Artificial Intelligence for Economics

Prof. Adway Mitra

Artificial Intelligence

Indian Institute of Technology Kharagpur

Week – 08

Lecture - 38

Lecture 38 : Computer Vision for Economics

Hello everyone. Welcome to this course on Artificial Intelligence for Economics. I am Adway Mitra, an Assistant Professor of Indian Institute of Technology, Kharagpur. So, like we are reaching the last week of this course. Today is the 38th lecture and today we are going to focus on some of the more specific or specialized topics of artificial intelligence applied to economics. Today's topic is going to be computer vision for economics.

So, like we will start off by giving you a brief background on computer vision and what are the typical tasks in computer vision. We will discuss a very important issue how we can do feature representation for tasks in computer vision and finally, we will also discuss about proxy variables and some use cases for which we will look into some recent research papers where these computer vision has been used for various economic tasks. So, first of all what is computer vision? Computer vision as the name suggests it is something which enables computers to see and what do I mean by that? It means we these are algorithms by which a computer can analyze visual data that is to say images. Now, as we all know image can be stored in a computer in the form of a matrix is usually a three dimensional matrix $m \times n \times l$ where m is the number of rows that is the height of the image n is the width of the image and l is the number of channels.

Now, what is that l equal to 1 means grace for gray scale images which is like typically a black and white image and l equal to 3 for colored images. Now, we say that an image is high resolution if m and n are high otherwise it is called a low resolution image. Now, what this is how it is an image is like x is represented as a matrix, but what are the contents of the matrix. So, if you can if you consider any particular pixel in the x matrix that is let us say small m small n and small k. Then

it is like every pixel is a number between is an integer between 0 and 255.

0 means I mean if it is close to 0 that means it is dark and if it is close to 255 that means it is right. So the 3 channels these as we most of us know anyway these are the 3 color channels the red, green and blue channels. So, for each of the channels every pixel or the pixel at every location has a value. So, for the green channel pixel value of 250 may mean it is a very strongly green pixel while a red value of let us say 12 at the same pixel may mean that its redness is very low. On the other hand if the pixel value of the blue channel at any given location let us say is 127 that means it is a moderate or lightish blue color and so on and so forth.

So now this way we can have images represented in a computer. Now what do we do with these images or what do we want the computer to do? the first thing which i may want the computer to do the most basic task of computer vision is image classification means that given a particular image we assign it to a class based on the objects which is contained in it now each possible object class may have an associated probability like as you can see in this case like it is there are like there are 3 sheep and 1 dog in this image. So, like we are that is a typical image classification we will classify this image as a sheep image with 60 percent probability a dog image with 30 percent probability and other possible class levels they also may have some probability because if the model is not too sure of like what objects are present in it. Now, so like this is like this classification is done by any kind of classification algorithm in the same way that usual classification takes place that is the image is somehow represented as a feature vector and then the like the feature vector is fed into some classifier which has been suitably trained. So, the classifier it can be either a neural network or a support vector machine decision tree random forest anything.

and it will give me like either it will give me one particular class level or it will give me a probability distribution over the class levels as you can see in this case. Now, the challenge here is from where these features will come from. So, like the success of the problem largely depends on how good features or feature based representation we can get for the image. So, why because the major challenge in case of the object recognition or or image image classification it is intra class variation or inter class similarity. So, as you can see like these are all images of calculators these are all images of checks board, but their visual characteristics are quite different that is as humans we can understand that these are all of these are

representing the same object, but there is that objects I mean the images themselves are considerably different from each other.

So, there is a lot of intra class variation. Similarly, there can be inter class variability also in some cases that is two things may they may not belong to the same class, but they can still look similar to each other. So, the task in this case is how to find image features which will be discriminative that is which will be able to which will be robust to such intra class variations as well as inter class similarities. So, that is one task another task in computer vision is that of object detection given an image find all occurrences of a particular class of objects. Now, usually we can put a bounding box around the location of that object as you can see in this case they are we are looking for human faces.

So, it is a face detection whenever there is a human face the model is putting a bounding box around it and may be along with it providing a confidence score also. The different instances of the same object they may be at different scales and orientation. So, like some of these faces like this face is a small one, this face is a large one. So, they can have different orientations also like this lady's face is like it is not straight it is inclined in a particular way while this man's face is upright. So, there can be significant amount of variations among the instances of the same object.

And the general approach here is you take any region of the image and you make a binary classification whether it contains your target object or not. Now for this of course you have to scan different regions of the image and you have to those scanning regions they should be different sizes also. Let us say in this given image I want to detect a dog. So, I may choose different candidates regions of the image. So, like say here this small box I can choose and I can classify it as either a dog or not dog.

So, this should be hopefully it will be classified as a not dog. Then this place I can select and I can classify it now just based on this small box in probably no algorithm will be able to classify it as a dog. So, it will still be called a non dog. Now, but like if there is a region here and around the face of the dog and which is like when we it may get classified as a dog. So, then like now that at some level dog has been identified then we may want to find the exact extent of the dog.

So, instead of considering small boxes like this I may scale myself up to bigger boxes and then try to classify these bigger boxes as dog or non dog. And finally, I may zero in upon this particular box as this is this box contains a dog while no other box of this size contains a dog. So, that is how object detection happens. Another related task is image segmentation that is divide the image according to the objects contained in it. So, all pixels inside a particular object class should have the same value.

So, and if there is so like we can have either semantic segmentation that is all pixels belonging to one class of objects they will be like clustered together and all pixels belonging to another object class they may be classified as in some other ways. So, like in this case like as you can see there are only two classes sheep and dog. In this case however, the every individual sheep will have to be like separated or segmented separately. So, this is semantic segmentation, this is instant segmentation. in this case like as long as they are sheep i am i am putting them as part of the same segment but in this case like different sheep also have to be separated out so this is a slightly more difficult problem but this image segmentation is essentially a pixel clustering problem The main clue here lies in the fact that pixels which are part of the same object they are likely to have a similar values except for the fact that they are also likely to be close to each other and so like we these two cues we can somehow use to do the pixel clustering.

Now as I said like in any of these image related problems the main issue is that of image representation. Now how do like I have the image which is represented as the matrix as we show, but from how do I represent it as a suitable feature vector which can be fed as input to a classification algorithm right. So, these pixel values they like one obvious way is like you have the like the image itself which is a collection of the pixel values you just flatten it in the form of a like a vector and provide it as an input. The problem these approach has two obvious disadvantages, the first disadvantage is that if you are it and making it as a vector then special information you are giving away that is 2 pixels which are neighboring in case of the original image they may no longer remain neighbors if you distort the image and make it into a vector. So, that problem of localization is of course, there another problem is that the pixel values themselves may they they are useful, but they may be too noisy or too inconsistent they may also contain some unnecessary details which are not relevant for our classification purpose.

And alternative is instead of focusing on the raw pixel values you extract the edges which are present in the image the corners that are present. may be textures or any other kinds of interest points and build a representation centered around these like these kinds of properties. But again finding these kinds of things is also not an easy task in any like in a image which is been taken anywhere in the world I mean unless it is taken in a very controlled setting it is very difficult for it is difficult to come up with an algorithm which can identify exactly where the edges of the image are present or that representation of the texture and so on. And secondly even if we find the edges or the corners etcetera how to represent them mathematically in the form of a feature vector that is still a difficult task. So, people in the image processing community had been working on it for a long times in most of the 2000s decade the image processing community was busy building various kinds of image features and descriptors that is and like these were primarily done using filters.

So, like filters are mathematical operations which are like which are carried out on the image matrix and like a typical filter will have high response in some places and low response in some other places. A filter basically looks for some localized spatial pattern if it is able to find such a pattern then it will return a high response else a low response. Say for there are filters which are good in finding corners or which are good in finding edges and so on. So, we can like we can convolve these images with those kind of filters and get the results. But then the question is how many different kinds of filters will I go for and like as I said like a filter is I mean it requires its own parameters.

like these are two examples of filters which are also known as masks that these colors they represent different pixel values so here i am when i specify that these colors i like what i am what i mean is that i am in the original image i am looking for a spatial pattern of pixels which has which is somehow somehow similar to this or to this. Now, these exact patterns these are something which I have specified according to my interest, but in most cases I will not know which what kind of features or what kind of I mean these local patterns I should be looking for. So, like a few I can specify by myself, but like in a if we have a large connection collection of images then there is also may be a very vast space of this kind of local features which are useful for me. So, how do I find that it is not an easy task. So, like we typically ever since the deep learning took over in 2012 the field of computer vision has been revolutionized.

not because not only because they can represent complex spaces for classification I mean even if the feature space is very complex with very difficult decision boundaries then neural networks can represent that because they because as we have discussed earlier also neural networks can represent highly non-linear decision boundaries. In addition to that, neural networks are also very useful in finding image features. That is ever since neural networks have made a lot of progress, we the image processing community no longer needs to invent new filters. Rather the most of these filter when we are considering a convolutional neural network what we are essentially doing is we are scanning an image with a mask and that is this convolution operation as we have seen earlier also it can be represented using a neural network. But where does this mask come from? So, like in large neural network each convolutional layer they involve multiple convolution operations which may with multiple filters on the same input.

So, as a result what we get is a creation of feature maps that is you do not have to fix one filter and look for it you can have a large number of filters. And these filters are actually they are learnt during the training phase itself that is when training the neural network what we are really doing is we are or what the neural network is really learning is it is finding what are the patterns it should be looking for. And accordingly like in the testing phase it is applying those same patterns or filters representing those same patterns and accordingly it is doing the classification. So, like this way like it allows us to have a like a multilayered description of each image which is known as a feature map. So, like for a single image of a dog.

So, what you are seeing here these are the feature maps at different stages of the convolution. So, now we can ask that how is this computer vision how is it useful for economics. So, like there are certain obvious questions can we measure the economic development in a region by analyzing some sort of imagery. So, like one source of imagery might be Google street views that is like Google street views we know are available on different streets. These are images taken from vehicles moving along the different streets of a city or a village or so on.

Now like they those images they capture the surroundings of a particular neighborhood. Now by like can a computer classify those images and identify which locations are like seem to be affluent and which seem to be more

impoverished and so on that is one task. another task might be like another good source of imagery is satellite imagery i mean google for google street view we have to go to every place and snap so many different photos so i mean every the coverage of a single photo will not be very large but when we like we have so many satellites which are encircling the earth so and they are taking images all the time so like if we can like an image taken from the sky will of course have a have a a huge coverage and by if we are able to analyze it then we can potentially get information about a wide region. The problem here is that these satellite images they will not directly I mean they may that is by analyzing them you may not directly be able to identify different characteristics of rich and poor regions. So, like these satellite images they are captured at different bandwidths some of these bandwidths are like the optical bandwidth that is they can be treated as natural images albeit they are taken from a different orientation namely from the top other at other bandwidths these like they cannot be treated as optical images at all that is they may have some very different I mean if we see them with their with our eyes we may not be able to make much out of it.

So, then how do we utilize the such satellite images to measure economic development and so on. So, like even if we cannot identify a specific structures from these images, we can still find some proxy variables. So, like there are certain indices like there for example, there is something known as vegetation index or NDVI. So, this it is basically a measure of the greenness in the different pixels. So, we can now understand that a greener region is or the pixel over region every pixel now corresponds to a particular region of the world which because it is taken by a satellite from above the world.

So, every pixel represents a region. So, if that pixel has a lot of greenness, it may mean that there is some green thing in the region around it for example, foliage. So, if a particular pixel has high value of this vegetation index, it may mean that there is some forest or agricultural land in that region. Similarly, we can have this built area indices also. Now a very interesting proxy variable for measuring economic activity or economic development is night time lights. So what you are seeing in the in this image is like the like how the region how the Indian region looks from the satellite so the like at night.

So you can see that some places are very bright and some places are quite dark. So, like the bright is like we can say these regions are I mean the it is bright

mostly around the cities where there are lots of lights which are which are on even during the night. And the places which are dark they are mostly the forested regions or the regions where there are relatively less human settlements and as a result there are less lights at night also. So, this night time light it is a very interesting proxy variable for economic development. Now, what we will do is let us look at a few research papers will not see them in detail, but we will just give you an essence of in these research papers how computer vision has been used for economic development.

So, the first paper as you can see are night time lights are good proxy for of economic activity in rural areas in middle or low income countries examining empirical evidence from Colombia. Colombia is a country in South America. So, like this is a map of the country and here like what you are seeing the these pixels they indicate the night time light. So, in the brighter pixels they probably represent cities or other human settlements the black are the forested regions where there are very less night time lights. So, let us see what they have to say the use of satellite imagery particularly night time lights has flourished in the last 20 years of socioeconomic studies.

The intensity of the lights captured through remote sensing is frequently used as a proxy of different socioeconomic indicators while some studies found a high correlation between night time lights light intensity and gross domestic product there have been inconclusive debate about the validity of this assumption that night time lights can serve as a good proxy for the economic development to sub national studies particularly in the rural areas of middle and low income countries. We test the suitability of night time lights from publicly available data sources for estimating the regional domestic product across like municipalities with different degrees of urbanization in Colombia. We use a series of cross sectional regression models to compare correlation between municipal RDP and luminosity from different sources in 2012. So, these are different like sources of night time light may be from different satellites or using different remote sensing technologies as well as multilevel regression models to estimate RDP time series from 2011 to 2018. Our findings reveal that all compared night time light products can serve as a good indicator of municipal RDP patterns.

while the one particular data source VI and VIRS. So, it is a type of remote sensing is imagery it represents the best model fit. So, and so, so like they this you

can say based on correlate simple correlations they have found there does exist a strong correlation between light night time luminosity and economic activity not only at large geographical regions like countries, but even very small local regions like individual municipalities. So, like this is the first paper we will discuss another paper which has similar aims this one poverty from space using high resolution satellite imagery for estimating economic well-being. So, like so, like here if you look at this image it is this is an optical image taken from the sky.

So, this is not like here as I said. Satellite imagery can be captured at multiple wavelengths. So this one has been taken at a wavelength which is within the optical region. So if we look at this image, it looks more like the Google Earth images which many of us are familiar with. So like you can identify structures like buildings, roads, trees, etc.

and also cars. So, what they are showing in this image is they have used some kind of neural network based car detection model to identify the different the cars and as we know vehicles or cars they can act as a proxy for economic development. So, coming to the abstract can features extracted from high spatial resolution satellite imagery accurately estimate poverty and economic well being. The present study investigates this question by extracting both object and texture features from satellite images of Sri Lanka. these features are used to estimate poverty rates and average expected consumption from small area estimates derived from census data from 1291 administrative units so an important task here is whenever you are making a like we are training a model to predict something of course we need ground truth data that is like in this case the ground truth data which is economic activity this has to come from some kind of census or something that has been carried out at very high spatial resolution let us say at a municipality level or taluk level and so on so so that like we the imagery we will have but how i can calibrate the imagery to the actual economic activity or economic attributes of that region so for that i need the as the those economic attributes as ground truth which can only be obtained from ground based surveys The problem here is that ground based surveys cannot be done like at very high like everywhere they can be because it requires a lot of manpower it requires I mean human resources it requires like infrastructure cost and so on.

So, we do not have that. So, like we can use like if we have it in some places like we can use them to calibrate a model and then that once the model has been

trained and calibrated then it may be it can be used in other regions also even though the ground truth may not be available there. So, here in they have specified what kinds of satellite imagery they have used. So, they have they consist of 55 unique scenes which are purchased from a digital globefrom the study sample area. Each scene is an individual image captured by a particular sensor at a particular time. So, images were acquired at by three different sensors which they have listed here and they have also specified which bands or which I mean which wavelengths they are considering the images and like.

So, these are some of the maps which they are they have of the so like some of the features which they have used. are also like these night time lights which we already discussed. So like this is a map of mean night time lights at the different like small regions of their study region. As you can understand they have like divided the study region into small cells and at each of the cells they have plotted the mean value of the night time lights as obtained from the satellites. and they are correlating them with the ground truth data which is like the true poverty rate and the population density.

And like here like they have also considered like other proxies for economic development let us say buildings. So, like they have they are they may they have used these object detection to like we have earlier also seen this car detection they have also done that is they have a list of things which they objects which they consider as indicators of economic development like cars buildings and so on. and they have like they have strained object detectors for each of them so they have run those object detectors on these satellite imagery and they have in different regions they have identified the density of cars the density of houses and so on from which they are like trying to identifying the identify the different economic activities so like these are the things like cars building densities agricultural lands paddy lands vegetation shadows and road variables and so on. So, like they have discussed in details how they do the feature extraction from these high resolution imagery like from the I mean the spectral data we I mean the I mean the the wavelengths and all that we already discussed. So, from those from the satellite imagery what kind of features they extract.

some of them will of course, be neural features used by the object detection models apart from that they may be doing some manual feature extraction also which they have discussed in details and then they have finally, come to the

question do high resolution satellite features actually explain the poverty gap that is like can we say that a more impoverished region like the they they are features are significantly different from a like a rich region and that is what and they answer in the affirmative that yes they are they are able to identify certain features from the satellite imagery which are like which are clearly distinguishable or which are clearly discriminative between the different I mean between regions having different levels of economic development. So, one more there are more studies like that this one focuses on using publicly available satellite imagery and deep learning to understand economic well-being in Africa. So, here they are like they have some asset wealth index again which is obtained by. So, they have certain measures or measurements of wealth like the quality of house floors, water supply, then phone, radio, TV, etcetera. and so so using these kinds of asset wealth they build their ground truth which again has to be come from surveys only and then on top of that they have their high resolution satellite imagery as you can see here they they have the night time light imagery they have the multispectral imagery and so on now based on every type of the imagery they can they have their their model can make some certain predictions related to the economic development and so like for this kind of imagery the night light imagery there there is a type of prediction.

So, the prediction what are they predicting they are predicting the wealth index as I already mentioned. So, like for different types of from different types of imagery they will be making different predictions the predictions may or may not match at some places, but they are basically having an ensemble of predictions from the different types of imagery and based on that they can they can either combine them or they can choose which which is the best i mean like the estimates from which kind of satellite imagery represents the ground truth the best and so on and so forth and they also have used various deep learning models using the i mean they are all based on cnn using certain specific architectures like resnet and so on and so like all the discussions are i mean the those details are all provided here which of course will not have the time to go into this is one more work. So, here we are looking into transfer learning from deep features for remote sensing and poverty mapping. This transfer learning this is an important aspect. So, transfer learning basically means that we train the model on one kind of data and then adapt it to another kind of data.

So, the like we like as I said to train these models we need a huge amount of

ground truth data which can come from surveys. So, we may not have such accurate survey data in all regions of the world. So, let us say I want to do this kind of analysis for the whole continent of Africa. but there are only some places where such surveys have been done. So, I collect the those survey data as well as the satellite imagery from that region learn to calibrate them using the suitable machine learning algorithm.

Now the now once the model is trained you apply it to other regions also to that is where you have only the satellite imagery and from that you make the predictions of the of the economic development. However, the features may from one region may be may differ to another region I mean the characteristic of the satellite imagery over the these regions which are more densely populated they may be quite different from the these regions of Africa which are more like which contain more of desert. So, the data distribution in a sense is changing from this region to this region as a result of which the decision boundary may also change. So, like these kinds of things are handled by a paradigm of algorithms known as the transfer learning and it is this which they have used to make sure that the models trained over some regions can be exported to carry out the same activity in other regions also. so similarly we have some more papers also along these lines so this is a set of references i mean i have there are lots of papers in on this topic but i have made these five recent papers for your reference so you are advised to go through all of them to get a fair understanding of how computer vision can be used in economics so with that we come to the end of this lecture in the last two lectures also we will be focusing on specialized topics like this so see you then take stay well and take care bye