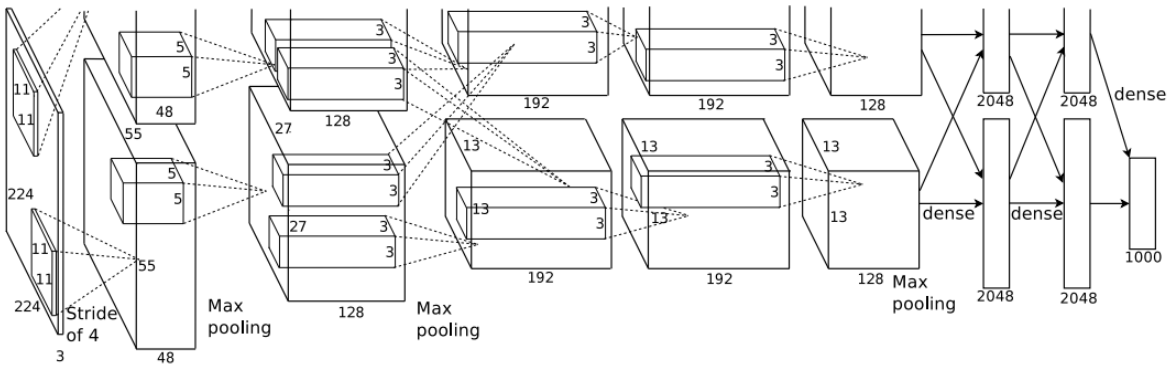
Name: Manoj Aryal
ID: 2020470025

AlexNet

AlexNet was the winning entry in ILSVRC 2012 and solved image classification task where the input is an image of one of 1000 different classes and the output is a vector of 1000 numbers. The input to AlexNet is an RGB image of size 256×256 and has 60 million parameters and 650,000 neurons.

AlexNet consists of 8 layers: 5 convolutional layers and three fully connected layers. ReLU is applied after all convolutional and fully connected layers. ReLU of first and second convolutional layers are followed by local normalization before the pooling. Dropout was used in the first two fully connected layers.



source: ImageNet Classification with Deep Convolutional Neural Networks (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton)

Fig: AlexNet Architecture

Convolutional Layers: Convolutional layers are the main building block of CNN models and most popular image classification model. For the AlexNet model 5 Convolutional layers were used and its structure are described below:

Convolutional Filters: In convolutional layers, multiple filters are taken to slice through the image and map them one by one and learn different portions of an input image. In AlexNet model, the first convolutional layer filters the 224×224×3 input image with 96 kernels of size. The second convolutional layer takes as input the (response-normalized and pooled) output of the first convolutional layer and filters it with 256 kernels of size 5 × 5 × 48. The third convolutional layer has 384 kernels of size 3 × 3 × 256. The fourth convolutional layer has 384 kernels of size 3 × 3 × 192, and the fifth convolutional layer has 256 kernels of size 3 × 3 × 192.

Strides: Strides is the number of pixel shifts over the input matrix.
The network architecture with the stride size is described below:

Conv1: Stride 4
Max Pool1 : Stride 2

Conv2: Stride 1
Max Pool2: Stride 2

Conv3: Stride 1

Conv4: Stride 1

Conv 5: Stride 1
Max Pool3: Stride 2

Max Pooling: The main objective of max pooling is to down sample input representation and reduce its dimensionality. It reduces the spatial size of a layer keeping just the maximum values. In Overlapping Max Pool, the adjacent windows over which the max is computed overlap each other. In AlexNet, the first two convolutional layers and the fifth layer are followed by Overlapping Max Pooling layer.

Padding: Padding refers to the number of pixels added to an image when it is being processed by the kernel of a CNN. AlexNet use zero padding or valid padding which helped to preserve the spatial size.

Fully Connected Layers: They are feed forward neural network and forms the last few layers in the network.
In AlexNet model, the output from the final Convolutional layer is flattened and passed through 3 fully connected layers.

ReLU: The Rectified Linear Unit is one of the most commonly used activation function and it returns 0 if the input is negative and returns the same value for any positive number.

In AlexNet, ReLU nonlinearity is applied after all the convolution and fully connected layers (not in the last fully connected layer).

The final fully connected layer returns 1000 classes of unscaled activations.

Report:

Methods/Shots	1	5	20	50
Softmax	0.0270	0.3060	0.6980	0.7620
Fine-Tuning	0.0210	0.2820	0.6390	0.7500
Metric-Based	0.015	0.013	0.022	0.016