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

Lecture - 01
Introduction to Visual Computing

Hello and welcome to you at another new subject and a new session and this is on Deep Learning. So, we would specifically we discussing on deep learning for visual computing. And a lot of you might have already heard about this buzzword, and I assume you may have heard and that is the reason why you got interested to enroll out for this course.

So, deep learning as such has been one is sort of what is recognized as one of the major technological realizations which has happened in starting off this whole millennium. And not quite an old as such in terms of when we look into the impact it has created, but definitely it has a very rich history which spans for almost a century, close to a century as of now. So, what happened of in the last half of a decade was quite interesting. So, starting around in 2010 is when this started around and later on picked up a lot of pitch. And somewhere around in 2014 is when a major number of discoveries and major challenges were being solved out using deep learning.

So, this is one such invention which is not just limited to impacting the software industry and AI driven solutions, so not just purely data and artificial intelligence driven solutions, but they impact us in a large number of ways. So, ranging from understanding speech on your phones to understanding, what queries you are asking and translating your sentences on to query languages, from their starting on to inferring from images tagging your photographs and even autonomous driving; driving of cars by themselves by just looking into camera feeds and views and even giving down predictive analytics on what is a good time to break, whether the car a head of you is now at a mode of breaking and then these systems can even realize stuff before you can do.

So, these are some of these applications which touch us on our day to day lives and from there it moves on to even a larger and more serious application and that is in the field of medical imaging and medical image analysis, where you would be using deep neural

networks in order to understand any infer about the characteristics of images and even discover newer kind of pathologies and how they are located.

So, given that this has been a major invention which came down to power a lot of solutions, given that all most for close to 2 decades we have been sitting on a large treasure trove of electronic data, but we did not have enough of compute power and as well as we did not have enough of a software technology, which includes the known mathematics the optimization solutions, the software implementation frameworks and the hardware to support all of these together which could have harnessed the power of this immense amount of data on which we were sitting.

And this data is not just limited to say hash tags, they are not just limited to your sales pattern analysis, they are not just limited to geo codes and geo tags or coordinates of how your movement is happening, they are also some things which include done much serious nature. So, they can be x rays there can be surveillance videos, there can be even human behavior analysis videos, then dashboard mounted cameras within cars which can even predict down driving behavior analysis and if there is something weird or something quite adversarial which is happening on the road while you are driving.

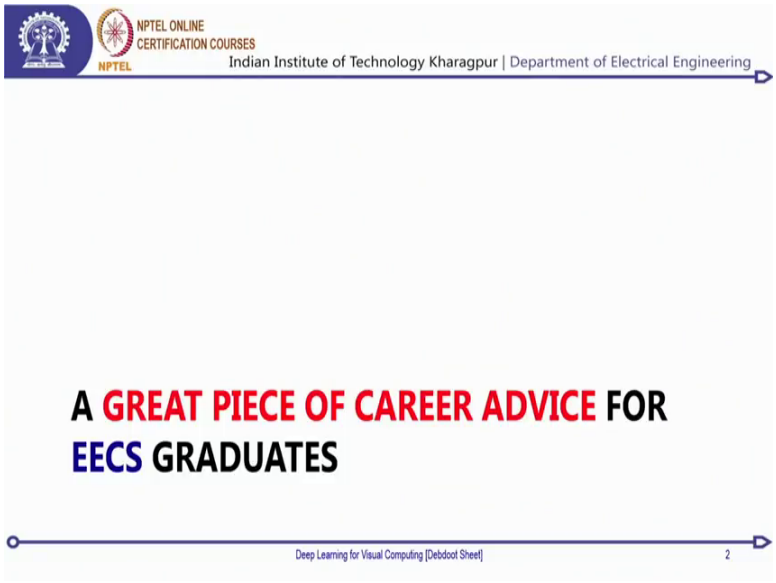
So, there have been a significant amount of implications which these technologies has brought down and what we are going to do is not just touch down on all of them because it is a large field actually. So, it this whole aspect of deep learning this has implications not just in visual computing, it also does impact medical image analysis it does impact natural language processing, it does impact business and finance pharmaceutical discoveries, petrochemical industries, a autonomous driving and a lot many more than we can even imagine at this point of time and by the time we will be finishing off with this course and as well I say within a year to go there would be even more significant discoveries and solutions coming down from the space of deep learning impacting all of them.

So, without keeping all of these too much of a mystery let us get into what we are going to study, so the particular course which we are going to do is called as deep learning for visual computing and I would be specifically focusing on the visual aspects of it.

We would be looking into visual data which will include images, they will include videos, they would include other sorts of visual data including say depth maps from

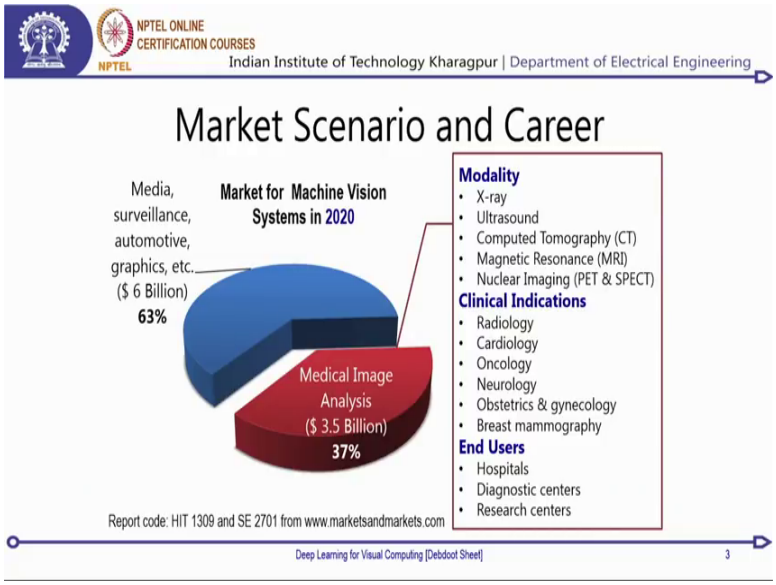
studio cameras, like your x box kinect which you use for playing your games they would also be including down medical images as well and they can be 2D 3D 4D. And an n dimensional sort of visual data which we can look into and given that what I generally pitch forward is that this whole thing is actually a great piece of career advice for EECS graduates.

(Refer Slide Time: 05:41)



Or graduates in the field of electrical electronics and computer sciences and the major reason I say is something based on the statistics which we know about the market scenario.

(Refer Slide Time: 05:54)



So, how it goes down impacting is something like this. So, within another 2 years or roughly I am assuming that most of you would be graduating out by that time or might have already been in the market. So, there if we look into it so this whole market is pitched down somewhere around 9 and a half billion us dollars and that does not just include the media surveillance automotive and graphics, but again a major chunk of it is related to something which impacts all of us very practically and that is medical image analysis.

Now given that this is not going to take away the job of doctors, but definitely make down the job of a clinician much more easier, as well as provide down a lot of affordable and accessible healthcare to the billions that is where we are pitched on to working it out. So, on 1 side of it if you look into it is almost like two-third of the market segment is what is related to media, so this media is in the sense of entertainment industry, a videography gaming industry. And all of them we have surveillance which is again a major taker in the field of visual computing, itself whereby you have analyzers running down on all surveillance cameras which are there you might see them at traffic junctions at ATM machines.

When you go down to draw your money at any of the entrance to institutions, banks, government offices, private offices, anywhere very well it is really becoming an essential part of observing people and I am trying to find out adversities who might be really hell

bent on making some nuisance over there in their premises; now from there we would be moving on to automotive which is again a major sectors. So, within auto motives if you really look into it then what comes down with automotive is autonomous driving, a major 1 driving assist systems are another one which whereby the driver is provided down with assistance. So, today you have your Google maps like behaviors which on your phone, they it also gives you not just the route prediction, but also gives you the time prediction based on now.

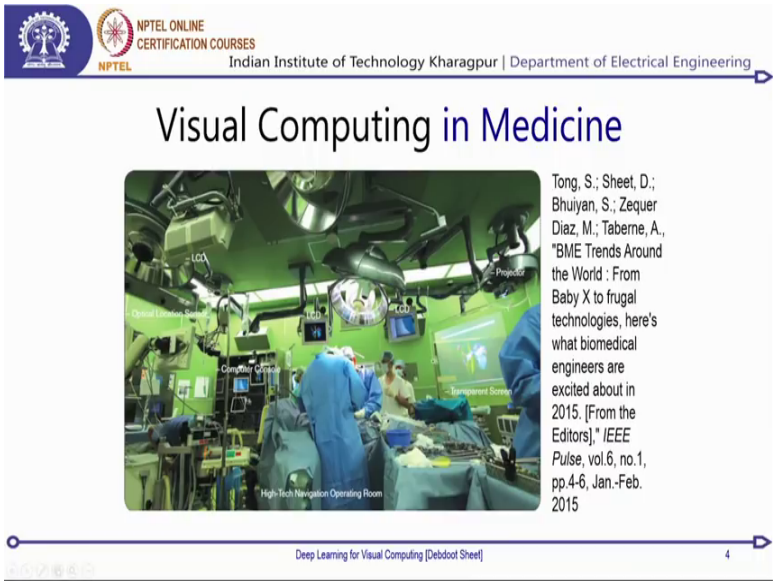
So that is something by making use of data of all the drivers ahead of you who had driven on the same route or on the same road which helps them getting it, then they are synchronized along with route maps and way layouts for different traffic and cities and how to integrate all of them and not just that where we are looking down; today there are certain number of cars which also provide a solution to braking assist, they provide solution on road condition on the drivers dashboard itself and that is about looking into visually inspecting and while we as humans can definitely understand a rich treasure trove of data.

But then we have a limited bandwidth and capacity to do it and it might be while we are focusing on driving the car, we are really cognitively loaded down for just the driving aspect of it and we are not introspecting on the aspects of the road or the car ahead of us or maybe the car which is ahead to ahead of us these kind of machine learning algorithms a lot of them are powered down by deep learning.

What is helping us blend on that and making our roads even safer, from there it is definitely graphics which is another major aspect which goes down. Now given all of this on the medical side of it what it really impacts is we can look into different kinds of modalities and they range down from ultrasound to x rays and as it goes on from there to clinical indicators, which include on radiology cardiology oncology and an these are the different clinical verticals, which basically get impacted or the diseases which will be and the end user for a majority of these are basically hospitals diagnostic centers.

And secondary research centers which may include even pharmaceutical labs, now given that we have been able to and we are striving on to make real big impacts onto this field and deep learning is definitely powering a lot of what we are doing today, where we are looking down in the future is something of this sort.

(Refer Slide Time: 09:59)



So, visual computing would definitely be one of the major impact makers in the field of medicine itself and this is from one of our earlier articles which we had published on the future of operating rooms and this is what the future of operating room looks like, though this is from a real one which is already on prototype test scale realized down for showing it out to the world how a futuristic operating room would look like; where you would have not just surgeons operating down on a patient.

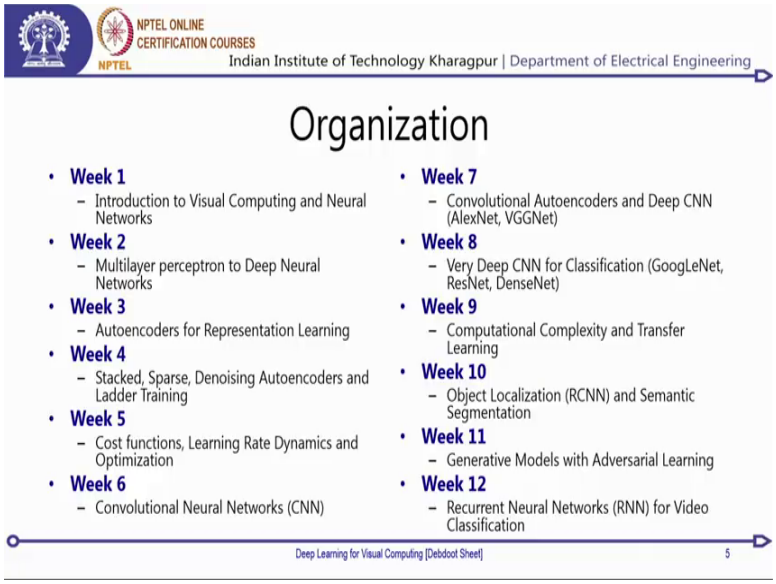
But you would have a lot of machines which will be assisting them and a major of these machines are all compute oriented which will be making use of this real time data of the operation, going on preoperative data in the form of CT and MRI scans of the whole patient over there, their projective anatomies which you can see on this transparent screen over here, and these are assistants which it would provide down. So, as and like you can get your road driving assistance on your cars, in the future we these systems would be helping surgeons provide them assistance on.

How to drive down a particular tool and a particular surgery, so at the end of the day this is the golden dream which all of us have and in trying to get down even much better quality of life and a quality of health also nobody wants to become sick in any way, but in case you fall sick in case there are accidents can we really ensure that there is a better treatment and a lot of it because this is a practice which is heavily vision guided, which

means that humans look into it with our eyes we see surgeons inspect visually and accordingly they take a decision and keep on operating.

So, the whole objective is can we combine all of this integrate that all the learning aspects of human into machines as well. So, that these can these machines can actually help in solving smaller pieces and blocks of the whole bigger task to be achieved and down the line as we are able to achieve all of these solutions what we would be seeing is a whole ecosystem and a product being built around that now given that, let us get into what we would be learning now.

(Refer Slide Time: 12:08)



So, typically this course is organized as a 12 week course and each week itself has it is own takeaways. So, we will start with the first week and that is what we are doing now. So, today I am introducing you to the course overview and what this whole aspect of visual computing is and on the subsequently we will be learning about visual computing aspects from the classical way of extracting features and the reason I would be covering that in one of the lectures.

So, that you understand what is the relationship between these classical aspects to why are we moving on to the deep learning aspect of it and where do we gain an edge in terms of moving from a classical aspect to a deep learning aspect and subsequent to that I will be introducing very simple neural networks and we would be having down lab sessions in between as well, where we would be making you go through not just a walk

through, but how to set up your own systems and machines on your end and how to even set up the HPC resources, so that you can get access to a much higher compute power on your side which we might not typically be available on your desktops and laptops itself.

So, from there we move on to week 2 where we would be starting with multi layer perception, which is the basic model for any kind of neural network and then eventually enter into a deep neural networks and over here we would also be. So, as we go down on each of these weeks while we will finish off some part of the theory, we would also be having a matched out lab session, all the lab resources codes and data and everything how to pack down data. These instructions will be leased on a separate git hub link which will be provided to you. So, you can make use of so as and when we come down to the lectures going down through the code walkthrough and programming implementations we will be getting.

Now into more of details of how to access it and we expect that these github links are going to remain permanent, that is what we have been doing for our earlier courses as well and so that you can get back at any point of time and relook into that. So, on week 3 I would be starting down with the classical representation of a deep neural network and called as an auto encoder, which is an unsupervised learning mechanism and used for doing something called as a representation learning and here we would be re looking back into what we do in this week's lecture with classical methods of vision using textures and try to draw an analogy between these different kind of textures to what we learned on with an simple auto encoding in the form of representations. So, from there going down to week 4 we would be doing stacked sparse denoising auto encoders and a particular concept called as ladder training, which really helps in training these networks much faster and without much of an issue as we would experience with other kind of a Deep neural networks.

So, from there we move on to week 5 and in week 5 we would be looking into cost a function, learning rate dynamics and optimizations and this is where how this whole learning happens. So, typically when we say that something has to learn, so there has to be certain aspects over there what is meant by learning and how does something learn. So, these are the concepts which we would be covering down in week 5, from there going on to week 6 we will start with the biggest celebrity of the lot and that is called as

a convolution neural network, which has been pitched today as one of the major inventions of this particular century if not the millennium at the least.

So, we start with the basic building blocks of a convolutional neural network, including explaining you why is it called as a convolutional 1 and then where does the neural network come from. So, finishing off with week 6 when we enter week 7 we would be doing some of the very classical architectures of Convolutional auto encoders and then enter into deep convolutional neural networks and get done with the examples of Alex Net and VGGNET.

So, when we will be doing down these examples on Alex Net and VGGNET, so these are basically on different people or different groups who had who were their inventor. So, we would be going down in a surgical fashion onto basically dissect out the different aspects of the network, what are the building blocks over there while these networks are being built what will be the computational complexity achieved, what will be the data complexity over there what will be the model complexity.

So, this basically means that how much time how many number of CPU seconds or say number of time clock cycles you would be needing in order to execute these networks, poor image what will be the size of the image this can deal with what will be the dependency between the image size or the input to the output size over there, as well as what will be the dependency in terms of the model weight, which is you are going to store all of this network everything together into one sort of file.

So, what will be the size of the file and the interesting part is that you can actually calculate these on pen and paper and that is what we would be doing in week 7 starting over there. Next we enter down week 8 and that is where we would go down to the recent advancements and these include down Google net to resent to dense net and these are some of the very deep networks which have been there.

So, these are the ones which have been proposed in the last 3 years itself and these are the winners of one of the major challenges called as image net challenge which we would subsequently come down. So, we would be surgically understanding each of them go down to their implementations and code level walkthroughs and how to reuse them if you would like to, from there we get down into week 9 into understanding computational complexity and this is where we revisit what we do in the earlier 2 weeks, in terms of

understanding how much of time it would require, but here into much more of a details and also try to get down into much deeper understanding of can we reduce the time in certain way and what can be the aspects of not to compromise on the quality, but yet get a reduced time implementation over there.

So, from there we would enter into another important aspect called as transfer learning and the whole aspect of transfer learning is that can I have a network which has been pre trained or which has been basically trained and in order to do a certain kind of a task and can I use it to do another kind of a task. So, if there is a network which has been trained to segregate bananas from oranges. Now can I use it to segregate say a thief from a normal person. So, this is a typical example of what I would put down as a transfer learning problem. So, not just these but even more complicated ones and how can we use them, so we will get down into domain adaptation and transfer learning over there in week 9 from there in week 10, so while all of these have been on classification problems in week ten we would enter down into something called as localization issues.

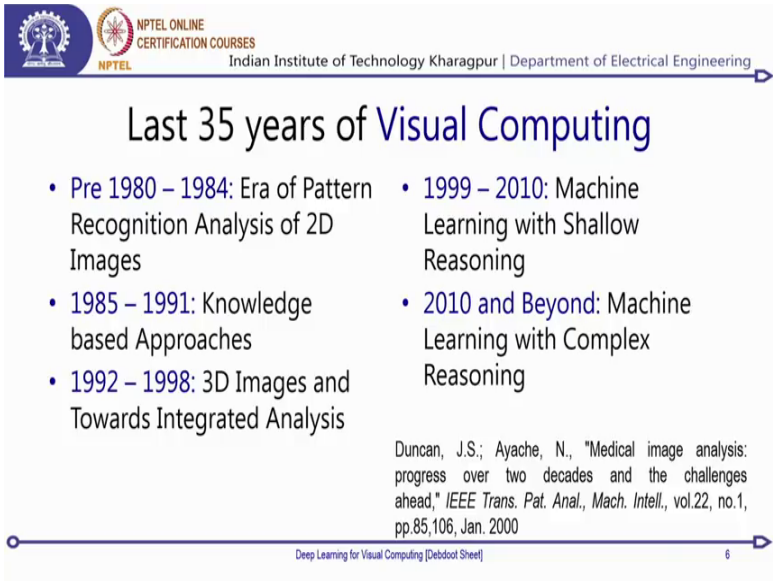
So, if can I localize down certain objects in the or in the particular image, so an image might have say a lot of fruits over there and a dog standing in the middle of it and if I just want to localize the dog, but not all the fruits then can I do it? The next part is semantic segmentation and this is about pixel to pixel wise labeling. So, can each pixel over there be labeled down in a given image, using a deep neural network.

So, we will be going down to through this important investigation as well and also walk down on the implementations, week 11 we would be starting down with another interesting discovery which has been coming down in the last few years and that is on generative models and this is about can we use these kind of deep networks in order to be generate images. So, if I write down a word say an apple with a black patch can it actually generate an image of an apple with a black patch on it for me and so we will come down to this really interesting aspect of generatic modeling and then using adversity of learning in order to do that and from there on to week 12, we entered down on sequence modeling with recurrent neural networks and for us in terms of visual computing problems the sequences are all video frames over there.

So, all frames of a video if you place it down, so they are basically images which acquire in a particular kind generated sequence and that is what we would be trying to do

with recurrent neural networks, now given that this is the organization on how this whole course is organized if we look a bit into the history then visual computing as such is not something new and it has a history which is of which is quite rich and then more than 35 years.

(Refer Slide Time: 20:35)



So, I do take down this interesting paper on medical image analysis though and this is from the journal from the transactions on pattern analysis and the reason I was doing down is because, this is one of the most comprehensive and while it does draw down correlates with medical image analysis, but nonetheless the visual computing is also almost of the same sort of running behavior which has been there.

So, around in 1980 is when the whole thing started down and a major reason for it starting known was that digitization of images was now possible. So, people had the line scanners invented by that and then you could be scanning down images and there was enough of a compute power, now with reprogrammable silicon based computers in order to do it and you could bring it down onto consumer grid applications and from there it started down with knowledge based approaches in the late of 1980 to early 1990 and from there from 92 to 98 is when 3D image analysis was becoming a bigger play, so can we analyze out volume.

So, big people when no more happy with pixels, but everybody was looking at voxels and bigger data more amount of data more complicated calculations and more space

requirements. So, can we really go beyond that bridge and once that was done. So, starting in the early of this millennium the major thing which happened was shallow reasoning was making a bigger way, so we were getting our machine learning based solutions into our consumer grade products.

So, you had cameras which could be detecting down faces you had cameras which would be shuttering down only if the face was smiling, no you had gaming consoles which had cameras which could track down your face and your hand movements over there and help you to interactively play video games. So, these were some of these small things which were happening there were interesting aspects in the field of medicine like cad systems, which were helping find out benign masses and segregate them from malignant masses on x ray images and not just that there were vessel segmentation algorithms we started coming down and they were major aspects which were good.

But all of these what was happening done was still with the limited compute power and the amount of reasoning was what is called a shallow reasoning, it was not something which is called as a complex reasoning as we say and that is something which started going since 2010 and beyond it and what this complex reasoning basically meant is that, say I give the camera I just have an image given down to the computer can it describe a semantic equivalent of the scene or can it tell me what should I do next that given this is the frame.

So, today's cars if you look at it, so there are a dash systems, driver assist systems which actually provide you an input on what should you be doing; if you look into your Google maps you would see an another interesting fact that if you are somewhere at the crossroads and you have 2 different ways to take and both of them will be leading to the same sort of a destination. But then one of is experiencing a heavy traffic whereas, the other 1 is where you will not be experiencing a lot of traffic jams and lags behind over there then it does give you a prediction that, this route b is 10 minutes shorter than route a and all of this is where it comes down the complex reasoning.

So, earlier it was just computing in the shallow reasoning it was just computing down what is the travel time estimated to be taken down, in complex reasoning it is actually telling you what to what you should possibly be doing. So, these are 2 different aspects which come down from shallow to complex reasoning, while Shallow reasoning will

deal with only one simple aspect of the problem, which is not so complicated even for humans to tackle down.

Whereas, for complex reasoning they it would be giving down an inference and a solution which is much more human like and what you would be expecting down from a machine to come down.

(Refer Slide Time: 24:19)

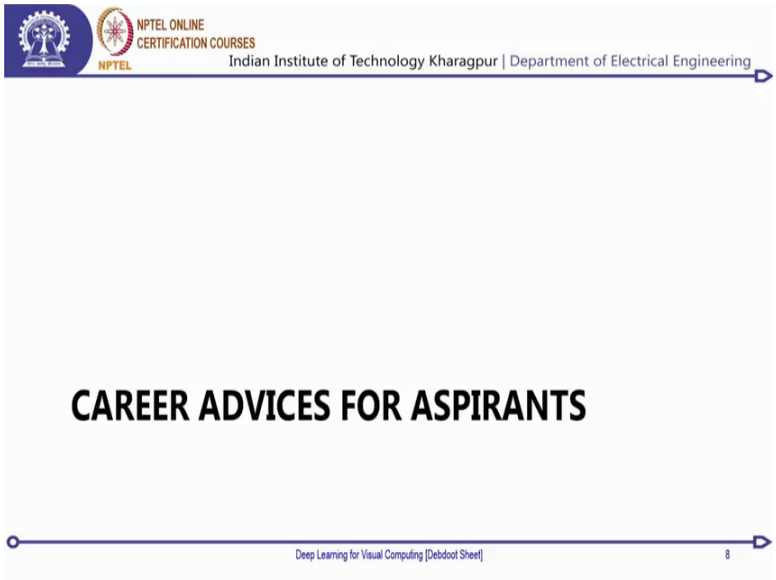
So, given all of that as we enter into 2018 and in this early part of it where we see major challenges in visual computing as one of them is activity classification within videos weather. So, what are the different actions being taken down within videos there are aspects of computational cameras and displays there is computer vision in sports and you might have been watching. So, as Indians we are quite crazy behind cricket, so you would be looking down into cricket matches or even now that recently we just finished off the fifa under 18 football tournament going down.

So, over there in case of sports you will be seeing down that a lot of computer vision applications and these are not just limited to tracking down crowds or anything, but there are individual player tracking switch goes on and this analytics is quite critical for sports analysts and even for budding sportsmen, who need to understand player strategies and team playing strategies over there.

So, computer vision does play a significant aspect over there, from there it goes down another quite rising topic of today which is called as virtual question answering. So, today if you give a image to a computer and then keep on asking questions it can keep on giving you answers in human like way. So, if I just say this is a typical system of where you have a VQA or visual question answering given down. So, likewise this is an image given down to the computer and what it asked is what the mustache is made of.

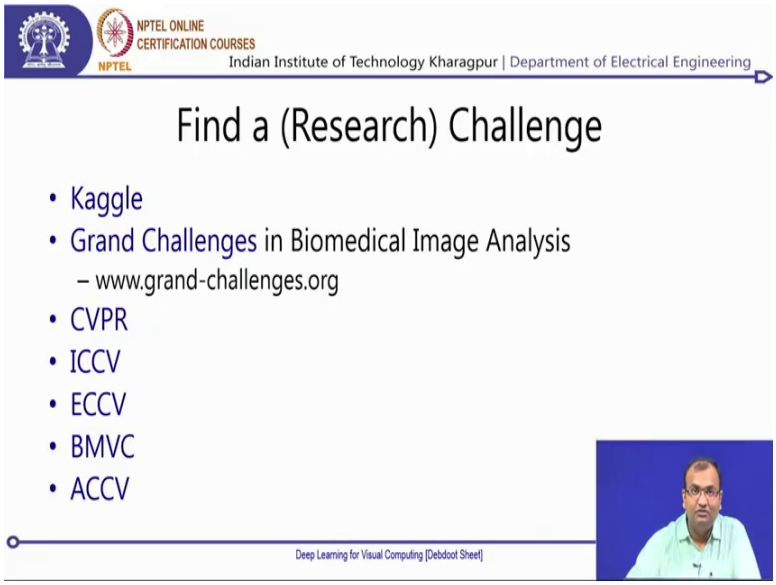
So, from there what it says out is that bananas, now this computer does understand where is the location of face and where is and the location of a mustache with respect to face and what should it be looking like and then it identifies what is the actual object which let is made off and then says out banana. So, this is the visual question answering or something you would be doing with the toddler yourself. So, from there goes down these aspects to what form was driving, which are real major impact creators as of today with major investments going down.

(Refer Slide Time: 26:25)



Now given that we do understand the aspect of visual computing as of today, now what I have is some sort of career advice for most of the aspirants over here.

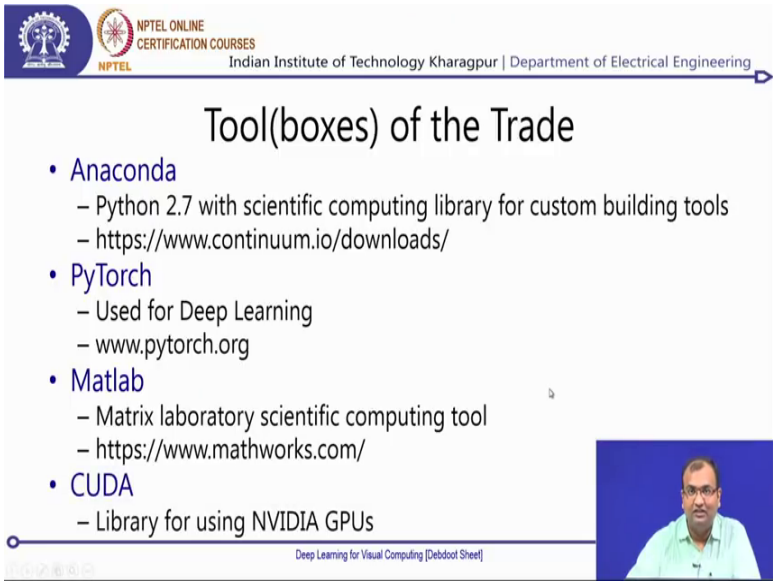
(Refer Slide Time: 26:31)



So, one is do definitely find out a research challenge over here. So, one way is grand challenges in biomedical image analysis is my personal favorite, where we put down at least on the field of medical image analysis we do love sharing our data and putting down challenges out over their problem statements and how what we really seek to solve along with clinical collaborators, the other place where you can really get down is on Kaggle and in Kaggle you will be getting known a lot of these region related challenges coming up recently.

So, the other place to read down and find out what are the recent ones are CVPR which are the international conference on computer vision and pattern recognition which is jointly organized by attribute computer society and the computer vision foundation, from there ICCV which is the international conference on computer vision, the European conference on computer vision, the British vision conference and the Asian conference on computer vision. So, these together are the major places where you can see down not just on vision, but a significant amount of contributions coming down in the field of deep learning and it is applications as well.

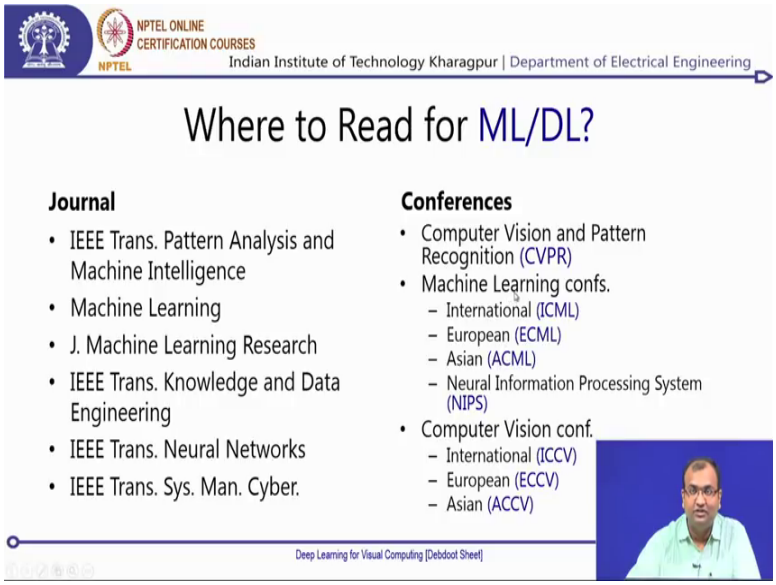
(Refer Slide Time: 27:42)



So, given all of this I would be getting down on to what are the toolboxes which you can be using over here. So, in order to achieve what we are going to solve, the major toolbox which we need is anaconda python. So, this is python 2.7 based which you would be using down through our lectures.

So, we are not using python 3 in any ways everything is 2.7 based, then for our deep learning aspects; we would be using pytorch which is a library compatible on python and has accelerations available for deep learning applications itself, from there mat lab is another interesting one which you can do for some tidbit experiments and initial part of learning, but we might not be using it together for most of it and cuda is used for accelerating your learning experiences with an NVIDIAGPU.

(Refer Slide Time: 28:38)



So, having given all of this where to read for machine learning and deep learning specifically you can go down through these journals, which include pattern recognition pattern analysis and machine intelligence transactions; then the journal on machine learning journal of machine learning research and nonetheless. So, not just journals but there are important aspects of looking down into conferences as well. So, CVPR is one of the major one for machine learning do definitely try to go through ICML ECML SEML and nips and for computer vision you can go down through ICCV ECCV and ACCV which I had underlined in the previous ones.

(Refer Slide Time: 29:16)

So, for listening and socializing more it is better to start going down to workshops and schools with the schools and summer schools.

So, international computer vision summer school is one of those ones then the machine learning summer school, the deep learning summer school are some of these interesting ones which happen now; on the Indian context it is its definitely worthwhile to go down to conferences which include ICVGIP and the NCVCRIGP. Now these are just the crux of what we have to do and some point as to where we go down, as we go through the rest of the lectures over here for this particular series, I would be putting down pointers on interesting aspects on interesting places to look down through your notes and then the book chapters which you would be doing.

So, with that stay tuned and geared up for the next lectures and as we come and keep on enjoying this.

Thank you.