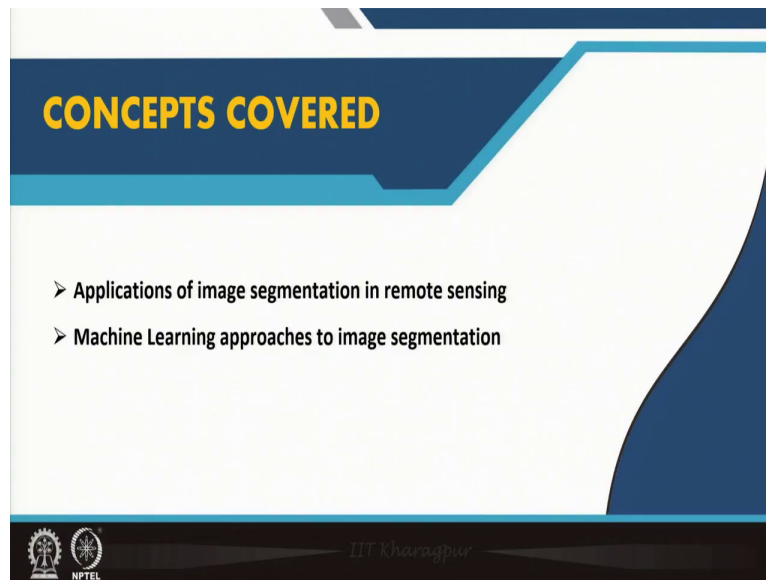


Machine Learning for Earth System Sciences
Prof. Adway Mitra
Department of Computer Science and Engineering
Centre of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur

Module - 04
Machine Learning for Earth Observation Systems
Lecture - 30
Image Segmentation for Remote Sensing

Hello everyone, welcome to lecture 30 of this course on Machine Learning for Earth System Science. We are still in module 4 where we are using the where we are studying how machine learning can be used for remote sensing for basically machine learning can be used for earth observation systems that is to say remote sensing. The topic of this lecture is Image Segmentation for Remote Sensing.

(Refer Slide Time: 00:51)



So, what we are going to cover today are different applications of image segmentation in remote sensing and how machine learning approaches especially deep learning can be used for such image segmentation.

(Refer Slide Time: 01:03)

Image Segmentation

- Image segmentation: divide an image into multiple parts
- Each part corresponds to one object/one class of objects
- Pixel-labelling task, adjacent pixels may have same label
- In remote sensing, standard application: LULC
- Classify each pixel according to type of land-use (eg. forest, building road, water body etc)

Challenge: In top view optical imagery, feature representation of classes difficult
In non-optical imagery, even more challenging!

Semantic Segmentation

Instance Segmentation

IIT Khuragpur

NPTEL

So, image segmentation is originally a task in computer vision or image processing if you want to call it. The task is as the name suggests to divide an image into multiple parts, but what like in if you want to do semantic segmentation, then basically the each part should mean like should be defined in terms of the objects present.

So, in other words, the every pixel of the image they can be assigned to certain or like we can imagine it as some kind of a clustering of the pixels, where two clusters where two image where two pixels that are associated with the same object can be considered to be as part of the same cluster. So, like if you look at this image for example, you can say that all, so, this is originally an image of a dog and three sheep running on a field.

So, here if you see its like the semantic segmentation shows all the pixels which are associated with the sheep, they are like coming in the same cluster as indicated by the blue color all the pixels that are associated with the dog are coming in another cluster as indicated by the red color and finally, the all the pixels which are associated with the field they are coming in another cluster which is the green color.

So, similarly apart from semantic segmentation there is instance segmentation also where each of the instead of considering all of these as the category sheep, we can consider this as individual

objects and basically we are doing something like individual object segmentation. So, it is essentially a pixel labeling task where each pixel has to be like either clustered or classified according to certain class labels if they are known beforehand, the catch is that the adjacent pixels are most likely to either belong to the same cluster or have the same label.

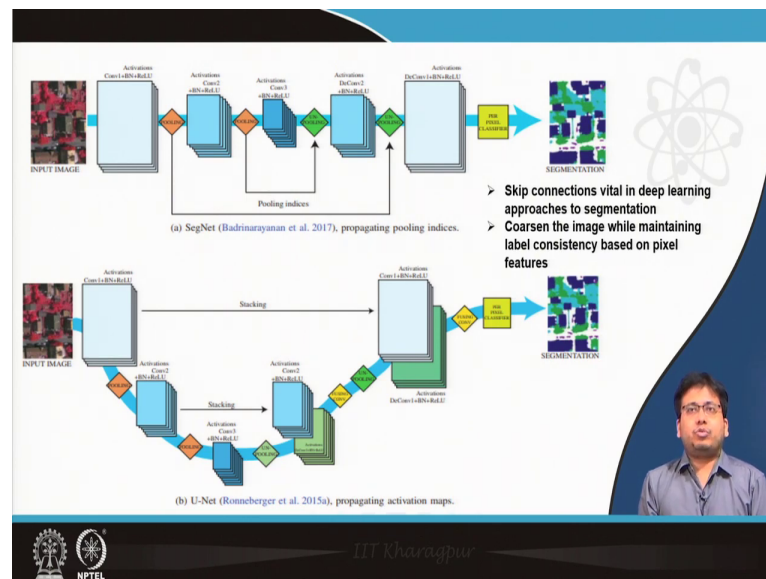
Because all the objects are known to be contiguous things, so, like usually two pixels that belong to the same object should come in the same cluster or they should have similar labels. Now in case of remote sensing one standard application of this kind of image segmentation is in LULC mapping that is LULC stands for Land Use and Land Cover changes. So, basically here the task is each pixel has to be classified according to the type of land use that is it like does that pixel cover a forest or building or a road or a water body or what.

Now, the challenge here is that in top view optical imagery, it is very difficult to obtain the feature representation of the different classes like when we are taking a lateral view like this we can see a clearer view of the or we get a bigger view of the object we are interested in so, that it becomes easier to recognize it.

But when we are taking the top view then only a small part of it is visible and from the top like we may not have enough information to identify what class it belongs to and like we earlier talked about different kinds of imagery or satellite imagery which we may get in remote sensing; hyperspectral, multispectral infrared and the SAR and so, on.

Now when it see if it is in non optical imagery then there this problem becomes even more challenging like partly because of the absence of or because of getting the ground truth also becomes very difficult in that case.

(Refer Slide Time: 04:45)



I mean if it is a RGB image which is in the visible spectrum, then we can expect a human assistant to actually label the different pixels. So, for the purpose of training the neural network. But if you are considering a SAR image which like which are human on seeing it will not be able to interpret it, then the task becomes even more difficult now.

So, like in the literature of computer vision several kinds of a algorithms or models have been developed for the task of segmentation. I mean segmentation of normal images like this one not necessarily for remote sensing, but they are still image segmentation approaches. So, its it might be possible to borrow them in the when we are dealing with remote sensing.

Now, this earlier this the like in the pre deep learning days this image segmentation often used to be done by graph based methods that is to say every pixel was considered as like as a node of the graph and then two pixels which like whose intensity values were similar like an edge was drawn between them and so on and then the task was to look for connected components in such a graph.

So, like this was largely the approach in the let us say around 2010 or so after that some latent variable based approaches also came where with each pixel one we considered a latent variable which indicated its cluster membership and then the like something like a Gaussian mixture model these kinds of things were done to estimate the or to infer the values of those latent

variables along with the constraint that adjacent pixels are likely to have the same values of the latent variable; somewhat similar to the graphical models which we had discussed earlier for extracting special patterns of Indian monsoon and so, on.

But now we have deep learning and now the idea is to like use these kinds of neural networks to for the purpose of image segmentation. Now two neural networks that have been two architectures that have been particularly successful for this task are the SegNet and U-Net. Now both of these SegNet and U-Net. So, which are shown here both of them are essentially based on the idea of like first downscaling I mean like reducing the spatial resolution and then upscaling that is in increasing the spatial resolution.

Now, in a CNN as we know the this kind of reducing of the spatial resolution can be done by pooling while the spatial resolution can be increased by unpooling now why do we do the pooling to reduce the spatial resolution?

Because basically the task is that of coarsening the image that is we want to throw away the excess information, we do not really care about what is the exact value of these pixels as long as they lie in the same object rather we would like to get a coarsened view of this image so, that like a pixel and so, that like in the coarse representation if two pixels have similar values we will understand that they probably belong to the same object and need to be placed in the same cluster.

So, basically the purpose of this coarsening is to throw away the local the local variations and preserve only the major or the global variations in the pixel values and of course, this is to all these while assuming that the different classes or the different clusters they have their characteristic pixel values.

If they did not have that is to say if they like if the there were two objects that two different objects that having the same color or having the same shape etcetera then we may not be able to separate them. Say for example, if we had like had a say a green snake on a green field then of course, we will probably not be able to separate because like on the basis of these RGB values there is there not sufficient information to separate it.

Now, like a once the pooling has been done, the image is reduced to a low resolution for the purpose I mentioned above, but then I want the output to be in the same resolution as the original image. So, I do the unpooling and which is aimed at increasing the spatial resolution back to the original resolution, but without restoring the all the details; however, when I am doing the unpooling I need to get some information from the previous steps.

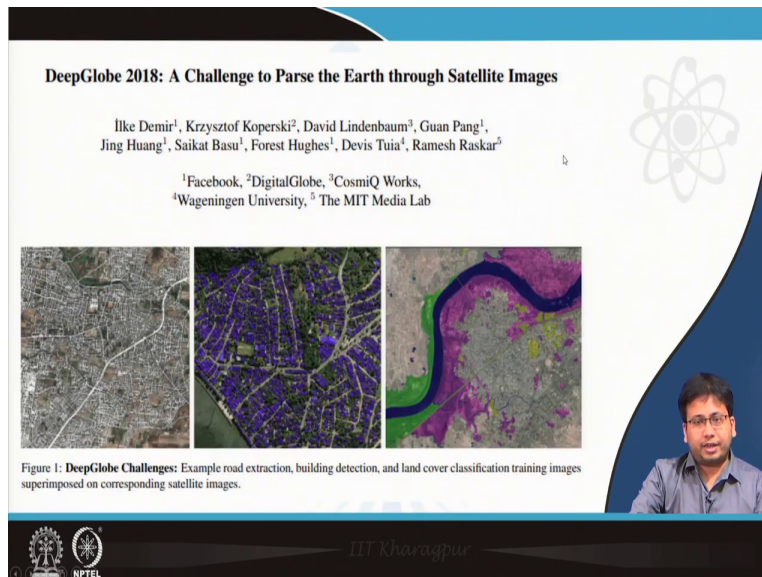
I mean I that is when I am coming back to going from the low resolution to a lower resolution to the higher resolution in a sense, I have to bring in some extra pixel values. So, how that will happen? So, for that to happen in a consistent way we need these kinds of skip connections. So, like as you can see these skip connections are present in U-Net also.

So, its like the pooling the downscaling is done in several we can say in several steps and in the upscaling is also done in the same number of steps and these skip connections basically connect the same levels. So, like as you can see like the second level of a pooling is like linked to the first level of unpooling; the first level of pooling is linked to the second level of un pooling and so on.

So, this helps to obtain an image like this which is in the of the same spatial resolution as the input, but like with the with this kind of clustering like with the pixels clustered. So, as you can see in the output image most of the pixels have like adjacent pixels have the same value indicating these kind of clustering, but this clustering is consistent with the input image.

So, like the example which we are seeing here is like the input is something like a city view taken from the top and the output is the land use land cover map of the city that is showing the buildings, the roads, the grassland areas etcetera in different colors which indicate different clusters.

(Refer Slide Time: 11:31)



The slide is titled "DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images". It lists the organizers: Ilke Demir¹, Krzysztof Koperski², David Lindenbaum³, Guan Pang¹, Jing Huang¹, Saikat Basu¹, Forest Hughes¹, Devis Tuia⁴, and Ramesh Raskar⁵. Below the names are the affiliations: ¹Facebook, ²DigitalGlobe, ³CosmiQ Works, ⁴Wageningen University, and ⁵The MIT Media Lab. A logo of a stylized atom is in the top right. Three satellite images are shown: a city street view, a city with purple blocks representing buildings, and a river with purple blocks representing land cover. A small video inset of a man is in the bottom right. The bottom of the slide features the IIT Kharagpur and NPTEL logos.

DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images

Ilke Demir¹, Krzysztof Koperski², David Lindenbaum³, Guan Pang¹,
Jing Huang¹, Saikat Basu¹, Forest Hughes¹, Devis Tuia⁴, Ramesh Raskar⁵

¹Facebook, ²DigitalGlobe, ³CosmiQ Works,
⁴Wageningen University, ⁵The MIT Media Lab

Figure 1: DeepGlobe Challenges: Example road extraction, building detection, and land cover classification training images superimposed on corresponding satellite images.

IIT Kharagpur
NPTEL

Now, we will see several examples of how this approach can be used to answer various interesting questions that on the basis of this remote sensing data. So, in there was one competition called DeepGlobe 2018 it was something like a challenge that aim was to parse the earth through satellite images that is satellite images are captured from different places all over the world as we see in Google earth and so on. The task was first of all road extraction that is given a city map like there is given a city image like this identify the major roads.

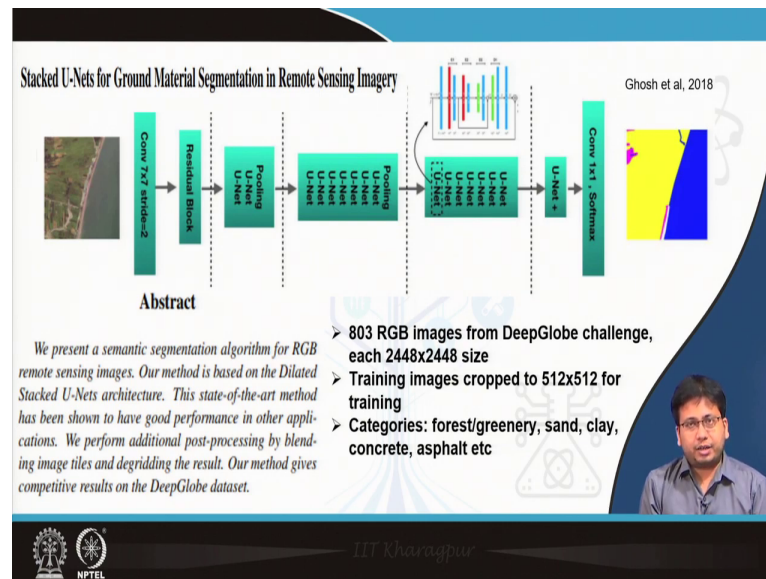
So, as you can right away understand there is a long highway kind of thing like this structure is probably something like a large highway, but there are also various other a long like avenues or busy roads or narrow roads lanes etcetera like. So, the aim here is to like build or identify the road network of the city, now the like this is this road will not come out immediately because that of the presence of many other things also in the satellite image all the buildings, the lakes, the parks everything.

So, how to I want to segment the image in such a way that I can like extract all the roads that is one challenge. The second challenge is building detection that is again now you are the same let us say we are seeing focusing on the same city, but we I want to identify each and every building. So, basically that is another image segmentation. So, like all these purple blocks you

can see here on this image if you take a closer look. So, these each of them are supposed to represent one building.

The third one is land cover classification that is I want to identify which pixels cover buildings, which pixels cover some fields or vegetation, which pixels cover river and so on and so forth.

(Refer Slide Time: 13:39)



And now we will. So, as a response to this challenge various researchers submitted like various machine learning models, deep learning architectures etcetera for solving specific parts of this problem. So, like one of these is related to this identification of ground material. So, the different ground materials might be say forest or greenery that is to say basically vegetation and sand, clay, concrete, asphalt etcetera that is the. So, these are different kinds of materials with which the ground is built. So, like can we from the remote sensing image can I identify.

So, for example, this is a remote sensing image satellite image. So, as you can see this is probably water, here we see vegetation along the water or along the coastline we see a sandy region, then apart from that there are some built regions which are like which have this which are made of concrete so and so on. So, like the output is to build this kind of a map where each region that is made of one particular material that is sure placed as one cluster and hence marked in one color yes.

So, this is again done using a u like something similar to a U-Net only the only thing is that instead of having a single U-Net architecture, there is a it is a stacked unit that is like there is one U-Net which produces an output then that same U-Net the output of that U-Net is then again passed to another U-Net then further another one more round of U-Net and so on.

So, you can call it as something like a nested unit architecture. So, the like one U-Net is suitable for one round of image segmentation. So, if we stack it. So, we can still further still further segment the different image these things ok.

(Refer Slide Time: 15:44)



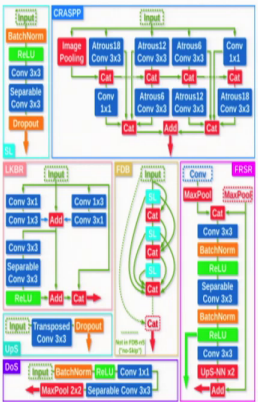
So, like these are some of the results which they have obtained. So, in different parts of the images of the satellite images if we focus on them they are getting the different spatial maps and so on.

(Refer Slide Time: 15:56)

SkyScapes – Fine-Grained Semantic Understanding of Aerial Scenes

Abstract

Understanding the complex urban infrastructure with centimeter-level accuracy is essential for many applications from autonomous driving to mapping, infrastructure monitoring, and urban management. Aerial images provide valuable information over a large area instantaneously; nevertheless, no current dataset captures the complexity of aerial scenes at the level of granularity required by real-world applications. To address this, we introduce SkyScapes, an aerial image dataset with highly-accurate, fine-grained annotations for pixel-level semantic labeling. SkyScapes provides annotations for 31 semantic categories ranging from large structures, such as buildings, roads and vegetation, to fine details, such as 12 (sub-)categories of lane markings. We have defined two main tasks on this dataset: dense semantic segmentation and multi-class lane-marking prediction. We carry out extensive experiments to evaluate state-of-the-art segmentation methods on SkyScapes. Existing methods struggle to deal with the wide range of classes, object sizes, scales, and fine details present. We therefore propose a novel multi-task model, which incorporates semantic edge detection and is better tuned for feature extraction from a wide range of scales. This model achieves notable improvements over the baselines in region outlines and level of detail on both tasks.



Azimi et al, 2019

IIT Kharagpur

NPTEL

Another task is like was that of this is for SkyScapes this is especially for aerial scenes over a city. So, understanding the complex urban infrastructure with centimeter-level accuracy is essential for many applications from autonomous driving to mapping infrastructure monitoring and urban management.

Aerial images provide valuable information over a large area instantaneously nevertheless no current data set captures the complexity of aerial scenes at the level of granularity required by real-world applications. To address this we introduce SkyScapes an aerial image data set with high highly accurate fine-grained annotations for pixel-level semantic leveling.

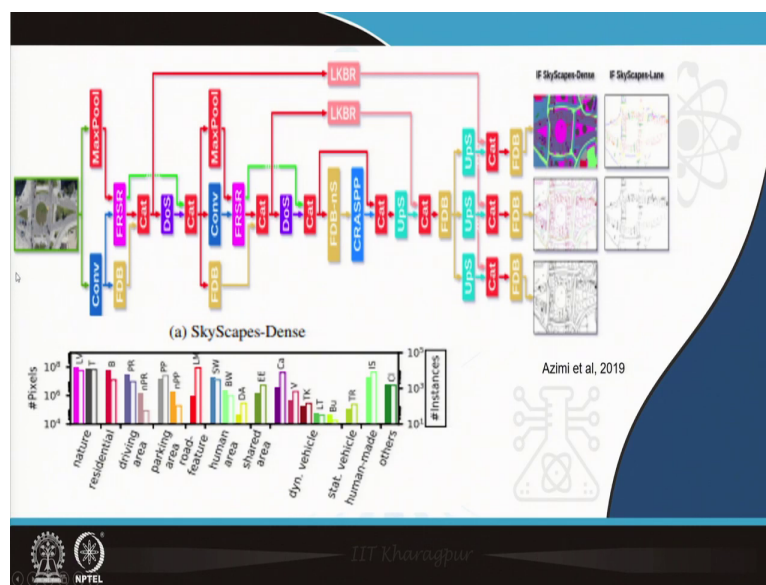
SkyScapes provides annotations for 31 semantic categories ranging from large structure such as building road and vegetation to find details such as 12 subcategories of lane marking. We have defined two main tasks on this data set dense semantic segmentation and multi class lane working prediction. We carry out extensive experiments to evaluate state of the earth segmentation methods and SkyScapes.

Existing methods struggle to deal with the wide range of classes object sizes scales and fine details present. We, therefore, propose a novel multitask model which incorporates semantic edge detection and is better tuned for feature extraction from a wide range of scales.

This model achieves notable improvement over the baseline in region outlines and level of details of both tasks. So, like this is what the model architecture basically looks like its a very complicated model with various modules in the model and so, like. So, basically this is the image and like which is based on a series of sequence of convolution image neural networks like this.

So, like. So, as I said its a multi multitasking framework in which in different like. So, these are the different modules for the multitask thing.

(Refer Slide Time: 18:18)



And so, the overall thing this is what it looks like. So, this you can see this FDB LKBR FRSR etcetera written here. Now, these are all modules and these modules they are individual their specific networks are actually shown here. Like LKBR branches are later like fused together and similarly the FRSR this is again a like a sequential a sequence of several convolutions and so on and finally, there like there is the there are some skip connections also then there are the these other modules. So, all these modules like.

So, basically they make like a very detailed pipeline in which all these modules are deployed some in serial and some in parallel there are all these skip connections etcetera etcetera and finally, the output is like given an input image like this which is like which is obtained from a

city like which is basically an image captured an aerial image captured over the city, the output is that like this kind of a segmented map.

So, here you can see the different like as suggest the different colors they indicate, different like different categories of things in the city. So, these green things these are the road, these pink things these are the green regions here the like you can say some kind of parks and then there are these buildings which are shown in the sea green color etcetera etcetera and so, for the different these are the different classes which they are looking for the nature, the residential areas, the driving areas, parking areas etcetera.

So, like each of like in the output each of these categories are to be indicated by some kind of by like is to be represented as a cluster or indicated by a color and that is what they have done. Now because there are so many categories and sub categories that they need this kind of a like a high very involved structure with so many like sequential and parallel connections, skip connections etcetera and many of these connections like there are lots of modules hidden between them as shown here.

(Refer Slide Time: 20:51)



So, this like also the another important thing here is that. So, the training data was collected by a helicopter over the city of Munich which is in Germany total took a 16 images. So, these were all

like very high resolution images as you can see like 5616×3744 this is obviously, at a much higher resolution than the images which we normally deal with when the natural images which we take on our cameras and so on and you can as you can see here the resolution is like 13 centimeter per pixel.

So, every pixel is this in this image it covers a $13\text{cm} \times 13\text{cm}$ area. So, you can imagine how high resolution or how high quality these images are that is the like even from a high altitude, it is being able to capture almost $1\text{cm} \times 1\text{cm}$ like map of the whole thing.

But the idea is that they are they have taken this image over Munich because there are only 16 images like they can actually and also because Munich is a planned city, where we actually have a detailed map of which thing is located in which region, its possible to do the manual annotation and then the model can be calibrated according to it by providing the output labels.

But once the thing has been once the model has been developed that is all the parameters have been learnt and so on the next step is to deploy it in other parts of the world where we have the imagery, but not necessarily the labels. So, they have in this paper they discussed how the model which they have learnt on the basis of the Munich data is actually applied on the different on many cities in different parts of the world in different countries, different continents and so on.

And they find that the proposed approach they it gives the results which are very similar to the ground truth in each of the cases and much it the results of the proposed SkyScape net this is much closer to the ground truth than alternative approaches to image segmentation.

(Refer Slide Time: 23:04)

SenIFloods11: a georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1

Accurate flood mapping at global scale can support disaster relief and recovery efforts. Improving flood relief efforts with more accurate data is of great importance due to expected increases in the frequency and magnitude of flood events due to climate change. To assist efforts to operationalize deep learning algorithms for flood mapping at global scale, we introduce SenIFloods11, a surface water data set including raw Sentinel-1 imagery and classified permanent water and flood water. This dataset consists of 4,831 512x512 chips covering 120,406 km² and spans all 14 biomes, 357 ecoregions, and 6 continents of the world across 11 flood events. We used SenIFloods11 to train, validate, and test fully convolutional neural networks (FCNNs) to segment permanent and flood water. We compare results of classifying permanent, flood, and total surface water from training a FCNN model on four subsets of this data: i) 446 hand labeled chips of surface water from flood events; ii) 814 chips of publicly available permanent water data labels from Landsat (JRC surface water dataset); iii) 4,385 chips of surface water classified from Sentinel-2 images from flood events and iv) 4,385 chips of surface water classified from Sentinel-1 imagery from flood events. We compare these four models to a common remote sensing approach of thresholding radar backscatter to identify surface water. Results show the FCNN model trained on classifications of Sentinel-2 flood events performs best to identify flood and total surface water, while backscatter thresholding yielded the best result to identify permanent water classes only. Our results suggest deep learning models for flood detection of radar data can outperform threshold based remote sensing algorithms, and perform better with training labels that include flood water specifically, not just permanent surface water. We also find that FCNN models trained on plentiful automatically generated labels from optical remote sensing algorithms perform better than models trained on scarce hand labeled data. Future research to operationalize computer vision approaches to mapping flood and surface water could build new models from SenIFloods11 and expand this dataset to include additional sensors and flood events. We provide SenIFloods11, as well as our training and evaluation code at: <https://github.com/cloudstreet/SenIFloods11>.

Bonafilia et al, 2020

IIT Khargapur

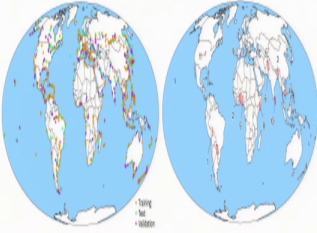
NPTEL

Similar there is another task here this is the like a this is talking about identifying a flood water. So, like we know that like inundation take when flooding takes place especially in urban as well as rural areas. So, the task here is take from the satellite imageries identify where all floods have happened. So, this is. So, there are two tasks here one is the bigger task is water detection and the second is within the regions which have been which are found to have water, then to separate them like either the like either the stagnant water or the I mean the; I mean the permanent water or the flood water.


That is of course, there are rivers and seas which always have water. So, they are not to be considered as flooding, but then there are also like other regions where which usually do not have water, but at the at a particular moment they may be having water so, that is an example of flooding. So, for that purpose also they have developed this kind of a neural network.

(Refer Slide Time: 24:24)

ID	Country	S2 date	S1 date	Rel. orbit	Orbit	VH Threshold
1	BOL	2018-02-15	2018-02-15	156	Descending	< -20.44
2	GHA	2018-09-19	2018-09-18	147	Ascending	< -22.81
3	IND	2016-08-12	2016-08-12	77	Descending	< -21.56
4	KHM	2018-08-04	2018-08-05	26	Ascending	< -23.06
5	NGA	2018-09-20	2018-09-21	103	Ascending	< -21.94
6	PAK	2017-06-28	2017-06-28	5	Descending	< -19.56
7	PRY	2018-10-31	2018-10-31	68	Ascending	< -19.94
8	SOM	2018-05-05	2018-05-07	116	Ascending	< -21.06
9	ESP	2019-09-18	2019-09-17	110	Descending	< -25.13
10	LKA	2017-05-28	2017-05-30	19	Descending	< -21.69
11	USA	2019-05-22	2019-05-22	136	Ascending	< -22.62



Bonafilia et al, 2020



IIT Kharagpur

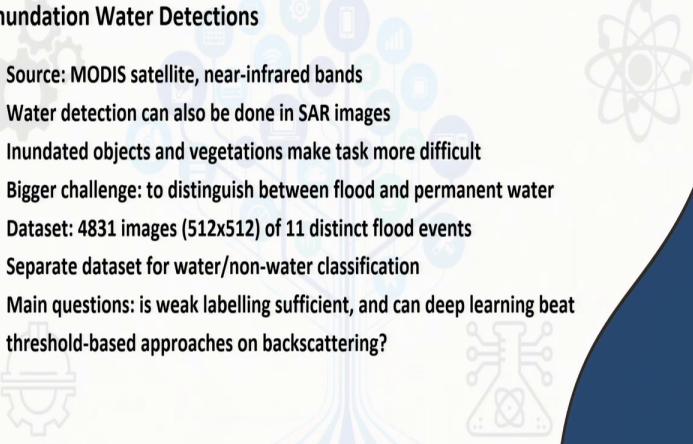
NPTEL

So, here actually the aim is not so much to come up with a new architecture. In fact, they have not done that they have they are focusing on fully rather simple fully connected neural network only, but the questions they are asking is different. So, usually this kind of flood water detection is done using some kind of a like based on some kind of threshold of something known as backscattering.

(Refer Slide Time: 24:43)

Inundation Water Detections

- Source: MODIS satellite, near-infrared bands
- Water detection can also be done in SAR images
- Inundated objects and vegetations make task more difficult
- Bigger challenge: to distinguish between flood and permanent water
- Dataset: 4831 images (512x512) of 11 distinct flood events
- Separate dataset for water/non-water classification
- Main questions: is weak labelling sufficient, and can deep learning beat threshold-based approaches on backscattering?



IIT Kharagpur

NPTEL

So, like the source of the data is from the MODIS satellite where images are captured in the NIR which is the near infrared bands and so, now, water detection can also be done on SAR images. Now the challenge here is like these inundated objects and the identifying the flood water is often difficult because the water is all not its not that the flood water is always present as water many other things are immersed in it also I mean object objects various objects may be floating in it or like suppose there is a building around which water has accumulated.

So, like that. So, as separating the water from the building that might often be a difficult task and so on and even bigger challenge is challenge is to distinguish between flood water and permanent water. So, the data set which they have obtained. So, there is again this time also like there is a challenge like this.

So, they have on the basis of that they have captured like a geo reference data set which has 4831 images each one each image is 512×512 resolution and of these are of different 11 distinct flood events and so on this they have to train their model and now the they also have a separate data set for water versus non water classification.

So, once as I said earlier also the one task is water versus non water classification for this they have built neural network based on some data set where only the water has been like where the aim is just to classify each pixel of whether it is water or it is non water and then based on that they have to build another neural network which like for the especially for the water pixels it has to classify whether it is permanent water or flood water.

So, again the it is seldom possible to do it based on one individual pixel, it is necessary to take the context into account, by context I mean the spatial context. So, like if it turns out that many so, many other adjacent pixels also have water then probably it is it looks like permanent water. But if there are some like various non-water things present in it, then it might be flood water.

Now if we have a time sequence of this such data then like identifying flood water might become more easy because flood water is not present all the time, but permanent water is by definition present in all it should be present in all the images. So, the main question here are of this particular study that we showed these are twofold. One is that, can deep learning at all beat the

threshold-based approaches for this flood water detection? So, as I said that there is this like the approach called a backscattering.

So, the by the way this map actually shows the different locations from which the water has data set has been created.

(Refer Slide Time: 27:55)

Dataset	Permanent Water		Flood Water		All Water	
	Om	Comm	Om	Comm	Om	Comm
Sentinel-1 Weak	.0660	.1354	.1190	.0997	.1124	.0997
Sentinel-2 Weak	.1209	.0534	.2684	.0778	.2482	.0778
Hand-Labeled	.0945	.1519	.1352	.1055	.1297	.1055
Permanent	.1485	.1064	.2440	.0633	.2340	.0633
Otsu Threshold-VH	.0540	.0849	.1510	.0849	.1427	.0849

Table 3. Performance on the hand-labeled test set of 10 flood events (all besides Bolivia) of models trained on each dataset in terms of omission and commission error for the water class.

Dataset	Permanent Water		Flood Water		All Water	
	Om	Comm	Om	Comm	Om	Comm
Sentinel-1 Weak	.3181	.0715	.3950	.0695	.2787	.0695
Sentinel-2 Weak	.0588	.0467	.2155	.0414	.1575	.0414
Hand-Labeled	.0669	.0904	.2148	.0813	.1518	.0813
Permanent Water	.3427	.0420	.5704	.0340	.4073	.0340
Otsu Threshold-VH	.0129	.0508	.3756	.0525	.2480	.0525

Table 4. Performance on the hand-labeled test set of flooding in Bolivia of models trained on each dataset in terms of omission and commission error for the water class.

Now its like a traditionally this is done by backscattering by measuring the by or putting a threshold on quantity known as backscattering which can be measured on the basis of remote sensing imagery. So, the one question is, whether deep learning can beat that method and the second is weak labeling possible?

That is that is to say do I need to label each pixel by saying that this is flood water, this is permanent water, this is not water etcetera or is it enough to have weak labeling that is for every like image I just say that this contains flood water without specifying exactly which content which pixels have.

And so, these are some of the results which they have shown and. So, they have measured this omission and commission errors which are basically equivalent to precision and recall and they have shown that the results are actually mixed; its not that like they the proposed neural network does well all the situation in all case. In fact, they find that as far as detecting permanent water is

concerned, its really the this threshold-based backscattering method that often does well at least in terms of omission error.

But though for commission error I think this some proposed neural networks can be effective; however, for flood water the threshold based approach is I mean for separating flood water this threshold based approach is often found to be lacking and in such situations the neural network may do a better job so, that is the output.

(Refer Slide Time: 29:42)

A novel Deep Structure U-Net for Sea-Land Segmentation in Remote Sensing Images

Pourya Shamsolmoali, Masoumeh Zareapoor, Ruili Wang*, Huiyu Zhou, Jie Yang

Abstract—Sea-land segmentation is an important process for many key applications in remote sensing. Proper operative sea-land segmentation for remote sensing images remains a challenging issue due to complex and diverse transition between sea and lands. Although several Convolutional Neural Networks (CNNs) have been developed for sea-land segmentation, the performance of these CNNs is far from the expected target. This paper presents a novel deep neural network structure for pixel-wise sea-land segmentation, a Residual Dense U-Net (RDU-Net), in complex and high-density remote sensing images. RDU-Net is a combination of both downsampling and upsampling paths to achieve satisfactory results. In each down- and up-sampling path, in addition to the convolution layers, several densely connected residual network blocks are proposed to systematically aggregate multi-scale contextual information. Each dense network block contains multilevel convolution layers, short-range connections and an identity mapping connection which facilitates features reuse in the network and makes full use of the hierarchical features from the original images. These proposed blocks have a certain number of connections that are designed with shorter distance backpropagation between the layers and can significantly improve segmentation results whilst minimizing computational costs. We have performed extensive experiments on two real datasets Google-Earth and NIRS and compare the proposed RDU-Net against several variations of Dense Networks. The experimental results show that RDU-Net outperforms the other state-of-the-art approaches on the sea-land segmentation tasks.

I. INTRODUCTION

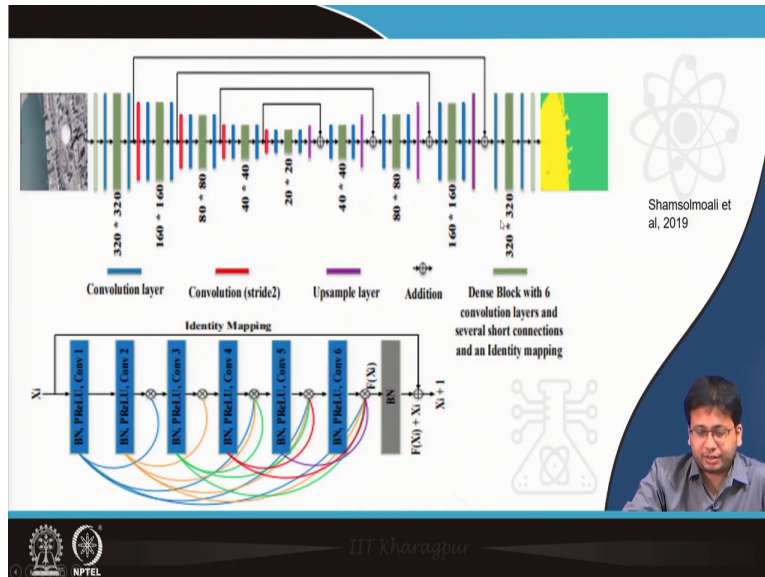
Machine vision is a technique of electronics field which widely applied in modern Remote Sensing Imagery. Remote sensing image segmentation, especially sea-land segmentation, has an important function in numerous fields such as coastline extraction [1] and maritime safety [2]. But up to now, feature extraction in remote sensing images, especially in a crowded scene, is a challenging task for sea-land segmentation. Continuous efforts have been made in the field. For instance, contrast to traditional thresholding segmentation models, Xia et al. [3] presented a model in which gray intensity features and local binary pattern features are combined. Ma et al. [4] and Liu et al. [33] presented a sea-land segmentation hierarchical model which reduced the computational costs. These approaches increased the segmentation accuracy; however, the models are not stable due to the compound intensity and texture distributions. There are some items like inland water, ships, and islands, which can confuse the algorithms and affect the segmentation results in high-resolution remote sensing images. Hence, the established classification models need to be improved. Akbarizadeh [54] proposed a model called KWE, to extract texture features by using wavelet transform; that forms a feature vector composed of kurtosis values of wavelet energy

LIT Kharagpur

NPTEL

And then finally, there is one more paper which is about a Sea-Land segmentation from remote sensing images. So, this is also based on U-Net.

(Refer Slide Time: 29:55)



So, the task here is to separate like. So, this is again as you can see this is again has a U-Net kind of structure with multiple skip connections like this and so, like they one important aspect or one important addition they have made on top of it is something known as identity mapping.

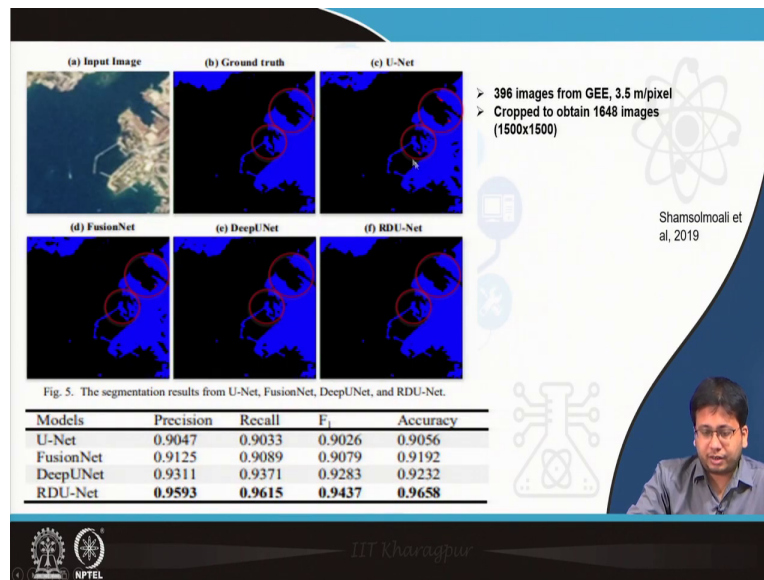
That is in the last block of U-Net they have basically made a dense block which has six convolutional layer and it has these kinds of short connections. So, like which they are calling as identity mapping. So, this identity mapping is again nothing but this kind of skip connection which are used and once again to maintain the kind of consistency. So, there are first of all there are these large skip connections across the different convolution layers and so on.

And then here within the same like within the same we can say the same resolution there are further skip connections. So, here a like every convolution is succeeded by some kind of a pooling that is to reduce the resolution as you can see till it reaches some kind of a base resolution, this 20×20 after which the resolution is again increased step by step.

So, this happens by pooling and like unpooling and from the and as discussed earlier also there are these skip connections from one level to another and then after they get reached the back to the original resolution they have another round of this kind of like what is they are calling as a

dense block. So, here there is no pooling there is only a sequence of convolution operations. And then across these convolutions also there are the skip connections.

(Refer Slide Time: 31:46)

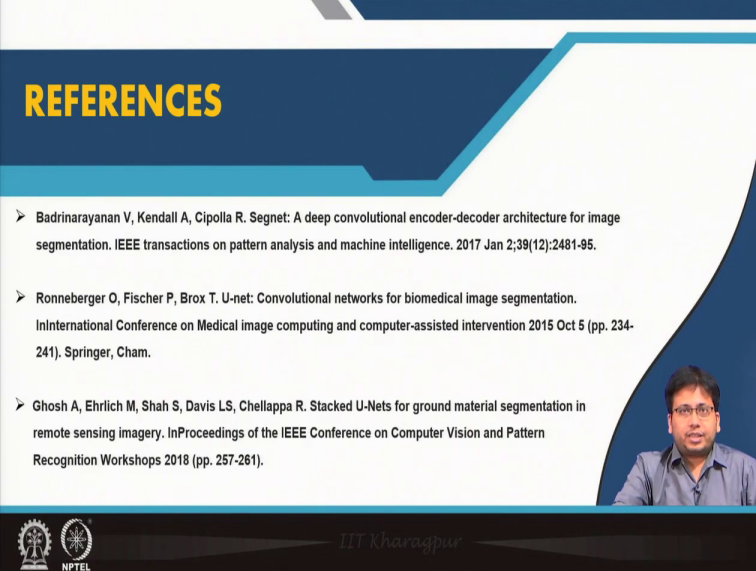


And so, this is like used for the mapping the like the land the land-sea classification. So, the every pixel here is to be segment I mean these kinds of images which are mostly over the coastline, they have to be segmented to specifically show which pixels are the ground and which or the land, and which pixels are the sea.

And as you can see these coastlines are often very complicated that is its not a straight coastline they are very much broken and you can see some inlets of sea into the land some inlets of land into the sea and so on. So, it turns out that. So, this is the ground truth and it turns out that the proposed method which contains all the. So, the so, originally there is the U-Net and then there is what they are they have proposed here as the residual dense U-Net.

So, that is U-Net containing this additional layer which I just discussed and it turns out that like providing this dense layer this actually performs somewhat better than U-Net on this particular task.

(Refer Slide Time: 32:53)



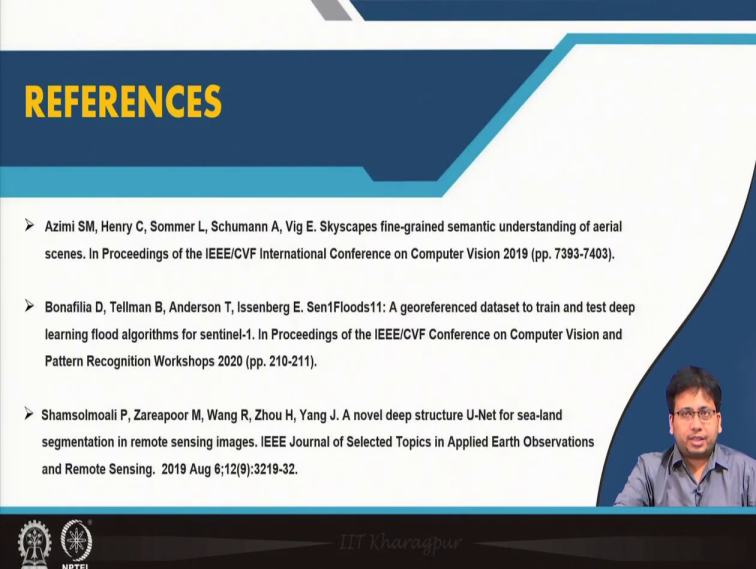
REFERENCES

- Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2017 Jan 2;39(12):2481-95.
- Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention* 2015 Oct 5 (pp. 234-241). Springer, Cham.
- Ghosh A, Ehrlich M, Shah S, Davis LS, Chellappa R. Stacked U-Nets for ground material segmentation in remote sensing imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 2018 (pp. 257-261).

IIT Kharagpur
NPTEL

So, these are the references for like for the different image segmentation problems that we discussed today.

(Refer Slide Time: 32:58)



REFERENCES

- Azimi SM, Henry C, Sommer L, Schumann A, Vig E. Skyscapes fine-grained semantic understanding of aerial scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 2019 (pp. 7393-7403).
- Bonafilia D, Tellman B, Anderson T, Issenberg E. Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* 2020 (pp. 210-211).
- Shamsolmoali P, Zareapoor M, Wang R, Zhou H, Yang J. A novel deep structure U-Net for sea-land segmentation in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2019 Aug 6;12(9):3219-32.

IIT Kharagpur
NPTEL

So, basically the message is that like this the concept of image segmentation this has multiple applications in remote sensing and like these various neural networks and deep learning models

starting from unit, but with various sophistication added to them as specific to the different tasks they help to improve the performance for those specific tasks. So, that brings us to the end of this lecture.

Thank you and we will continue our discussion on how machine learning can be used for remote sensing in the previous lectures. So, till then bye.