Understanding of Convolutional Neural Network (CNN) — Deep Learning

Prabhu Raghav · Follow

5 min read · Mar 4, 2018

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 x w x d(h = Height, w = Width, d = Dimension). Eg., An image of 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image.

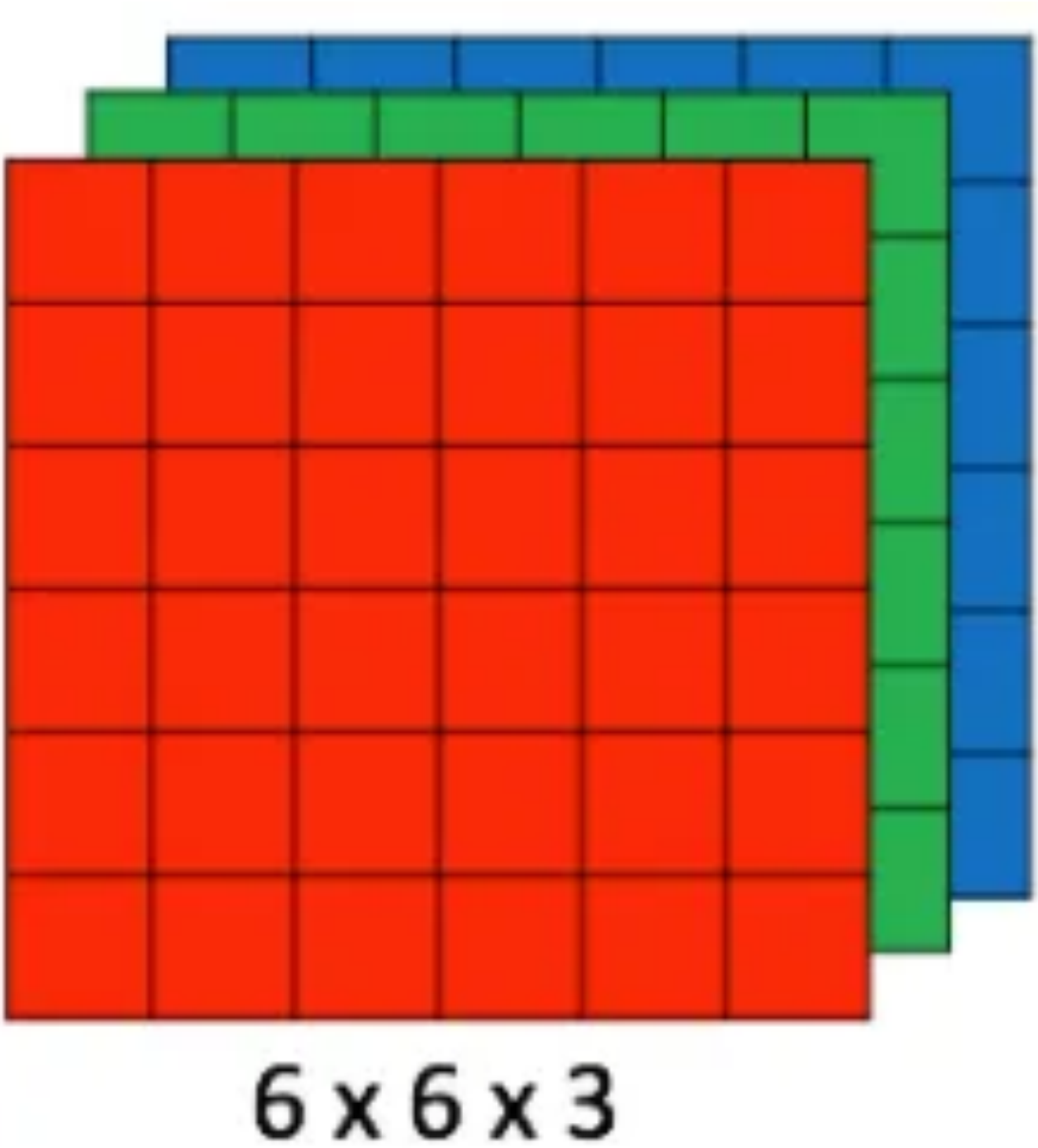


Figure 1 : Array of RGB Matrix

Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernals), 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.

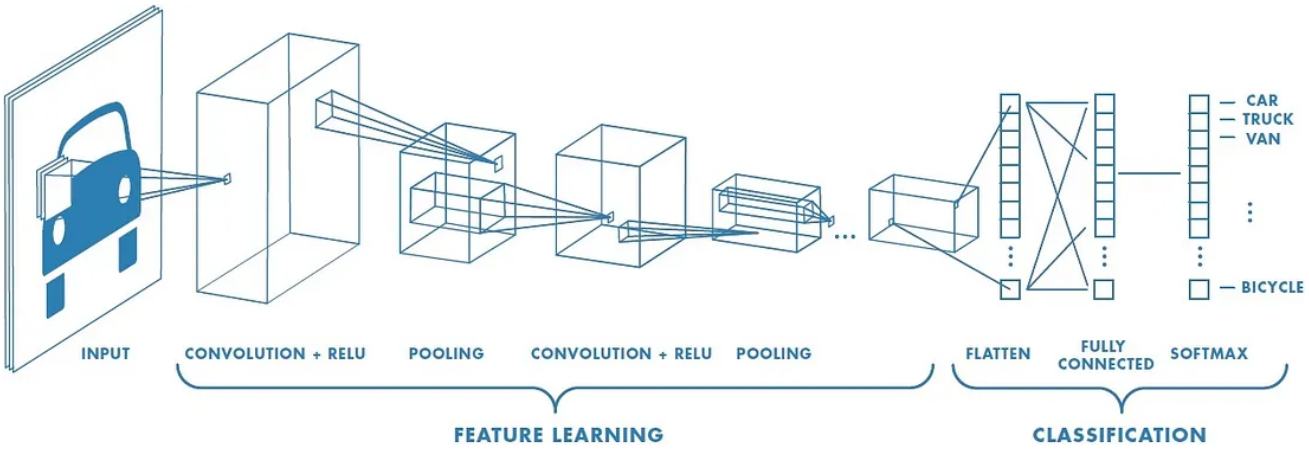


Figure 2 : Neural network with many convolutional layers

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**

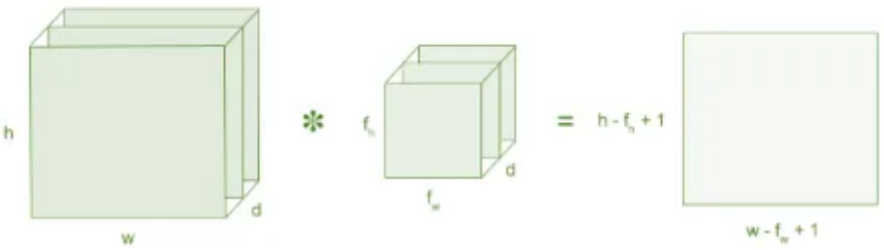


Figure 3: Image matrix multiplies kernel or filter matrix

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

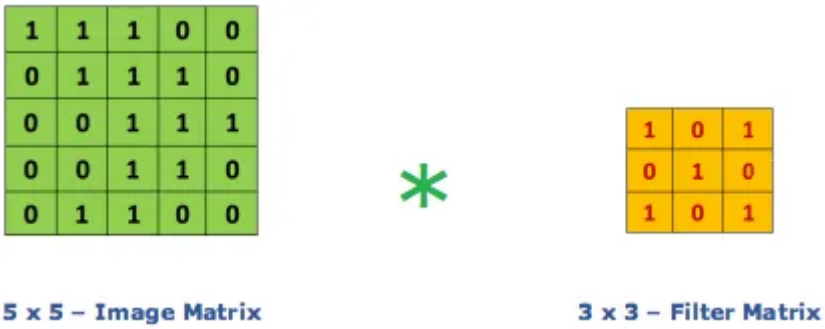


Figure 4: Image matrix multiplies kernel or filter matrix

Then the convolution of 5 x 5 image matrix multiplies with 3 x 3 filter matrix which is called "**Feature Map**" as output shown in below

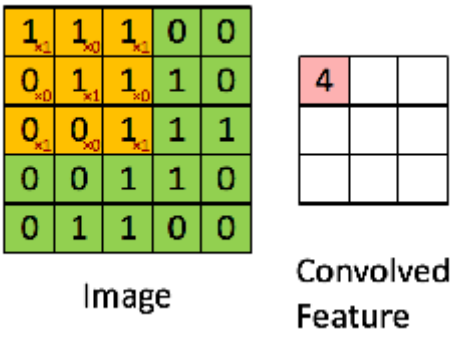


Figure 5: 3 × 3 Output matrix

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).

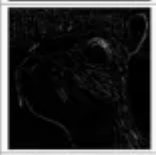
Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
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.

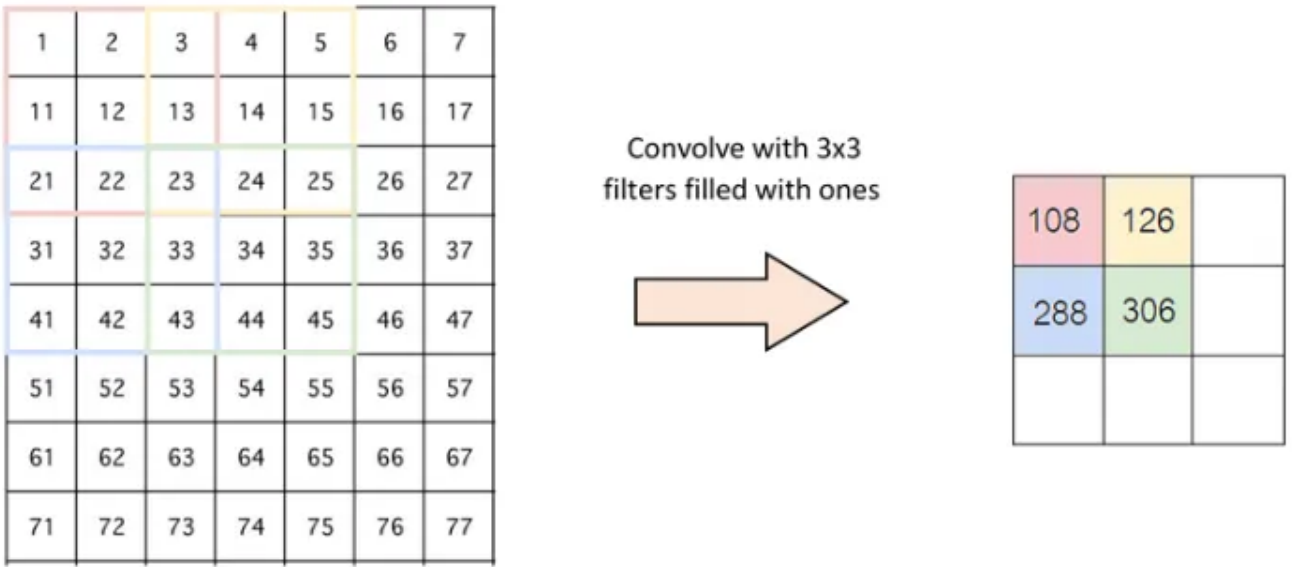


Figure 6 : Stride of 2 pixels

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.

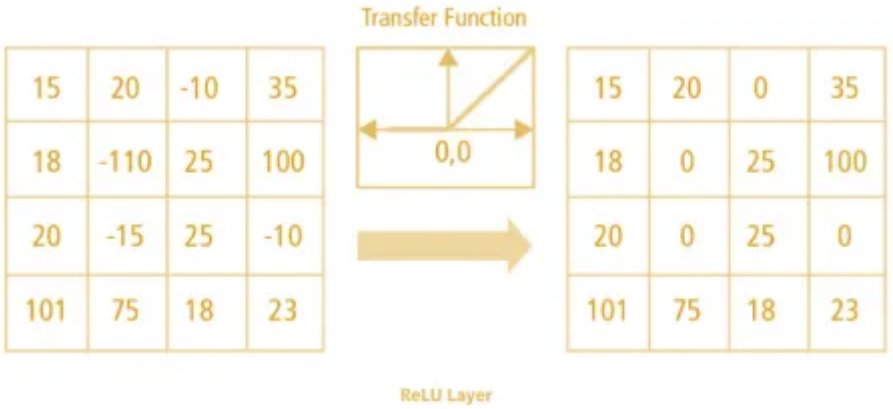


Figure 7 : ReLU operation

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.

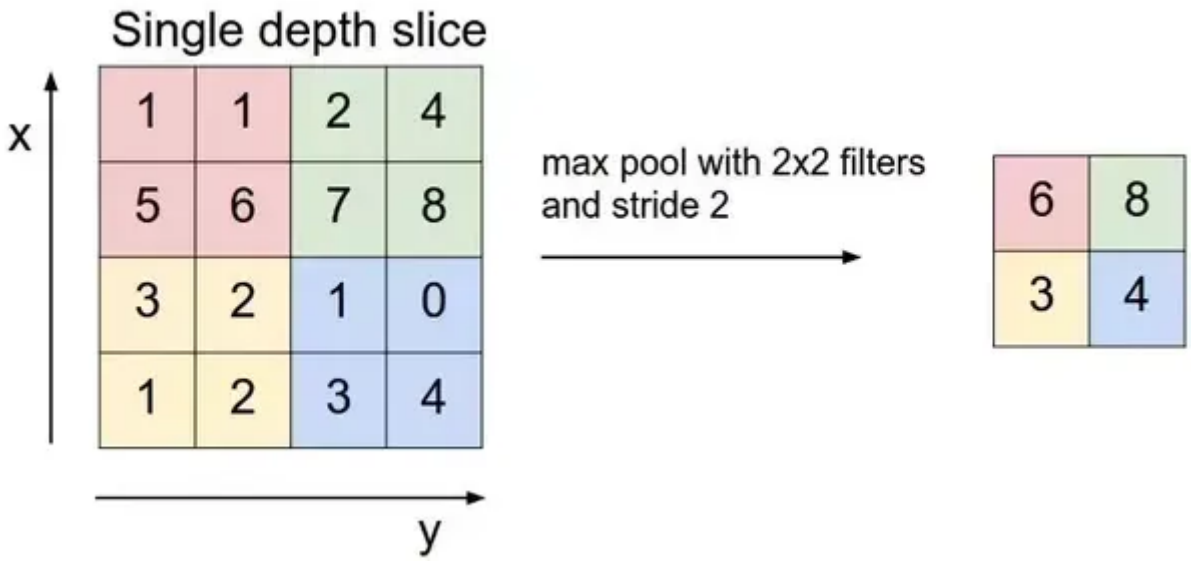


Figure 8 : Max Pooling

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.

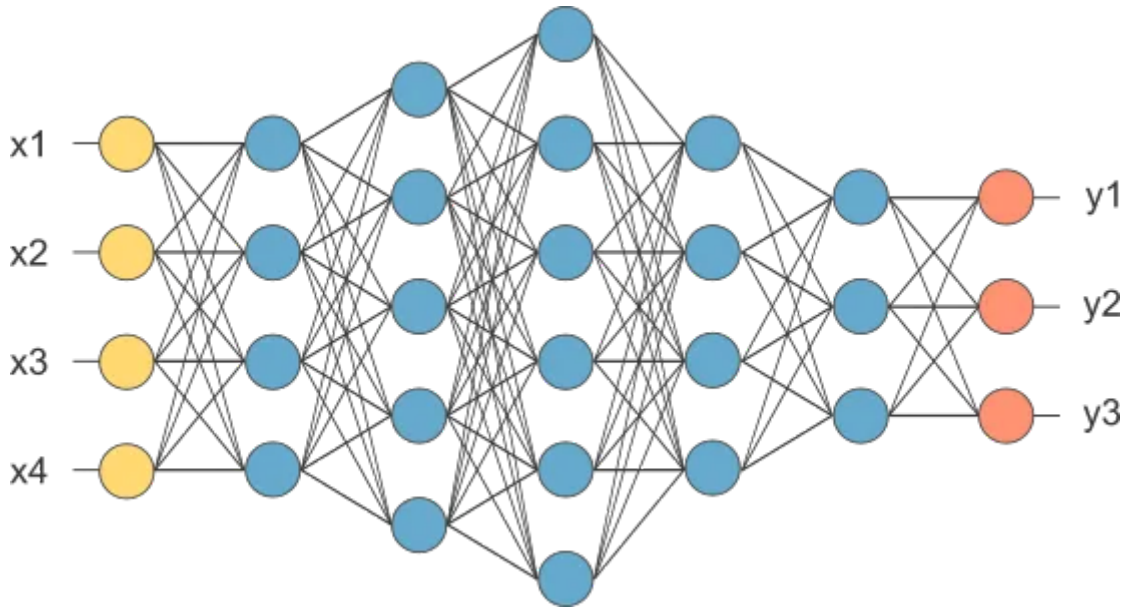


Figure 9 : After pooling layer, flattened as FC layer

In the above diagram, the feature map matrix will be converted as vector (x1, x2, x3, ...). 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.,

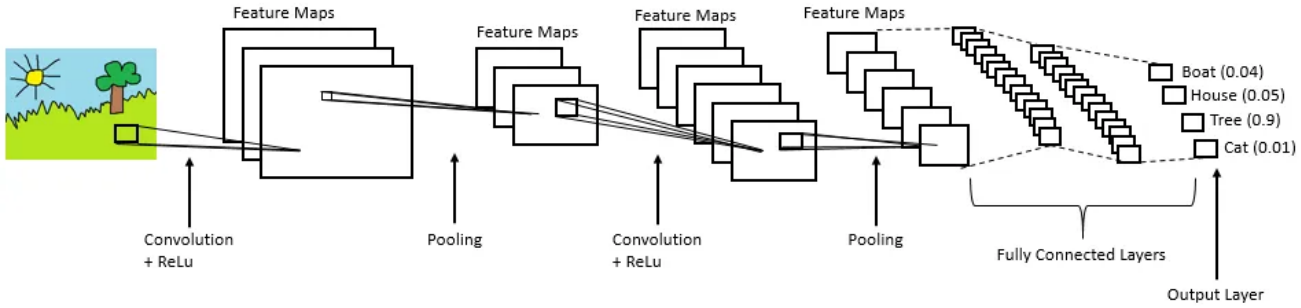


Figure 10 : Complete CNN architecture

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.

Machine Learning | Cnn | Convolution Neural Net | Image Recognition

Neural Networks

Follow

Written by Prabhu Raghav

1.9K Followers · 183 Following

SuperAgentX - https://www.superagentx.ai/ OpenSource Agentic AGI Framework — Decisionfacts.ai, Deeplore.io

Responses (43)

What are your thoughts?

Respond

Saliya Ekanayake
over 6 years ago

Really nice explanation. Thanks!

16

Reply

Hamid Haghdoost
about 6 years ago

Thank you for your clear and fluent tutorial, I enjoyed that and I learned the base of CNN in less than 10 minutes :)

14 Reply

veena tapaswi
almost 6 years ago

neat explanation... thanks

6 Reply

See all responses

More from Prabhu Raghav

Name: Crema Coffee Roasting Company
Address: San Jose, CA,
Features: Coffee Roasting

Name: Nirvana Soul
Address: 315 South 1st Street, San Jose, CA 95113
Features: Internet Access, Outdoor Seating

Name: Philz Coffee
Address: 118 Paseo De San Antonio, San Jose, CA 95112
Features: Takeaway, Coffee Shop, Safety Mask

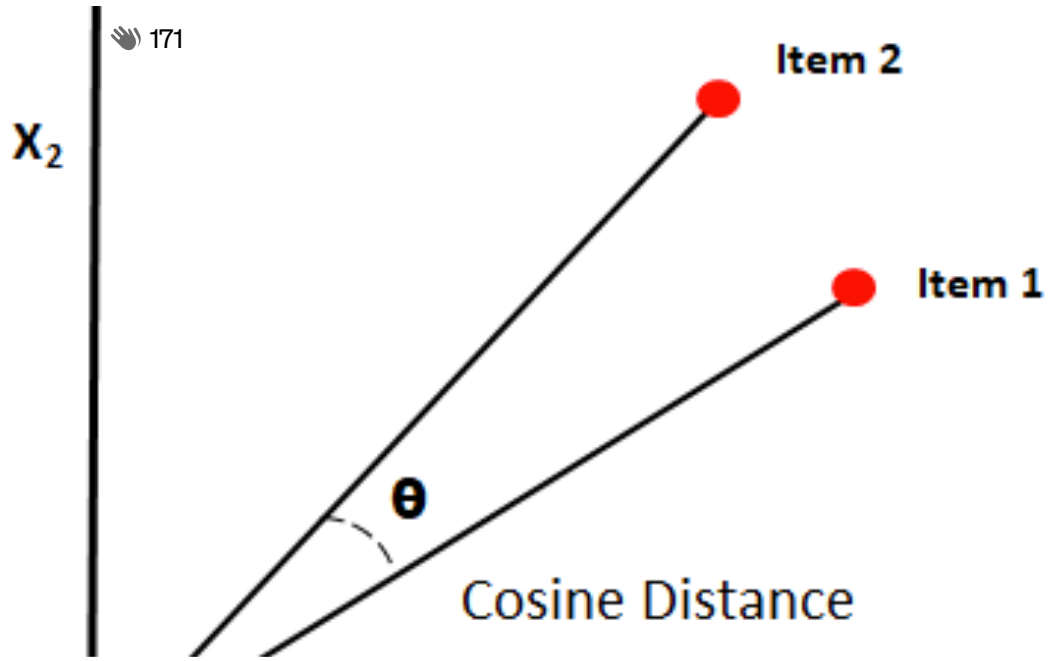
Name: Voyager Craft Coffee
Address: 87 North San Pedro Street, San Jose, CA 95110

Name: Academic Coffee
Address: 499 South 2nd Street, San Jose, CA 95113

Prabhu Raghav

Master Agentic AI: A Beginner's Step-by-Step Guide with SuperAgentX — Tutorial Series (Part 1)

Hello Everyone, Welcome to the Agent AI Tutorial Series — Part 1! 🚀



In Towards Data Science by Prabhu Raghav

Understanding NLP Word Embeddings — Text Vectorization

Processing natural language text and extract useful information from the given word, a sentence using machine learning and deep learning...

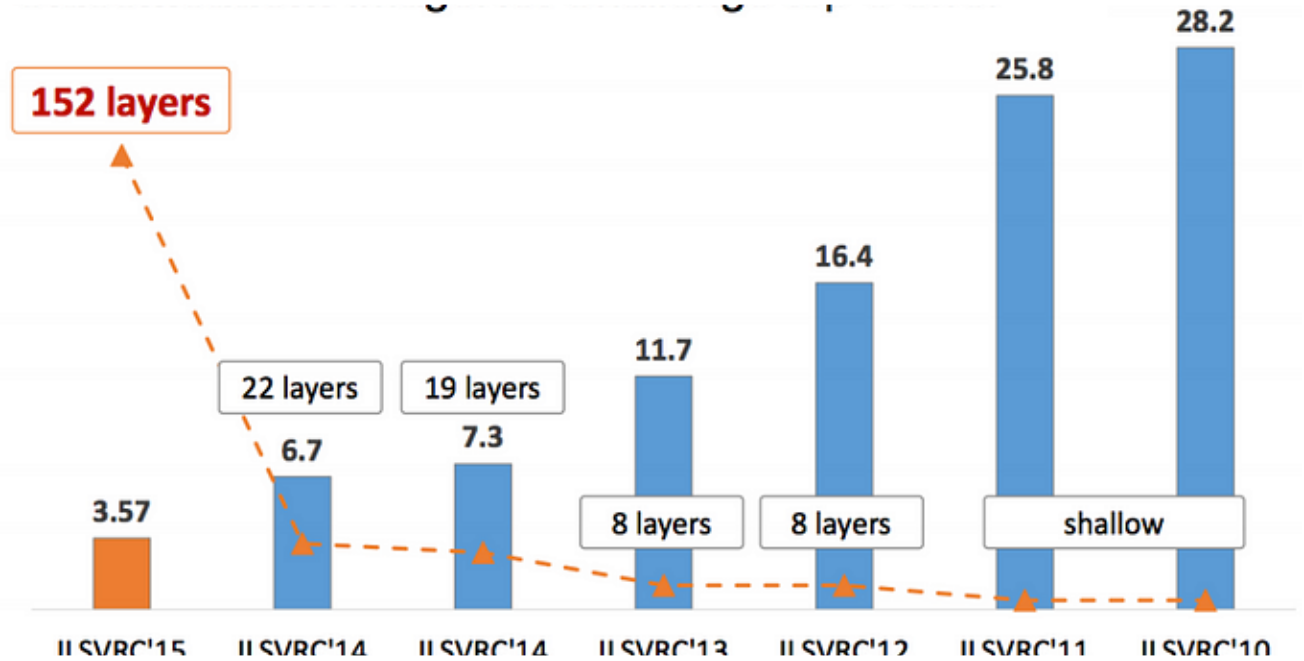
Nov 11, 2019 477 3

In Towards Data Science by Prabhu Raghav

Linear Regression Simplified - Ordinary Least Square vs Gradient Descent

What is Linear Regression? Linear regression is a statistical method of finding the relationship between independent and dependent...

May 15, 2018 2.2K 13



Prabhu Raghav

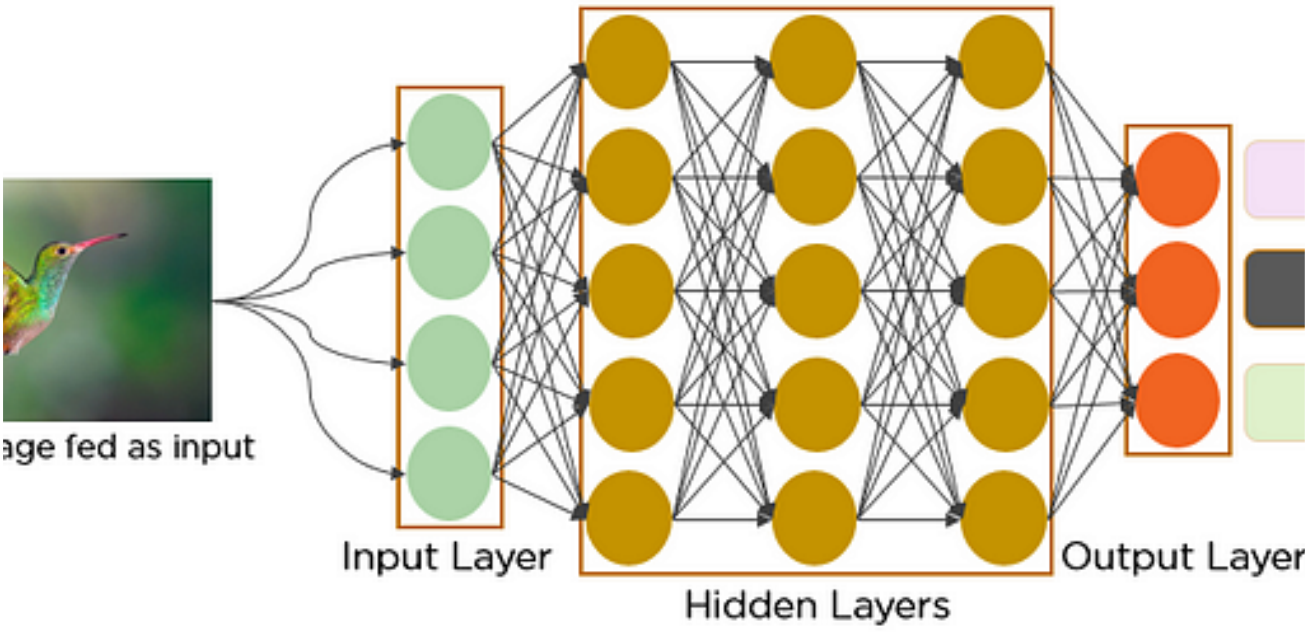
CNN Architectures — LeNet, AlexNet, VGG, GoogLeNet and ResNet

In my previous blog post, explained about my understanding of Convolution Neural Network (CNN). In this post, I am going to detailing about...

Mar 15, 2018 348 3

See all from Prabhu Raghav

Recommended from Medium

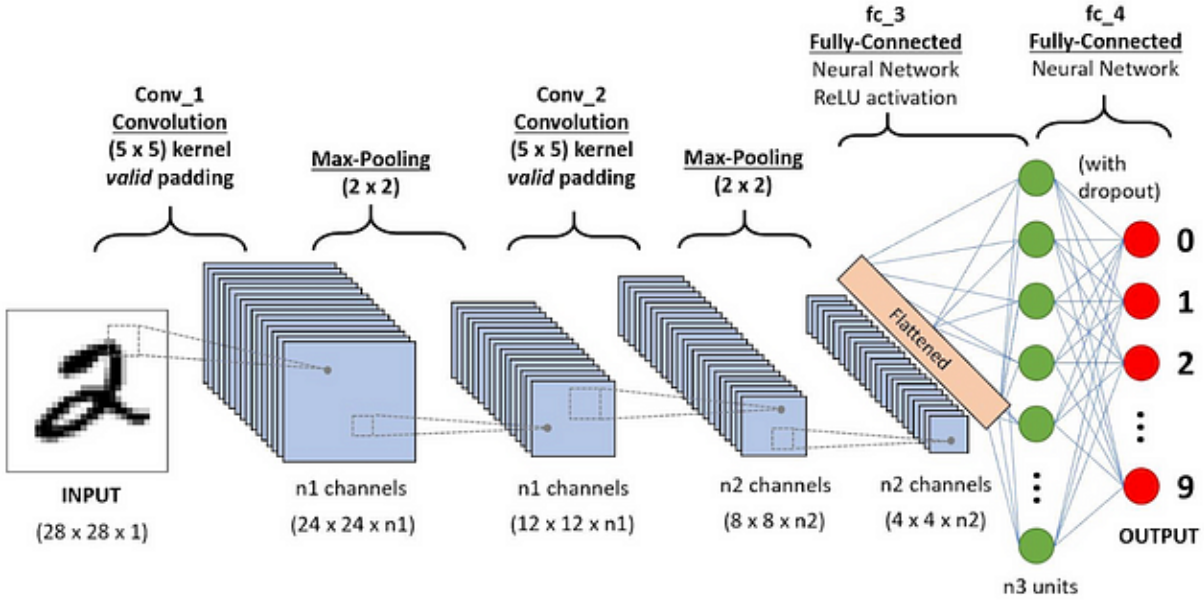


Rishabh Singh

Convolutional Neural Network (CNN) — Part 1

When classifying images, traditional neural networks struggle because each pixel is treated as an independent feature, which misses...

Nov 6 100 2



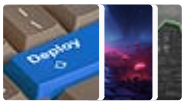
In Lumos by Luqman Zaceria

Understanding Convolutional Neural Networks (CNNs)

Hey everyone! We're going to explore one of the most influential and powerful tools in the world of deep learning: Convolutional Neural...

Aug 5 73

Lists

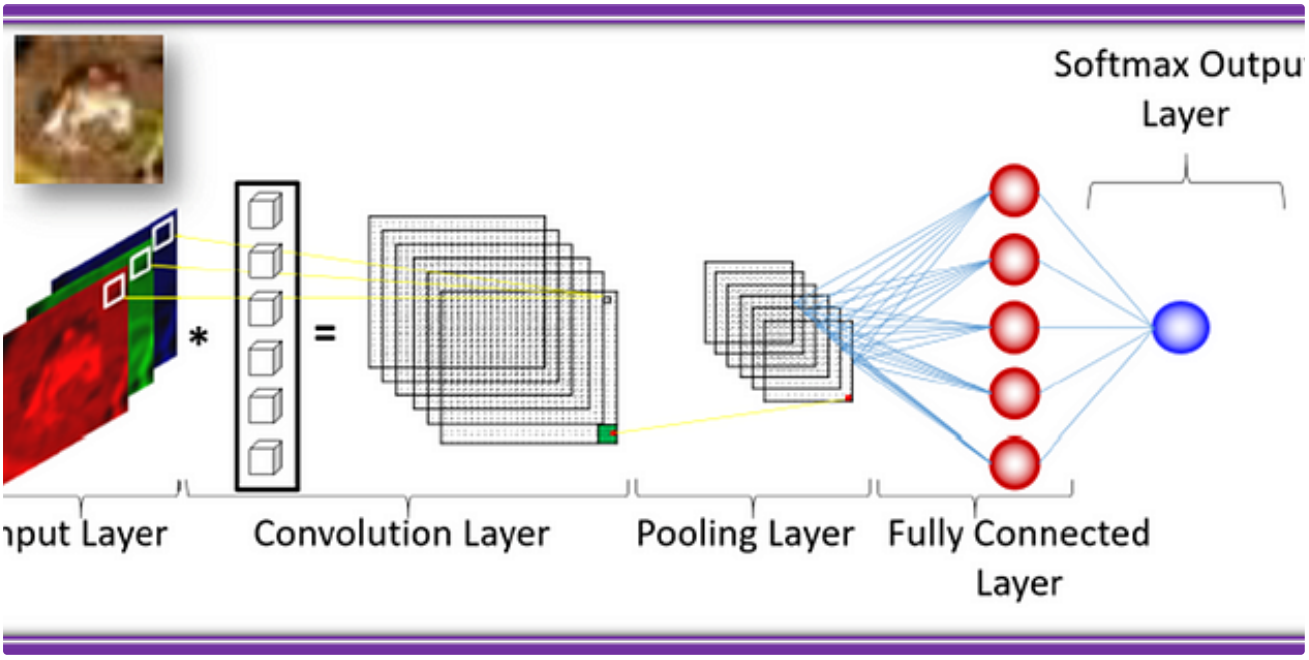


Predictive Modeling w/ Python
20 stories · 1718 saves

Practical Guides to Machine Learning
10 stories · 2090 saves

Natural Language Processing
1856 stories · 1484 saves

The New Chatbots: ChatGPT, Bard, and Beyond
12 stories · 523 saves

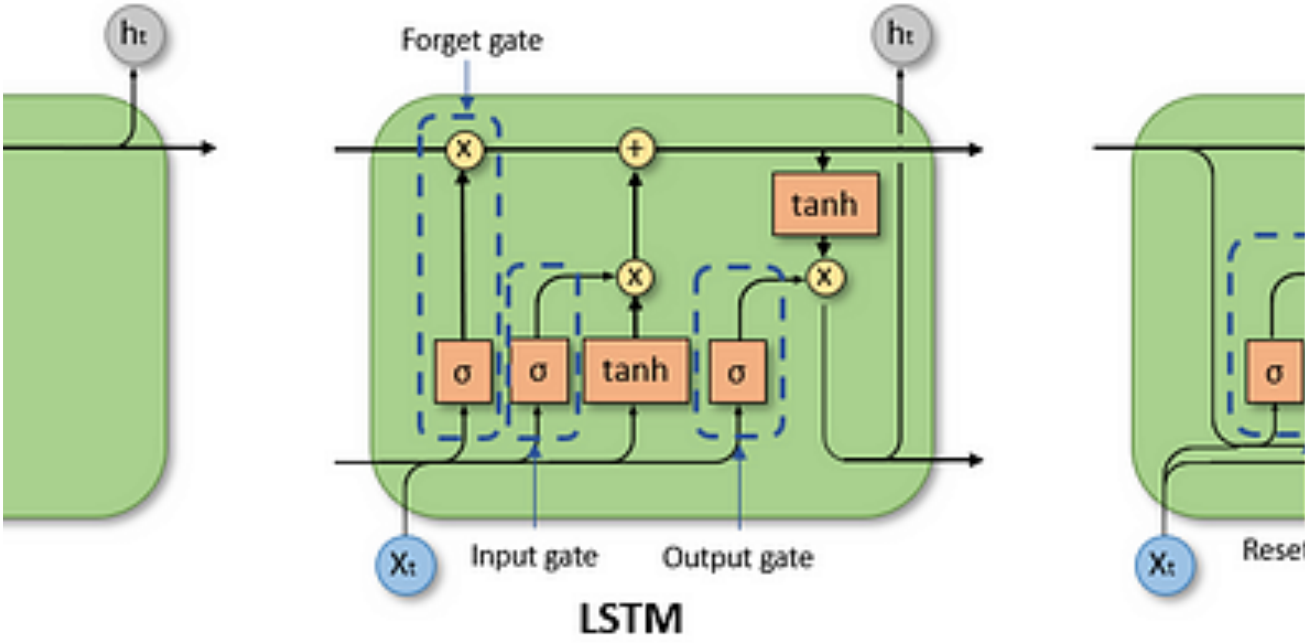


In Python in Plain English by Jyoti Dabass, Ph.D.

Friendly Introduction to Deep Learning Architectures (CNN, RNN, GAN, Transformers, Encoder-Decoder...

This blog aims to provide a friendly introduction to deep learning architectures involving Convolutional Neural Networks (CNN), Recurrent...

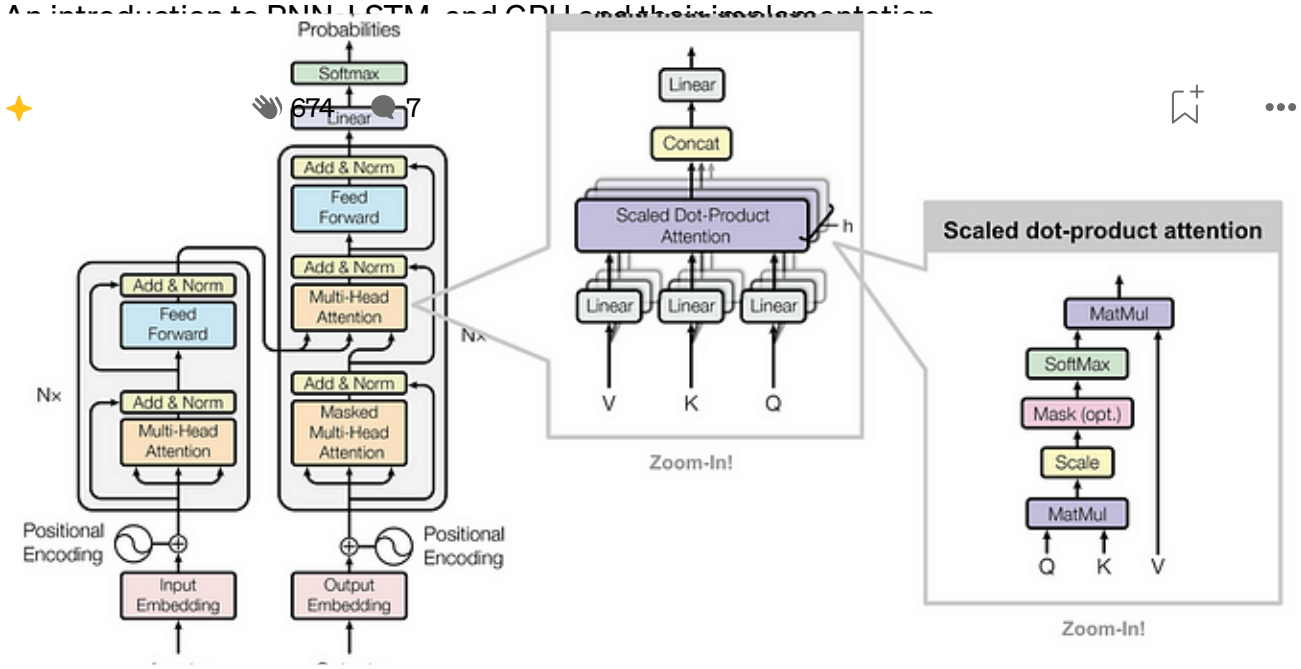
Apr 1 1.3K 13



In Towards Data Science by Jonte Dancker

A Brief Introduction to Recurrent Neural Networks

An introduction to RNN, LSTM, and GRU, and their implementation



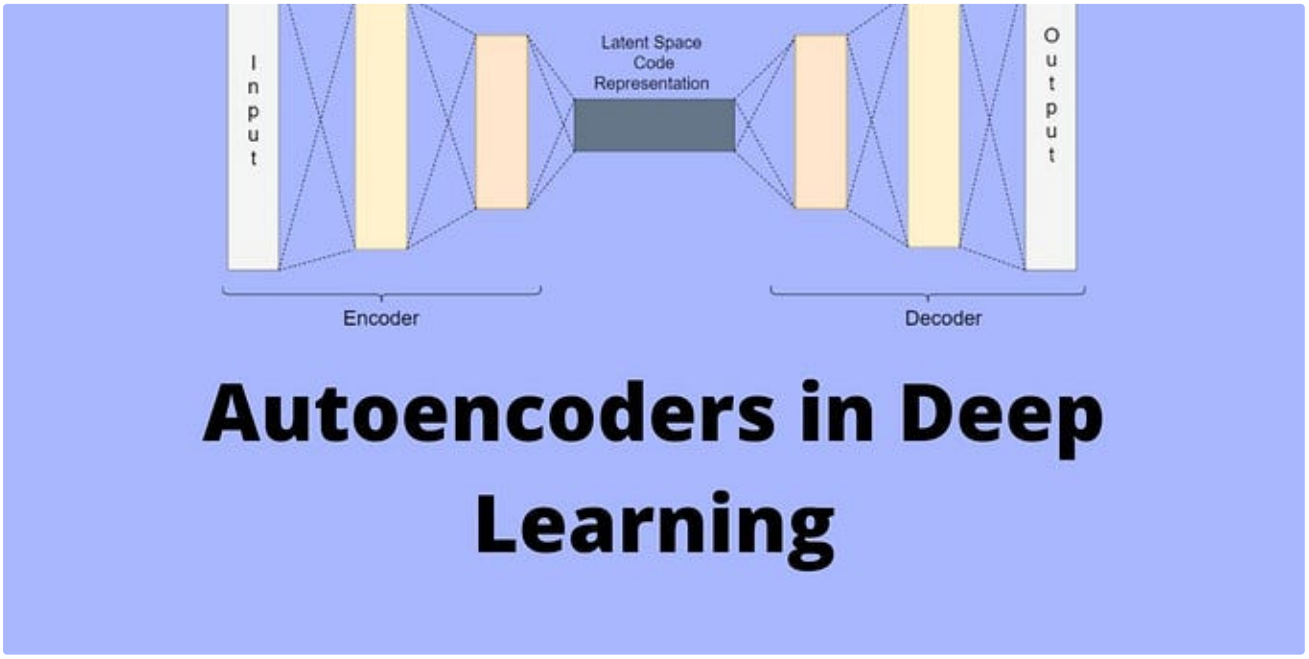
674 7

LM Po

Self-Attention and Transformer Network Architecture

The introduction of Transformer models in 2017 marked a significant turning point in the fields of Natural Language Processing (NLP) and...

Oct 18 34 1



Jo Wang

Autoencoder

An autoencoder is a type of neural network used primarily for unsupervised learning.

Oct 21 4