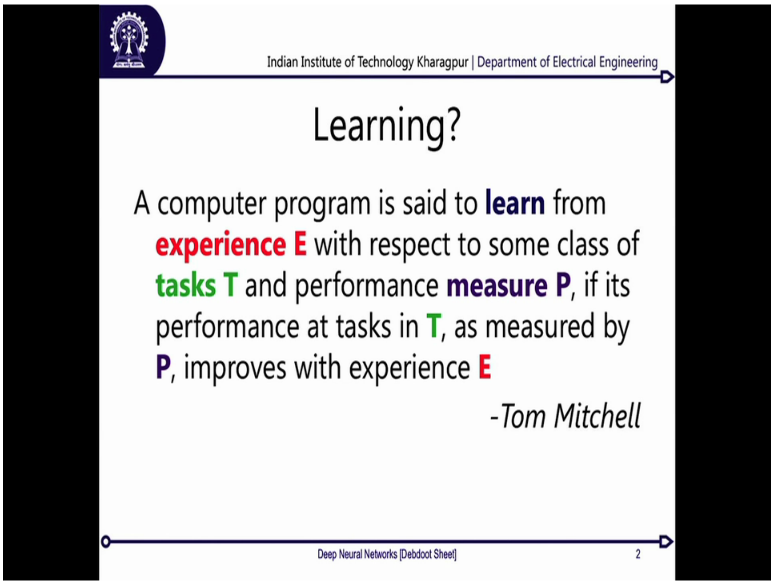
Introduction to Medical Imaging and Analysis Softwares
Professor Debdoot Sheet
Department of Electrical Engineering
Indian Institute of Technology Kharagpur
Module 3
Lecture No 14
Deep Learning for Medical Image Analysis

Welcome to yet another exciting topic on Deep Neural Networks and this is where we are going to continue from our previous discussions on Neural Networks and that was about using a very simple Neural Network and then we use just 100 neurons over there to train and then classify these microscopic images of WBCs into whether they are leukemic or not leukemic. Now here I am going to speak about something which is weight upcoming in the community and actually has got the community by a buzz and that is called as Deep Learning and neural networks are in fact, one of the major runners in this whole fact whole family of deep learning engines in which we are going to work. But before I start down I would come down to this basic definition of learning.

(Refer Slide Time: 1:12)



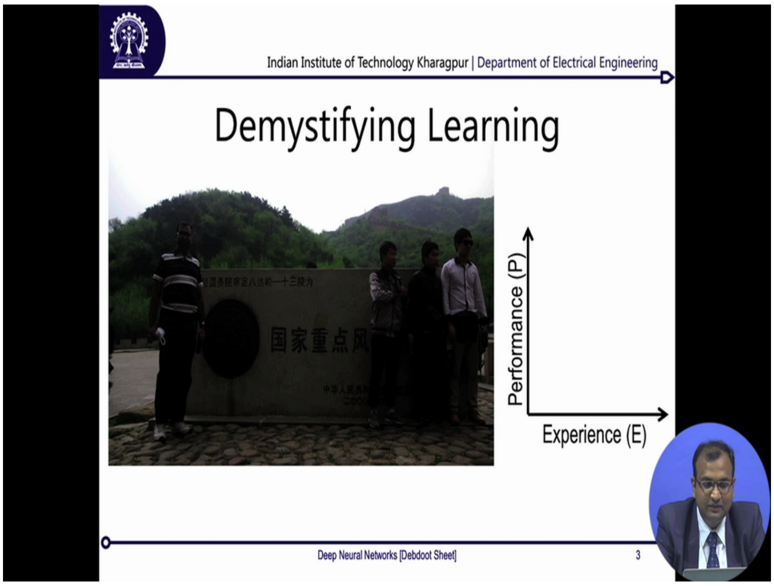
When we say learning it is not by definition that there should be a set of features and then I should be able to classify something and do it, this is one of the misconceptions people have there might. So this can be a regression problem as we had seen down for decision trees and random forest as well. For a very simple stuff let us define what this learning is and the definition which Tom Mitchell puts down is quite simple. He says that a computer program is said to learn from experience E with respect to certain class of task T and performance measure P.

Now you would be saying that it is learning only if its performance at a task T as measured by P is going to improve with experience E. So you need to know one thing that since we have already done that basic class on neural networks, one thing in mind is that this experiences the number of epochs and the number of times you are looking at the same data. Now as you were looking at the same data as the whole neural network was looking at the same data, it was trying to look at the data and find out what is the error and then again back propagate the error, this was the looking back phenomena on it.

Now, as it was updating itself what you can see is that it was doing the same task which was just classifying the data, but the measure of performance which is P, which is at cost function over there was increasing as it was looking at more number of samples over a longer duration of time, the same samples over multiple number of times or a larger number of samples. So that is where we said that the neural network was learning because this definition was being satisfied over there.

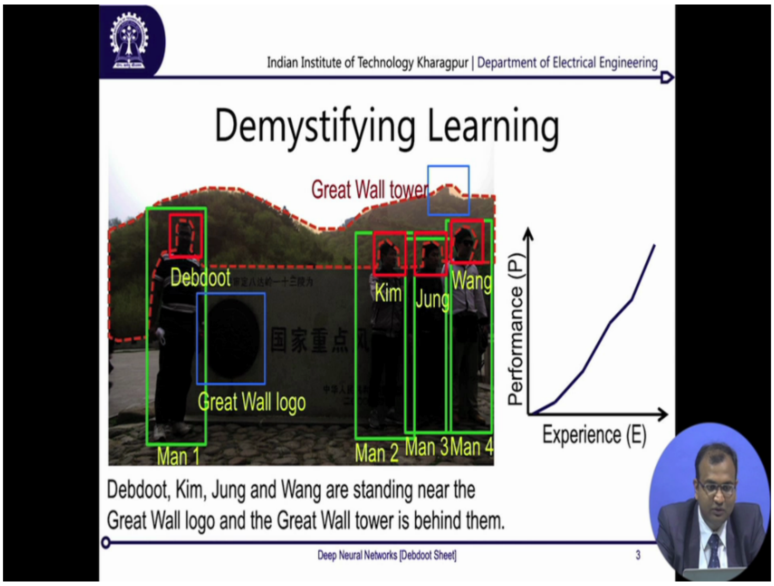
Now over here, as per the definition and by its own basic concept, it is nowhere said that you need to have features over there, nowhere we are saying that you need to be able to classify. Now that is where we get into the actual problem. Now if we look into learning as its own paradigm, I would take a very simple example from very scenic images, I am not going to take down very specific medical images over there because when we do have a much wider understanding for them.

(Refer Slide Time: 3:39)



Let us let us look at a very simple problem of seen image classification and on that let us say that this is one of the images, which I am going to take. Now you would often be seeing just a if you have not been to this particular place then what you would be seeing is just 4 people over there somehow if you are looking into much more detail you would see me if I standing on as one of the persons in that photograph as well and there is something written is Chinese and it is some sort of a forest and mountain, this is what you would be seeing over there.

(Refer Slide Time: 3:58)

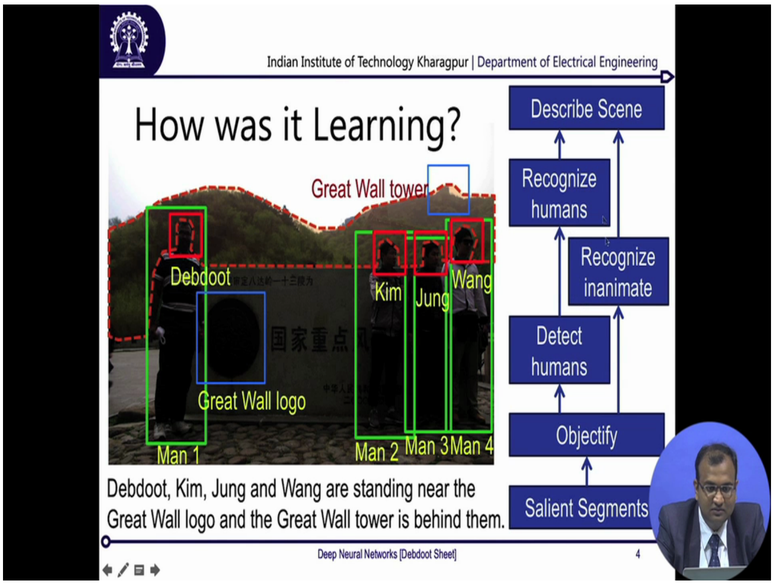


Now if a computer is given down with this simple task of, can it interpret this particular image, then how is it going to start? Now, what it would be doing is something of this sort, it is going to break this whole image into some small-small regions, ok. Now, in the whole time it is not seeing just one image may be it is seeing millions of those images and over time it is learning what to do. So it is going to do something like it is going to break it down into multiple number of such small blocks and then identify what each of these blocks mean over there and then from there it is going to look into certain things which is knows over there as having seen something earlier and then finally it is able to look into the whole stuff and identify people and then some sort of create a sentence.

Now, if you look at this curve over here, the fun which you see is that as you are increasing your experience, which is the more the number of images you are looking through your performance is going to increase, which is sort of say accuracy in identifying objects so the number of objects you can identify. So over epochs or over the whole dataset this is increasing and this is when we say it is learning.

Now all if these names are basically not real so they are just indicator names in any way. So I do not know whether they these were the real names or not. But this was just to make it as a much more compiling problem and give it a much more personal flavor. Now, let us look at what it was doing when it was trying to learn and that is about how it was learning. So you have this whole image, so the first thing your computer program runs to do is, divide it into number of segments which are called at Salient Segments or these are the segments which define a certain object present over there.

(Refer Slide Time: 5:48)

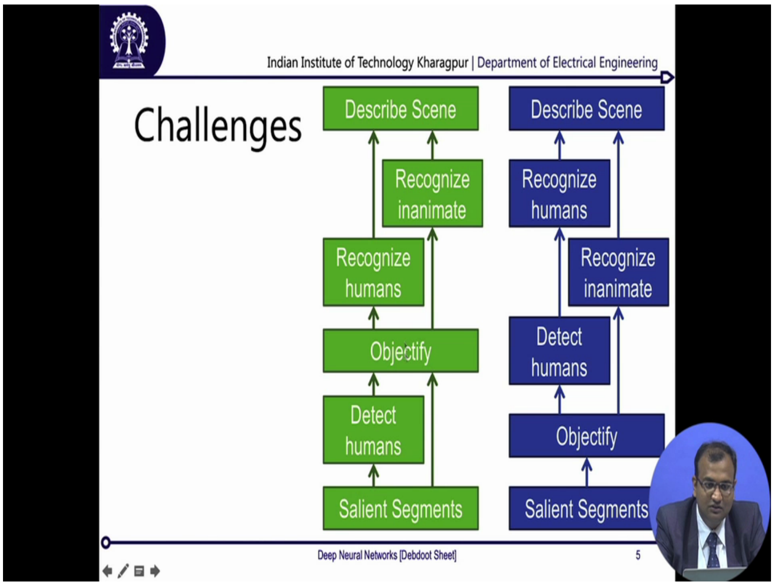


Now on each of these Salient Segments it would try to objectify whether it can identify certain objects present in those Salient Segments over there. Now from the objectification it will try to detect what those objects are and say it is able to detect humans over there by seeing that there are human like faces. And then say on the other side of it, some of these segments which were not properly detected although they were objective and they were not detected, they might be going down onto another pipeline where the purpose of this pipeline is to recognize these inanimate objects.

Now from there it can recognize these humans if there is a large database over which it has looked at all the faces of those humans. And from there it can actually go on to describe this whole seen by using this complete information. Now this is a way in which a computer program is going to do and inherently if you look at to this whole paradigm in which it is solving, you would be seeing that this is a hierarchical framework in which it is solving, this is necessarily hierarchy, so you need to first find out the Salient Segments and then you would be objectifying, then some parts of this objectified things you would be detecting humans on some of them you would be recognizing inanimate and you come down eventually to a place where you are able to close the whole pipeline and able to describe the scene.

Now the question is, we do like agree on one part that this is a hierarchical process in which any sort of a vision problem is being solved today and any kind of a computer vision problem, any sort of a medical image analysis in respect of computer vision is also being solved today, we agree to this point over here. Now the question is this whole thing unique or can there be non-uniqueness to this hierarchy as well, which is what we need to answer.

(Refer Slide Time: 7:29)



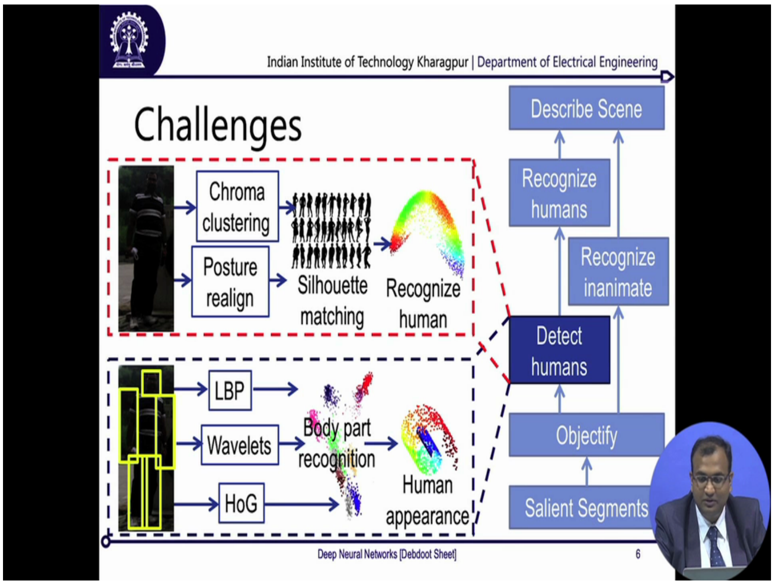
Now let us do a very simple experiment over there. Let me just take a clone of the same network and try to reshuffle around certain of these blocks and then again reconnect all of them. So once I have done all of this what you would see is that in both of them on the green pipeline as well as on the blue pipeline, you actually have the same number of blocks, the same blocks doing the same things, but what I have done is the order in which they are going to do this is where I have brought in a difference, ok. Now this is also a valid one, this is also a valid one, both blue and green are valid ways of solving the problem. The question comes is which one do I choose; this is something which is going to plague the community for quite long as to which pipeline.

So there can be person A who says that I will be go pipeline green, there can be person B who says I can go by pipeline of blue and you would keep on debating for say time immemorial as to who is correct and there might not possibly be a solution and that is how we see in recent days that somebody or else has a different way of finding it out and nobody is very sure about whether they are hierarchy is an absolute hierarchy in some way of doing it.

Now, in terms of an industrial problem this pose is a very major challenge and that challenge is that since I am not able to define which is a very unique pipeline in the way of solving it out so I will have to invest a lot many more man power in order to optimize which of them is going to be my best product outcome over there. So typically the cost factor associated with developing software for medical image analysis in this case would keep on increasing by the number of such replicable model, which we are going create, such identically similar performing pipelines.

So all though the codes will be same which had been developed in each of them but then, there will one not be much of a code reuse going down across them. There will be a team of people who will solve out on the green pipeline and there will be a team of people solving on the blue pipeline and I do not know there can be infinite such number of pipelines actually none of us know. So this is going to escalate the cost if my final objective is to come up with the best software to solve this particular problem. Now, this is one problem, this is one challenge, the question is, is this the only challenge which is going to give me (()) (9:53) now that will not happen, because if you look into it say let us look into one of these blocks which is about detecting humans.

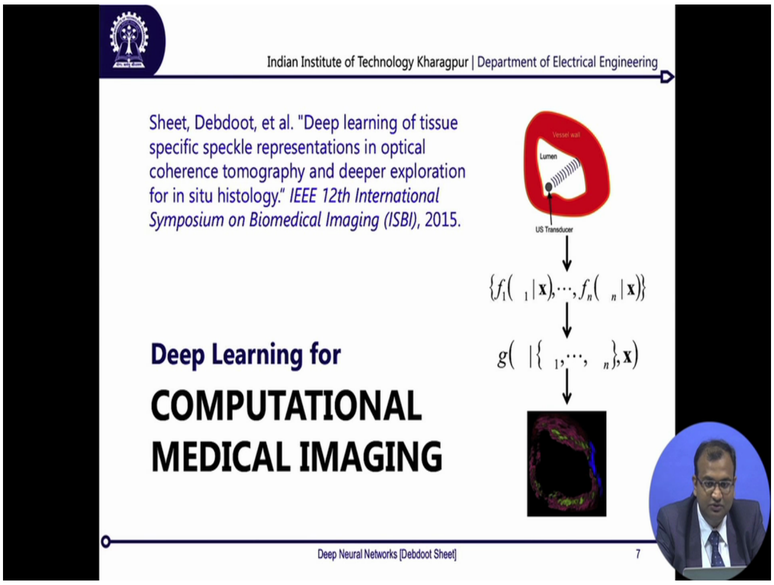
(Refer Slide Time: 10:03)



Now one way of detecting human may be I take a patch of that Salient object on that I find out certain features, so may be local binary patterns, wavelets and histogram of oriented gradients. Now I use all of them, I plot them onto a body part recognition manifold, which is going to recognize my body parts, so there is a head, there are two arms and there are two legs. Now based on that it can project it onto some sort of a human appearance manifold and from there it will recognize that that is a valid human. This is one way of detecting that there were humans present on the image. There can also be another way of detecting it out and so certain different portion might say that what I would do is, I would do a Chroma clustering and Posture realign and that would lead to finding out some sort of a Silhouette over there.

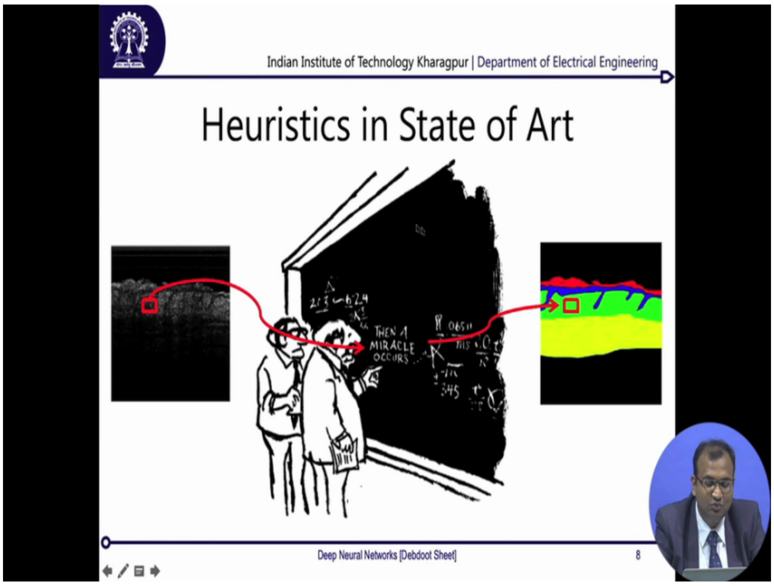
So portion in different poses, how they would like and then you can have a library of this Silhouette images and you just match down with the Silhouette. Now, once you have matched down with these Silhouettes then you can actually project it this matching distance onto certain sort of a manifold and then be able to recognize humans. Now you see that even for detecting humans there can be the first block, there can be the second block and both of them may be equally good and equally bad and people would still just be keeping on arguing us to which one they use.

(Refer Slide Time: 11:27)



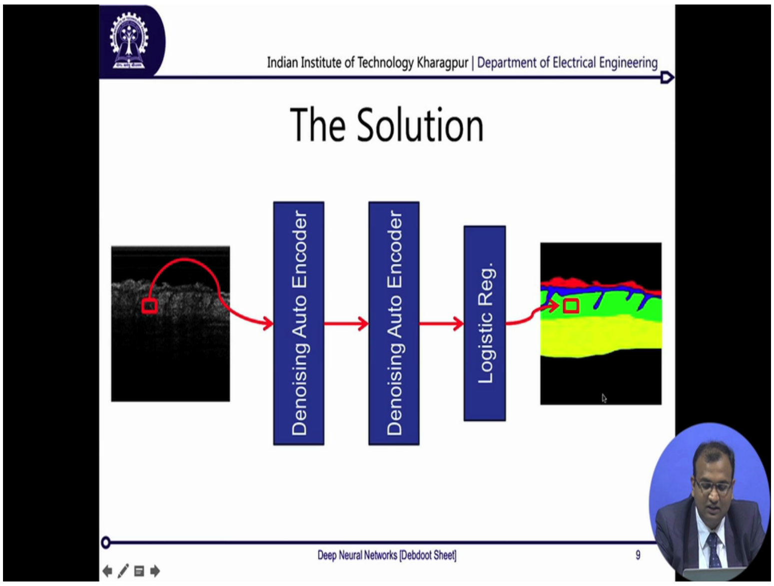
Now this is a major problem and these are actually very big problems, which are plaguing the community today, and that is where deep learning comes to play. Now I would draw your attention onto problem which we solved a few just two year ago and what this whole problem was that say you have some sort of a computational model, which finds out so this is about computational imaging as we had discussed in the previous once. So where you would be looking into different natures of tissues present by just looking into the say ultrasonic signal or optical special signals over there. Now, over here we would be having one model to extract out features, which is again guided by certain good knowledge about the physics part of how signals get propagated. Then we will have another model just to map these features extracted onto these tissue types over there.

(Refer Slide Time: 12:30)



Now the question comes is, can I have one single model which can do all of this or in a very simple terms there is a lot of mathematics which goes into this whole place and for a lot of people how this would appear is as if within that mathematics there is some sort of a weird miracle which happens and this is solving my whole problem. If I do not want to do this can I get down on an actual network over there, which will be solving the same problem without me having to bother that what sort of mathematics is going on.

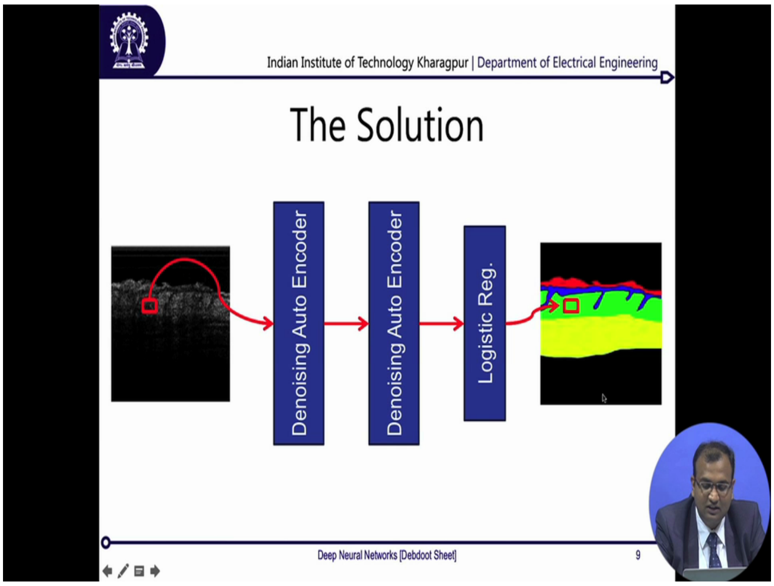
(Refer Slide Time: 12:54)



And in fact it is not so hard to imagine because we have a very simple solution to this one as well. So what we propose down was something like this it is called as a auto encoder solution. So this in the subsequent slide you will be having a view of what they mean down. But if you look over here say I take down one small patch from this image and then I am going to feed it through this small block, which is called as a Denoising Auto Encoder. Whatever comes out from the output I again feed it into another block, which is also a Denoising Auto Encoder. But this block takes an input from this block only it can this block cannot take an input from here; this is what you need to remember. And then you keep on passing this through over here and this thing is called as Logistic Regression layer, which is very similar to sigmoid transfer function actually.

And you are able to find out probability of different classes of tissues, so over here I basically have 4 different classes of tissues which are marked down on this image which is an OCT image of the skin. So this is my whole pipeline which I would be solving and how this whole pipeline and inherently assuming that this pipeline, somehow learns by itself to solve all of this complex mathematics which we were thinking of as a miracle which is happening in the earlier case.

(Refer Slide Time: 14:00)



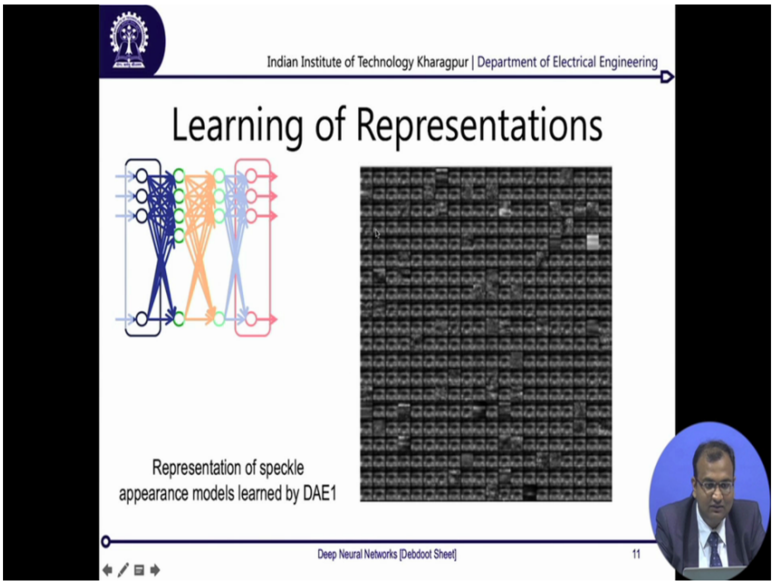
Now if we look over here what we design is a network very similar to a neural network, but it is with more number of hidden layers over there and that is the point from where comes the term called as a deep neural network. So you increase the number of such hierarchical representation layers under the assumption that on each of these hierarchical mappings you would be mapping down one small extra feature, extra amount of information onto your network and that will help you in solving a much complex problem onto a much simpler solution over there.

So if you look into this particular network over here, it appears very similar to a multilayer perceptron. So you basically have 1-1 layer of single perceptrons over there, you stack on these layers and you keep on continuing and that is how you are going to get this whole network. The challenge obviously exist in how you are going to train this network and that is where a lot of people will come into with their expertise and we will also be solving a subsequent problem, where I would show you how we can figure out the way of taking care of these problems as well.

But the challenge is it once I have somehow been able to train this network, do individual representations over here make any sense, which is say I have this first neuron, second neuron, third neuron these are say there is a patch of 5 cross 5, so there will 25 pixels over there and I am just ordering each of these 25 pixels over there. Now all of these 25 pixels connected to say this first node over here, so there will be 5 cross 5 weight matrix which is also going to connect to this first node, for the second node also I will be getting a 5 cross 5 weight matrix which is going to connect, for the third node also I am going to get it. Now the point is are all of these weight matrix going to be the same or are they going to be different. Now if they are different how will the network make sure that they are different, these are questions which we need to answer.

On top of that, so this is from where I am going to map down to my first neuron, similarly I am going to map down to all the hidden neurons over here. Now from these neurons I will again be having my weight matrices connected down to each of these neurons over here. So these are going to be some sort of a dependent analysis of all the representation we are learning over here. Now that is where comes in the beauty of a deep neural network actually using a set of very small operations in order to solve a very complex function. So let us look into what that meant.

(Refer Slide Time: 16:32)

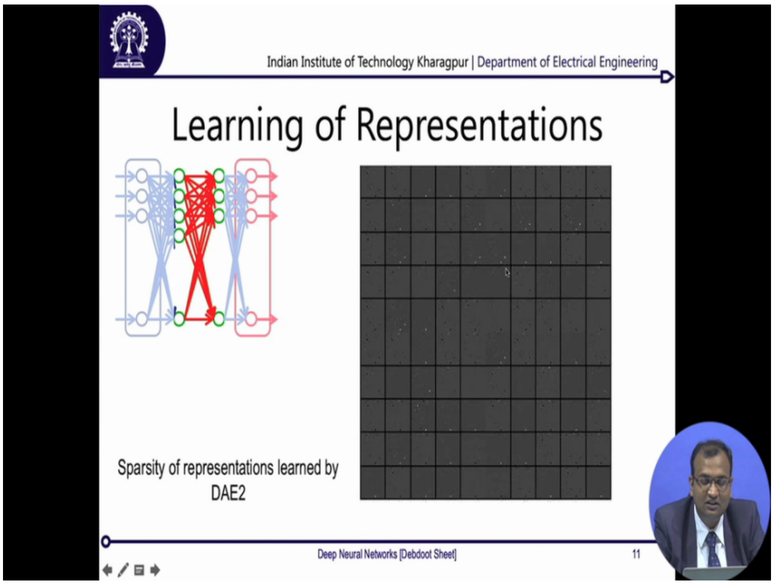


So in terms of learning my representations what I would do is say I have this neural network which I have learned down over here. Now from that let us look into the first representation over there. So what I have is basically a matrix of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, so it is a 20 cross 20 matrix, which means that I basically have 400 nodes over here, each of this patch is again made up of a set number of points, so over here this is basically 21 cross 21, which is the total number of nodes I have over here, ok. Now, each of them if you can if you would zoom into them you would be looking that they somehow look like wavelets, they have some sort of a definition and some of them are inverse form of it, this one looks like more of a Mexican hat kind of a wavelet.

This looks like an inverse of that particular wavelet and there are some ones which are perfectly like flat regions like this one or you can have a perfect white one which is a decisive or unity multiplier over there as well. So there are multiple of them and a lot of them actually look very similar to each other. So they might be different, they might be similar as well, but what comes out at the end of here is that here whatever it learns each of them is basically some sort of a correlation kernel which it is learning.

So your feature extractors are getting learnt over here itself, without me telling what kind of feature extractions to take place. So your network is just doing classification by taking in an image, but one image it is going to extract out different features, so these feature extractions are what is happening over here. Now let us look into this second layer over here what happens over there.

(Refer Slide Time: 18:23)



Now on this average grey value is basically a commodity of 0, that has an intensity of 0 wherever its white small dot over there they are all bright spots which are corresponding to very high positive value and the black dots over here they correspond to a very low very high negative value over there. Now what this does is this particular matrix will be mapping all of this, so there are 20 cross 20 or 400 such nodes, which are mapped over here. And here we have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, so there is a 10 cross 10 or 100 such nodes over here. So each of them is a 400 is a 20 cross 20 or 400 patch element. Now, each of them is giving a weight to each of such correlation kernels, which we had learnt or some sort of features which we were going, so this is weighted combination of features which is getting learnt in this second layer.

So similarly if I keep on looking at the third layer, fourth layer, fifth layer they would just be looking into a hierarchical combination of how we can learn down these features one by one. And this is where I meant to say that a deep neural network itself can synthesize the hierarchy from the data without you having to explicitly say what how to synthesize this hierarchy, and at the same time it would also be doing a classification problem for you.

So the hierarchy gets synthesize as well as how you are going to combine each of these features at different levels of hierarchy all of them are getting synthesized together. Now, if we get back to our old problem on that seen classification, where there were 4 people and everybody and then we had this parallelism that there can be 2 different possible hierarchical combinations in which we can solve the problem.

And the question which had in was in our mind is that if we have this such two different combinations, then which one do I choose, I just cannot arbitrarily choose one of them, there should be some basis of choosing it down. Now, if I say that today we have techniques, where you can just give a simple network over there and let the network decide the order in which it can choose a hierarchy on that making much more of exciting proposition for us because we will not have to bother anymore with feature engineering, we will not have to bother anymore with classification margins and stuff, we can just feed in images on the input side of it and just get a classification over there under the fact that we can somehow code this whole neural network this can be computationally solvable and tractable, the gradient should be existing, there should be a error backpropogation throughout it and we are able to solve it.

(Refer Slide Time: 21:04)

We would definitely be making lot more promising results onto it and this is one result particular result which I would like to show. So this is about Optical Coherence Tomography OCT image which we had studied about so this is an OCT image of a skin of a mice and this is about skin of a mice which is healing so this there was certain wound in which the skin was pealed up and the mice is now healing and over here you would see that epithelium is not yet clearly formed on over here some part of the epithelium is formed and the full thickness of the skin is not yet restored because typically it is about this order of thickness, which is roughly about 1 millimeter thick, roughly about 1 to 1 and a half millimeters thick but this is roughly 0.5 millimeter thick, so it is still in a healing state.

Now, on this one this was a ground truth by looking into the histology which was marked down by histologist and this is the sort of predictions which we are able to get down over there. Although they look very shabby but given a point that the whole network was actually trained with very healthy samples where it was a full thickness skin of 1 and a half to 2 millimeters on which these labels were obtained and each single layer was perfectly present, so your (()) (22:04) formation was also in a much more clear way. So here we have been able to reproduce a similar way over here. So you can get a very clear and distinct epithelium you get the Papillary Dermis coming down perfectly, the Dermis and the Adipose appear very perfect over here. And here we never used a single sort of feature extraction over there, what was fed to the network was just these small patches and what it was asked to predict was just this tissue levels over here.

And the network learns by itself, what sort of features to extract and what hierarchy should it be following in order to get to this particular problem solution. So this is the beauty of deep neural networks and with that I would conclude the first part about deep neural networks. In the subsequent lecture you would be studying about the history of deep neural networks and how they have been working the way they are working and a survey of certain examples, where in the recent past we have had great successes in solving medical image analysis problem with deep neural network.

And following that I would also be entering into showing you the solution to using a Deep Neural Networks in order to solve the same ALL classification problem which we were doing by the classical way of first extracting features and then doing a classification. So here we will start with the point that can we just put in the image itself as it is and solve the whole problem where it gives me a classification coming down to it. So let us be excited for the next part of the lecture until then, thank you.