Hi Sir. We have studied 5 primarily used Image Networks :

1)Alexnet

2)Resnet

3)VGG

4)Mobilenet

5)Googlenet

We dug a bit into CNN :

After convolving, the convolved feature the output of the previous layer in the NN, will be the input for the next layer. If the pixel in the image is the same as the filter, the pixel in convolved feature will be very high, if there’s nothing related to the **filter**in the image, then the pixel will be very low in the convolved feature, which is the next layer.

1. Alexnet:

\* Alex net wass the most used CNN model mainly used for computer vision, born out of the 2010 Imagenet challenge.

\*The network comprises 5 Convolutional (CONV) layers and 3 Fully Connected (FC) layers.

\*The activation used is the Rectified Linear Unit (ReLU).

Relu activation function is used instead of functions like Tanh, which were commonly used previously, to add non-linearity.

It accelerates the speed by 6 times at the same accuracy. Relu is applied after every convolutional layer and fully connected layer.

\*Alexnet uses dropout instead of regularisation to deal with overfitting. However, the training time is doubled with the dropout rate of 0.5.

\*Overlap pooling is used to reduce the size of the network. It reduces the error rates tremendously.

\*The architecture is comprised of eight layers in total, out of which the first 5 are convolutional layers and the last 3 are fully-connected. The first two convolutional layers are connected to overlapping max-pooling layers to extract a maximum number of features.

\*The final output layer is connected to a softmax activation layer, which produces a distribution of class labels.

1. Resnet :

\*ResNet-18 is a convolutional neural network that is 18 layers deep.It is the winner of ILSVRC Image Challenge 2015..

The logic we use in resnet is that we bypass the input to first layer to be output of last layer of the model,then network should be predict whatever function it was learning before the input is added to it.

\*Here, f(x)=x , and f(x) + x =h(x) is the general activation function in a deep neural network.

\*The intuition behind resnet is simple : Adding many layers to the neural network won’t always result in the best learning of the identity function.

\*Resnet solves the problem of vanishing gradients, where the loss function can easily become zero by using chain rule. This results in weights never updating their values and therefore no learning being performed .With resnet the gradients can easily flow to skip connections backwards from later layers to initial filters. While increasing network depth, Resnet avoids negative outcomes. So we can increase the depth but we have fast training and higher accuracy.

1. VGG-16 :

\*Vgg-16 is a simple architecture model,since its not using much hyper parameters.it always uses 3\*3 filters with stride of 1 in convolution layer with pooling layers 2\*2 with stride of 2 .

\*Large fliters are large-kernel sized fliters. VGG was the winner of the 2014 ILSVRC.

\* The first two layers have 64 channels of 3\*3 filter size. Then after a max pool layer we have two layers of 256 filter size. This is followed by max pooling layer which is same as previous layer. Both VGG and Alexnet output feature maps with the same size in last pooling layer, but the neurons of VGG-16 cover the receptive field with larger size.

In contrast to VGG-16, AlexNet retains more unrelated background information in last convolutional layer, which often disturbs the final prediction. and Hence the VGG predicts better than AlexNet.

\* VGG is very computationally expensive, it took 2 weeks to run on an Nvidia Titan GPU. Even though they offer vast improvement over alexnet by having multiple smaller sized kernels the time complexity plus along with high storage requirement reduces its application in certain areas. VGG-16 is the first runnerup in the ILSVRC 2014.

4) MobileNet :

\* It is an architecture designed for mobile devices,to build light weight deep neural networks. They are the latest CNNs used in mobile architecture.

\*Hyper parameters used is the optimal trade-off between latency and accuracy.

\*Depthwise separable convolutions,a form of factorized convolutions are used here instead of normal convolutions. They are followed by a pointwise convolution. In MobileNet, the depthwise convolution applies a single filter to each input channel. Mobilenet has a reduction of number of parameters significantly and thereby reduces the total number of floating point multiplication operations which is favorable in mobile and embedded vision applications with less compute power.

\*However, it is sometimes perceived that this leads to a bit of a loss in terms of accuracy. Overall, these are great exclusively for mobile computing.

5) Googlenet:

\*Googlenet is the winner of the ILSVRC 2014. The error rate of around 6.6 % was very close to human level performance, which is unmatched.

\*The network used a CNN inspired by LeNet but introduced an innovation callled an inception module.

\*Googlenet uses batch normalization and image distortions . It relies on few very small convolutions in order to reduce the number of parameters significantly. Googlenet is comprised of 22 layers of neural networks, but with highly reduced parameter values.

Because of this, the training time of Googlenet is much better than like VGG. Size of Googlenet is also around 100mb, almost 5 times smaller.

\* The inception layer aims to cover a bigger area, but also keeps a fine resolution for small information on the images. Googlenet can convolve in different sizes from the most accurate detailing .

\* Then, multiple inception modules aree piled. Even the training is slightly different in GoogleNet, as most of the topmost layers have their own output layer. Thus the model converges faster, as there is a joint training as well as parallel training for the layers itself.