Deep Learning
Prof. Prabir Kumar Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture – 34
Convolution

Hello, welcome to the NPTEL online certification course on Deep Learning. In last few lectures, we have talked about one tool of deep learning which is auto encoder and you have seen that we have talked about the various versions of the auto encoder, like under complete auto encoder, sparse auto encoder, denoising auto encoder and so on. And we have also talked about that how the auto encoders can be trained or pre-trained layer by layer. And then of course, you have to have a final round of training which is end-to-end training.

What we have seen in case of auto encoder is that the auto encoder represents the input signal in a compressed domain, where the dimensionality of the compressed domain representation is significantly lower than the dimensionality of the input raw data. And while doing so, the auto encoder tries to capture the salient features or important features of the input data and it tries to discard, any sort of redundancy which is present in the input data.

And we have also seen that how the auto encoder output compares with the output of principal component analysis. And we have observed that under certain cases, the principal component analysis output of the principal components of the input data and the encoding output which is given by the auto encoder, they almost converge.

We have also talked about some applications of the auto encoder, like we can use auto encoder for classification purpose, where the decoder part of the auto encoder is not used. So, once the auto encoder is properly trained, then the encoding half of the auto encoder that gives you the compressed domain representation or it gives you the output which are salient features of the input data. And using those salient features, now you can go for classification or understanding of the input data.

And the various applications of such classification can be say image segmentation, where every pixel of within an image can be classified into one of the given classes. So, once

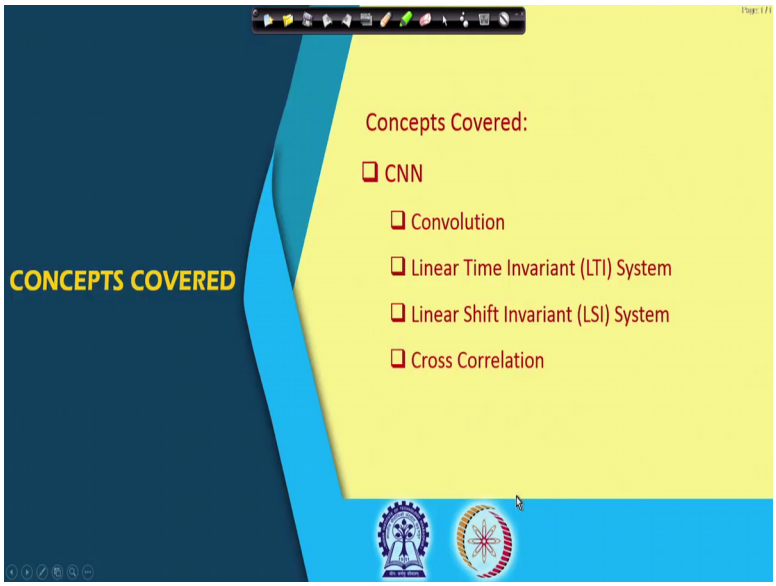
every pixel in the input image is leveled with a class, then the collection of all the pixels which are labeled to the same class forms a segment. So, segmentation is one of the applications of such auto encoder.

Similarly, we have also said that auto encoders are widely used for anomaly detection. So, how is it used for anomaly detection, when you train the auto encoder you train the auto encoder with all the inputs which are normally inputs, they are not abnormal; so as the auto encoder is trained with those normal inputs. So, now any instance of the normal input if it is fed to the auto encoder, the auto encoder will be able to reconstruct that input faithfully.

But if any input or any part of the input image or the video which the auto encoder has not seen before that means, while training such inputs were not fed to the auto encoder. So, those parts of the image of those parts of the video the auto encoder will be unable to reconstruct properly, as a result the reconstruction error in those areas will be quite high. So, based on this where whenever you find that the reconstruction error is very high, we can assume that those are the areas which are abnormal areas that means, such sequences or such data does not normally occur in normal data. So, auto encoders can also be used in abnormality detection.

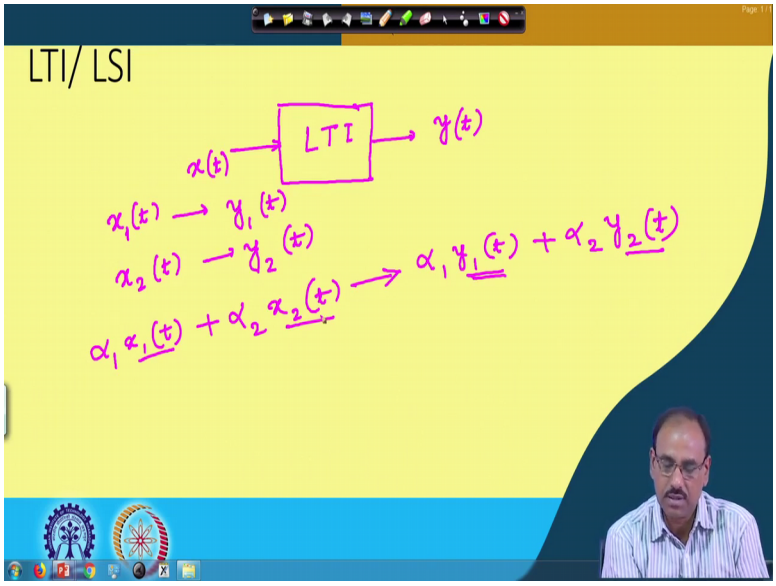
Today we will start discussing on another very very important tool of deep learning which is convolutional neural network or in short these are known as CNN. So, we are going to talk about CNN or Convolutional Neural Network.

(Refer Slide Time: 05:01)



So, when I talk about this convolutional neural network we will touch upon the topics of convolution and while we talk about convolution we will also try to see what is meant by Linear Time Invariant System or LTI system, we will also talk about what is Linear Shift Invariant system or LSI system. And then we will also discuss about a very very similar concept which is known as cross correlation, and we will also try to highlight that what is the difference between a convolution operation and a cross correlation operation.

(Refer Slide Time: 05:41)



So, let us first try to discuss what is convolution. So, convolution is actually an operator. So, given a linear time invariant system or a linear shift invariant system; the convolution operator actually tells you that if I have an input signal to a linear time invariant system, then what will be the response of the linear in time invariant system to that input signal that means, what will be the output of the linear time invariant system if an input x is given to that linear time invariant system.

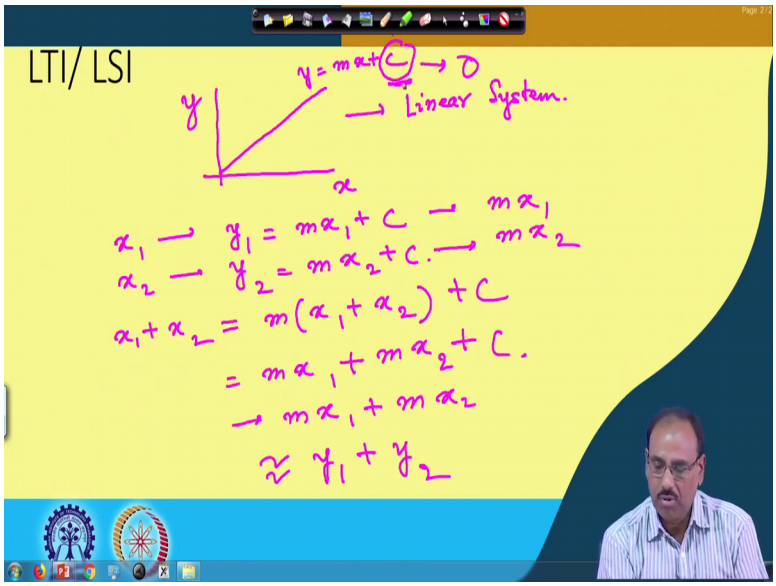
Now, all the linear time invariant systems are actually characterized by a particular characteristics of that linear time invariant system which is known as impulse response, I will come to that a bit later. So, first let us try to see what is a linear time invariant system or a linear shift invariant system. So, given any system let me just draw it as a box, so this is my system which I am terming as Linear Time Invariant or LTI system.

So, if we give an input signal say x t to this LTI system, the output of it will be given by signal y t. So, if it is linear, so when I talk about linear time invariant, you find that there are two concept; one is linear, one is time invariant. So, the system is linear if the linearity property holds true for that particular system that means, given an input signal x t or say x 1 t, the output signal is y 1 t that is the response of the system to an input signal x 1 t.

Similarly if we feed an input signal say x 2 t, suppose the output of the system is y 2 t; so given x 1 t, the output is y 1 t; given x 2 t, the output is y 2 t. Then if the linear if the system is linear, then I must have or the system must satisfy a property that given an input say alpha 1 x 1 t plus alpha 2 x 2 t, the output of the system or the response of the system to this input alpha 1 x 1 t plus alpha 2 x 2 t must be alpha 1 y 1 t plus alpha 2 y 2 t.

So, we find that what is y 1 t, y 1 t is the response of the system when the input is x 1 t; similarly y 2 t is the response of the system when the input is x 2 t. So, when given x 1 t, the output is y 1 t and given x 2 t output is y 2 t. So, if the system is linear that indicates that if I give an input which is alpha 1 times x 1 t plus alpha 2 times x 2 t; the output of the system must be alpha 1 times y 1 t plus alpha 1 times y 2 t. So, this property must holds true hold true if the system is a linear system. Now, suppose a system is having characteristics something like this.

(Refer Slide Time: 09:41)

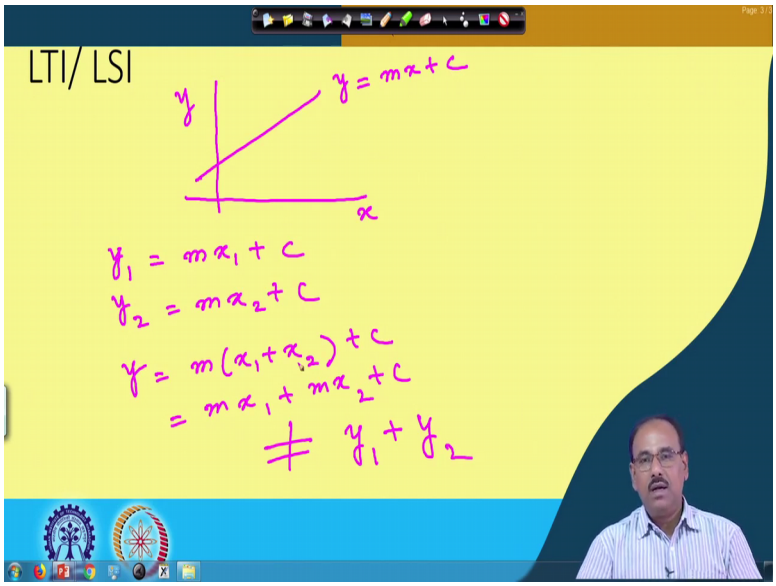


So, this is my input x, this is my input y. Obviously, this is a linear time this is a linear system why, because if the slope of this system suppose this equation is given by y is equal to m x plus C. Now, given an input x 1 output of the system will be y 1 is equal to m x 1 plus C; similarly given an input x 2; output of the system will be y 2 is equal to m x 2 plus C.

So, now if I give an input which is say x 1 plus x 2, the output of the system must be m into x 1 plus x 2 plus C which is m x 1 sorry, yeah. So, given an input x 1 plus x 2 output of the system is m into x 1 plus x 2 plus C, which is m into x 1 plus m into x 2 plus C.

So, when I have an equation of this form, here you find that value of C is equal to 0 right, as this straight line passes through the origin so value of C is equal to 0. So, given value of C equal to 0 this y 1 is nothing but m times x 1 and y 2 is nothing but m time x 2. And here this x 1 plus x 2 is nothing but m's times x 1 plus m times x 2, so which is nothing but y 1 plus y 2. So, if this constant C is equal to 0, the system becomes a linear system.

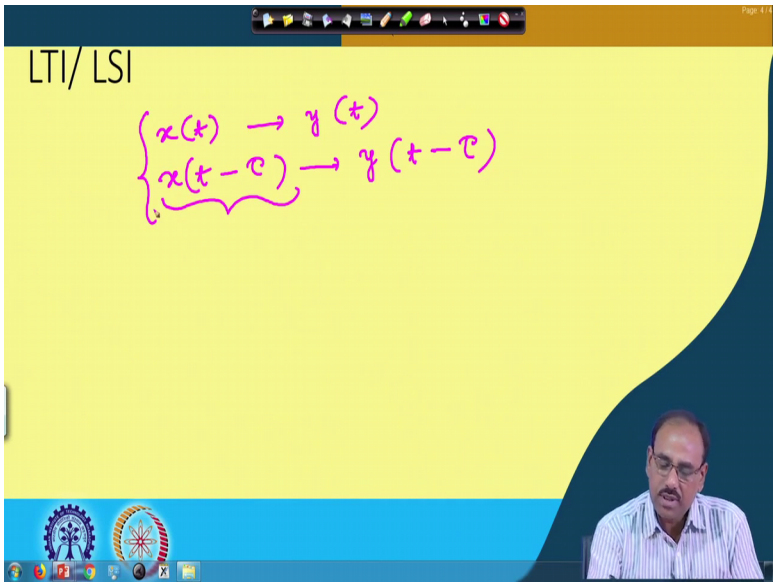
(Refer Slide Time: 12:19)



Now, if this constant C is not equal to 0, so in that case the characteristics of the system will be something like this, this is x, this is y and my transfer function here it is given by y is equal to m x plus C where C is nonzero. So, over here given an input x 1, y 1 will be m x 1 plus C; y 2 will be m x 2 plus C.

And now if you give input as x 1 plus y 1, x 1 plus x 2, then my output will be say y which is m into x 1 plus x 2 plus C which is nothing but m into x 1 plus m into x 2 plus C, which is not equal to y 1 plus y 2; because y 1 plus y 2 is m into x 1 plus m into x 2 plus twice C, which is not same as this. So, though a system having these type of transfer function appears to be a linear system, but actually it is not a linear system, because when I feed x 1 plus x 2 as input to the system, my output is not the sum of the responses when you feed x 1 and x 2 separately ok, so this is not a linear system.

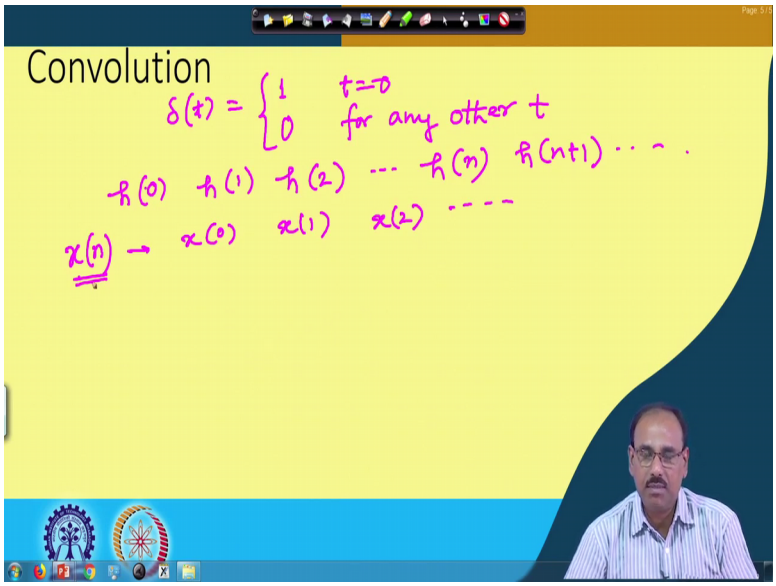
(Refer Slide Time: 13:57)



Now, what is time invariance? Time invariance means that given an input signal say x t, the corresponding output is y t. Now, if the input is delayed by some amount say tau, so instead of t I put it as x t minus tau, so the input signal is delayed by a delay tau; correspondingly the output should also be delayed by the same amount. So, if y t is the output given an input x t, then y t minus tau must be the output when the input is x t minus tau. So, this is what is time invariant that is, if the input is shifted, the response will also be shifted by the same amount. So this we talk about linear time invariant system, when we are in the time domain that means my input signal is a time domain signal it varies with time.

But when we talk about images, images are special domain signals where the intensity value or the color value varies over speed over space. So, there instead of talking about a time invariant system we talk about space invariant system, because our signal is a special signal it is not a time domain signal. So, now given this idea about the linear time invariant system and linear space invariant system, let us see what does the convolution mean.

(Refer Slide Time: 15:39)



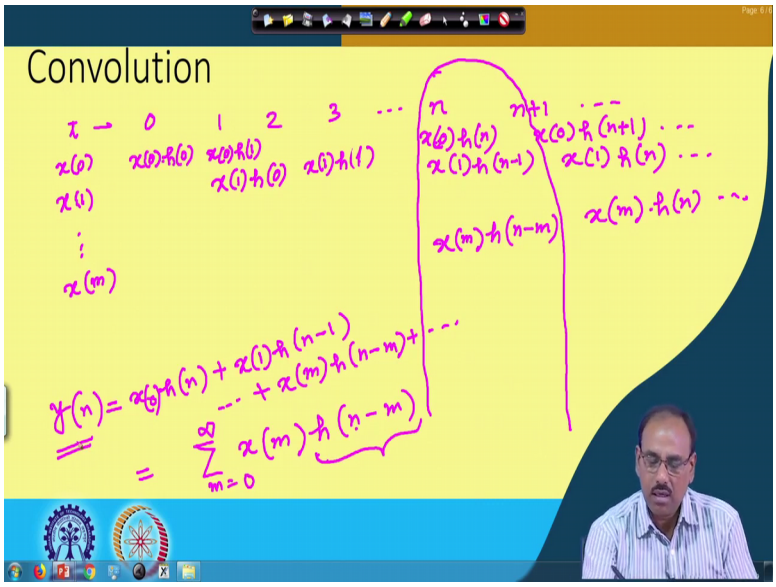
So, as we said that an linear time invariant system is completely characterized by its impulse response. So, for that let us first define what an impulse is. So, an impulse is defined like this sign impulse delta t is equal to 1, when t is equal to 0 and this is 0 for any other value of t, for any other t or t is not equal to 0. So, only at t equal to 0, delta will be delta t will be equal to 1 and for any other value of t where t is not equal to 0, delta t will be equal to 0 ok, so this is what is an impulse.

So, given a linear time invariant system the response of the system to an impulse is what is known as impulse response. Now, given any such impulse to a linear time invariant system, the impulse response let us consider in discrete domain at time t equal to 0, so impulse response is h 0; at time t equal to 1, it is h 1; at time t equal to 2, h 2 so on; at t equal to n, it is h n, then h n plus 1 and it continues like this. So, this is the impulse response of a linear time invariant system.

Now, given any signal say x n entering from 0 to whatever the number of samples we have we are talking about discrete signals, this x n can be represented as sum of scaled impulses, say this x n is nothing but x 0, x 1, x 2, it continues this way. Now, what I can assume is, because here I have the value of x 0 only at t equal to 0. So, I can consider this as a product of amplitude of x 0 times delta t, this one can be the product of the amplitude of x 1 times delta 1, because delta 1 will be equal to 1 at t equal to 1 and it will be equal to 0 for any other value of t and so on.

So, I can consider my input sequence which is my signal as a scaled version or some of scaled delta functions. So, assuming this and considering the property of linear time invariant systems, I can represent the output sequence like this.

(Refer Slide Time: 18:35)



So, let us put as say on this side I have t equal to 0, 1, 2, 3 say t equal to n, t equal to n plus 1 and so on. So, you are feeding x 0 at t equal to 0, so with this x 0, we said that the impulse response of the system is given by h 0, h 1, h 2 and so on.

So, when we are feeding an input x 0, sample x 0 at time t equal to 0 the output will be given by x 0 times h 0; at t equal to 1, it will be x 0 h 1; similarly here it will be x 0 h n; here it will be x 0 h h n plus 1 and so on. Similarly, you are feeding input x 1 at time t equal to 1; at time t equal to 0, I did not have x 1. So, it is I can consider that this is a delayed input which is delayed by a time equal to 1.

So, with x 1 my output will be something like this, so here it will be delayed. So, I will have x 1 into h 0, which will be available at time instant t equal to 1. Similarly here it will be x 1 h 2, which will be available at sorry, x 1 h 1 which will be available at time instant 2. Here it will be x 1 h n minus 1 which will be available at time instant n, here it will be x 1 h n which will be available at time instant n plus 1 and so on.

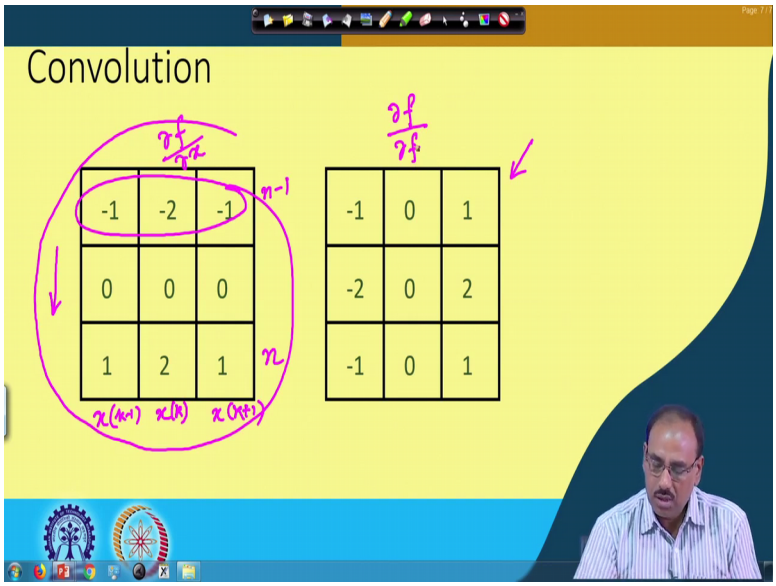
So, considering this way when I have an input signal x m, input sample x m at time instant n my output will be x m into h n minus m, here it will be x m into h n and so on.

So, considering this you find that if I want to find out, what will be my output at t equal to m which is nothing but y n. So, output at t equal to n which is y n is given by x 0 h n plus x 1 h n minus 1, continue like this plus x m h n minus m plus so on.

So, I can simply write this as sum of x m h n minus m, where m will vary from say 0 to infinity if I have an infinite sequence of input samples. So, what it gives apparently it appears that your impulse response h is flipped, because I have h minus m; it is flipped shifted to n, then multiplied point by point to x m and then added from m equal to 0 to infinity and that is what gives you the output at y equal to n.

So, when you compute convolution, the convolution is computed by flipping your impulse response and then point by point multiplication of the with the input signal and then sum them up that is what gives you the convolution output at time t equal to n.

(Refer Slide Time: 23:19)



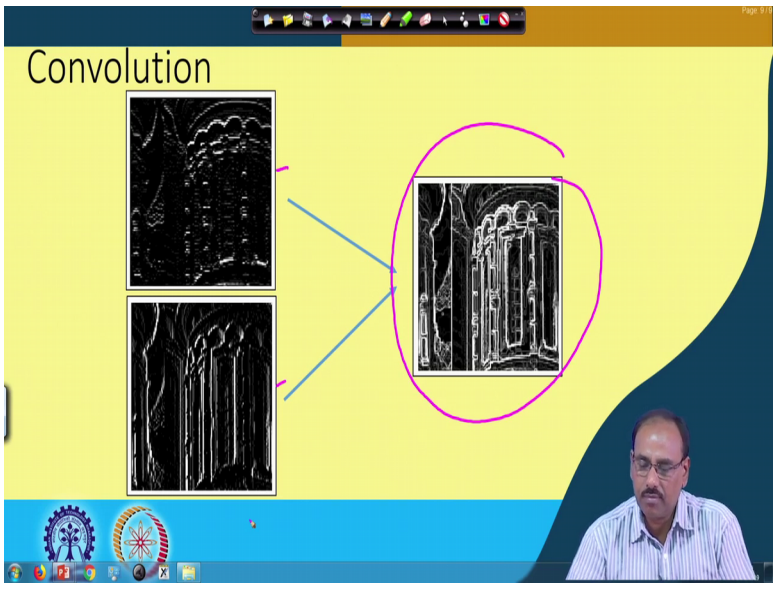
So, let us see that what does this convolution actually gives you physically. So, let must let me assume that in a special domain signal in a two-dimensional case, suppose these are the impulse responses of a linear space invariant system, because we are talking about images. So, these are the two responses impulse responses of a linear time linear space invariant system and when your input signal or an image is convolved with these two what output do we get.

Now, before that let us see that what is the characteristics of these two impulse responses. You find that in the first case this particular impulse response, come to this last row what it does; it does an weighted sum of the inputs of 3 consecutive pixel values in a particular row. So, suppose here I have a pixel location say x say k minus 1, this is x k and this is x k plus 1. So, this row will simply perform x k minus 1 plus twice x k plus x k plus 1, so this makes an weighted sum of the 3 pixels.

Similarly, this row also makes an weighted sum of the 3 pixels on two rows above it and then negate it and then finally this entire operator will give a sum of these two that means, this operator gives you the difference of weighted sum of the pixels belonging to say this is nth row and this is the n minus first row. So, it gives you the difference of the weighted sum of these 3 pixels in the nth row and these 3 pixels in the n minus first row or effectively these gives a derivative in the vertical direction.

So, if I put this vertical direction as x, so this is an input signal as f, so this gives you a derivative in the vertical direction which is given by del f del x. In the same manner you can find that this template gives you a derivative in the orthogonal direction, so which is del f del y. So, these are the physical meaning of these two convolution operators. So, when you convolve your input image with this convolution operator, what is the kind of output that we get.

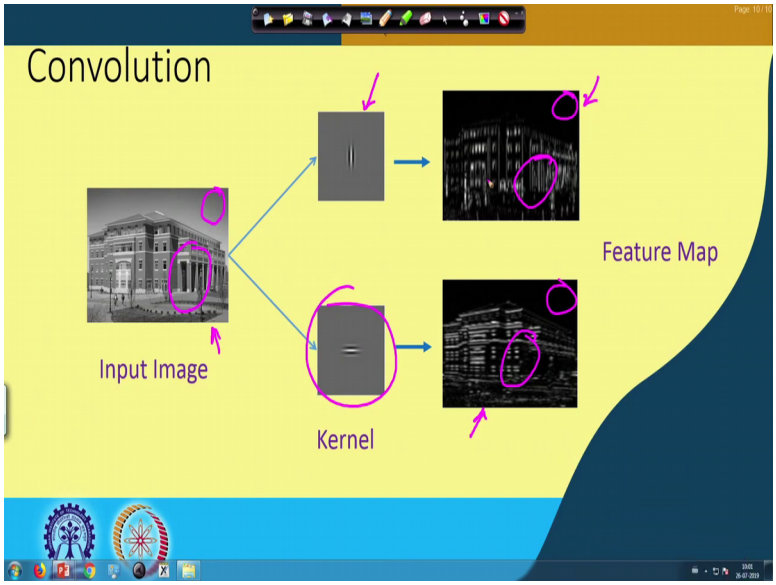
(Refer Slide Time: 26:01)



So, given an input image of this form, this is the derivative in the vertical direction and this is the derivative in the horizontal direction. So, obviously when you take the derivative in the vertical direction, all the discontinuities or changes in the intensity in the horizontal direction of the edges which are present in the horizontal direction, they will be highlighted. Whereas, if you take the derivative in the vertical direction, then all the lines which are if you take the derivative in the horizontal direction, all the lines which are vertically oriented they will be highlighted, so that is what you find in this case here, you are getting all the lines which are mostly horizontal and here you are getting all the lines which are mostly vertical.

And if I combine these two the outputs of two such operators, then this is the edge map that is what I get. So, in this final edge map what I am saying, if I combine these horizontal edges and vertical edges, I get the complete edge map. So, this edge map actually tells you some features of the input image, because in many cases the images are the information in the image is actually contained in the edge map ok. And using this, your recognition or understanding of the image will be very convenient.

(Refer Slide Time: 27:43)



In the same manner if I take say another template, say for example here, if I input is input image is something like this and this is one of the convolution templates, then the corresponding output will be this one; where this input image is convolved with this

convolution kernel. Similarly, if I take this convolution kernel and with this convolution kernel you convolve this input image, this will be the output.

So, here again you find that these two outputs which are known as feature maps, they gives you the salient features of the structural information which is present in the input image. Here you find that here we have an uniform region, so here in both of them they have become 0. So, where there is no structure in the input image, in the feature map I do not get any information. Whereas, in all these areas where you have structures over here, over here; where we have some structural information, in the feature map I have important information's in those areas, so this is what is convolution.

So, what you said is convolution is actually the response of a linear time invariant system or a linear space invariant system to the input signal, which is presented to that LTI or LSI system. And we have also said that LTI or LSI systems are completely characterized by its impulse response. So, when I have an LTI system characterized by its impulse response or an LSI system characterized by its impulse response, then given an input signal the corresponding output of the response of the system will be simply convolution of the input signal with the impulse response of that particular system.

So, today we will stop here, next day we will talk about the other concept which is cost correlation and computationally these two are more or less similar. And in most of the cases, people get confused between what is the convolution and it what is the cross correlation. So, in order to clarify that in our next class, we will we will discuss about cross correlation and we will also try to find out that where they are similar and where they are de-similar.

Thank you.