Alex Net

- Assumptions made by CNNs:
  - Locality of pixel dependencies: Pixels in a neighbourhood are likely to be correlated than faraway pixels.
  - Stationarity of statistics: Similar features have similar distributions and can be detected by the same type of filters.
- Alexnet consists of 5 convolutional layers and 3 fully connected layers.Though parameters of convolutional layers comprise only a small fraction of total no. of parameters, they have a significant role in the network's performance.
- Data was taken from Imagenet. Mean image subtraction was the only pre-processing method used to bring down the image size to 256x256.
- Relu function works faster than tanh(x) or sigmoid(x) and achieves lower error in less time which is very useful in case of large datasets.
- LRN(Local Response Normalisation): A neuron at a position(x,y) is normalised with a few neurons of other layers at the same position from layers in its neighbourhood.It imitates the *lateral inhibition* phenomenon observed in neurology where a neuron's activity is reduced by surrounding neurons.
- Overlapping pooling (stride<filter_size): It was observed that overlapping pooling made overfitting difficult.
- Architecture:
  - LRN layers follow only the 1st and 2nd convolutional layers.
  - Max pooling layers follow the LRN layers and 5th convolutional layer.
  - Relu activation is used for the output of each convolutional block.
  - Softmax function was applied to the output of the eighth layer.
- Two types of data augmentation were used which did not store the images in the GPU:
  - Random patches were extracted from the image and their horizontal reflections were also used for training.While testing, the average prediction of the augmented images was presented.
  - PCA (Principal Component Analysis): The intensities and colours of the RBG pixels are modified. It depicts the property of object that its identity is independent of the colour and intensities of the image pixels.
- Dropout regularisation was applied.As a neuron cannot depend on nearby neurons for learning, it ends up learning better features as its presence is random in dropout.
- Momentum and weight decay were also used.
- A technique used to decrease cross validation error was to decrease learning rate by factor of 10 when error stopped decreasing.
- It was found that the network detected objects which were away from the center of image correctly.