

## Models

### Transfer Learning Used:

ImageNet transfer learning has been used on the following pretrained models.

#### Optimizers used:

- Adam

Adam is the replacement algorithm for stochastic gradient descent for training deep learning models.

Adam combines the best practices of the AdaGrid and Rmsprop algorithms to provide an optimization that can handle sparse gradients on noisy problems.

#### Rmsprop

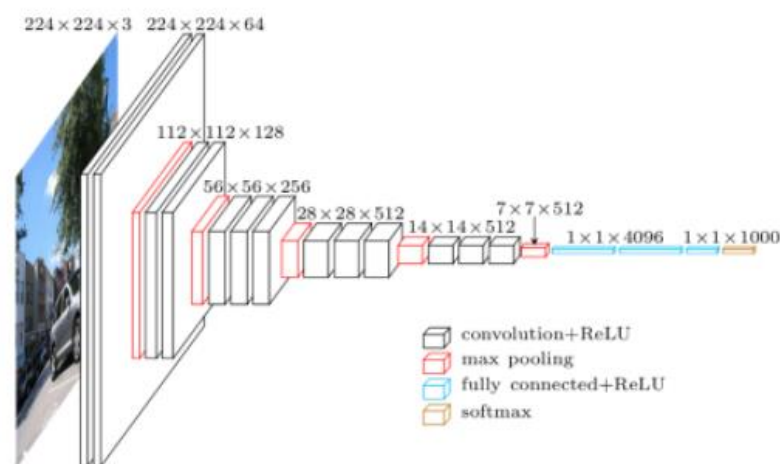
Rmsprop was developed as a stochastic technique for mini-batch learning.

RMSprop deals with the above issue by using a moving average of squared gradients to normalize the gradient. This normalization balances the step size (momentum), decreasing the step for large gradients to avoid exploding, and increasing the step for small gradients to avoid vanishing.

### VGG16:

- It has 16 layers.
- A CNN model used for image classification.
- Pretty large network and has 138 million parameters.

- Follows the arrangement of convolution and max pool layers consistently throughout the architecture. In the ends, it has 2 fully connected layers and a SoftMax layer for output.  
(Use of SoftMax: it assigns probabilities to multi-class problem.)



Architecture of VGG16

## Max-pooling:

This is nothing but a process of taking max value out of window space.

Let's say we have image of  $n \times n$  matrix. And our kernel is of size of let's say  $3 \times 3$  ( $n > 3$ ). So when we hover our kernel on image matrix, the points that it covers and the max value out of them will be taken as the output for the max-pooling value.

Convolution:

This is a process of image processing where we multiply kernel elements with the elements of the image matrix in reverse order.

Formula:

$$G[i, j] = \sum_{u=-k}^k \sum_{v=-k}^k h[u, v] F[i - u, j - v]$$

### **Resnet50 Model:**

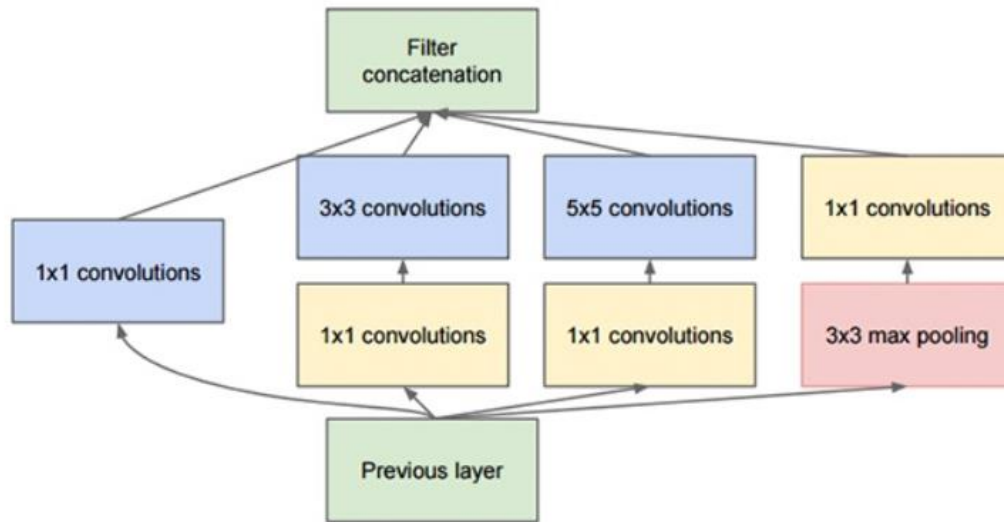
- This is 50 layered model.
- When we go deeper into neural networks, that time we have good probability of the accuracy getting saturated and the degradation might happen at a faster rate.
- Thus, Resnet or Residual Net comes into picture.
- Here, instead of simply mapping the values as: -

$X \rightarrow Y$  with function  $H(X)$ , we use a residual function: -

$$H(X) - x = F(X)$$

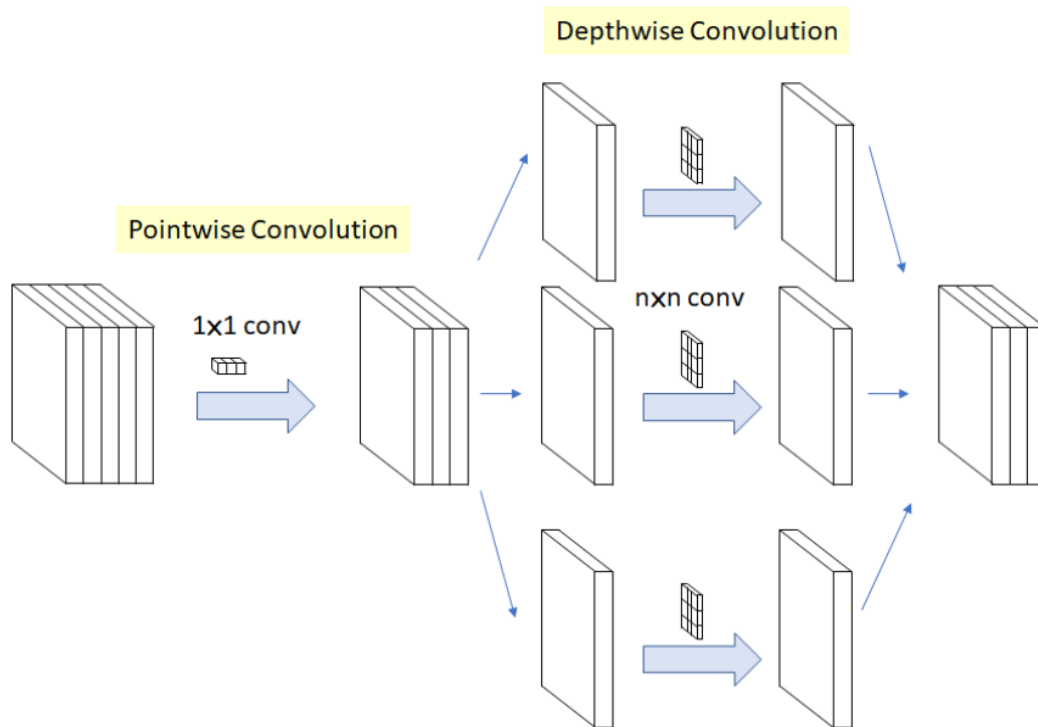
It is very easy to come up with a solution like  $F(x) = 0$  rather than  $F(x) = x$  using stack of non-linear CNN layers as function. So, this function  $F(x)$  is what the authors called Residual function.

## InceptionV3 model



- Inception v3 is a widely-used image recognition model that has been shown to attain greater than 78.1% accuracy on the ImageNet dataset.
- The third edition of Google's Inception CNN.
- The goal of the inception module is to act as a “multi-level feature extractor” by computing  $1\times 1$ ,  $3\times 3$ , and  $5\times 5$  convolutions within the *same* module of the network — the output of these filters is then stacked along the channel dimension and before being fed into the next layer in the network.

## Xception model



- This is an extension of the Inception model.
- The modified depth wise separable convolution is the **pointwise convolution followed by a depth wise convolution**. This modification is motivated by the inception module in Inception-v3 that  $1 \times 1$  convolution is done first before any  $N \times N$  spatial convolutions.