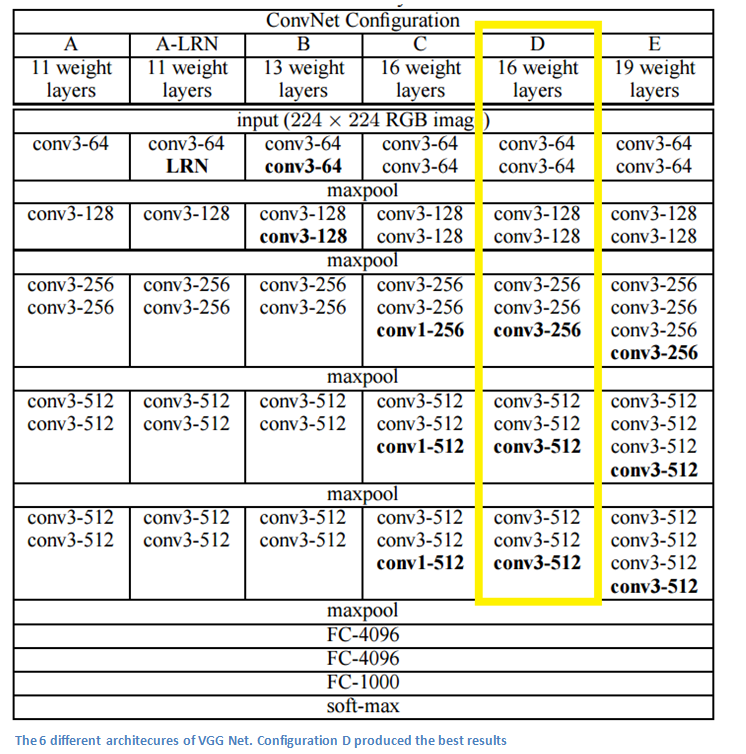
Tools for Machine Learning:

# VGG Net (2014)

Simplicity and depth. That’s what a model created in 2014 (weren’t the winners of ILSVRC 2014) best utilized with its 7.3% error rate. Karen Simonyan and Andrew Zisserman of the University of Oxford created a 19 layer CNN that strictly used 3x3 filters with stride and pad of 1, along with 2x2 maxpooling layers with stride 2. Simple enough right?



**Main Points**

* The use of only 3x3 sized filters is quite different from AlexNet’s 11x11 filters in the first layer and ZF Net’s 7x7 filters. The authors’ reasoning is that the combination of two 3x3 conv layers has an effective receptive field of 5x5. This in turn simulates a larger filter while keeping the benefits of smaller filter sizes. One of the benefits is a decrease in the number of parameters. Also, with two conv layers, we’re able to use two ReLU layers instead of one.
* 3 conv layers back to back have an effective receptive field of 7x7.
* As the spatial size of the input volumes at each layer decrease (result of the conv and pool layers), the depth of the volumes increase due to the increased number of filters as you go down the network.
* Interesting to notice that the number of filters doubles after each maxpool layer. This reinforces the idea of shrinking spatial dimensions, but growing depth.
* Worked well on both image classification and localization tasks. The authors used a form of localization as regression (see page 10 of the [paper](http://arxiv.org/pdf/1409.1556v6.pdf) for all details).
* Built model with the Caffe toolbox.
* Used scale jittering as one data augmentation technique during training.
* Used ReLU layers after each conv layer and trained with batch gradient descent.
* Trained on 4 Nvidia Titan Black GPUs for **two to three weeks**.

**Why It’s Important**

VGG Net is one of the most influential papers in my mind because it reinforced the notion that **convolutional neural networks have to have a deep network of layers in order for this hierarchical representation of visual data to work**. Keep it deep. Keep it simple.