The paper is about very deep convolutional networks (up to 19 weight layers) for large scale image classification. This work investigates the effect of convolutional network depth on accuracy in large-scale image recognition, resulting in improvements on prior-art configurations and generalization to other datasets. As we know, Convolutional networks (ConvNets) have been successful in large-scale image and video recognition due to large public image repositories and high-performance computing systems. The ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has served as a test bed for a few generations of deep visual recognition architectures. This paper addresses the depth of the network by adding more convolutional layers, which is feasible due to the use of small (3\*3) convolution filters in all layers.

The input to ConvNets is a fixed-size RGB image, which is preprocessed by subtracting the mean RGB value from each pixel. The image is passed through a stack of convolutional layers with filters with a small receptive field. Spatial pooling is carried out by five max-pooling layers, three Fully-Connected (FC) layers, and a soft-max layer. All hidden layers are equipped with rectification (ReLU) non-linearity, but none contain Local Response Normalization (LRN) normalisation. The ConvNet configurations evaluated in this paper differ only in depth, with 11 weight layers in the network A and 19 in the network E. The number of weights is not greater than a more shallow net with larger conv. layer widths and receptive fields. The ConvNet configurations used in the top-performing entries of the ILSVRC-2012 and ILSVRC-2013 competitions use very small 3\*3 receptive fields throughout the whole net, which are convolved with the input at every pixel (with stride 1). This reduces the number of parameters and makes the decision function more discriminative. Additionally, three non-linear rectification layers are included, making the decision function more discriminative and decreasing the number of parameters.

The ConvNet training procedure follows which involves optimizing the multinomial logistic regression objective using mini-batch gradient descent with momentum. The learning rate was initially set to 102, and then decreased by a factor of 10 when the validation set accuracy stopped improving. The initialization of the network weights is important, as bad initialization can stall learning due to instability of gradient in deep nets. To circumvent this problem, the weights were sampled from a normal distribution with the zero mean and 102 variance. Combining the outputs of several models by averaging their soft-max class posteriors improved performance, and was used in top ILSVRC submissions in 2012 and 2013. Best-performing single model achieved 7.1% error.

This work evaluated very deep convolutional networks (up to 19 weight layers) for large-scale image classification. It was demonstrated that representation depth is beneficial for classification accuracy and that state-of-the-art performance can be achieved using a conventional ConvNet architecture with increased depth. Models generalize well to a wide range of tasks and datasets, confirming the importance of depth in visual representations.

**Paper Name:** Very Deep Convolutional Networks for Large-Scale Image Recognition

**Authors Name:** Karen Simonyan & Andrew Zisserman

**Published On:** 10 April, 2015