTO & Data Scientist, Architect, Full Stack Developer — Tech Enthusiast, Learner

Nov 23, 2017

## Mathematical Base for Machine Learning, Deep Learning & AI — Data Science

Mathematics is a base in our day to day life. It helps to develop our logical reasoning and analytical skills. It is essential in every aspects. ....



Machine Learning 5 min read



Share your ideas with millions of readers.

[Write on Medium](#)

Nov 16, 2017

## Data Project — 1 using Pandas

In my previous blog, I have explained the steps to be followed to process and analyse data. In this blog, let us see it in action!!! I have used a small data set from open dataset — India for this exercise. The data is all abo...

```
V file. Cleaning JSON file
a into a CSV file"""
json_data

[])

# filename
file_header=[field_header]

r[0].values[0],field_header[1].int(values)
r[1].int(values[1]),field_header[1].int(va
```

Data Science 4 min read



Nov 15, 2017

## Data Science — Fundamentals

During at the time of exploring the “How to start” data science, there were so many diversions and deviations for me before getting on to the



were so many diversions and deviations for me before getting on to the track. Various courses (Syllabus) and expertise's blogs helping me to...

SCIENCE

Machine Learning 3 min read



...

May 8, 2017

## Make Your Chores in Autopilot Mode — Automation

Automation is the key area and it has become ubiquitous. Geeks, we call as smart engineers are used to automate their regular routines and jobs. The automation implementations are pervasive in all domains. Some of...



Prabhu

1.5K Followers

[deeplore.io](#) — CTO & Data Scientist, Architect, Full Stack Developer — Tech Enthusiast, Learner

Raspberry Pi 2 min read



...

Love podcasts or audiobooks? Learn on the go with our new app.

[Try Knowable](#)

Follow



Related