So, let me just go back to the last few slides of the previous day. So, I think the course content, I mean somehow there is a slide on that it does not seem to come up, but you people know what it is. So, I will just start from there. Okay, so the way right we are going to go about doing it is to give you a quick review of deep learning which we will start today, wherein we will start with you know MLP. Okay, to start with. There will be some overlap for those who have done a DL course on imaging or who are doing.

Okay, there will be some amount of overlap, but then hopefully right once we are through with this deep learning after that you will not have any overlap at all. And then the contents to follow are very different, but because you know deep learning is a modern vision course, so deep learning has to be introduced. So, right, I mean there are very few of you. Okay, so I am not so worried.

So only those, for those of you who have already done all of this, just do the vision. That is all. Okay, but anyway this will be a quick thing, not the very detailed way written which we normally do a deep learning course. Okay, and then after that we go to CNNs which again right you must have heard about, mainly you know these are used for actually images and then eventually just these three topics right, not exhaustive. Okay, so this is RNNs right, which again is something that is interesting especially if you are dealing with videos and so on.

Okay, so then the next thing right is, then the next is, so the way we structure the course is that right, so we do a quick review of deep learning then followed by low level vision. So low level vision right, if you see, okay what we would like to see is you know how do you sort of extract edges and so on. This again, you know this is a traditional way of doing things and these days there are also modern ways of doing it. There is something called an edge net and all right. So there are also deep network ways of doing it but we will also study how these things are typically used to be done and you know what kind of, what kind of right you know theory was there right in those days that helped you understand what edges are and so on.

So edge extraction is one part and you know it is very important because based upon edges also you can say so many things right. I mean you might wonder what do we do with edges but the fact is that you know even cartoons and all you do not have shading, you do not have any shading information. Somebody just draws an edge sketch right and then you are able to make out who they are. Typically of course it is for famous people right, not like we can identify everybody from that but we know that there is so much information that lies in edges and there are also other kind of uses of edges especially from a vision perspective right which we will see later. Then you have feature extraction which is again something like if you wanted to build like we said last time right, triangulation when we talked about triangulation we said in a stereo you will have two images and then a scene point right will create an image on the left and then it will have an image on the right and you have to establish a correspondence right.

You have to know that this scene point is that in the other image. In order to do that you have to have something that helps you tell that right. Usually you might be able to say but then for an algorithm to be able to do it, it should be able to find out something, some features right that is able to match right. First of all it has to arrive at what it thinks are robust features right that is why we have different kinds of features right. There is not just one type of feature out there.

There is I mean the most common is probably SIFT which you must have heard about scale invariant feature transform that is the most common used to be right. But even today people use it. It is not like these things have you know are completely you know out of the picture or something just that they were quite I mean they were rampantly used right at one point of time and now I think you know people are looking at deep network based features you know which the network itself figures out as to what is an appropriate feature for a particular task and so on. But like I said last time you can still have you know physics injected stuff right in the sense that you can also bring these things inside your network if you wish to right depending upon how it helps your cause right. There is no and in most of these things there are no straight answers okay.

You have to really try it out. I mean nobody will be able to tell you that hey you do this right you are going to write that is the good FHU results. So this deep learning is like that right. I mean it is also an art really kind of more an art. I mean you may know the problem.

You may know what you want but then you know it may take a while for you to figure out how to actually make that happen right and that comes only by practice. So this feature extraction is like that right. So you need to be able to not only you know do a detection what is called a detect and match right. So a detect features in an image and then match it in the other image right. So similarly so it is like saying the left image you have a bunch of features in the right image you have a bunch of here but now you have to match them because only after matching you can do the triangulation if that is the thing of interest.

You can also do it for stitching images and all right. You must have seen Mosaikin right where people you know create a panorama. So there again right you have to before you do the panorama you have to sort of align the images right. You cannot just put them you cannot slap one on top of the other just like that right. You have to align.

You need those features to be able to align. So the features are very important right. They can give you 3D information. They can give you 2D transformations depending upon what you are looking at right. And then this line detection right.

This again is something that which we will talk about vanishing points and vanishing lines and so on. Then blob detection mostly do with circular kinds of things right that you want to probably say detect. And again when what would be useful we will show some examples

right when what can be useful and so on. And filtering in those days right it used to be Gebauer kind of filters and simple filters right that would give you orientation information and so on. And now of course you know people more or less talk in terms of even now it is well known that even in a CNN and already initial layers actually learn orientations and so on.

I mean that is well established. So these edges are like edges are like you know low level features right. So it could come directly by using any of the standard techniques. But these days we do not do that so much. So what we do is we let the deep network figure out and it turns out that what it figures out is something similar to what our own visual system also does in the sense that we also have at a very preliminary level right we actually extract to see edge information.

We are very sensitive to edges by the way. And so right so that way people have understood that edges have a lot of important information. But those by themselves are not enough. They have to kind of they have to be grouped together in a particular way. They have to right they have to sort of what you call interact in a particular way in order to make higher inferences right.

That is what happens in a deep network. Initial layers may learn something which is created of very very basic right. I mean you may not be able to say much except for the fact that except for the fact that you know there are these edges. But then how they come together it will happen in the you know mid layers then you get more complex features. Then they sort of interact in a sort of a nonlinear way to give even more complex features and eventually you get to the task right that you want to solve.

Then geometry right so after low level vision so it is like a quick review of deep learning followed by low level vision which is like you know basic features filtering and so on. Then after that right we come to geometry right which is what I sort of you know I sort of hinted at right in the last class which is like stereo or it could be structured from motion. So you could have single view geometry or you could have you know two view geometry something like a stereo or you could have multiple multi view geometry right something like structure from motion or you could have photometric stereo where like I said last time you do not move what? What you do not move in photometric stereo? You do not move the camera exactly right only the light source is changed right. You could either move the light source or you could have light sources you know fixed it somewhere and then you turn on one at a time right and then you get different different shading information and that is what is in fact exploited in order to you know get depth map and of course you know then this calls for a restricted kind of an environment right where you have to control all this I mean you cannot do it in direct sunlight and all so different challenges for different problems. Then mid-level vision is a little higher right where you start you do not just talk about edges and so on I mean there is some information there is some labeling right that goes on right so you are able to tell for example the first one is about image segmentation that you

are able to segment people versus you know aircraft or vehicle or something right or for example there is a retinal image right that is I think it is called a fundus image right at the bottom and then you want to be able to I think those must be the veins or something right so I want to segment out the veins again right depending upon whatever application one is actually looking at and then you can have optical flow.

So optical flow again is very interesting right I mean it is like saying that each pixel if it were to move independently of the others right and you want to know by how much it moved and which way it moved it is like saying what offset rate would you apply to each pixel in order to be able to get to the second image that is roughly what the optical flow is. And that is not the same as computing a homography like I said last time. Homography is a very is more straight forward optical flow is not so straight forward right but these are all things that give you a lot of information about the scene and that is why we call the mid-level vision. Tracking I think last time itself I showed you a few examples then retrieval and again we have to see okay as we go along we will try to cover as much as we can or uncover as much as we can right that is a I like to use the word uncover rather than do covering up. So we will try to uncover as much as we can right along the way and we will see right I mean how it kind of pans out we may not be able to do everything that I am showing here but to the extent possible we will try to cover as much as we can.

Then we come to high level vision so here you are talking about recognizing people object recognition people recognition or captioning right for example these are all high level tasks right I mean you are sort of giving a description of an image you can also extend it to the case of this one a video right does not have to be restricted to we can do video captioning so all these are like high level tasks right where you know just as we humans would do right so somebody gives me a picture like this right suppose I see this picture then I will probably say right what is being said there two children playing in the water so something as close to that right you want to come to it does not mean that these are all solved okay let us now let us also understand that it is not like everything is solved okay people are still attempting various kinds of things here and it is still these are all many of these are still open research problems but face recognition to a certain extent is done in the sense that if you are a cooperative subject it is kind of done but if you are uncooperative then no then the whole thing is wide open and also if you are if it is a distant identification that is also not solved for example if you say that I am flying a drone and then I want to identify who is on the road you know that and all is not done okay so face recognition is good I mean if you are very close to the camera not too far away and they are giving a proper pose right or do not do not give too much of pose variations and all one can do one can do a decent job right so that way this Viola Jones and all is actually a traditional method okay and these days there are these you know deep network based approaches and similarly a natural language has what is called the bag of words representation and so on and this is classification right so some of these we saw last time right I mean you can do classification where you simply say what is there in the image you can also localize right put a bounding box around it so then that becomes a higher sort of you know this one and then you can do a semantic

segmentation where at a pixel level you say right whether this pixel to what class label should it actually belong to and so on right no not really that is what I am saying right I mean on a high level used to be done even even using using old methods but these days a lot of it is actually driven by you know right deep networks so it is not like high level vision emerged after deep learning no so it is Viola Jones was already there bag of words was already there so all these representations were already there okay but I think the way it has taken off right so those were still very very there was kind of limited playground but now I mean you know people are doing fantastic things but like I keep saying right one also has to look at you know how much of data has gone in right I mean these are all data hungry methods right so lot of it is also there are the amount of data that people have been uploading on the net right and then along with that the see GPUs and all that help computation I mean those all a bunch of factors coming together in order to enable these kind of things to happen and the fact that people started showing fantastic results determine that's why but even today right what I find is you know if you change a data set right there is there's still trouble I mean it's not like for example right you try to do I don't know how many of you have heard about the term super resolution it's like saying that you know you have low quality pictures and then you want to you want to you want to build high quality pictures out of that now you can do that right you can use a use a know deep network train it now the training means that I mean you have to have nice pairs right you have to have pairs where you have a very high quality guy and then you have a corresponding low quality pair this is the simplest case that you can think of and you don't have to go far right so you can take this high resolution guy you can apply whatever transformations you want so typically you will down sample it right and again it what kind of model you use it is all learning that right so what might happen is if you try to try to create pairs like that and then you train a network we show this low resolution chap and then you say that that is the high resolution counterpart that you have to produce it will produce so you do this over a bunch of images right thousands of images but tomorrow right if I give you give you give you a low resolution image and if you can actually feed it to this particular deep network what do you think will happen it will actually produce a high resolution image but then what will it be bounded by it will be bounded by exactly this law that you applied initially right so what it will try to do is how will this image how would it how would the high resolution image have looked like whose down sample version is this that's how that's all it can learn it and nothing more because that's what you taught it but you but you know what there are cases where for example your low resolution image could have artifacts in it for example I don't know how many of you have seen scanning electron microscope images and all there you will have what are called you know charging effects so you'll have suddenly you know suddenly a bright spot somewhere so if you have not shown that kind of thing that bright spot will still remain in the in the in the high resolution there because all that it knows is there is a high resolution guy if you down sample this you should get this and if that if that spot is not in the high resolution you cannot see it in the low resolution so it will it will nicely reproduce it but you don't want it whereas in a in a typical scanning electron microscope if you did it at a high resolution you won't see that effect at all the charging effect won't be won't be there right so again so that's what I'm

saying right so it's not like everything is you know everything is everything is hunky-dory right so you can have problems it's not like that everything is solved right there is still a long way to go okay despite all the hype around all of this right let's be let's kind of stand firmly to the ground right and understand that there are limitations there are of course right good things that we have to learn but at the same time let's not assume that our deep learning everybody says deep learning and therefore right without deep learning life is zero there is nothing like that right and there is more there's a lot to be done so that way it's good right otherwise we will all be jobless right.