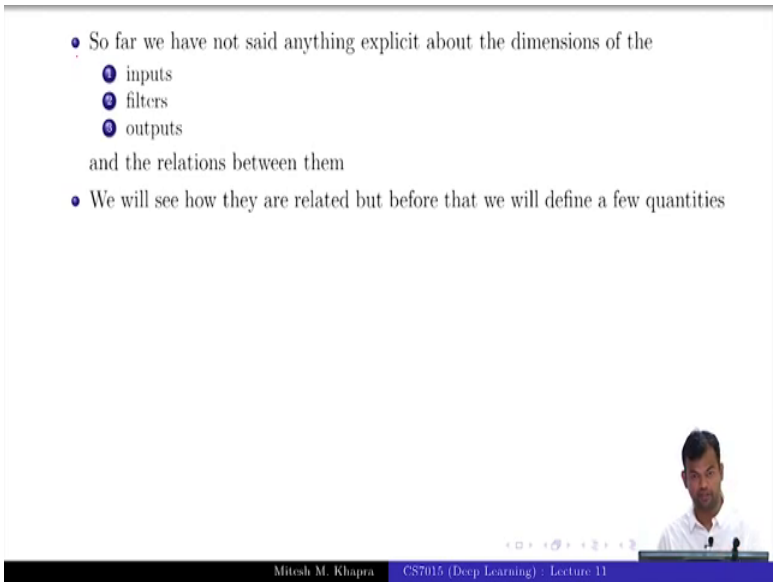
Deep Learning
Prof. Mitesh M. Khapra
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 11
Relation between input size, output size and filter size

Ok. So, now we will go to the next module where we will try to learn the Relationship between the input size, the output size and the filter size ok.

(Refer Slide Time: 00:21)



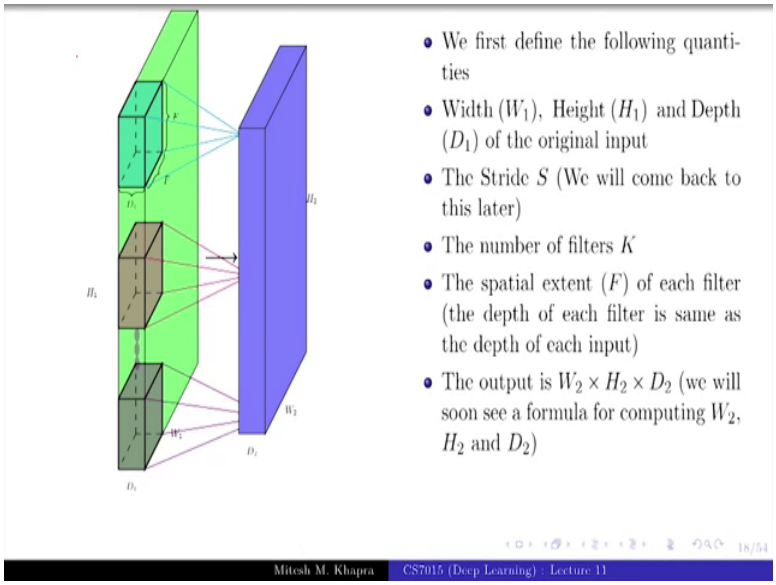
So, so far we have not said anything about the dimensions of the inputs I have just been very vague that its M cross N and also for the filter I have just said 3 cross 3; 5 cross 5 and so on. And in fact, I am not even told you what the dimensions of the outputs are? Except that I be got some intuition that it seems that the output dimension is smaller than the input dimension right.

So, let us look at these in more detail and see what do these different outputs look like. Why do I care about these things? What do the inputs and the output sizes look like? Because these are your matrices that you will be dealing with this tell you these tell you how many parameters you are going to have.

These tell you what is the size of the memory that you need to load this entire network into your memory and so on it. So, that is why this computation is very very important

and I will have some more things to say about it towards the end of this lecture or some lecture ok.

(Refer Slide Time: 01:06)



So, if you first define the following quantities. So, we have the input which has a width W 1, height H 1, and depth D 1. So, if you are looking at the original image then the depth is going to be three in most cases it is going to be RGB ok. There is something known as the stride S, I will come back to it later on.

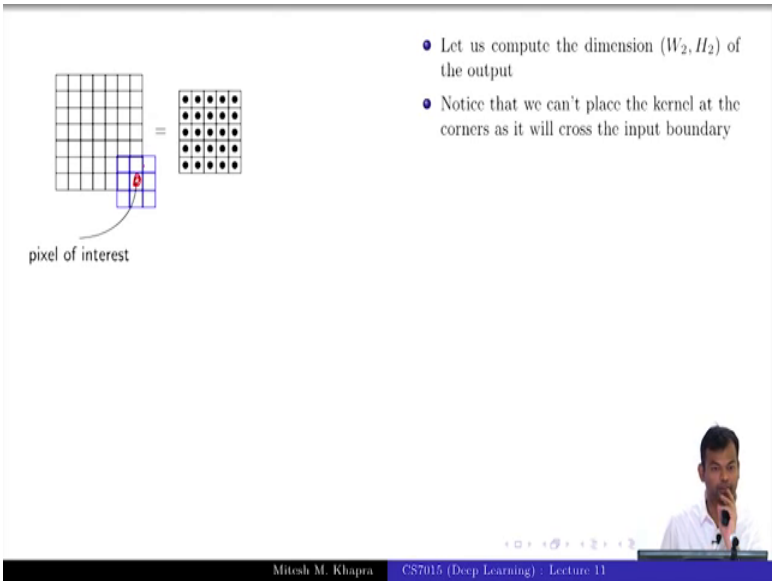
But I am just defining it here, or others just introducing it here. And then you have number of filters. So, I said that every filter that you apply you are going to get one feature map which is two dimensional. If you have K such filters, you will get K feature maps each of them is 2D right.

So, we will have something known as number of filters K and then you will have something known as the spatial extent of these filters. So, that is the number 3 cross 3; 5 cross 5 which I have been saying. So, that is known as the spatial extent I am going I am going to refer to it as F ok, and we are going to always assume square filters.

So, it is always going to be F cross F is that fine ok, and the depth of the filter is one more thing which I need to worry about. But I have already said that I am going to assume that the depth of the filter is the same as the depth of the input ok.

Now, the output is again a volume which is of size W 2, H 2, and D 2. And my quest is to find out how do I compute these W 2s, H 2s, and D 2s that is what I want to figure out. It is not very difficult, but I just want to do it properly, so that is what the setup is right. So, we have defined the sizes of everything on the input and the filters now you want to look at how do we get the output sizes ok?

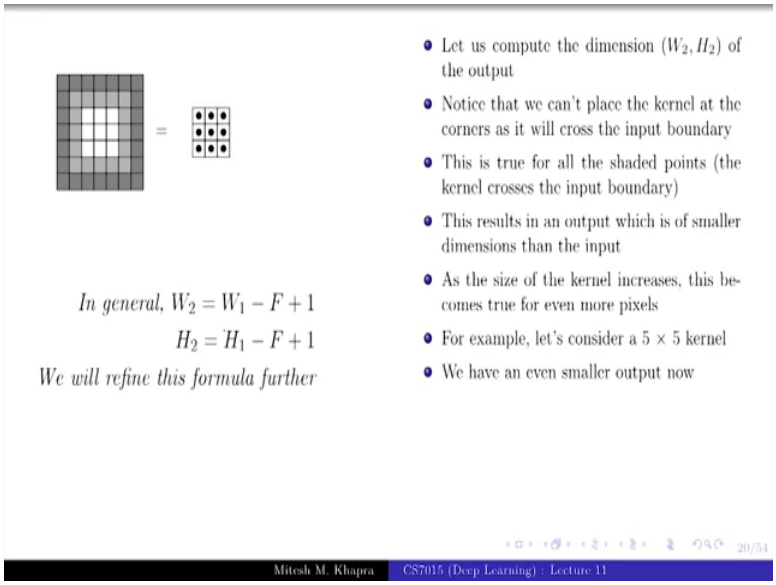
(Refer Slide Time: 02:39)



So, let us compute this for one example. So, this is my original image, so I am looking at a two dimensional input which I believe is 7 cross 7. And I am applying a 3 cross 3 filter to it. So, every time I slide the filter I will get one pixel in the output and I got the entire feature map.

Now it is obvious that the size of the output is less than the size of the input why is it so? Because there are certain pixels at which I cannot place the filter why? You will go out of the boundary right. So, I cannot if this is my pixel of interest, I cannot place my filter there. Because then the filter will go outside the image and I do not know what the average to come to how to compute the average right those values are undefined do you get that? Ok.

(Refer Slide Time: 03:26)



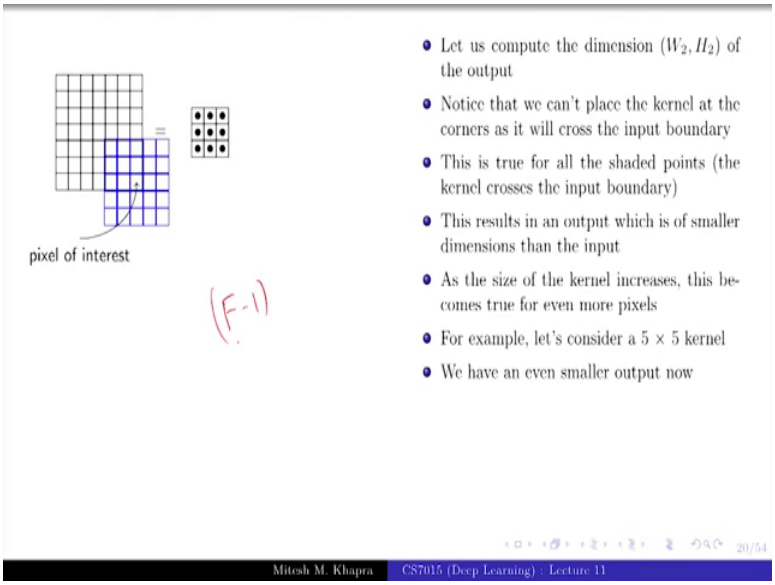
So, in general for let me see for the 3 cross. In fact, this is true for all these pixels which have been shaded or any of these pixels I cannot place the filter because you will go outside the boundary. So, now, for a 3 cross 3 filter what is the reduction in the size of the output compared to the input? The width decreases by 2 and the height decreases by.

Student: 2.

2 right. So, can I be bold enough and say that the width and height decreases by not yet ok.

So, let us see if we had a 5 cross 5 kernels ok. Then which are the places at which I will not be able to place the filter? These two shaded boxes and both these boxes I cannot place the filter because the filter will go outside the image.

(Refer Slide Time: 04:13)



So, now can you tell me how many? So, in this case the size reduce is by how much? The width reduces by.

Student: (Refer Time: 04:20).

No the width reduces by 4 sorry and the height reduces by?

Student: 4.

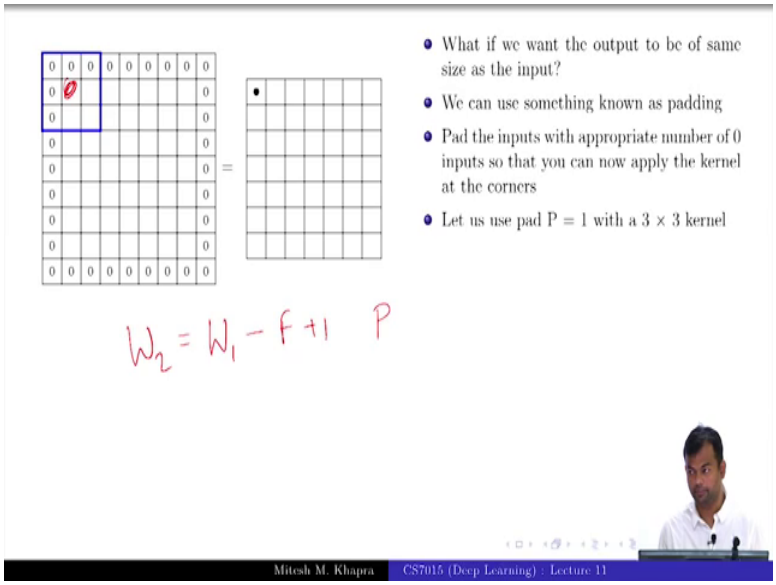
4. So now can I say that the width and height actually reduce by F minus 1, where F is the size of the filter is that ok? How many of you get this? So, you did not get this? No. You did not get this? Ok. So, in the 3 cross 3 case you see that that is one row and one column on each side left and right which I cannot apply ok.

So, let us focus on the columns. So, columns give me the width right. So, when I have a 3 cross 3 filter there are two columns which get subtracted, because these are the boundary columns. When I have a 5 cross 5 filter how many columns get subtracted? 2 on the left hand side, 2 on the right side is a total of 4.

If I have a 7 cross 7 filter; 3 on the left hand side, 3 on the right hand side, so total of six. So, you see the relation it is always F minus 1 right. So, your new width and height is always going to be W 1 minus F minus 1 which is W 1 minus F plus 1 is that ok.

Everyone gets this and same for the height the width and height are going to be symmetric. Because the filter we have chosen to be symmetric it is F cross F ok.

(Refer Slide Time: 05:28)



Now, but what if we want the output to be the same as the input? What do we do? Padding ok, you can use you know something known as padding. So that means, now I will have a boundary of 0's. So, I am saying that this is my original image and outside it there is a black border or a white border I do not know whether 0 stands for black, or white it is embarrassing, but.

Student: Black.

Black ok. So, it is a black border outside the image ok. And now I am going to take an average that way.

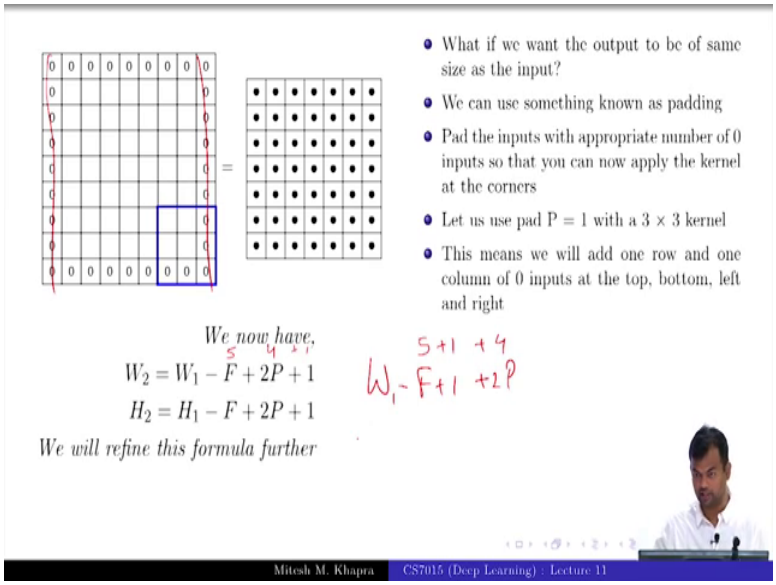
Now, this was the pixel earlier on which I could not place the filter. But now I can artificially place on it assuming that it there is a black boundary around it. So, now what would be the output size?

Student: (Refer Time: 06:08).

Same as the input. So, now can you tell me? So, I have W 1, I have F and now I have something known as P also ok. So, I know that w output rather w 2 is the output is 1.

Now when I add the padding what would the formula be? 2 P is that fine? Everyone gets that? Ok.

(Refer Slide Time: 06:33)



So, now if I have a 5 cross 5 filter and if I want the output size to be the same as the input size. What is the padding that I should use?

Student: P.

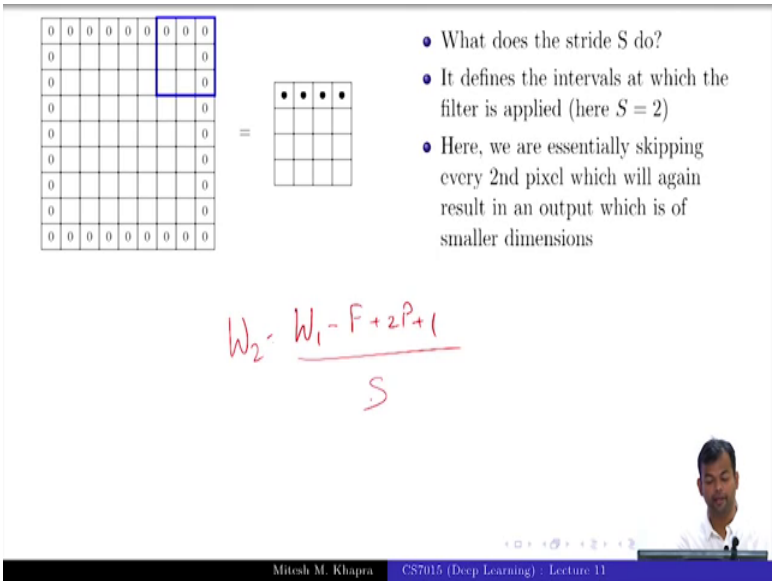
P is equal to 2 right that is clear from the example that we saw that there were two shaded columns and rows which were problematic. So, I need to do a padding of 2 and then if I substitute in this formula you can just see right. So, 5 plus 4 plus 1 right, so you will get back W 1 is that fine is that ok. How many if you get this? How the formula works with the padding? How many of you do not get this please raise your hands? You do not get this ok.

So, remember in the 3 cross 3 case there was one column on the left hand right which was a problem. So, when I say a padding of one I add one column to the left; one column to the right; one column to the bottom and one column to the top.

And that is exactly the problematic region in the 3 cross 3 case right. So that means, this was my original formula ok. Now the new width is going to be plus 2 times the padding which I am going to use. Because I have used one padding here, and one padding here right. Now, in the 3 cross 3 case that is ok.

Now in the 5 cross 5 case how many columns are problematic? 2 so; that means, I have use a padding of 2, when I say a padding of 2 I add 2 on all the sides. So, now again if you substitute in this formula so you would have W 1 minus 5 plus 1 plus 4. So, that will give you back W 1 right. So, that is how it is right.

(Refer Slide Time: 08:08)



Now, the question is you have if you have taken care of filter size and padding. Now the other thing that we need to look at is stride ok. So, the stride defines the interval at which the filter is applied. Now what do I mean by that?

So, remember that stride is basically a step right the same definition as it is applies to walking rate. So, right now what we were doing is we placed the filter at as this pixel at the center, then this pixel has the center, and then this pixel has the center. Instead of that I could take a that means, my stride is 1, I am taking one step at a time. So, if I do a stride of 2 what would happen?

Student: (Refer Time: 08:48).

I will apply two alternate pixels right. So, this is how I will move. So, then what would happen to my output size if my stride is 2?

Student: (Refer Time: 08:57).

What would happen?

Student: (Refer Time: 08:59).

It will become half ok. So, what would the formula be then? So, so far my formula was now if I have a stride of 2 what would the formula be?

Student: (Refer Time: 09:12).

They divide the whole thing by 2.

Student: (Refer Time: 09:16).

By S right. If it was S was 3 then this would have become one-third roughly right. If S was 4 this would have become one-fourth. So what should I divide by?

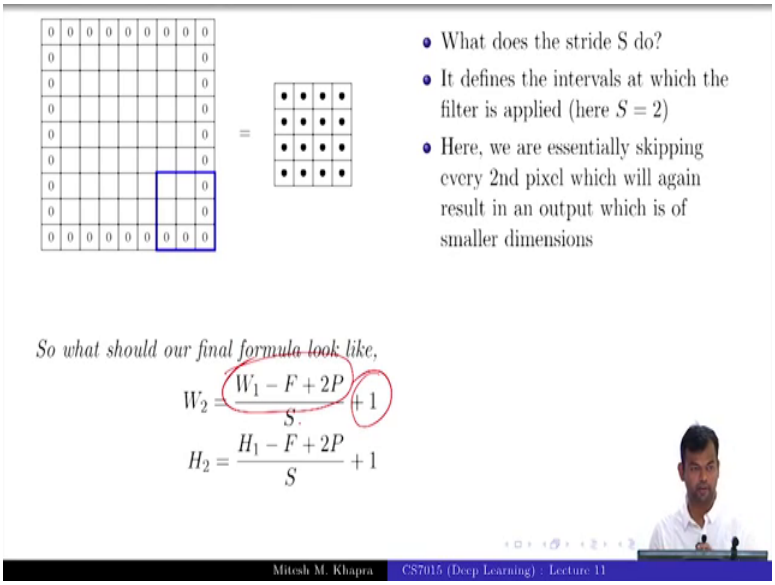
Student: S.

S. So, I should divide the whole thing by S.

Student: (Refer Time: 09:32).

S ok.

(Refer Slide Time: 09:36)

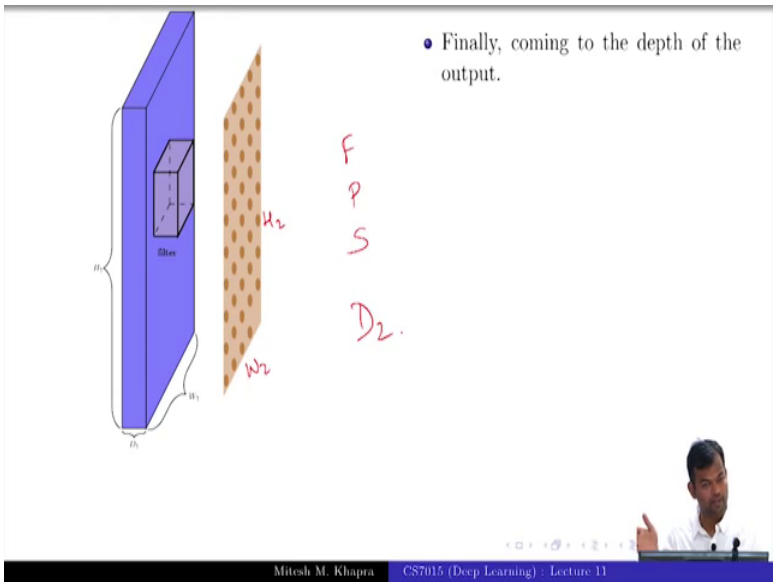


So, it turns out that is not exactly that, but you get the intuition and you can work out the formula. So you do not divide this by S and you will figure it out, it is easy to see because of some ceiling and flooring and things like that. So, you can go back and check

that out and basically you could just think of this that, this was your original weight in the absence of stride, or rather than the stride was one.

So, now if you are going to take longer strides you have two account for that if you take a stride of 2, stride of 2 your width is going to become half; you should take a stride of 4 your width is going to become one-fourth is that fine? Do not worry about this additional plus point you can go home and figure it out.

(Refer Slide Time: 10:06)



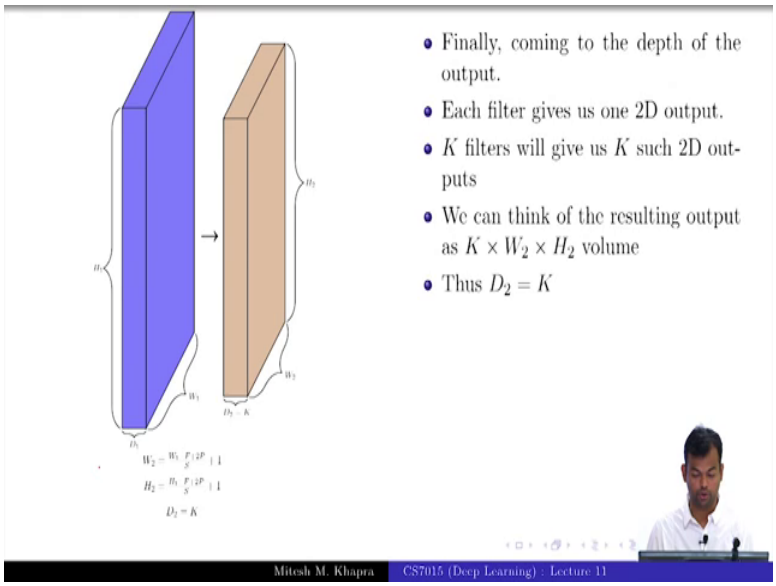
Finally coming to the depth of the output what would the depth of the output be? So, let me just see right now all our formulas were W 2, H 2 in the presence of filter padding and size stride sorry but we never had a formula for D 2, so that is what I am asking.

Now what is the depth of the output? Same as the every filter is going to give you one two dimensional output. If you have K filters.

Student: K.

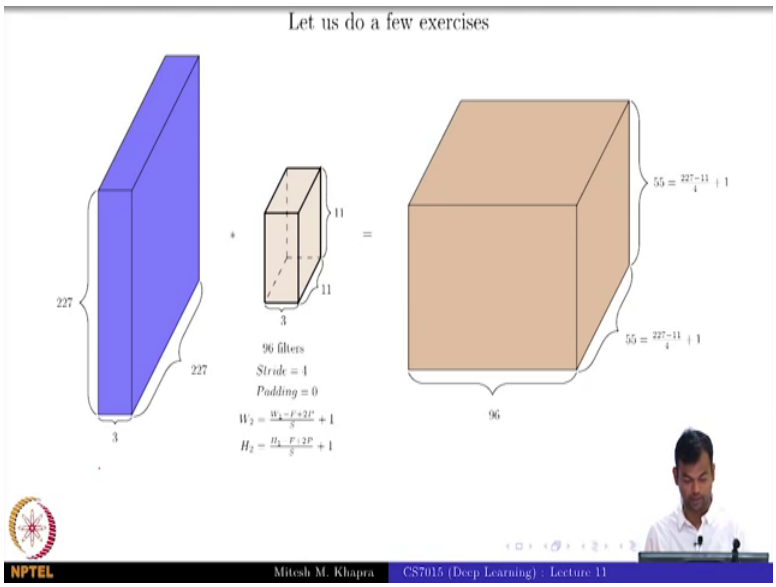
You will get K two dimensional outputs. That means, the depth of your output is K right.

(Refer Slide Time: 10:38)



So, the depth is very simple it is just equal to K the number of filters that you have. So, I want you guys to note down these three formulas.

(Refer Slide Time: 10:48)



Now, let us do some exercises. So, this is my original input which is 227 cross 227 cross three; I have decided to apply 11 cross 11 filters and I am not going to tell you the depth of the filter, because it is going to be the same as the input ok.

And I have 96 such filters I have decided to use a stride of 4 there is no padding. Can you tell me what is the output volume going to look like? What are the dimensions of the output volume? Ok. So, D 2 is simple. What is D 2? 96. What is W 2?

Student: 55.

55. H 2 you guys have the last class fine. So, similarly you can do it for so actually the computation which I had that this was just not some random computation. This is actually the first convolution layer from Alexnet right, so one of the popular architectures that we are going to cover later on right.

So, this is what Aalexnet does at its input. It takes the RGB input and gives you a volume of this slice this side and then there is something else with this volume right. So, we will see that later on. There is one more exercise you can do it later on, I do not want to do it now ok.