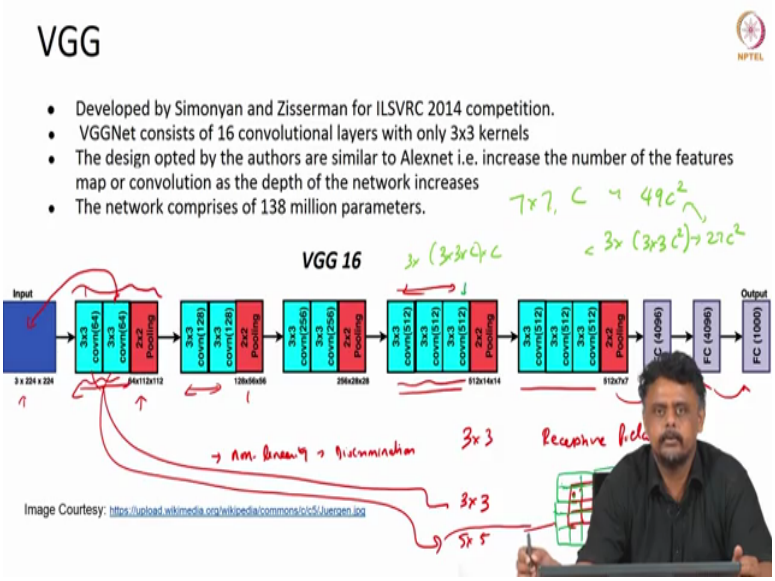
Machine Learning for Engineering and Science Applications
Professor Dr. Ganapathy Krishnamurthi
Department of Engineering Design
Indian Institute of Technology, Madras
CNN Architecture VGG Net
Part 2

(Refer Slide Time: 00:14)



We will look at the VGG or network specifically the VGG 16, so VGG stands for the visual geometry group at Oxford University. So this particular network was entered into the 2014 image net challenge ok, so it had 16 weight layers ok 16 layers with this one we are looking at a 16 weight layer they also had another version which 19 million 19 at 19 weight layers so very this the design of this is very similar to LeNet and AlexNet ok except that they have made this a little bit more systematic, so we will see what the systematics is here.

So we saw one of the in the earlier videos we saw that it is best that as you go deeper into the network we increase the size of the depth or the number of feature maps increases so it becomes wider network becomes wider ok, so that that was incorporated here they also stuck to one filter size 3 cross 3 this is the smallest filter size smallest meaningful filter size, so 3 cross 3 gives you I can also do one cross one collision but of course we need a receptive field so 3 cross 3 is the small they stuck to 3 cross 3 filtered throughout the (lay) all the layers.

There are a lot of parameters of around 130 to 140 million parameters depending on which network you are looking at once again we will just walk through the network briefly and then we will see what are the advantages given by this network, so if you see the input again same

as 3, 3 channel RGB image size 224 cross 224 so the first layer has to 3 cross 3 convolutions in succession followed by a max pooling layer which reduces the size of the feature maps to 1 cross 112 cross 12 and each of these 3 cross 3 convolution layers at 64 feature Maps followed by another set of convolution and max pooling so on and so forth till we get to the these 3 2 layers which have again 3 cross 3 convolution but this time 3 in succession 3 of them in succession.

If you see here followed by again a max pooling and here of course these are rasterize 10 made into a fully (con) connection only fully connected layers here these two and the output is a 1 of 10 1 of 100 classification. Now this as if you look at this we will see that this is a block it is one of the earliest networks to introduce this kind of concept when you see another network which most contacts now use this.

This is a block of convolutions, so basically instead of your layer is now been replaced by yeah a bunch of convolutional layers ok a sequence of layers is now used as one layer. So in between these two convolution layers there is no max pooling but non linearity is still there, so 3 cross 3 non linearity and then 3 cross 3 non linearity ok that is the sequence ok in this case 2 3 successive nonlinear piece of or are applied following the convolutions.

So what does this provide in terms of an advantage so if you look at a succession of 3 by 3 convolutions one is the non-linearity leads to greater discriminate discrimination ok seen that we have looked at other classifiers so if we have way nonlinear features we expect that at some point the class has become linearly separable that is but then how do we in crop this here incorporate this here network is to have a succession of nonlinearities hopefully giving rise to better discriminatory power in the network.

The second advantage is in terms of the receptive field ok, the authors also mentioned this that if you look at this particulars what we can just come here look at this for instance if you look at these 2 3 cross 3 convolution ok so this is 1 convolution layer with the 64 feature maps 3 cross 3 convolutions again 64 3 cross 3 followed by another set of 64 3 cross 3 convolutions ok.

Now if you look at the receptive field of this layer of this block so the first one as a receptive field of 3 cross 3 ok so then new size to 64 feature maps right, so now if you look at the second if you look at the this is the first one if you look at the second one ok the filter kernel size is 3 cross 3 ok but the receptive field on the original image so we are looking at the

receptive field of this particular layer on the input side ok so the receptive layer field for this one is 3 cross 3 and for the second it becomes 5 cross 5 it is not too hard to see because if you look at the if you look at the second convolution the feature maps in the second convolution layer ok if we have 3 cross 3 ok.

So if you look at this particular look at this particular activation in the feature map then it is a get itself is a result of a 3 cross 3 convolution, so this itself is a result of a 3 cross 3 convolution right, so you have 1, 2 and 3 here right so this particular feature map in the second 3 cross 3 convolution layer itself is the output after 3 cross 3 convolution from the previous layer, so this is looking at 3, 1 2 let us say so then we see that all around we will have to add one more row and column ok.

So it is receptive field on the input is actually 5 cross 5, so similarly if you look at the succession of 3 convolution layers so the third the final receptive field for this part the receptive field of this particular convolution layer would correspond to a 7 cross 7 and you can verify that on your own. So what does this provide in terms of you know computational gain that is what the authors claim or the authors claim this what it that is the computational gain you get from this is that if you want to do a success or a succession of 3 cross 3 convolutions 3 of them the number of computations involved to produce one element in the output feature map is much lesser when compared to if you just do a 7 cross 7 convolution ok.

So this is the and at the same time you have a larger receptive field on input sometimes it is desirable and in many cases is desirable to have a larger receptive field. so that you get more of the context in the image into your activation maps but doing a much larger convolution using a much larger convolution means that the number of competition increase, so one advantage of using a succession of 3 by 3 filters convolution layers instead of just using a filter with a larger receptive field is the savings in the number of parameters.

So for instance if you had used 7 cross 7 filters on C channels then the number of parameters would be 49 C squares on the other hand if you are use if you would use a succession of three convolution layers to get the same receptive field with 3 by 3 filter sizes then the number of filter sizes would be about 27 C square, ok. So this thing gives rise to it is a decrease in the number of parameters actually this might not be very obvious but you have to do the computation to figure out because if you if you have C input channels then we for so how do

we get this number if you have C input channels then the size of each filter would be if you use a 3 cross 3 the number of elements in which should filter would be 3 cross 3 times C, right and then if you have C output channels then we have one more C outside that is the total number of parameters ok.

Then we have a section of three layers with 3 cross 3 filters so if 3 times that ok, so this is how we calculate the number of parameters in when you use a succession of 3 of course the similar calculations for 7 cross 7 ok. So VGG net by using the succession of convolution layers with very small filter sizes but of course the number of parameters is much higher than the AlexNet but the results were much better which is about 7 percent in the top five 7 percent error rate in the top five which was which is right in fact that was not the winning it was not the winning entry but it won in other categories in the challenges he may localization etcetera but the error rate was still pretty good for shorts around 7 percent, ok.

So the winning network was actually the inception network or Google net from Google we will see much more about it in the next videos.