

Machine Learning for Earth System Sciences
Prof. Adway Mitra
Department of Computer Science and Engineering
Centre of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur

Module - 04
Machine Learning for Earth Observation Systems
Lecture - 27
Object Detection in Satellite Imagery

Hello everyone. Welcome to lecture 27 of this course on Machine Learning for Earth System Science. Today we are going to begin module 4 and the topic of this module is Machine Learning for Earth Observation Systems. So, earth observation systems are is the network of satellite, sensors and all the measurement devices that are that have been deployed extensively over land, over ocean and also in the air space through space-borne vehicles.

And these devices, they keep on sending a huge volume of data of various forms to the ground stations. So, these huge volumes of data they can be like, these can be in situ measurements of say temperature or maybe other properties of the land or of the sea or it could also be like various photography captured from the air. There are satellites, radars and the all these kinds of devices which like which operate, like which basically capture imagery.

These imagery are need not be the kind of images that we are familiar with in computers, but they are like they have their own technology by which they send in some signal and then as that signal is reflected back they try to capture that reflected signal. And on the basis of the intensity of the reflection, they form some kind of a like image like structure and based on such reflections like it may be possible to like make some measurements about the like about the field on which they are focusing.

So, what we are going to do in this module is, we will see how like a useful information can be captured or can be extracted from this vast volume of data that is being generated by earth observation systems. The major focus will of course, lie on the satellite and radar imagery and because these are images, like we will see if the parallel progress that has taken place in the

domain of computer vision, image processing etcetera; if some of the algorithms used in these fields can be applied in this, like for in the case of remote sensing imagery.

And many of the algorithms that have been developed off late for computer vision or image processing, where the task is to like look at normal images which are let us say captured by human beings and put up on social media and things like that. Like much of contemporary computer vision is about interpreting those kinds of images as we have may have discussed in some of the earlier lectures.

And the most successful methods used in computer vision are deep learning models such as convolutional neural network and so on. So, the aim of this module is will be to explore how these methodologies or how these models may be used to extract meaningful information out of the remote sensing imagery.

(Refer Slide Time: 03:44)

CONCEPTS COVERED

- Deep Learning architectures for object detection
- Challenges of object detection in satellite imagery
- Adapting object detection frameworks for remote sensing

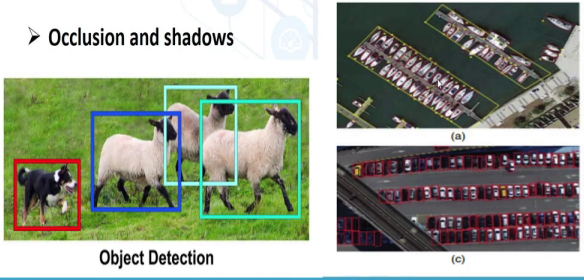
IIT Kharagpur
NPTEL

So, the basic concepts that we are going to cover today is deep learning of architectures for object detections, challenges of object detection in satellite imagery and adapting object detection frameworks for remote sensing.

(Refer Slide Time: 03:56)

Object Detection as a Challenge

- Shape, scale, orientation, other properties of object may vary
- Representation of target's visual properties challenging
- Occlusion and shadows



Object Detection

NPTEL

IIT Kharagpur

So, object detection like we know is a like is one of the standard tasks in computer vision. So, let us say this is a standard image in which like as you can see there are there is a field on which 3 sheep and 1 dog are moving, so that the target is to detect the different objects that are present in the image. So, in this case like you can see that there are actually two class, like apart from the background grass there are 2 classes of objects or we can say broadly we can say there is one class of object the animals.

Then within the animal class there are again subdivisions; this dog and sheep and then for sheep again there are three instances. So, first there is like the broad level category of objects, then there are like in or more fine grained categories of objects and then for each category of objects there are there can be multiple instances. And now, so this is a normal image that we that computer vision typically deals with.

Now, if you look at these image, these are aerial images, they may be taken from some helicopter or something like that over as you can see over a port or over a parking lot. So, this port has like lots of boats parked one at a time. So, the task might be first of all to detect the port, that is like or to detect this see the series of vehicles that are parked here or it can also be to detect the individual boats.

Like for example, I may be interested in counting how many cars are there are currently parked in the parking lot, is there overcrowding; is there space to park some more vehicles and so on. So, to do this in an automated manner, I need to like actually detect each individual car and so that I can count their number.

So, in any object detection problem, the challenges are shape, scale, orientation and various other properties of the objects which may vary from one image to another. So, like in this case like we can see that the dog is here and its size is something like this, but in another image like if that image may focus entirely on the dog. So, the whole image can be focused around the dog only, in which case its shape and scale will be very different or orientation.

In this case like we are seeing their side views, but we could also see their front views. Their front views like these objects could look very different from like when viewed from the front or from the top.

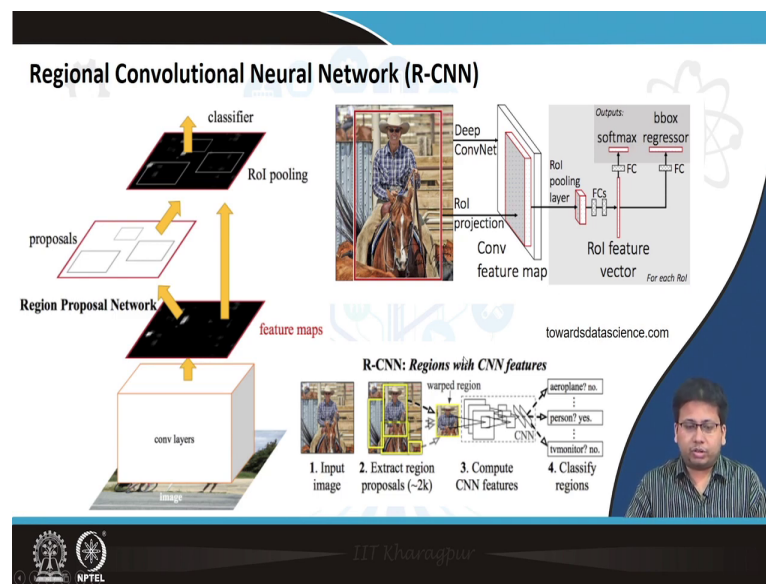
So, like all these properties can vary from image to image and to learn a representation of a particular class of objects such as dog, sheep, boat or car, one has to build a representation that is either robust to multiple to different shape, scales, orientations etcetera or like we have to have an ensemble of models like for like that your ensemble of representations one for one orientation another for another orientation and so on.

So, but representation of these visual properties is itself a challenging task. So, earlier like in classical image processing computer vision, like the it like we used to have features inspired by signal processing now, of course, we do not do that. So, those features are sometimes now referred to as hand crafted features. Now, computer vision community has moved away from those features, nowadays they use mostly neural features which are automatically learned by neural networks.

So, that has removed the feature representation problem, but the neural features that are learnt, whether they are actually able to represent all the like variations of the properties of the object that is an open question. And it may vary from one application to another. And then of course, there are some secular challenges which are irrespective of the method you used.

So, these are like occlusions, shadows etcetera. There by occlusion I mean that the thing or the object which you are trying to detect that may be partially or fully covered by another different object or there might be shadows which like, if a shadow of a tree falls, then this white sheep that can be look blackish. Or then that can kind of these kinds of occlusions and shadow they can really mess up with any object detection methodology.

(Refer Slide Time: 08:43)



Now, coming to some of the rather well-known methods for object detection which have been used off late in computer vision. So, one is the R-CNN, the regional convolutional neural network. So, in a like when we are doing the object detection in any image, the most like one of the biggest challenges is first of all to localize where the object may be. And secondly, and next to understand whether the object is there or not.

So, like here is an image like let us talk about this image. So, here we are let us say we are looking for a person. So, the person as you can see is located in this part of the image. But how do I know that it is really this part that of the image that I should focus about on? Why not say this part or this part or this part or any other part. So, what R-CNN does is, first of all it passes the image through a series of convolutional layers just like any CNN. So, that we have a like a neural representation of the like of the image.

The next important thing which happens in R-CNN which is different from other neural networks is the regional proposal network. So, it proposes something like a regionalization of the image, like a person is located here other things are located. Or may I mean not necessarily person, but the target class is located here and the other things or other objects may be located in something else or the target object may be located here other objects may be located somewhere else.

So, these are all proposals and each proposal needs to be evaluated, whether that kind that particular regionalization has any makes any sense or not as far as this image is concerned. So, what the region proposal network does is, first it proposes a regionalization, then on each of the regions that it has proposed the object detector works or which is basically something like a classifier.

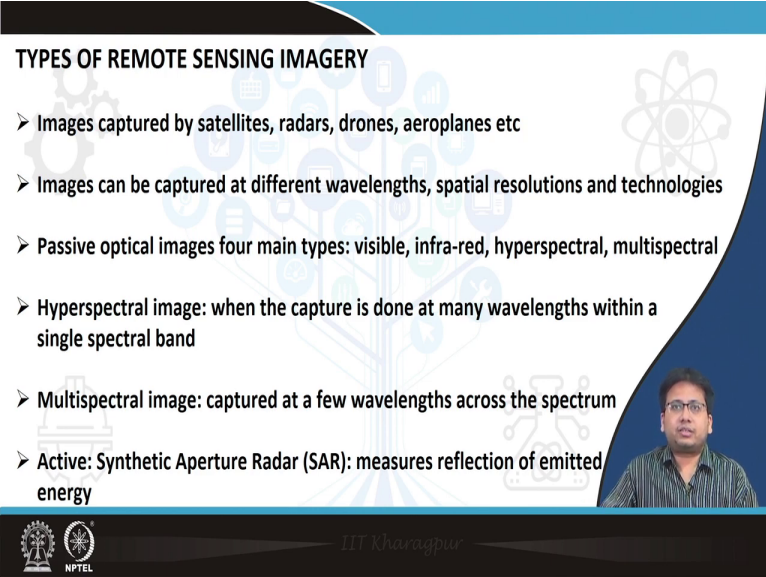
It the classifier tries to classify each region of the image, whether it belongs to the target class or not. That is if this is this region of the image, if this is a region proposed then the classifier will try to classify this region as a person or non-person, it will say hopefully it will say non-person. Then and that prop like we may choose to reject that proposal, then another proposal will focus on this region and then it will like, again we will try to classify is this person or not person. Again, it will say not person so we will reject it.

Then we may focus on this region, this region might be proposed by the network and here the classifier will run and then the it will say that yes it is a person. So, like we can say it will be accepted. Then similarly like other overlapping regions involving the person they may be evaluated I mean classified and like then all these things like all the regions that have been accepted they will somehow be pooled together.

So, there is an RoI pooling layer. So, region of interest RoI stands for region of interest. So, the different regions of interest will be evaluated and then pooled together to form a single representation and that thing is next going to be like pass through another neural or some other neural layers so then so that the final output will be one the that is the output whether I mean the binary output, whether the target class is present or not that is in this case a person is present on the image or not.

somehow be all combined together to form a final like detection or I mean to say to find the final bounding boxes or not.

(Refer Slide Time: 14:33)



TYPES OF REMOTE SENSING IMAGERY

- Images captured by satellites, radars, drones, aeroplanes etc
- Images can be captured at different wavelengths, spatial resolutions and technologies
- Passive optical images four main types: visible, infra-red, hyperspectral, multispectral
- Hyperspectral image: when the capture is done at many wavelengths within a single spectral band
- Multispectral image: captured at a few wavelengths across the spectrum
- Active: Synthetic Aperture Radar (SAR): measures reflection of emitted energy

The slide features a blue and white color scheme with a background of faint icons related to remote sensing. A small video inset in the bottom right corner shows a man with glasses speaking. The bottom of the slide includes the IIT Kharagpur and NPTEL logos.

And now this is normal object detection. Now, let us talk about remote sensing. So, the same object detection problem we want to use not on ordinary images which we capture with our cameras, but which are captured by captured through remote sensing. So, now if we want to use object detection on the remote sensing imagery, like first of all, let us look at what are the kinds of remote sensing imagery. So, images are captured they may be captured by satellites, radars, drones, aeroplanes, and so on.

There are there can be more sources of remote sensing data. Now, these images we they can be captured at different wavelengths, spatial resolutions and as well as using technologies. So, the wavelength like as we know there is the electromagnetic spectrum; a part of it is visible to the human eyes, other parts are not visible.

By that I mean that there are certain wavelengths and at certain wave like if the images are captured within some wavelengths, we can say that they lie within the visual range that is by looking at those images we can understand that ok this is a car, here is a car, here is a tree, here is a human being etcetera etcetera.

But in other wavelengths we may not be able that is by looking at them we may not be able to make much sense of it. So, there are basically 4 major types, the visible images, infra-red images, hyperspectral and multispectral images. So, the visible image as we know like it sometimes also can be called as RGB images. So, there is a red channel, green channel and blue channel. So, these are actually different wavelengths or a set of wavelengths within the visual, visual range of the spectrum.

But then there are there is infra-red images which is like of course, which means at a lower frequency than the then the red wavelength so to say. And then there are hyper spectral and multispectral image. So, a hyper spectral image is something is such a an image which where the capturing is done at many wavelengths or we can say at a like a series of wavelengths which within a single spectral band.

So, let us say there is that the entire spectrum can be divided into several blocks and we focus on one particular block just for saying let us say like 30, 30 to 50 let us say this is one block. So, within that there are there is there are so many wavelengths are possible. So, we choose, that is we do the image capturing at like at a very large number of wavelengths within that block. That is called and the set of images we get like that is called a hyperspectral image.

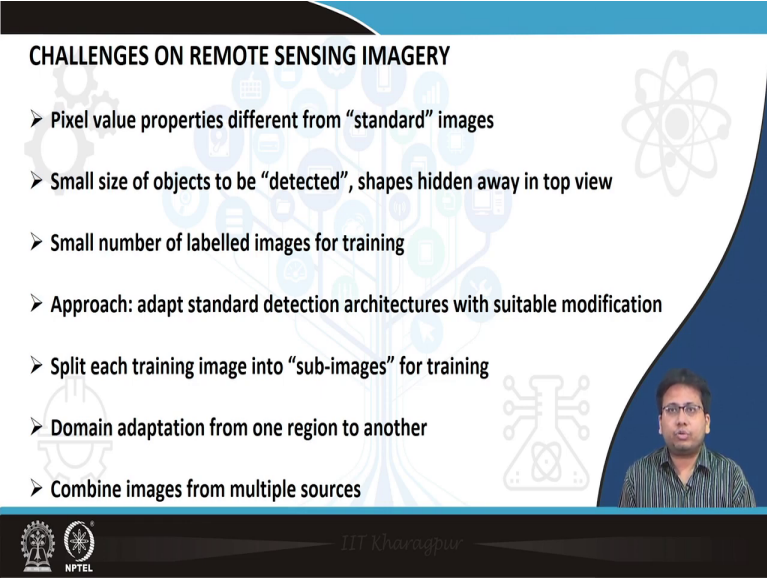
So, like its can be represented as a cube, that is at each wavelength we are getting a 2D image, but then there is a long sequence of images like that, each at a different wavelengths. So, taken together they form something like a cube. Then there is multi spectral images so this is captured at a few wavelengths, but that are across the spectrum. So, like unlike the hyperspectral images, where all the wavelengths are within a particular block here the different wavelengths are spread across the spectrum, but they are relatively few in number.

So, multispectral and hyperspectral we can say they are complementary to each other and so but now all of these can be some these are sometimes referred to as a passive sensing. Now, there is something called active sensing also, where like the active in the sense that you the device actually sends some signal and then like measures how much of that signal is reflected.

So, in the passive case, the device does not send any signal, it merely like tries to capture how like whatever light etcetera is reaching it. Now, in case of like a among active sensing there is

this synthetic aperture radar or SAR technology which basically measures the reflection of the emitted energy.

(Refer Slide Time: 18:50)



CHALLENGES ON REMOTE SENSING IMAGERY

- Pixel value properties different from “standard” images
- Small size of objects to be “detected”, shapes hidden away in top view
- Small number of labelled images for training
- Approach: adapt standard detection architectures with suitable modification
- Split each training image into “sub-images” for training
- Domain adaptation from one region to another
- Combine images from multiple sources

The slide features a blue header and footer. The footer contains the IIT Kharagpur logo, the NPTEL logo, and the text 'IIT Kharagpur'. A small video feed of a man with glasses is visible in the bottom right corner of the slide content area.

Now, like when we are doing remote object detection or remote sensing or for that matter any other problem on remote sensing imagery, these are some of the problems we face. So, first of all the pixel value properties might be different from standard images. So, like we like we know that in standard images like there is an R-channel, G-channel, B-channel each of the like pixels they have a value between 0 to 255, which somehow indicates how red they are or how green they are or how blue they are.

If it is a grayscale image it means either how black it is or I mean how dark it is or how bright it is and so on. In this case like these pixel properties they may be different, depending on the sensing technology use the spectrum used and so on. And as a result the way like the responses of the different filters and so on which we are use like which we use in standard imagery, if I mean they those filters may not be very suitable for like in this domain or on these kinds of imagery.

Then when it comes to object detection, so first of all there is a small size of the objects to be detected, because if they are taken from the top like if you want to do a ship detection or car

detection from the top; obviously, it will be of small size and their shapes may also be hidden away, if we are taking a top view; this. Now, also the another problem is that there is a small number of labeled images which are available for training.

Because these images are difficult to capture, I mean or even if they are not difficult even if the modern technology can capture lots of these images, but the big challenge is labeling them, that is who will actually sit and do the labeling of all the images that is this image contains such and such objects and then localize those. So, that is a very tedious task.

So, some of the approaches which are taken to alleviate some of these problems. So, sometimes, what we do is like we like on the standard detection architectures which we talked about earlier R-CNN, YOLO etcetera, we just make some suitable modification for the domain and use them. Another thing that is often done regarding the problem of small number of labeled images. Now, we take a reasonably small size of imagery, but then each image we split into many sub images and use them for training.

So, instead of capture like a normal neural network for used for computer vision, they are usually trained on millions of images, but in this case it is of course, very difficult or impossible to collect millions of images and annotate them. So, what is done is manageable number of images let us say a few 100s or at most a few 1000 images are collected, they are labeled and then those labeled images they are split into multiple sub images and each sub image is used for training purpose.

And now sometimes also domain adaptation is used from one region to another. And finally, we combine the images from multiple sources to get some kind of a unified representation, that is, like we may go like use the images of the same phenomena or the same region captured by let us say three different satellites or a satellite and radar and so on.

And then somehow, we combine all those together to now the reason for doing that is that the satellite image may have certain properties, like which the radar image may not have or the hyperspectral image may have certain properties with the multispectral image may not have and vice versa. So, like if we do the detection separately on the different images, maybe in some

images it will or like the images from one source it or one technology it might be the object may be easier to detect, in the other it may not be so easy to detect.

So, what we do is we like do something like fusion. So, either we treat all the images from the different sources independently do the detection in each of them and bring like and then combine their results somehow or we combine the images themselves from the different sources and do something like a whatever something known as image fusion and then from the that is build an unified representation for all those images. And then on that unified representation we do the object detection or whatever we want to do.

So, like these are some of the concepts which we will see in detail in the coming lectures. So, in the next 3 or 4 lectures, we will see some detailed applications of the like of the things we that we discussed today. So, in so, we will follow this up in the coming lectures. So, till then bye.