

# Introduction to The Architecture of Alexnet

[ADVANCED](#)[COMPUTER VISION](#)[DEEP LEARNING](#)[VIDEOS](#)

## Objective

- When we talk about the Pre-trained model in the Computer Vision domain, Alexnet comes out as a leading architecture.
- Let's understand the architecture of Alexnet as proposed by its authors.

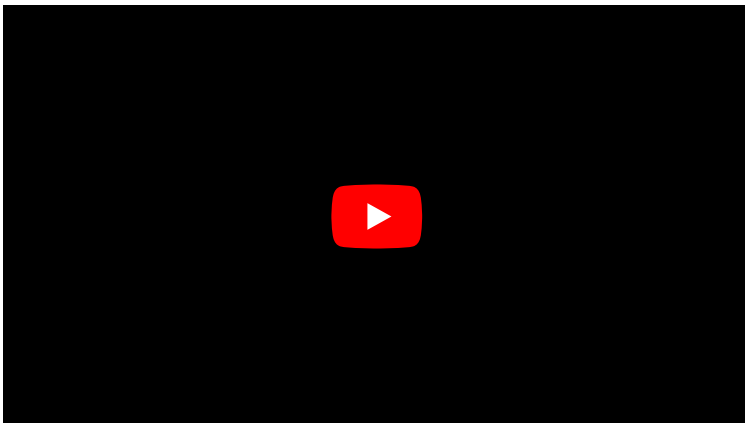
## Introduction

Alexnet won the Imagenet large-scale visual recognition challenge in 2012. The model was proposed in 2012 in the research paper named [Imagenet Classification with Deep Convolution Neural Network](#) by Alex Krizhevsky and his colleagues.

*In this model, the depth of the network was increased in comparison to Lenet-5. In case you want to know more about Lenet-5, I will recommend you to check the following article-*

- [The Architecture of Lenet-5](#)

*Note: If you are more interested in learning concepts in an Audio-Visual format, We have this entire article explained in the video below. If not, you may continue reading.*



The Alexnet has eight layers with learnable parameters. The model consists of five layers with a combination of max pooling followed by 3 fully connected layers and they use Relu activation in each of these layers except the output layer.

They found out that using the relu as an activation function accelerated the speed of the training process by almost six times. They also used the dropout layers, that prevented their model from overfitting. Further, the model is trained on the Imagenet dataset. The Imagenet dataset has almost 14 million images across a thousand classes.

Let's see the architectural details in this article.

## Alexnet Architecture

One thing to note here, since Alexnet is a deep architecture, the authors introduced padding to prevent the size of the feature maps from reducing drastically. The input to this model is the images of size 227X227X3.

Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map
Input	-	-	-	-	227 x 227 x 3
Conv 1	96	11 x 11	4	-	55 x 55 x 96
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96
Conv 2	256	5 x 5	1	2	27 x 27 x 256
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256
Conv 3	384	3 x 3	1	1	13 x 13 x 384
Conv 4	384	3 x 3	1	1	13 x 13 x 384
Conv 5	256	3 x 3	1	1	13 x 13 x 256
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256

## Convolution and Maxpooling Layers

Then we apply the first convolution layer with 96 filters of size 11X11 with stride 4. The activation function used in this layer is relu. The output feature map is 55X55X96.

In case, you are unaware of how to calculate the output size of a convolution layer

output= ((Input-filter size)/ stride)+1

Also, the number of filters becomes the channel in the output feature map.

Next, we have the first Maxpooling layer, of size 3X3 and stride 2. Then we get the resulting feature map with the size 27X27X96.

After this, we apply the second convolution operation. This time the filter size is reduced to 5X5 and we have 256 such filters. The stride is 1 and padding 2. The activation function used is again relu. Now the output size we get is 27X27X256.

Again we applied a max-pooling layer of size 3X3 with stride 2. The resulting feature map is of shape 13X13X256.

Now we apply the third convolution operation with 384 filters of size 3X3 stride 1 and also padding 1. Again the activation function used is relu. The output feature map is of shape 13X13X384.

Then we have the fourth convolution operation with 384 filters of size 3X3. The stride along with the padding is 1. On top of that activation function used is relu. Now the output size remains unchanged i.e 13X13X384.

After this, we have the final convolution layer of size 3X3 with 256 such filters. The stride and padding are set to one also the activation function is relu. The resulting feature map is of shape 13X13X256.

So if you look at the architecture till now, the number of filters is increasing as we are going deeper. Hence it is extracting more features as we move deeper into the architecture. Also, the filter size is reducing, which means the initial filter was larger and as we go ahead the filter size is decreasing, resulting in a decrease in the feature map shape.

Next, we apply the third max-pooling layer of size 3X3 and stride 2. Resulting in the feature map of the shape 6X6X256.

## Fully Connected and Dropout Layers



After this, we have our first dropout layer. The drop-out rate is set to be 0.5.

Then we have the first fully connected layer with a relu activation function. The size of the output is 4096. Next comes another dropout layer with the dropout rate fixed at 0.5.

This followed by a second fully connected layer with 4096 neurons and relu activation.

Finally, we have the last fully connected layer or output layer with 1000 neurons as we have 10000 classes in the data set. The activation function used at this layer is Softmax.

This is the architecture of the Alexnet model. It has a total of 62.3 million learnable parameters.

## Frequently Asked Questions

### Q1. What is the use of AlexNet?

A. AlexNet is a pioneering convolutional neural network (CNN) used primarily for image recognition and classification tasks. It won the ImageNet Large Scale Visual Recognition Challenge in 2012, marking a breakthrough in deep learning. AlexNet's architecture, with its innovative use of convolutional layers and rectified linear units (ReLU), laid the foundation for modern deep learning models, advancing computer vision and pattern recognition applications.

### Q2. Why AlexNet is better than CNN?

A. AlexNet is a specific type of CNN, which is a kind of neural network particularly good at understanding images. When AlexNet was introduced, it showed impressive results in recognizing objects in pictures. It became popular because it was deeper (had more layers) and used some smart tricks to improve accuracy. So, AlexNet is not better than CNN; it is a type of CNN that was influential in making CNNs popular for image-related tasks.

## End Notes

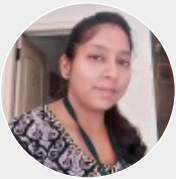
To quickly summarize the architecture that we have seen in this article.

- It has 8 layers with learnable parameters.
- The input to the Model is RGB images.
- It has 5 convolution layers with a combination of max-pooling layers.
- Then it has 3 fully connected layers.
- The activation function used in all layers is Relu.
- It used two Dropout layers.
- The activation function used in the output layer is Softmax.
- The total number of parameters in this architecture is 62.3 million.

So this was all about Alexnet. If you have any questions let me know in the comments below!

---

Article Url - <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-the-architecture-of-alexnet/>



**[Shipra Saxena](#)**