# AlexNet - ImageNet Classification with Deep Convolutional Neural Networks

Ankan Sarkar (210050013)
Atharva Tambat (210070014)

September 11, 2022

## 1 Architecture

The CNN has 8 learned layers, the distribution being:

- 5 Convolution layers, some being a composition of convolution, pooling and normalization
- 3 Fully Connected layers

To provide a deeper insight into the model structure, the following image has been included.

| Layer | Output / Neurons (FCN) | Feature maps | Kernel size | Stride | Padding |
|---|---|---|---|---|---|
| Input Layer | 227 x 227 x 3 [Input] | - | - | - | - |
| Convolutional #1 | 55 x 55 x 96 | 96 | 11 x 11 | 4 | 0 |
| Max Pool #1 | 27 x 27 x 96 | 96 | 3 x 3 | 2 | - |
| Norm #1 | 27 x 27 x 96 | 96 | - | - | - |
| Convolutional #2 | 27 x 27 x 256 | 256 | 5 x 5 | 1 | 2 |
| Max Pool #2 | 13 x 13 x 256 | 256 | 3 x 3 | 2 | - |
| Norm #2 | 13 x 13 x 256 | 256 | - | - | - |
| Convolutional #3 | 13 x 13 x 384 | 384 | 3 x 3 | 1 | 1 |
| Convolutional #4 | 13 x 13 x 384 | 384 | 3 x 3 | 1 | 1 |
| Convolutional #5 | 13 x 13 x 256 | 256 | 3 x 3 | 1 | 1 |
| Max Pool #3 | 6 x 6 x 256 | 256 | 3 x 3 | 2 | - |
| Fully Connected Layer #1 (FC6) | 4096 | - | - | - | - |
| Fully Connected Layer #2 (FC7) | 4096 | - | - | - | - |
| Fully Connected Layer #3 (FC8) | 1000 | - | - | - | - |

AlexNet architecture properties table

## 2 Features

The following are the novel features introduced in this architecture:

1. **ReLu Nonlinearity**: Traditionally, *tanh* or *sigmoid* function was used for activation function, but they suffer from the disadvantage of saturation at high and low values. ReLU function does not saturate in the positive region and it's also computationally very efficient.

2. **Local Response Normalization**: Denoting by $a^i_{x,y}$ the activity of a neuron computed by applying kernel $i$ at position (x, y) and then applying the ReLU nonlinearity, the response-normalized activity $b^i_{x,y}$ is given by the expression:

$$b^i_{x,y} = a^i_{x,y} / \left( k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^\beta$$

3. **Overlapping Pooling**: A type of pooling where the stride of the pooling kernel is less than the dimensions of the kernel itself. According to the author, this reduces the chances of overfitting.

4. **Data Augmentation**: Using this, chances of overfitting reduces by providing a dynamic versatility to the data distribution. The author used *horizontal flip* and *random cropping*.

5. **Dropout**: The network randomly drops certain number of neurons with probability $p$(taken as 0.5 in the paper) so as to make each neuron learn some generic features of the data distribution.

## 3 Edge

The model had a number of successes, namely

1. Unlike the prior models, AlexNet could be applied to high resolution images of the $ImageNet$ dataset.
2. Direct image input could be sent to the classification model.
3. The convolution layers could automatically extract the edges of the images and, subsequently, the fully connected layers could learn these features.
4. The model achieved low classification errors without having overfitting due to Dropout, Data Augmentation & Overlapping Pooling.
5. The model was an innovation which optimized for GPUs and cut down on training times while improving performance.

## 4 Drawbacks

Given the successes of the model, it still had certain issues due to which subsequent models had an even less amount of error.

1. As more layers using certain activation functions are added to neural networks, the gradient of the loss function approaches zero, making the network hard to train. This problem is known as $Vanishing\ Gradient$. To avoid this, $Xavier\ Initialisation$ is used to initialise the weights.
2. The depth of a neural network is a crucial component. AlexNet falls short on this leading to being surpassed by $VGGNet$, $GoogleNet$ & $ResNet$.
3. The size of the filter used was $5 \times 5$, which was pretty large. It could be replaced by two $3 \times 3$ filters. The benefit being the lower $3 \times 3$ kernel is shared more frequently than the $5 \times 5$ one.

## 5 Conclusion

AlexNet, introduced in the year 2012, revolutionized deep learning approaches. The model solved the task of image classification through deep CNN and efficient computation resources. AlexNet won the 2012 competition with a top-5 error rate of 15.3% compared to second place top-5 error rate of 26.2%. AlexNet was able to recognize off-center objects and most of its top 5 classes for each image were reasonable. Thus, AlexNet paved the way for deep Convolution Neural Networks.