VGG:

=> VGG stands for Visual Geometry Group. It is a standard deep convolutional neural network (CNN) architecture with multiple layers.

=> the "deep" refers to the no. of layers with VGG-16 or VGG-19 consisting of 16 and 19 convolutional layers.

=> the VGG architecture is the basis of ground-breaking object recognition models.

=> developed as a deep neural network, the VGGNet also surpasses baselines on many tasks and datasets beyond ImageNet.

⇒ It is now still one of the most popular img recognition architectures.

VGG 16:

⇒ The VGG model, or VGG Net, that supports 16 layers is also refferred to as VGG 16, which is a convolutional neural network model proposed by A. Zisserman and K. Simonyan from the University of Oxford.

⇒ This model achieves almost 92.7% top-5 test accuracy in ImageNet. ImageNet is a dataset consisting of more than 14 million img belonging to nearly 1000 classes.

⇒ It is one of the most popular model submitted to ILSVRC-2014.

⇒ It replaces the large kernel-sized filters with several 3x3 kernel-sized filters one after the other, thereby making significant improvements over AlexNet. The VGG16 model was trained using Nvidia Titan Black GPUs for multiple weeks.

⇒ the VGG Net-16 supports 16 layers and can classify img into 1000 obj. categories, including keyboard, animals, pencil, mouse etc. Additionally, the model has an i/p size of 224 by 224.

VGG Architecture:

=> VGGNet are based on the most essential features of CNN.

=> the VGG network is constructed with very small convolutional filters.

=> the VGG-16 consists of 13 convolutional layers & 3 fully connected layers.

Input:

=> the VGGNet takes in an img i/p size of 224 x 224. for the imgNet competition, the creators of the model cropped out the center 224 x 224 patch in each img to keep the i/p size of the img consistent.

Convolutional layers:

=> VGG's convolutional layers leverage a minimal receptive field, i.e 3x3, the smallest possible size that still captures up/down and left/right.

=> moreover, there are also 1x1 convolution filters acting as a linear transformation of the input.

=> this is followed by a ReLU unit, which is a huge innovation from AlexNet that reduces training time.

=> ReLU stands for rectified linear unit activation function; it is a piecewise linear function that will o/p the i/p if +ve; otherwise, the o/p is zero.

=> the convolution stride is fixed at 1 px to keep the spatial resolution preserved after convolution.

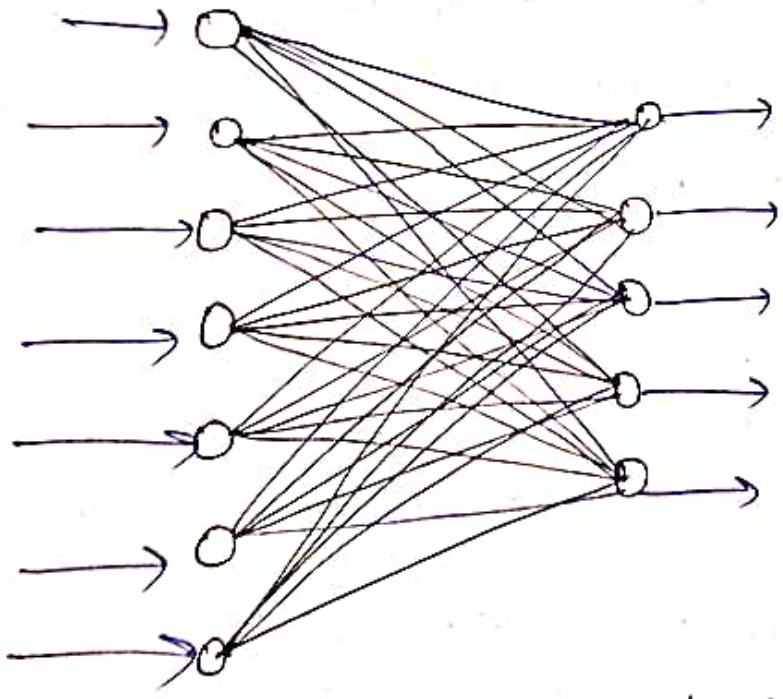
Hidden Layers :

=> All the hidden layers in the VGG network use ReLU. VGG does not usually leverage Local Response Normalization (LRN) as it increases memory consumption and training time.

=> moreover, it makes no improvements to overall accuracy.

Fully-connected Layers :.

=> the VGGNet has 3 fully connected layers. Out of the 3 layers, the first 2 have 4096 channels each, and the third has 1000 channels 1, for each class.



fully connected layers.

VGG-16 architecture:

=> the number 16 in the name VGG refers to the fact that it is 16 layers deep neural network (VGG net).

=> this means that VGG16 is a pretty extensive network and has a total of around 138 million parameters.

=> even according to modern standards, it is a huge network

=> this architecture's simplicity is what makes the network more appealing. just by looking at its architecture, it can be said that it is quite uniform.

=> there are a few convolution layers followed by a pooling layer that reduces the height and the width. If we look at the no. of filters that can use, around 64 filters are available that we can double to about 128 and then to 256 filters. In the last layers, we can use 512 filters.

complexity and challenges:

=> the no. of filters that we can use doubles on every step or through every stack of the convolution layers.

=> this is a major principle used to design the architecture of the VGG16 network.

=> one of the crucial downsides of the VGG16 network is that it is a huge network, which means that it takes more time to train its parameters.

=> because of its depth & no. of fully connected layers, the VGG16 model is more than 533 MB. This makes implementing a VGG networks a time-consuming task.

=> The VGG16 model is used in several deep learning ing classification problems, but smaller network architectures such as GoogleNet and SqueezeNet are often preferable.

=> In any case, the VGG Net is a great building block for learning purposes as it is straightforward to implement.

Performance of VGG models :-

=> VGG16 highly surpasses the previous versions of the models in the ILSVRC-2012 & ILSVRC-2013 competitions.

=> VGG16 result is competing for the classification task winner (GoogLeNet with 6.7% error) and considerably outperforms the ILSVRC-2013 winning submission clarifai.

=> It obtained 11.2% with external training data and around 11.7% without it. In terms of the single-net performance, the VGGNet-16 model achieves the best result with about 7.0% test error, thereby surpassing a single GoogLeNet by around 0.9%.

VGG-16 Architecture of a VGG-16 model