



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Coimbatore – 35.



DEPARTMENT OF BIOMEDICAL ENGINEERING

UNIT 3

VGG 16, Inception and ResNet

Some common architectures:-

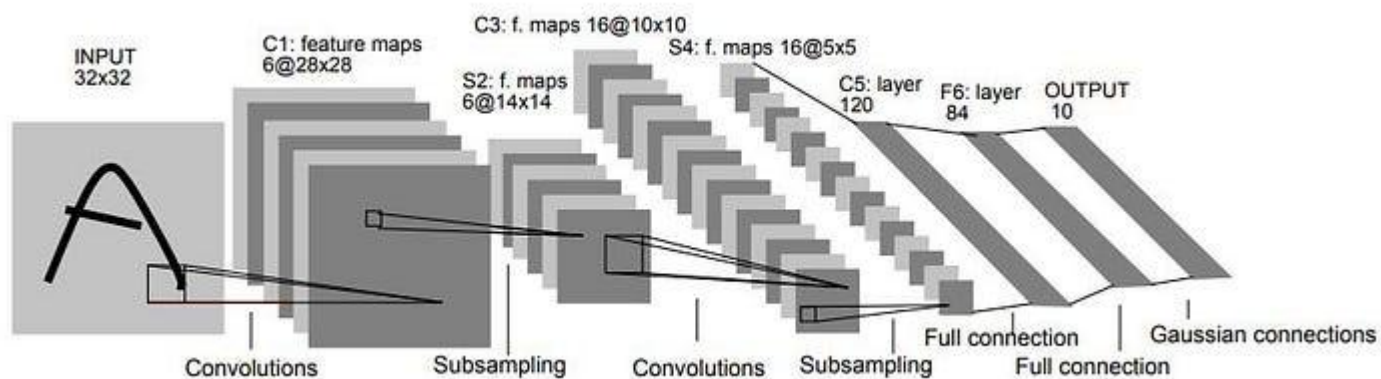
1. VGG 16
2. Inception (GoogLeNet)
3. ResNet

In this part we will talk about the first 3 architectures, which can be regarded as the classic ConvNet architectures.

LeNet-5

LeNet-5 is a convolutional neural network proposed by [Yann LeCun](#) in 1989. It was one of the earliest ConvNet architecture and has had a dominant influence over the coming architectures.

Structure



LeNet-5 Structure (source:- yann.lecun.com)

LeNet-5 comprises 2 sets of convolution and max-pooling layers, followed by a flattening convolutional layer, then two fully-connected layers and finally a softmax classifier.

Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	32x32	-	-	-
1	Convolution	6	28x28	5x5	1	tanh
2	Average Pooling	6	14x14	2x2	2	tanh
3	Convolution	16	10x10	5x5	1	tanh
4	Average Pooling	16	5x5	2x2	2	tanh
5	Convolution	120	1x1	5x5	1	tanh
6	FC	-	84	-	-	tanh
Output	FC	-	10	-	-	softmax

Summary Table for LeNet-5

Parameters

~60,000 parameters

Applications

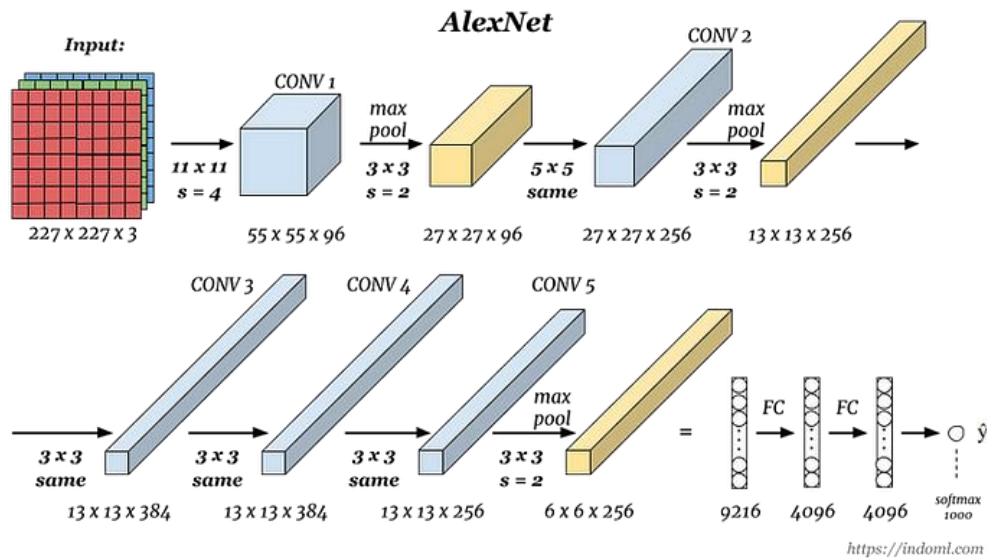
The driving application of this architecture was to recognize simple handwritten numerical digits and was prominently used for recognition of handwritten zip codes in US Postal Services. It can produce >98% accuracy on the **MNIST** dataset after only 20 epochs.

***MNIST**- [Modified National Institute of Standards and Technology](#) database is an extensive database of handwritten digits that is commonly used for training various image processing systems. The database is also widely used for training and testing in machine learning.*

AlexNet

AlexNet is a convolutional neural network, designed by [Alex Krizhevsky](#) in 2012. AlexNet is considered the most influential neural net architecture and as of 2020 has been cited over 65,000 times.

Structure



AlexNet Structure (source:- indoml.com)

AlexNet comprises of 5 Convolutional layers with 3 Fully Connected layers followed with a softmax layer. From these 5 Convolutional layers, 3 layers have a Max Pooling Layer.

Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	227x227x3	-	-	-
1	Convolution	96	55 x 55 x 96	11x11	4	relu
	Max Pooling	96	27 x 27 x 96	3x3	2	relu
2	Convolution	256	27 x 27 x 256	5x5	1	relu
	Max Pooling	256	13 x 13 x 256	3x3	2	relu
3	Convolution	384	13 x 13 x 384	3x3	1	relu
4	Convolution	384	13 x 13 x 384	3x3	1	relu
5	Convolution	256	13 x 13 x 256	3x3	1	relu
	Max Pooling	256	6 x 6 x 256	3x3	2	relu
6	FC	-	9216	-	-	relu
7	FC	-	4096	-	-	relu
8	FC	-	4096	-	-	relu
Output	FC	-	1000	-	-	Softmax

Summary Table for AlexNet

Parameters

~62 million parameters

Application

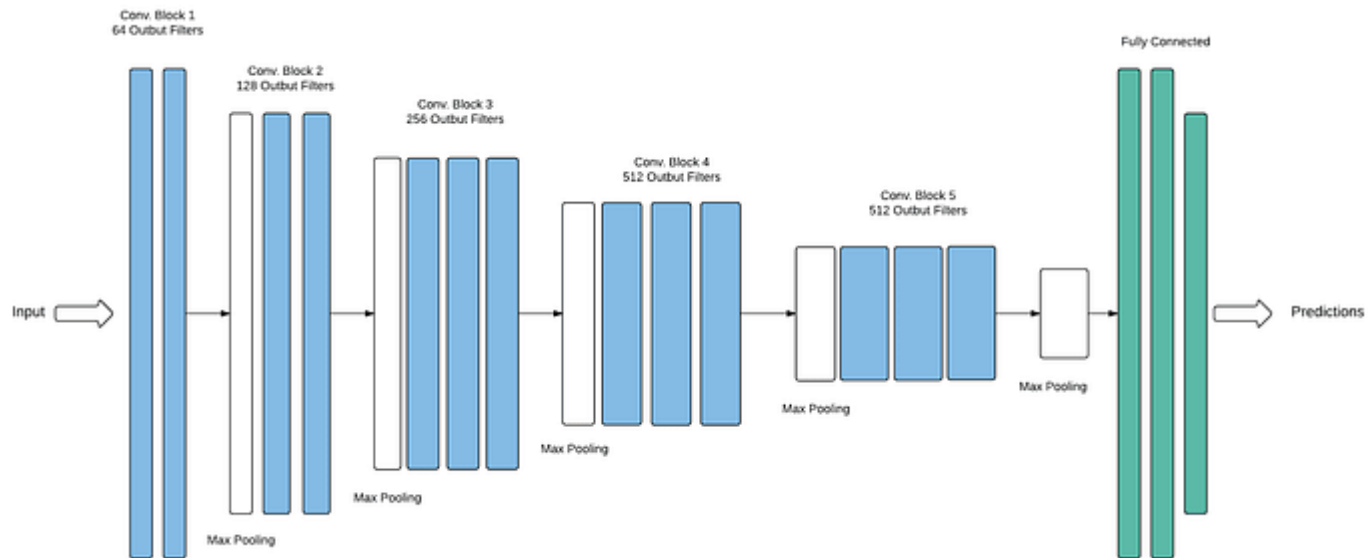
AlexNet can perform Image Classification on images and is famous for winning the 2012 **ImageNet LSVRC-2012** competition by a large margin (15.3% VS 26.2% (second place) error rates).

ImageNet- [ImageNet](#) is formally a project aimed at (manually) labeling and categorizing images into almost 22,000 separate object categories for computer vision research. Models are trained on ~1.2 million training images with another 50,000 images for validation and 100,000 images for testing.

VGG-16

VGG Net is a convolutional neural network invented by **Simonyan** and **Zisserman** from **Visual Geometry Group (VGG)** at **University of Oxford** in 2014.

Structure



VGG-16 Structure (source:- codesofinterest.com)

The VGG-16 comprises 13 Convolutional Layers with a Max-pooling Layer every 2–3 layers followed with 3 Fully Connected Layers and finally a softmax layer. What puts this ConvNet above others is continuous use of same convolutions with a fixed filter and stride and always using the same padding and max-pool layer of 2x2 filter of stride 2.

	Layer	Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	224 x 224 x 3	-	-	-
1	2 X Convolution	64	224 x 224 x 64	3x3	1	relu
	Max Pooling	64	112 x 112 x 64	3x3	2	relu
3	2 X Convolution	128	112 x 112 x 128	3x3	1	relu
	Max Pooling	128	56 x 56 x 128	3x3	2	relu
5	2 X Convolution	256	56 x 56 x 256	3x3	1	relu
	Max Pooling	256	28 x 28 x 256	3x3	2	relu
7	3 X Convolution	512	28 x 28 x 512	3x3	1	relu
	Max Pooling	512	14 x 14 x 512	3x3	2	relu
10	3 X Convolution	512	14 x 14 x 512	3x3	1	relu
	Max Pooling	512	7 x 7 x 512	3x3	2	relu
13	FC	-	25088	-	-	relu
14	FC	-	4096	-	-	relu
15	FC	-	4096	-	-	relu
Output	FC	-	1000	-	-	Softmax

Summary Table for VGG-16

There is a variant of this network called VGG-19 which follows the same pattern but with 16 Convolutional Layers and 3 Fully Connected Layers.

Parameters

~138 Million parameters

Application

VGG-16 is considered to be one of the most excellent vision model architecture till date, having scored a tremendous **92.7%** accuracy on **ImageNet** in 2014.

References

<https://www.jeremyjordan.me/convnet-architectures/>

https://en.wikipedia.org/wiki/Convolutional_neural_network#History

<https://engmrk.com/lenet-5-a-classic-cnn-architecture/>

<https://en.wikipedia.org/wiki/AlexNet>

<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

<https://engmrk.com/alexnet-implementation-using-keras/>

https://www.robots.ox.ac.uk/~vgg/research/very_deep/