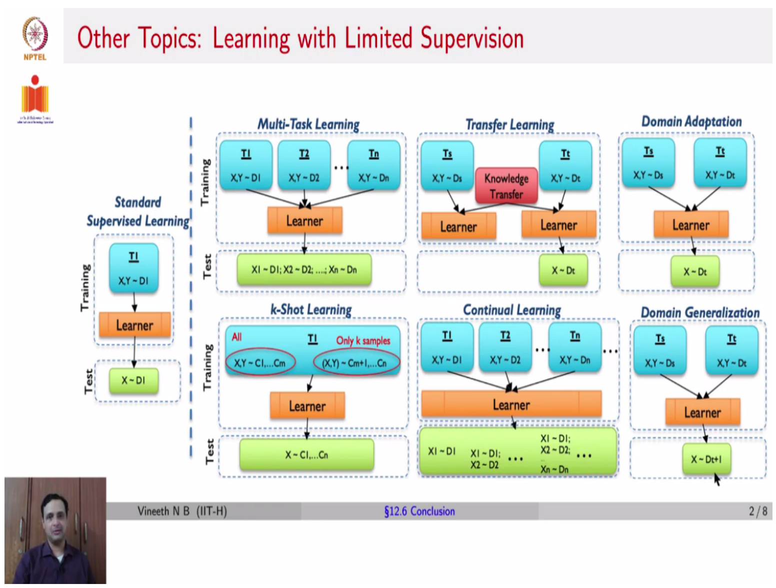
Deep Learning for Computer Vision
Professor. Vineeth N Balasubramanian
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad
Lecture No. 79
Course Conclusion

After these 12 weeks of lectures, which you hopefully learned something valuable from, we come to the conclusion of this course. In these last few minutes, I would like to share a few topics that we did not get a chance to cover in this course, as well as look at some emerging areas and applications of Computer Vision, which are worth looking into.

(Refer Slide Time: 00:44)



One topic, which, which we briefly touched upon earlier this week, but has many faces, facets and dimensions is learning with limited supervision. In the standard supervised learning setting, you have a training set, and you train a learner, and you get a model which is used on the test set. So, now, you can adapt this paradigm to several other paradigms, some of which we have seen, some of which we have not.

Multitask Learning is a setting where you train a model using a learner using data from multiple different tasks to learn a single model that can then classify all the tasks across the image at the test time. An example of multitask learning could be given a face image, you may want to recognize the identity of the person, the expression of the person, the pose of the person, so on and so forth, using a single model, and this would come under Multitask Learning.

The second kind of a setting is Transfer Learning, which we have seen earlier in the course, where you have a source domain data, where you have trained a model. Now, the knowledge of that model is used to train a different model on a different target dataset to be able to perform well on the target domain. We saw fine tuning as one of the most preferred and used approaches in Deep Learning for Computer Vision but there are other methods that one can use for transfer learning too.

A variant or a generalization is of this idea is also called domain adaptation where in the source domain, you do not necessarily learn a model, but you directly use the source data to train a model which performs on the target domain. Domain Adaptation, when you see the diagram, it looks similar to Multitask Learning, but there is a subtle difference. In Multitask Learning, your goal is to do well on the on each of these tasks at test time. Whereas, in domain adaptation, your goal is to do well on the target domain alone. The source domain is intended merely to provide that adaptation or extra information from another domain.

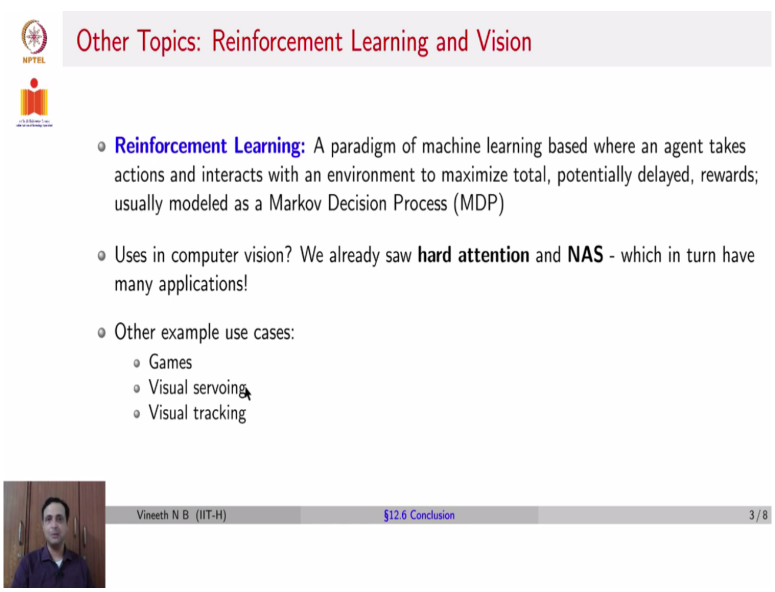
K-shot learning is a generalization of Few-shot or Zero-shot learning. We have already seen that where we have a set of classes for which you have a lot of data, a set of classes for which you have only k samples, where k can be 0. We now have to train a model to be able to do well on all the classes, we have seen a few examples of methods here. Another setting in the context of learning with Limited Supervision is Continual Learning, which has become popular in recent years, where you once again have multiple tasks similar to Multitask Learning, the only difference being is that the tasks could keep coming incrementally over time, that universe of tasks may not be limited.

Now, you want to train a model that works on all the tasks seen so far. Initially, you have data from task 1, you train a model, it should work on other data from task 1, then data for task 2 comes, you update the model. And the model now has to work on data from task 1 and task 2. Similarly, so on and so forth, you get data from task T_n, you now have to ensure that your train updated model works on all of your data from task 1, task 2 through task n. This especially gets challenging when you do not have access to the data from the previous tasks.

The phenomenon of Catastrophic Forgetting, we briefly spoke about this in the lecture on Few-shot and Zero-shot learning, becomes even more prominent here, when you do not have access to data from task 1 or task 2, when you come to the task T_n but you still have to perform well on them. Neural Networks in the Vanilla versions suffer from Catastrophic Forgetting in this context. Addressing Catastrophic Forgetting is one of the most important objectives of Continual Learning methods.

The last setting we will talk about is called Domain Generalization. You could consider this as an extension of domain adaptation. You have multiple domains, in this case, say two T_s and T_t. You train a model but now the model has to work on a third domain, or a different domain, D_{t+1}. This is why we call this domain generalization. All of these are very contemporary topics in Deep Learning for Computer Vision, to be able to go beyond supervised learning in general, which we understand can be solved by engineering networks and training them and evaluating them in a comprehensive manner.

(Refer Slide Time: 06:43)

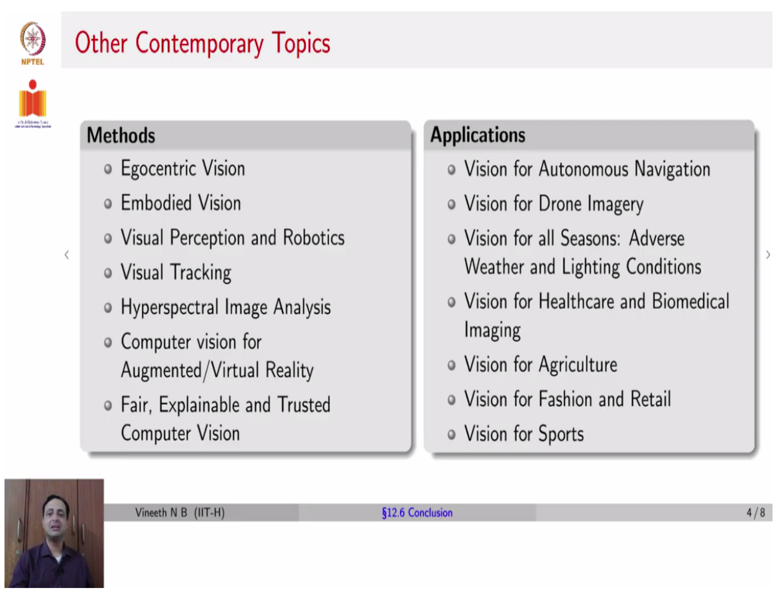


Another topic that we did not have the scope to cover in this course an important one is the intersection of Reinforcement Learning and Computer Vision. Reinforcement Learning, as you may know, is a paradigm of Machine Learning similar to supervised and unsupervised, where an agent takes actions and interacts with an environment to maximize potentially delayed rewards. And this is usually modeled as a Markov Decision Process.

Where is Reinforcement Learning used in Computer Vision? We already saw a couple of examples, we saw the use of Reinforcement Learning we at least briefly spoke about it in hard attention, where one cannot back propagate. We said we use Reinforcement Learning based algorithm. There is also NAS that we just saw, where Reinforcement Learning based methods are used to search for architectures.

Other use cases are in games, for example, in Atari Games, to automatically play the game, one needs to understand the visual elements on the screen, as well as use Reinforcement Learning to give a path of actions that the agent has to take to attain a specific reward. Visual servoing where a motor is moved, to be able to capture or achieve a specific purpose is done using Reinforcement Learning. Visual tracking can also be done using Reinforcement Learning, there are many more applications. But these are a few sample use cases of Reinforcement Learning in Vision.

(Refer Slide Time: 08:36)



More generally speaking, other contemporary topics, in terms of methods could be Egocentric vision, where the camera is placed on the person who is traversing an environment and not necessarily placed as a CCTV or a web camera or anything like that it is placed on the person.

So, the problems are associated with how one views the world around an individual using camera feed.

You could look an example application for this could be based on something like google glass a wearable glass, or could be very useful for developing technologies for individuals with visual impairments, where the problems have to be seen from that individual's, egocentric perspective, when we say egocentric, we mean from the viewpoint of that individual. You can also have Embodied Vision, which becomes similar to Egocentric vision, where we talk about the vision technologies being embodied in a human. We also have the intersection of Visual Perception and Robotics, which is a very useful and a very contemporary topic.

Visual Tracking, Hyperspectral Image Analysis where we try to understand images of different spectra, for example, ultraviolet images, or IR images, so on and so forth. A lot of the methods that we discussed so far directly can be applied to hyperspectral images. It may just be a difference in the number of channels in the input, so on and so forth. However, studying images across the visible beyond the visible spectrum is an important area of research for various problems, such as nighttime vision, or astronomy, or many other such domains.

Further, doing Computer Vision for augmented and virtual reality is also becoming an important topic. While it could have been for games or the entertainment industry, these days it is also becoming important for remote interactions of any kind, be it Instructional, Educational or simply a family interaction, doing Computer Vision on augmented and virtual reality settings is becoming an important topic. And finally, having fair, explainable and trusted models, continues to be a very important topic.

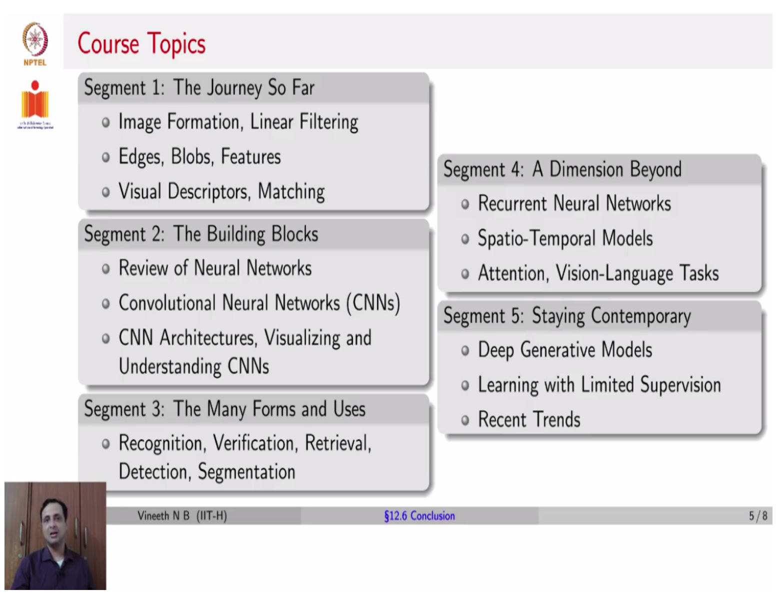
We saw various methods for explaining the decisions of a CNN, but one can go beyond and also talk about fairness of these models, to de-bias these models from any biases that may be there in the dataset, and help humans develop trust in these models. On the application front some of the applications that are quite popular, and have received a lot of attention in recent years, are vision for autonomous navigation, self driving cars, this is a very important topic that has a lot of investment from the industry. Companies like Uber, Tesla, Google itself, many other companies investing a lot of effort in this direction.

Vision for drone images. Drones are becoming increasingly popular for various tasks ranging from security to Disaster management. How do you understand images that come from drones becomes an important task. Vision for all seasons, a lot of the datasets that are available are all based on daylight scenes. How do you now make these tasks work on adverse weather and lighting conditions becomes a very general context of Vision applications. Vision for Healthcare, and Biomedical Imaging continues to be an important topic.

Vision for Agriculture has received a lot of traction in the last few years, to help Deep Learning models and Computer Vision models for Precision Agriculture, to understand Plant Phenotypes, to be able to spray pesticides, to be able to understand how much harvest you would get out of a patch of land, and so on and so forth. Similarly, Vision for Fashion and Retail, trying to use a virtual try on to see how a person looks in a particular dress, or simply using Computer Vision, such as an Amazon go in a shopping environment to automate the shopping experience is an important dimension of applications in vision.

And finally, an application of a huge commercial interest, which is Vision for Sports. In many different sports, Computer Vision is being used these days to understand biomechanics of humans, perhaps for rehabilitation, understand strategies on a football field, understand movements of a particular player for analysis, either could be for an opponent, or for a coach of that player, so on and so forth.

(Refer Slide Time: 14:20)



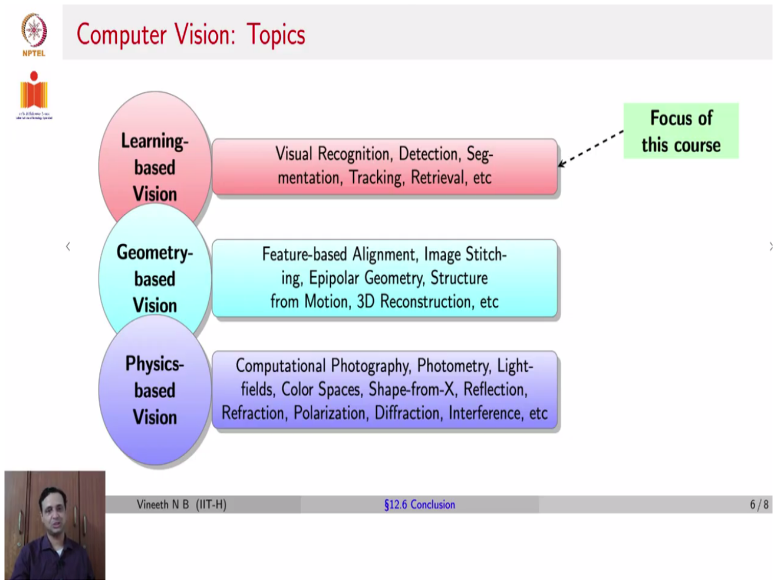
To conclude, let us revisit the outline of this lecture of this course that we started with. So we had 5 different segments. We started with trying to understand Image Formation and Linear Filtering. We then talked about Edges, Blobs and Features. We then went into Visual Descriptors, Feature Matching, so on and so forth. And then moved from Traditional Computer Vision to Deep Learning for Computer Vision, where we started with a review of Neural Networks and backpropagation went on to the building blocks of Neural Networks for Computer Vision, which are convolutional Neural Networks.

We talked about various different convolutional Neural Network Architectures, how to visualize and understand how they work. Then we also looked at the many forms and uses of CNNs, ranging from recognition, verification, detection, segmentation. We talked about the changes in Architectures, the changes in Loss Functions, and so on. Then we moved on to adding a dimension to the input, going from images to videos, or time series data.

In this context, we talked about Recurrent Neural Networks, and variants such as LSTMs, we at least briefly discussed Spatio-Temporal Models, as well as Attention and Vision Language tasks. And in the final 2-3 weeks, we discussed a lot of contemporary topics such as Deep Generative models, learning with Limited Supervision, and recent trends such as Pruning, Adversarial Robustness, Neural Architecture Search, and so on.

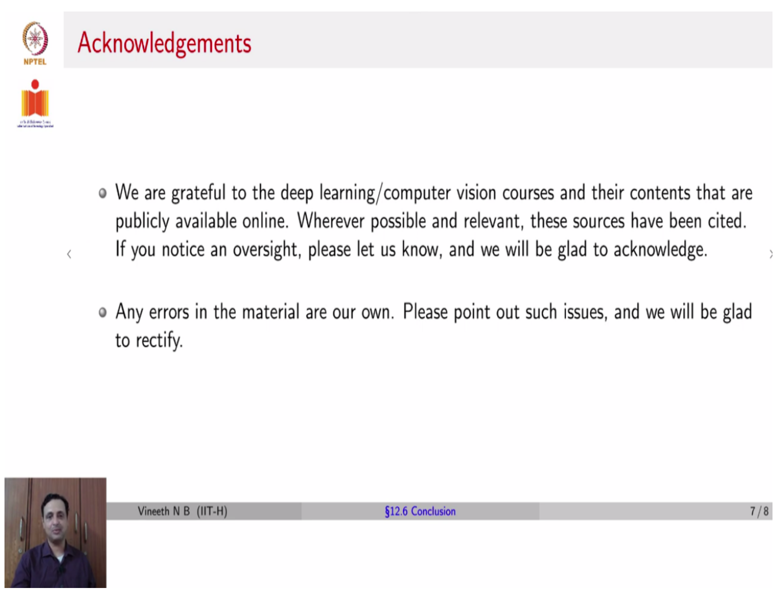
Hopefully, this course gave you a good introduction, and a good grip on the topics related to contemporary Deep Learning Models for Computer Vision.

(Refer Slide Time: 16:22)



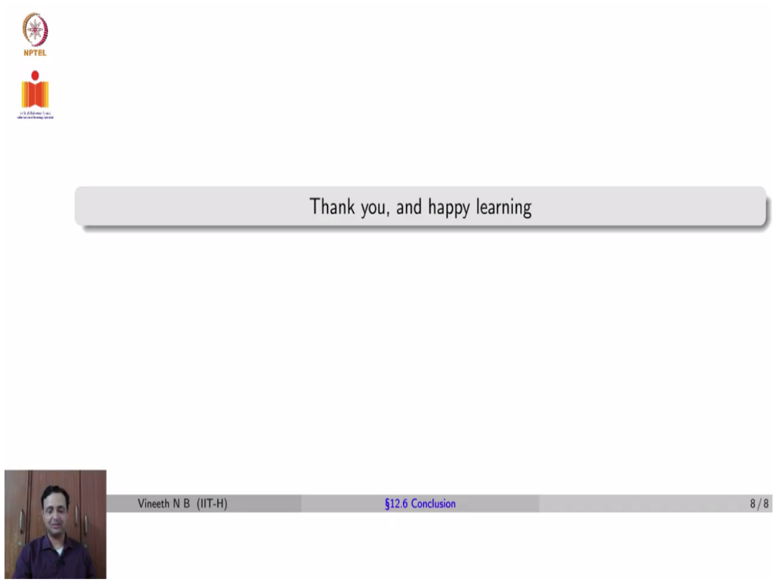
Let me recall that Computer Vision is far broader than the topics that we covered in this course. Recall that in the first lecture of this course, we said that one could divide Computer Vision into 3 parts, Learning based vision, Geometry based vision, and Physics based vision. For a large part, we focused on Learning based vision in this course. And you can go back and revisit the links that we provided for you if you would like to learn more about Geometry based vision, or Physics based vision.

(Refer Slide Time: 16:59)



With that, let me take the opportunity to acknowledge and thank all the resources, including Deep Learning and Computer Vision courses that are publicly available, which definitely inspired and helped create the content for this course, although it was content of the content of this course, was created exclusively. Definitely there was influence and content, adapted and perhaps borrowed from different sources. We thank all of these sources. And if there are any errors in the material, we will gladly take your suggestions and improve them as we move forward.

(Refer Slide Time: 17:35)



With that, thank you to each one of you for being a part of this course, participating in this course. Hope you learn something valuable by going through this course, through the lectures, as well as the quizzes and the assignments. Wish you Happy Learning and hope you get a chance to do more in this field. Thank you.

THIS BOOK IS NOT FOR SALE
NOR COMMERCIAL USE

(044) 2257 5905/08

nptel.ac.in

swayam.gov.in