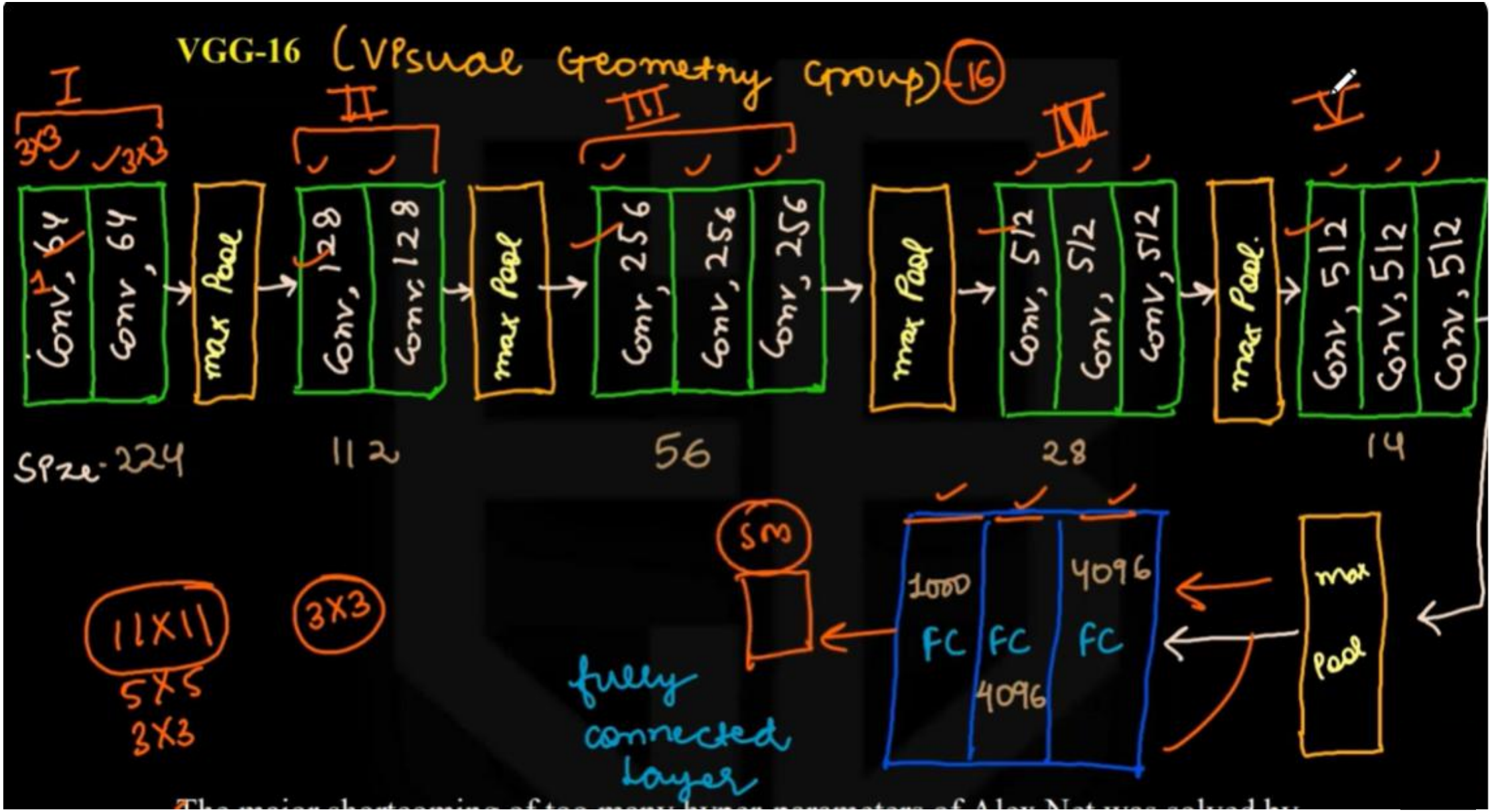
CONV NETS

VGG,INCEPTION NET,RESNET


VGG-16 (Visual Geometry Group) 16
I
II
III
IV
V
3X3
3X3
Conv, 64
Conv, 64
max Pool
Conv, 128
Conv, 128
max Pool
Conv, 256
Conv, 256
Conv, 256
max Pool
Conv, 512
Conv, 512
Conv, 512
max Pool.
Conv, 512
Conv, 512
Conv, 512
Size-224
112
56
28
14
11X11
5X5
3X3
3X3
SM
fully connected layer
1000
FC
FC
4096
4096
FC
max
Pool

The major shortcoming of too many hyper-parameters of Alex Net was solved by VGG Net by replacing large kernel-sized filters (11 and 5 in the first and second convolution layer, respectively) with multiple 3×3 kernel-sized filters one after another.

The architecture developed by Simonyan and Zisserman was the 1st runner up of the Visual Recognition Challenge of 2014.

The architecture consists of 3*3 Convolutional filters, 2*2 Max Pooling layer with a stride of 1. Padding is kept the same to preserve the dimensions.

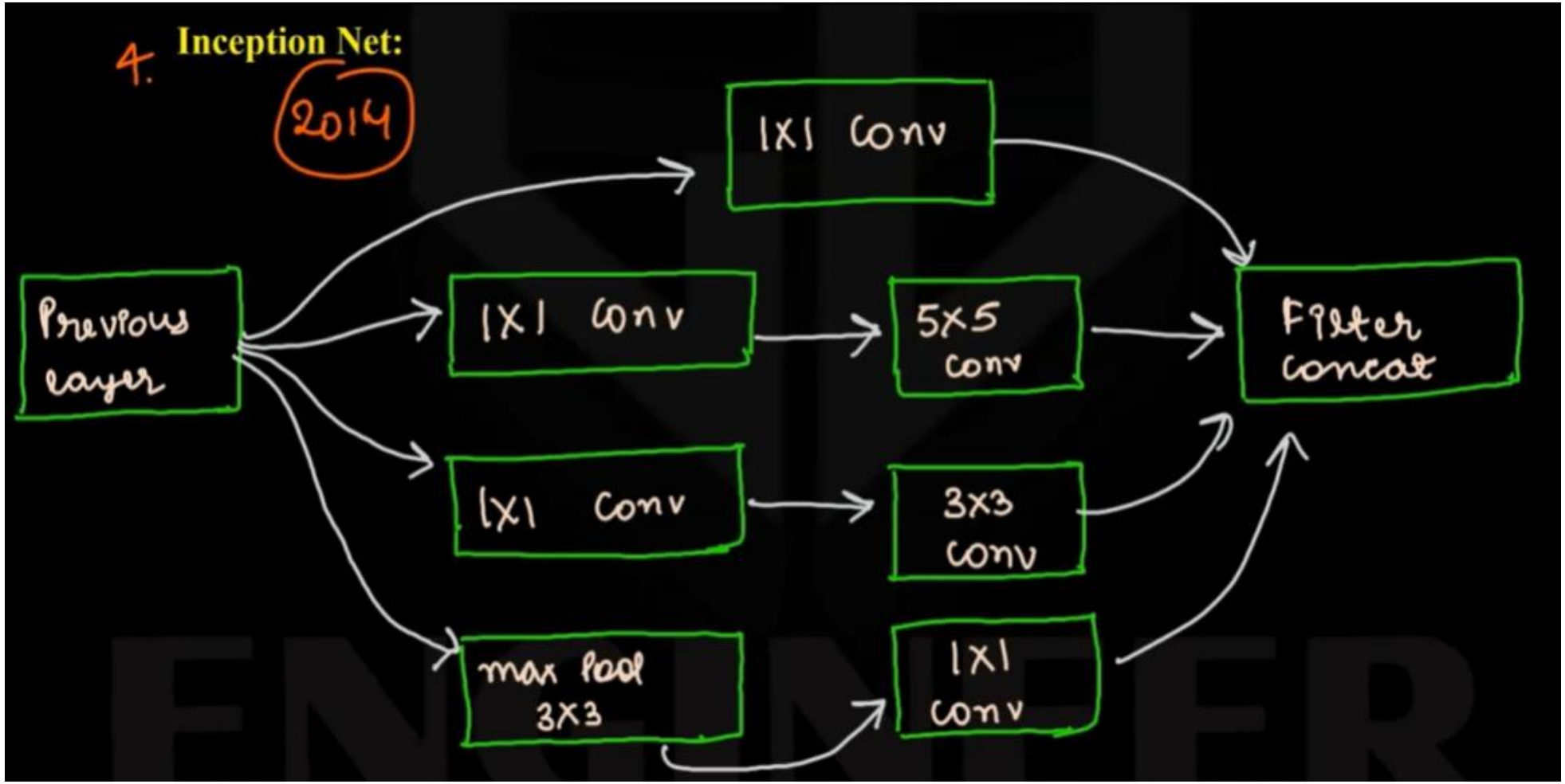
There are 16 layers in the network where the input image is RGB format with dimension of 224*224*3, followed by 5 pairs of Convolution (filters: 64, 128, 256,512,512) and Max Pooling.

The output of these layers is fed into three fully connected layers and a SoftMax function in the output layer.

In total there are 138 million parameters in VGG Net.

ResNet:

ResNet, the winner of ILSVRC-2015 competition is a deep network with over 100 layers. Residual networks (ResNet) is similar to VGG nets however with a sequential approach they also use “Skip connections” and “batch normalization” that helps to train deep layers without hampering the performance. After VGG Nets, as CNNs were going deep, it was becoming hard to train them because of vanishing gradients problem that makes the derivate infinitely small. Therefore, the overall performance saturates or even degrades. The idea of skips connection came from highway network where gated shortcut connections were used


4. Inception Net:
2014
1X1 Conv
Previous layer
1X1 Conv
5X5 conv
Filter concat
1X1 Conv
3X3 conv
max pool 3X3
1X1 conv

Inception network also known as GoogleLe Net was proposed by developers at google in "Going Deeper with Convolutions" in 2014. The motivation of InceptionNet comes from the presence of sparse features Salient parts in the image that can have a large variation in size. Due to this, the selection of right kernel size becomes extremely difficult as big kernels are selected for global features and small kernels when the features are locally located. The InceptionNets resolves this by stacking multiple kernels at the same level. Typically it uses 5*5, 3*3 and 1*1 filters in one go.