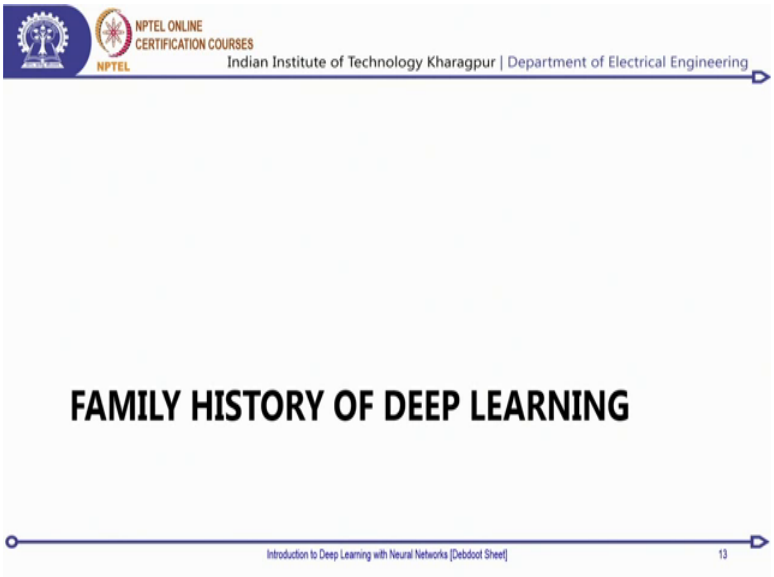
Deep Learning for Visual Computing
Prof. Debdoot Sheet
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture – 07
Introduction to Deep Learning with Neural Networks (Contd.)

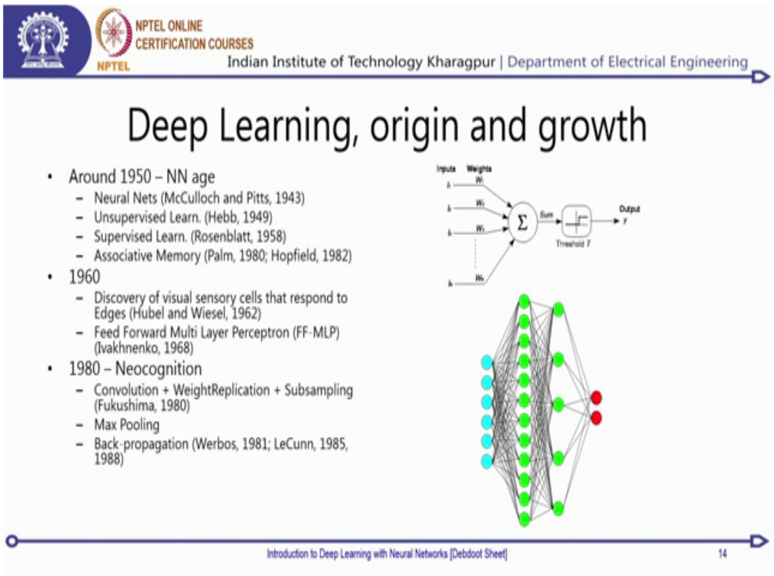
So, welcome to this next lecture, and then here is when we would be going down as a continuation of what we had done in the earlier one with introduction to deep neural networks, and as I said in the earlier class that we would be discussing on to the history of deep learning with neural networks as it has been evolving.

(Refer Slide Time: 00:38)



So, this is more of centered around the theme of family history of deep learning, and how these deep neural networks have been going around over there.

(Refer Slide Time: 00:45)



So, if you look into the origin and growth of these networks. So, as we had also done been discussing for quite some time and so, what these neural networks over here essentially are that they are not something new all though deep learning as such has come to a line right in just the recent past within even less than half a decade as of now, but then neural networks have been therefore, quite a long time.

And as we say that say some somewhere around 1950s is what is called as the edge of these neural networks, and this is when new the mathematical definition of a neural network and the basic perceptron as we see and what we had studied in the last weeks lecture was about.

So, these this mathematical model was what was proposed just around the edges of 1950s, and then eventually what it laid down was from a very simple model of Mcculloch and Pitts of neural network in 1943 through the unsupervised way of learning falling down a Hebbian, rule and then going on to supervised learning with the Rosenblatt perceptron in 1958. And then eventually after a lot of delay in from 1980 by pam and then Hopfield in 1982 was the associative memory concept and these are what lay down as precursors to what deep learning is today based on.

So, around in the time of 1916s there were some more interesting things which started happening. So, initially till around the year of 1915 what was going on is that the mathematicians were independently working, and then they were not at all they there was not much of an interdisciplinary interaction going down between different fields over there.

Now around in 1960 they started to be these inter disability collaborations between mathematicians who were working out on developing neural networks, and neuroscientists. So, this is the first time when you could see electrical engineers people of mathematics information theory, and then also neuroscience researchers coming down into together.

And the whole objective was can you find on whether this whole mathematical model of a neural network has some sort of an analogy or does provide a plausible explanation of how biological neurons, within say the human body or within living organism any kind of an living organizing organism does have the neural network and the neuro transmission pathway.

So, whether it was down to the same sort of a neuro transmission pathway over there in another living organism. So, that is what was going down in 1960, so the first one was by Hubel and Wiesel, and what this gives down is the visual sensory cells which respond down to edges and what they found out eventually had a very interesting culmination, because when we get down into those initial neural networks and trying to do down with digit recognition you would find out, that the first few layers or the first few hidden layers over there they would be what are responsive to more of edges, and complex patterns of edges and in fact, these discoveries of 1960 is did help us find.

That within our biological neurons within our vision system from our eyes, the first few things which we recognize that basically they just line like behavior straight lines curved lines or odd circular arcs these are the ones, which are the first level of behavioral recognition, which happens in order to make us recognize a particular object and then associate it to classifying it out.

So, the next one was a feed forward multi-layer perceptron and that is the standard multi-layer perceptron, which we are looking over here and which we had studied. So, in the subsequent lecture we would will be going down to a mathematical depth sent to them, then around 1960s came down what is called as a neo cognition. And these the first theories which were being proposed on with this kind of an association with neuroscientists in terms of understanding whole images. So, what they found out is apparently it turns out that these neural networks, as we were initially thinking are fully connected structures, but then within the biological system and within our bodies they are not fully connected, but they are sort of like what is called as a convolutional.

So, instead of so if you remember clearly in the first weeks lecture on neural network where I was writing on the mathematical model. So, you had an x into w. So, there is a each is a unique weight which is associated with one neuron and associates to another neuron, whereas what it comes down from this neo cognition perspective is that these weights over here are not a huge family of weights. So, each neuron does not have a unique weight, but it is basically a combination of weights which has a translational property.

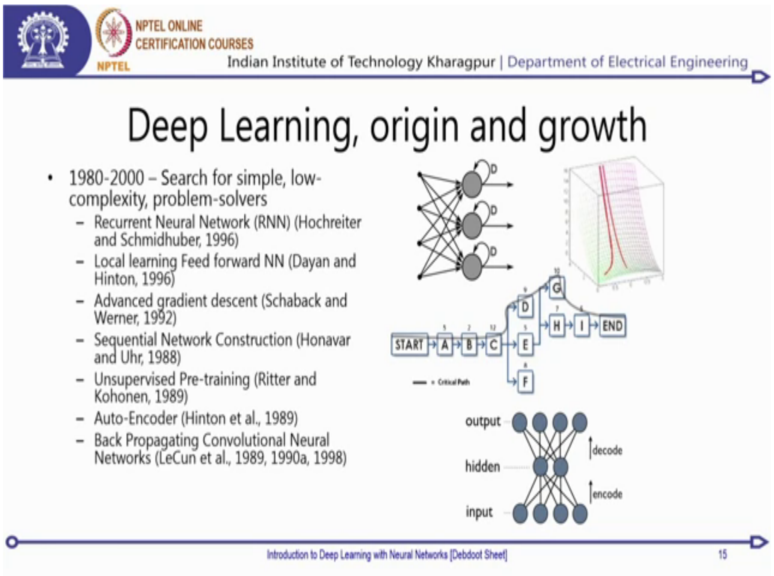
So, that would mean that you can operate these with a convolutional kind of operator. So, x is convolved with a weight matrix called as w, and the resultant is the convolutional some of this coming out over there, then we got down into something called as a weight replication which is across so, if my left eye has a certain sort of weight my right eye will also have a replica of those weights this is what it came down, and as we go into more understanding of these deeper networks, we would find out that weight replication within these kind of stereo networks or pairwise networks is again a common thing, which either you impose it implicitly or if even if you do not impose it implicitly, it would turn out that they would learn down directly naturally using all learning rules.

Now from there went onto this new discovery of what is called as a max pooling and was a very important concept as far as neural convolutional neural networks within deep architectures as of today go down, then came down this idea of back propagation or the learning rule. So, what we were doing down yesterday was that gradient descent over there, but then that gradient descent so you remember that we did take a derivative of the cost function with respect to the weights of the network.

Now when we try to solve this whole derivative over there you would see that there would be something for a multi layer perceptron that it will be going down across the different depth layers. So, from the final target output layer via the immediately next hidden layer to the next hidden layer and eventually coming down to the input layer itself.

And as this whole progresses along the depth from output to the input that is why it is called as a back propagation. So well, we will come down to the mathematics and more details of it in the subsequent layer. So, this is where the history was at this crucial learning rule on which hole of deep learning resides today is a discovery which was from 1985, and that is almost close to 30 years as of now.

(Refer Slide Time: 07:25)

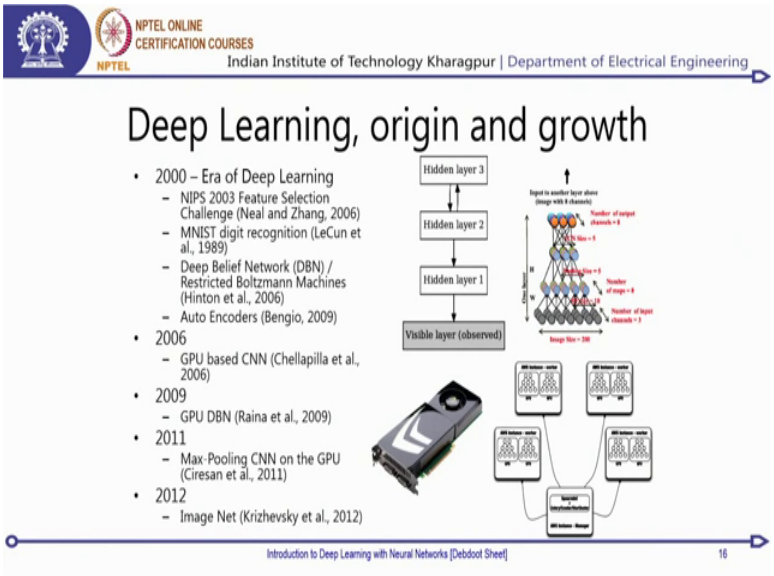


So, going down from there is more things which came down in 1980s to 2000 and this was a point where we had even more complicated problems. So, one of them was what is called as the recurrent neural network which started coming down around the 1980s to 2000 then came down the local learning within feed forward neural networks, and advanced gradient descents, then sequential led for construction which is quite critical. Because what happens is that when you have a complex problem to solve you would not like to solve it from start to end, but then go down by a certain route and then keep on solving it out one at a time.

So, it is like breaking down a bigger complex problem into through multiple number of smaller problems over there, then came down unsupervised pre training or what we would also be doing as auto encoders subsequently, and then as we go down in the next few lectures.

So, we will be initially starting with going from multi layer perceptron onto an auto encoder and then understanding what is the relationship between a multi layer perceptron and an auto encoder. And then that is what will be going down through back propagating convolutional neural networks as well. So, these very simple models are what construct down the basic building blocks of understanding a very deep neural network, and they were all which took place in 1980s to 2000.

(Refer Slide Time: 08:48)



From there at the start of this particular millennium in 2000 is what is more of heralded as the Era of deep learning, because all the theories which was developed before 2000 is.

Where you need a lot of compute power, and then around this time is when this compute power software libraries implementations and data sets. And you definitely need a huge amount of data as well. So, these data sets and everything is what started coming down and eventually around in 2000 we had enough of consumer grade compute power to get these working.

So, these too much of mathematics to be made it solvable within a human lifetime, today if you solve a deep neural network you can pretty much train a very complex model on challenges like image knit or something, within 1 to 2 days or maybe maximum of a week with within your computers water within your reach whereas, if you look at the year of early 2000 this would have taken more than a month strength or some of these problems were even what required training for over a year, and that was not a feasible engineering in an idea I mean very few people had resources to spare enough for this one, and that is one of the prime reasons why deep learning was still out of the reach of a lot of people and was not coming into the consumer space.

So, from there in 2006 some interesting things which happened was this advent of the GPUs and with NVidia and a lot of other partners strategically positioning their business around from just mired computer graphics generation or some of this mesh grid like solvers for multi physics, or physical simulations to getting down more of a compute centric thing, and getting down architectures of memory, interfacing data transfers, which are something which are analogous to support down this high bandwidth requirement, within neural networks for their implementation for data transfers, because if you clearly see I have one layer. And then via certain number of weights I connected to the other layer.

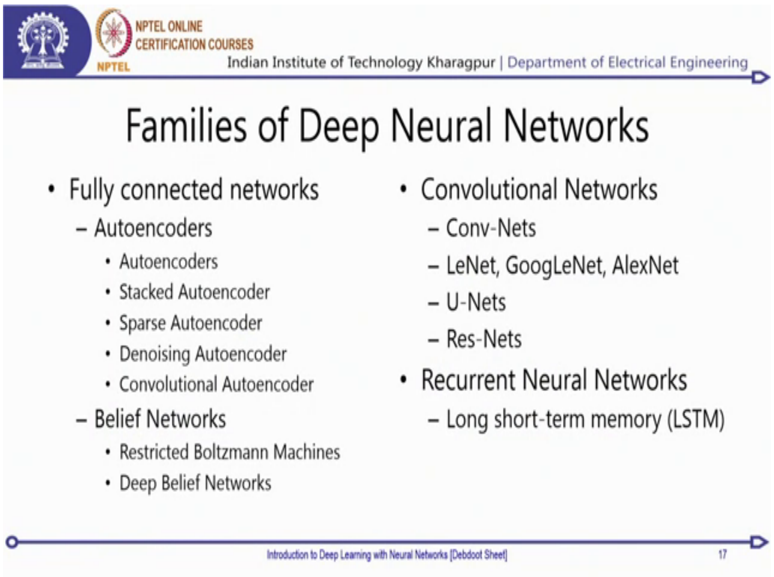
So, each of these layers are what are these what require certain memory the weights over here will also require certain memory, and this operation in order for it to happen it will require a lot of memory transfer. So, whenever I do a x into w, I would x 1 into w 1. So, there are 2 memory fetch operations, and then a product and then write to a memory.

So, for every one single operation there are 3 memory operations of read write which are going down over there. And this is from a very heavy volume RAM. So, basically your CPU to RAM access bandwidths need to be really higher, and then these getting better and better is what led down to the advent as of now. So, from there on 2009 to was a GPU implementation of deep belief network, and which was very crucial in in terms of being able to get down these belief networks working down, and then in 2011 came down the max pooling CNN so, on the GPU.

And this was with advent of certain critical architectures within the hardware itself that it led to much faster otherwise earlier max pooling I think which had to be done only on the CPU side of it. So, as we get down more into details you will get down.

Where these libraries accelerate and what are the hardware h constructs, which can be addressed and referenced down by the software libraries directly for the best access. And then in 2012 was the image net willing winning model by Alexander Gritsky Alex net of 2012 which is the one which so, this was the first deep learning model which was beating down any of the classical models for filling the image net challenge which recently closed down in 2017, and then got remodeled into others. So, this is more of the history and in the subsequent classes, we would be touching down on one single attribute of this history one single model, and then see how this has contributed in a big way to what pc deep learning as of today.

(Refer Slide Time: 12:51)



So, as we have gone down through the history, the next which comes down is a family of these deep neural networks now, these deep neural networks can typically be divided into 3 families as we call them.

So, one of them is the fully connected networks within this fully connected networks comes down the concept of auto encoders and so, they can be auto encoder stacked ordered (Refer Time: 13:08) parts denoising as well as convolutional. So, convolutional auto encoder is some sort of a relationship some sort of a hybrid, between a convolutional network and auto encoder itself.

So, what auto encoders do is typically what we will be studying in a subsequent, but to give you a very gist of it. So, if I have a pattern x I would somehow encode it through certain weights in order to get down the same pattern x as the output. Now essentially you would see that well it does turn out what is the use of all of this like whatever I put down in the input I get the same as output, but you see there are multiple uses of it one is you can do a denoising out of it.

So, if I have a noisy input side over here somehow I make this network. So, that it gives me a noise free. So, you can use it as a claim cleaning image, cleaning filter or denoising option, you can use it in order to find out a latent representation or a compressed version of whatever it is given on the input.

So, if my hidden layers keep on getting smaller and smaller than my input layer or my output layer. So, somewhere in between what I can do is if my input is some 1000 neurons, I can get down a hidden layer of hundred neurons. And if I am able to with through this network get down a 1000 neurons again back. So, it means that I can compress down 1000 pixels to 100 pixels. So, this is an image compression which it can solve out.

So, we will come down to those examples as well of how to get down an image compression as well running down with these neural networks, then the next one is what is called as a belief network. So, the typical one is a restricted Boltzmann machine, which is already known quite widely within the community. So, this is where you have some sort of a Boltzmann distribution being carried down. So, if I have an input and a output or I connected by a hidden layer, and this hidden layer you see which is which produces a outputs of Boltzmann distributed.

So, any variable state out in this inner layer is a Boltzmann distributed variable. So, given any input you can get an output or given so, and input outputs are not so, predefined over here it is just a pair of x and y. So, if you give a y it can also give you an x given that the hidden layer over here is Boltzmann distributed, and then when you stack them one on top of the other that is what leads to something called as a deep belief network.

So, this is where all inputs all outputs and all intermittent are ones are directly connected when you change all of these direct connections or a dot product like connection to a convolutional like connection, and then that would necessarily help you to get down a space invariance because now you can have non locality as well address down. Then these kind of networks are what is called as convolutional networks, and or again also briefly termed as convents. So, today what you would hear down as say GoogLeNet, AlexNet, LeNet, U-Nets, then Res-Net residual networks Res-Nets, these are all what

what rely predominantly on the first few operational layers in terms of convolutions itself, and are typically defined as convolutional networks.

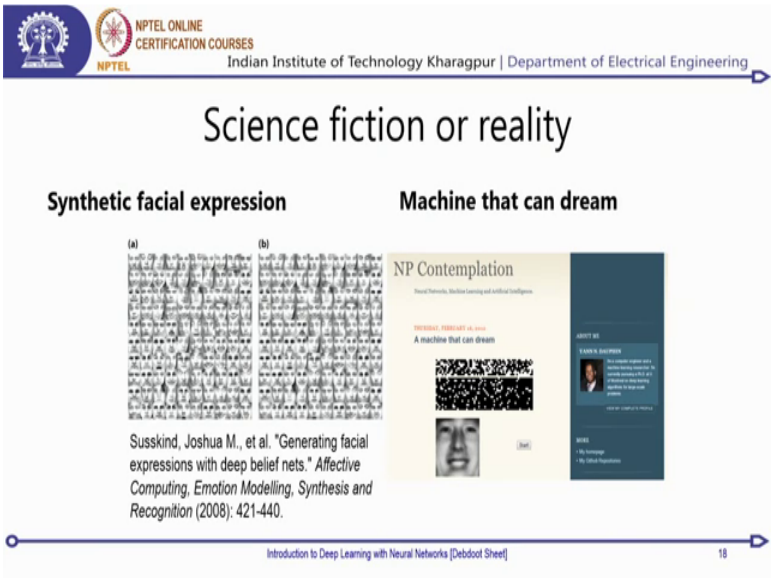
From there comes down the next version which is a time sort of a and a neural network which operates on the time space itself. So, and it is also called as a recurrent neural network. So, what happens is that the output of the neuron gets added down to the input of the neuron in the next time step. So, not in the same time step, if you I am processing down a sequence.

So, the first time stepping whatever is the output that output will be getting down, when I am trying to process down. The next in the time sequence data over there, this is very useful for doing. So, natural language processing say you want to do an error correction measure. So, you would see that often when you are typing on your smart phones, if you if you just write start typing a message after one alphabet it starts showing you a few alphabets or even words over there. And as you see as you keep on typing more alphabets more alphabets it keeps on getting better and better in you see closer to the exact word.

So, these are kind of things which are associated with recurrent neural network behavior. So, we will be getting down more and more details in to of them, but while we will not be doing this say sentence or word correction kind of behavior, we would be exactly using these recurrent neural networks for our video analytics problem, where it is frames which are not.

So, distinct lead different, but somewhat related, but come down and in a series sequence of time, and then can we used some sort of a recurs property between their object appearance across different frames as in a video in order to get down an analysis of a video or classify your video. So, that is broadly there are 3 families in which they are.

(Refer Slide Time: 17:43)



And today if you see so, this deep learning thing is no more sort of a science fiction which it was initially thought of to be. So, you can get down to these very interesting examples on. So, NP contemplation is just a web site over there is Google it and find it all, so what it does is that.

If you see over there it those black and white dots over there are basically some neuron outputs of a restricted Boltzmann machine. So, as it generates a Boltzmann distributed 0 or 1 0 or one kind of output over there. So, this is a perfect black bodies output, and you would get down a face corresponding to it. So, it is a bit creepy because just by doing certain black and white black and white or 0s and 1 sequences over there you can generate a whole human face, looking down and every time it does generate a different face coming down.

The other is a paper well given a face you can use these kind of deep neural networks in order to synthesize different facial expressions. So, as we go down into more of regulation modeling on deeper lectures, and then the next courses, and in the subsequent weeks which would be a bit later on we would be coming closer and closer to how do you even synthesize these kind of images using some simple black and white dots as well.

(Refer Slide Time: 19:07).

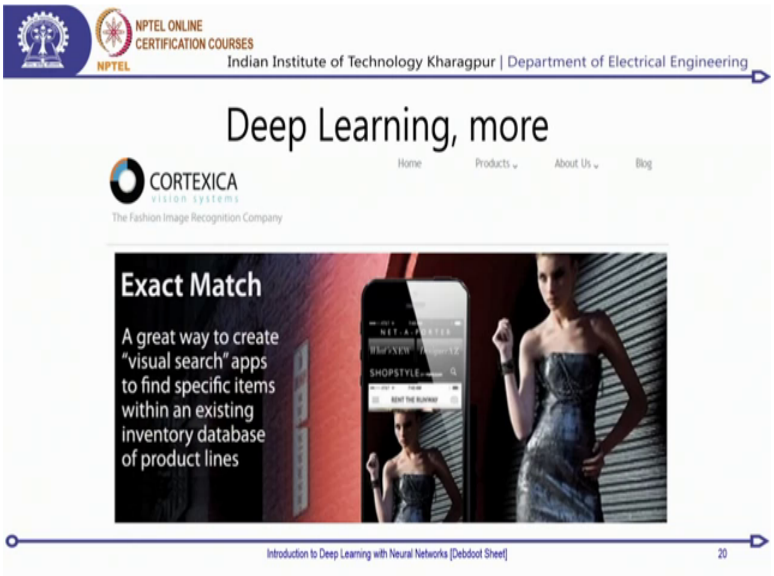
So, from there going on to application side of it where we stand today say a Facebooks face recognition or object recognition, whenever you upload an image it just says whether these people are present over there or whether it is you or not. So, that is what has been building up on top of the years of corpus you have built by tagging your individual faces.

So, in the initial days if you remember, that that is like almost close to a decade back when Facebook was starting up only. You could put down images and then you could draw a square box around those images, and then tag down your faces or your friends over there and that was helping them create a large corpus and eventually initially those boxes were all fixed size square boxes eventually they got into variable sized, but square boxes then it gone down into rectangular irregular shaped boxes, and then not necessarily a square aspect ratio box coming down, then you can annotate objects then and that is what help down in creating a lot of corpus of this supervised learning coming up.

So, eventually from there. So, baidu is large search engine within china, and this is where deep learning power suits retrieval engine. So, if you put down an image of a person and it fetches you all possible images of the portion and does not restrict only to photograph. So, they can be even hand sketched versions as you see in the last row over there that can be person who is tilting the head and some different poses, and then which is a really interesting part, because if you put down some object or a persons face you would like to get down that portion.

So, all other faces over there, this is a very critical search task which this particular kind of technology or deep learning is helping us achieve in a real time scenario. Then there is this this particular website on Cortica which what it do does is that it pulls down random stream of images from the, web and then it starts generating small captions over there or more typically it is like what is present in the image, and it can give you 1 or 2 words over there and this is what it does on the browser side there is nothing running on the server side it works on your browser side. So, and It was really a fun to watch out.

(Refer Slide Time: 21:11)



So, more about them is with this kind of products on fashion so. In fact, now even some of these is a big ecommerce companies like amazon also have launched it out, and that is about where you can take an image of somebody wearing a dress, and then it is it somehow searches and finds out through it is visual catalogs and gives you the product catalog category on their E store, and you can buy that sort of a dress.

(Refer Slide Time: 21:42)

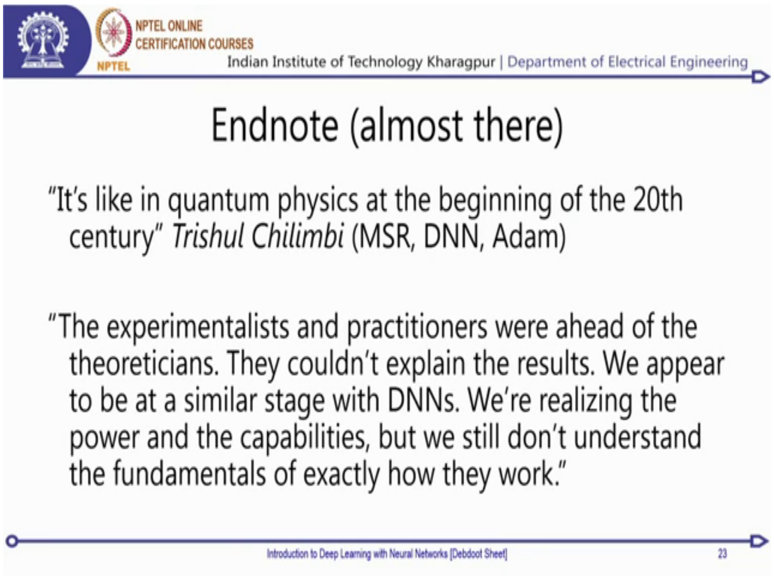
So, this is where it is going down on impacting the consumer space as well. So, from there you see a huge aspect of going it into self-driving cars, and then autonomous driving full enormous mobility and not much left behind is Microsoft thing.

(Refer Slide Time: 21:54)

So, somewhere in 2014 they started up getting this public release on what is called as Adam and today, if you see there is Microsoft Cortana for speech, and as your assistant for pc systems. So, they are like really building up huge in terms of it has it is own liabilities itself there is something called as a Microsoft CNTK which has a cool API, where you can give you can use that API within your website within your apps, anything which you are developing, and what this can do is given an image it can say whether it is a male female what is the edge of the male.

These this kind of intuitive information and also we have done even from spatial expression to give down expression analysis, whether the person is angry or some sort of emotional dependence whether he is happy he is smiling these kind of things. So, this is what is becoming increasingly deep learning powered AI as of today, and that is where it is going, but the challenge with this is even bigger and that is where we are almost at the end now.

(Refer Slide Time: 22:53)

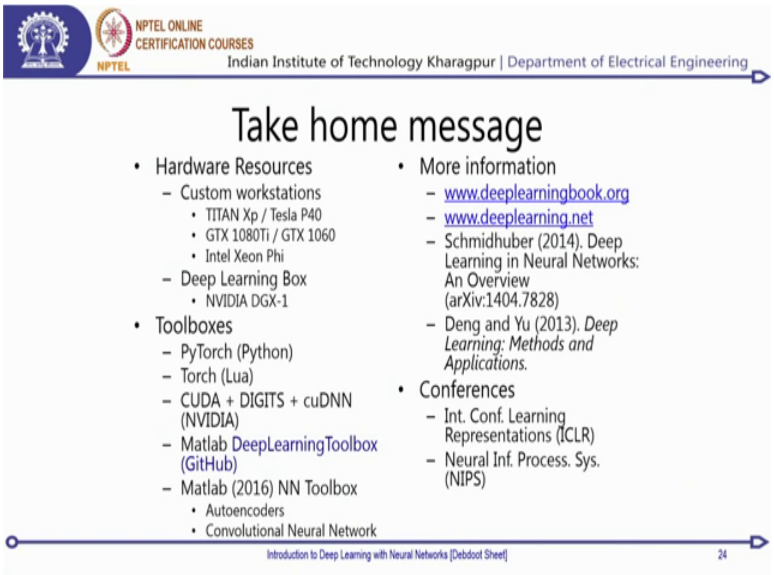


So, as Trishul Chilimbi had put it down was an interesting observation was that this whole thing of deep learning is quite like quantum physics at the beginning of the 20th century, and the reason behind this was more of this that that experimentally and based on practitioners and software coders, these experiments have been far ahead of it is time because we are getting down more results better results coming down, but the problem is that there is another group of people who are theoreticians and who come down to this aspect of explained a I.

Which is drawing an explanation as to why is this particular model working and that is something which we are still not at a point to understand heavily, we know some of these explanations, but not all of these explanations. And that is a prime research which is the major challenge within deep learning and learning with deep neural networks as of today. So, as we go down through this lectures where I will be covering a substantial part of what works I will also be working down on why it works, and what are possible explanations of why this particular deep neural networks can do.

And expecting that you can also build up newer architectures on your own side by going through this whole route, or if you have a different kind of a candidate search which is you have some n different number say 3 or 4 different architectures, then how can you choose out which is architecture which is most suited to solve a particular problem in hand, and that is what we will be doing from theory to experiments and eventually that that is the whole objective of this course itself.

(Refer Slide Time: 24:32)



So, finally, as we come to an end I have a few take home messages for you. So, one is you do require hardware resources and what you can do is get down any of these, and then which are really good except for let is say custom workstations you would need some sort of GPUs to build out, and in India it is easier to get down a GTX 9 80 t I or GTX 10 60 and from NVIDIA, and then this can help you create down a machine.

So, eventually so, we have one session in which we would be unwrapping and unboxing a machine and show you different parts. So, that you can get a hardware set your place while we also have done the clustered axis given down for participants of this course. So, that you can get down access to an HPC, or you can; obviously, buy down this say D J it is one DEV box from NVIDIA, which let us come at a quite premium price maybe just for a few institutional purchases, but not for much of personal things, then on the tool boxes all of these are open source as of now.

So, you can use any of this other than the 1on Matlab for which you will definitely have to pay for the licenses, but the Matlab neural network toolbox since 2016 does support auto encoders and convolutional neural networks as well. So, if you want to read more about it go to this website on deep learning book which is now also available as a printed book, itself from MIT press you can get this book and that is that is what we will be using as a major reading material over here, other than whenever there is something else.

I would be putting now pointers to those exact materials and to follow down on conferences it is NIPS and I see large which formed on the major corpus of what we provide today, and disseminate in terms of newer research in the field of deep learning. So, with that we come to an end of this particular lecture on the introduction to deep learning. So, in the next class I would be getting down started on what will be the toolboxes and toolkits of how to get started, and eventually we will get down into writing down the main math of a multi layer perceptron and getting into an auto encoder, and then subsequently going into coding exercises for auto encoders as well. So, when that this comes to an end and.

Thanks.