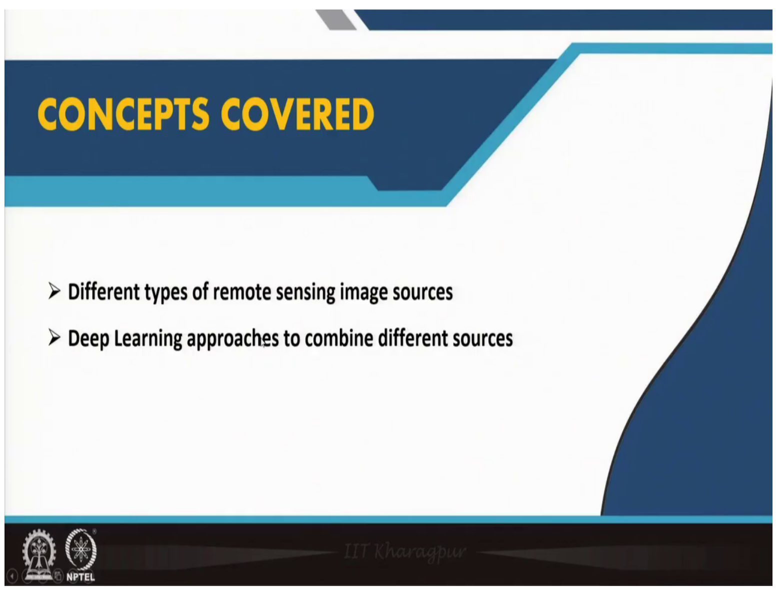
Machine Learning for Earth System Sciences
Prof. Adway Mitra
Department of Computer Science and Engineering
Centre of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur

Module - 04
Machine Learning for Earth Observation Systems
Lecture - 29
Image Fusion from Multiple Sources for Remote Sensing

Hello, welcome to lecture 29 of this course on Machine Learning for Earth System Science. We are right now in module 4, where we are discussing Machine Learning for Earth Observation Systems. In today's lecture we will be focusing on Image Fusion from Multiple Sources for Remote Sensing.

(Refer Slide Time: 00:43)



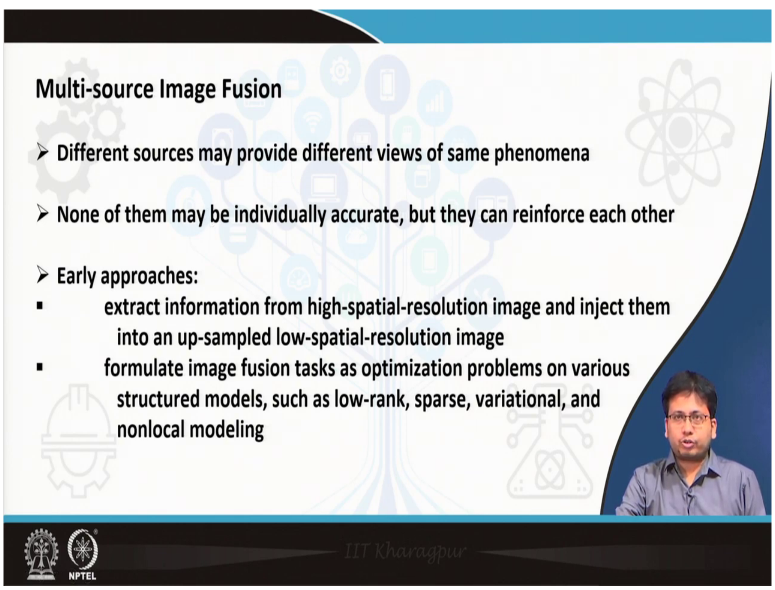
So, we have already discussed that there are different sources of remote sensing and different kinds of remote sensing imagery are available, say there are satellites, radars, and other kinds of devices or maybe flying devices like drones, even aeroplanes, helicopters and so on which acquire images.

And, then these images they can be captured in different parts of the spectrum. There can be visible range images or optical images which are also known as RGB, then there are infrared images, hyperspectral, multispectral images and so on. So, suppose we have different images of different technologies obtained from different sources like this.

And, now how can we fuse the like these different images? Like maybe it is possible that all of the images are captured over the same region in the same geographical region, but because of the different technology each of them we will have we will be measuring the same quantities in different ways. So, how like how can we reach some kind of a consensus between the measurements from these different sources?

So, in this lecture we will discuss how machine learning and deep learning can be used for this purpose.

(Refer Slide Time: 02:02)



So, like the important thing is to remember that the like the different sources are basically providing same different views of the same phenomena. Say for example, rainfall is taking place. Now, there may be space-borne device which is taking an image of that from the top, there might be some ground devices which are taking the images of that from the side from the profile views and so on.

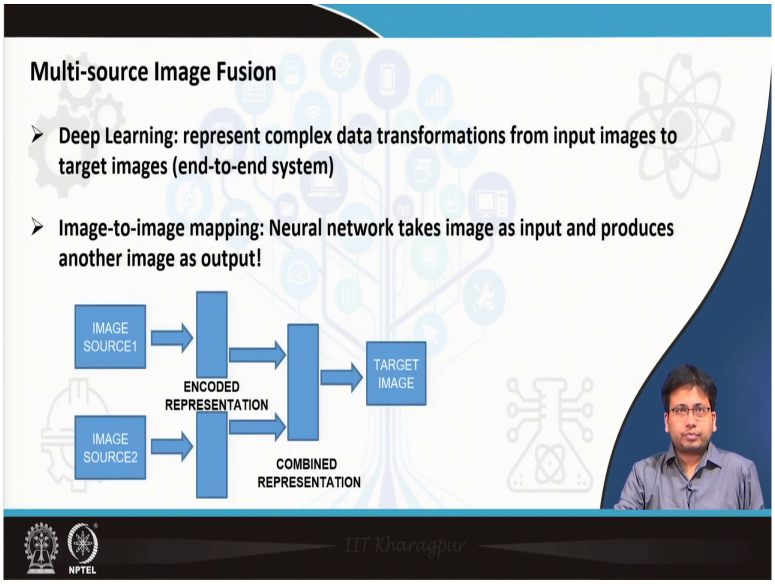
They or even like when we are let us say we are carrying the carrying out surveillance over a region then like different sources of images or I mean different sources of remote sensing they will be capturing the remote the their images in like using different technology or, so, some of the images can be low resolution, but at, but they can have high spectral bandwidth. There might be other images which have high spatial resolution, but low spectral we can say spectral resolution and so on.

Even if the resolutions are the same, then also if different technologies are used then the measurements of the quantity that we are interested may be different in different cases. So, we can say that none of them are individually accurate, but like if we take an ensemble of these images, they can reinforce each other, so that we can reach at some kind of consensus which will be accurate, with somewhat like the idea of ensemble classification in machine learning.

So, some the early ex approaches were like to extract information from the higher from one source of image which might have certain desirable property like high spatial resolution and somehow inject those properties into an into a another image which may not have that property. Let us say it is a low resolution image but, it may have some other desirable properties.

And, so, this extraction of information which I just mentioned this was like. So, this information which I am talking about this may have been represented in different ways say such as something like a low-rank matrix or a sparse matrix or some like other models with very specific structures. So, you can say that some representation of this of the source image was learned and that representation was passed on to the target image to like to we can say to reinforce it or to improve it.

(Refer Slide Time: 04:42)



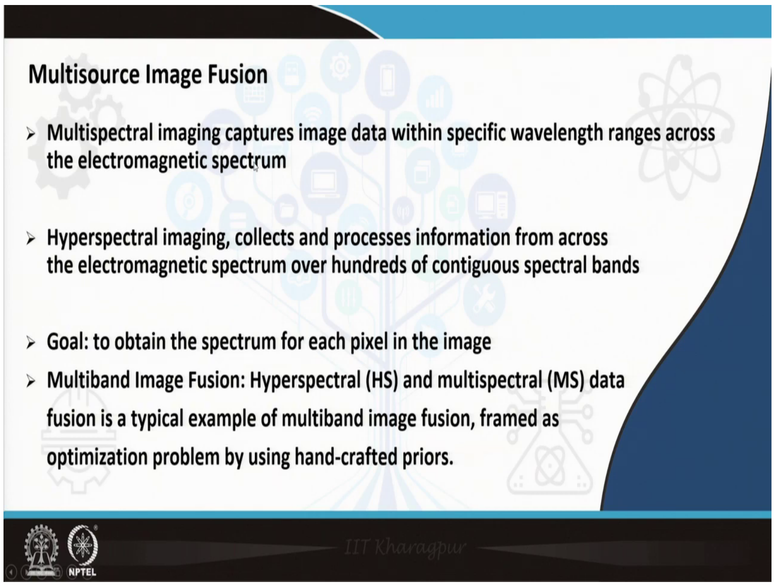
Now, the way machine learning based or deep learning based approaches for the same problem works are is as follows. So, we may have some input images which are like let us say there are two image sources – source 1 and source 2. So, what neural network does is for both source 1 and source 2, it creates something like a com like an encoded representation. So, both of these are neural networks which need not be a similar or identical.

What is the exact architecture of these two networks they may depend on the nature of these two sources. Now, once the we have these representations we combine them to get something like a fused representation and then that. So, this again may be this kind of. So, we can say this might be a simple concatenation or it can be some kind of complex stacking with another level of non-linearity which and that non-linearity may be imposed by a another neural network.

Like, here you might remember the architecture which we had discussed for the lightning prediction, the LITNet architecture which we had discussed in a couple of lectures back. So, there also there were two neural networks each of them were producing some kind of intermediate or encoded representation. Those two were being somehow fused to get a combined representation and then from that combined representation a target image is being created.

An alternative will be just to train a neural network to like to using image source 1 as the input and image source 2 as the output, that also like you may want to do in some other kind of situations.

(Refer Slide Time: 06:31)

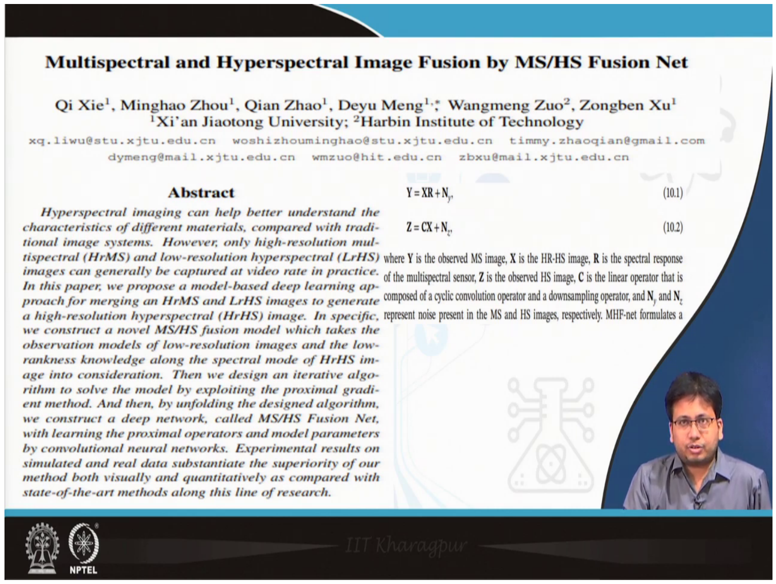


So, now let us consider what are the different kinds of sources of which we may want to fuse. So, for example, multispectral image so, like as discussed earlier also multispectral image captures image data with specific wavelength which ranges across the electromagnetic spectrum.

On the other hand, hyperspectral image it collects and processes information like within a particular band within a particular spectral band, but it captures thousands of it can it captures the imagery at hundreds of wavelengths at that within that particular band. So, we can say hyperspectral and multispectral imagery they have like complementary strengths.

Now, the goal in this case is to obtain the spectrum for each pixel in the image. So, let us say that we have a hyperspectral image and a multispectral image of the same region and we want to like fuse them and we want to get what is known as like a multiband image.

(Refer Slide Time: 07:44)



So, here is one paper which does that multispectral and hyperspectral image fusion using MS/HS fusion net. Hyperspectral imaging can help better understand the characteristics of different materials compared with traditional image systems. However, only high resolution multispectral and low resolution hyperspectral images can generally be captured at video rate in practice.

So, as I said these like as far as resolution is concerned they I mean special resolution is concerned these two technologies hyperspectral and multispectral, they have complementary strength. The like one of them the multispectral image it has we can say low spectral resolution that is it is captured at only a few wavelengths, but it has high a spatial resolution.

But, hyperspectral imagery on the other hand it is it like it is captured at multiple or it has high spectral resolution because it is captured at multiple wavelength, but low spatial resolution because that is to say each pixel covers a larger area. In this paper we propose a model based deep learning approach for merging and HrMS and LRHS images to generate a high resolution hyperspectral image or HrHS. That is, the aim is to take the best of both worlds, high spatial resolution as well as high spectral resolution.

In specific, we construct a novel MS/HS fusion model which takes the observation models of low-resolution images and the low-rankness knowledge along with the spectral mode of HrHS images into consideration. Then we design an iterative algorithm to solve the model by exploiting the proximal gradient method. And, then by unfolding the designed algorithm we construct a deep network called MS/HS fusion net with learning the proximal operators and model parameters by convolutional neural networks.

Experimental results on simulated and real data substantiate the superiority of our method both visually and quantitatively as compared with state of the art methods along with this line of research. So, like basically the aim is to build a neural network as I mentioned which they are calling as the fusion net which learns the like the various parameters and operators through CNNS, the Convolutional Neural Network.

So, the like the basic thing is. So, let us say the basic image model is as follows. So, let us say Y is the multispectral image and Z is the hyperspectral image. So, Y and Z both of them are observed and let us say that X is the target image which we are interested in, the high resolution hyperspectral image.

$$Y = XR + N_y$$

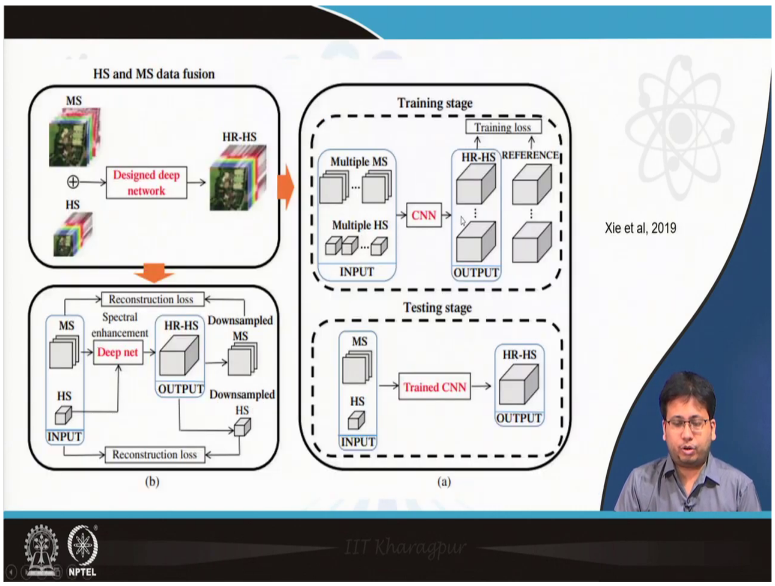
$$Z = CX + N_z$$

So, we can say that both the observed images Y and Z both of them are obtained let us just assume that X is the fundamental thing, high resolution hyperspectral and both the observed ones Y and Z both are obtained by some kind sort of a projection operation on this original image X. So, like if we do XR, so, R is some kind of a spectral response operator which creates the observed multispectral image that is to say it preserves the spatial resolution, but like somehow merges the different bandwidths etcetera sorry, I mean merges the different wavelengths etcetera so that we get a an image which has low spectral resolution.

The C on the other hand is a linear operator which is which does something equivalent to like coarsening of an image so that when it operates on this X which is the high spatial resolution image we get an output Z which has the which maintains the same spectral wave spectral resolution, but reduces the spatial resolution. So, that may be like Y and Z these are we treat these Y and Z as the observed variables and X is the unobserved variable.

So, basically it this thing is reduced to something like an inverse problem where from Y and Z we have to somehow estimate X. So, now the also the important thing is R and C these two things these are not known to us. These are some is like some like we have considered these as some kind of projection matrices, but we do not know what they are and on top of that the noise are always also present.

(Refer Slide Time: 12:53)



So, this is the general architecture. So, like the input is the MS and HS images both of them are simultaneously provided as inputs to a deep network and whose output is going to be the HrHS image. So, for like. So, basically the once the HrHS output is provided like then the important thing is to for any learning problem the important the most important thing is to construct the loss function which basically measures what is the quality of the output obtained.

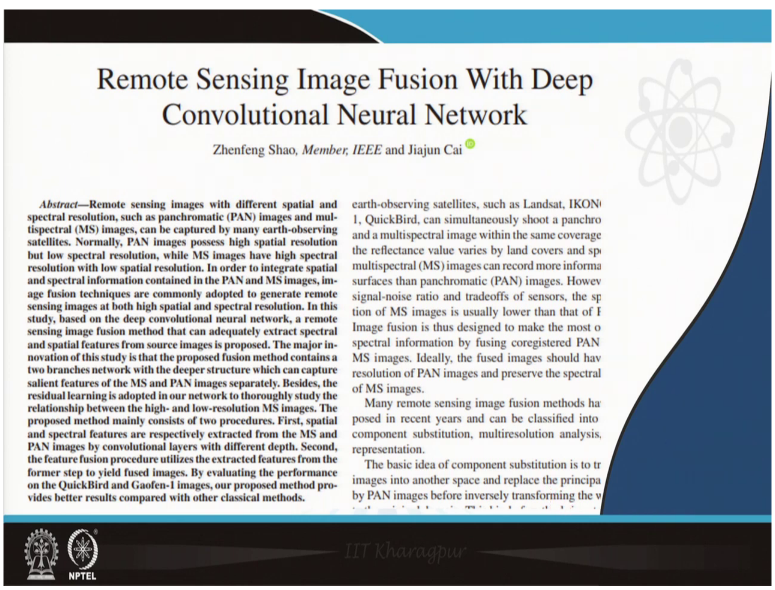
So, in this case the output image is expected to be HrHS so, which is supposed to be the best of both world that is it should have the same spatial resolution as the multispectral image and the same spectral resolution as the hyperspectral image. So, with the this output, so, basically we need something what is known as a reconstruction loss so, just to measure these quantities.

So, this output image first it is down sampled I mean spatially downsampled to the same spatial resolution as the hyper spectral image and then they are compared to the input to see if the that if the spectral properties are maintained or not.

On the other hand, they are also downsampled in the sorry, I mean to say they are downscaled in the spectral domain so as to get the same resolution as the multispectral image so that both the multispectral image input multispectral image and the output one are now reduced to the same spatial resolution and then they are they are compared whether the same spatial properties are maintained or not.

So, these both of these two like constitute our reconstruction loss. The reconstruction loss like this is a concept which we have seen earlier in case of auto encoders also, it basically means the output which is obtained how close it is to the inputs. So, this is how this is measured.

(Refer Slide Time: 15:16)



And so, like for both the in the architecture it is basically a convolutional neural network based architecture the details of which we are not going into here, but you can see it in this paper by Xie et al. So, like basically this is how it works. So, the CNN obviously, it will like the general architecture is along these lines only. So, in one domain with thus convolutions will take place in the spectral domain and in the other one the convolutions will take place in the spatial domain.

And, then of course, like they will have to be like the we will reach our something like a common representation which will be fused like which will be treated to by another neural network to get the combined output yeah. So, note that the input and the like both the inputs we can consider them like each image is basically a combination of several image. So, this is like here you can see this is like a high resolution image which we can assume is present in several channels.

Just like a normal optical image it has three channels R, G and B. So, this also can be considered to be a single high resolution image that is in several channels and each of these it is a low resolution image as indicated by the low size the small size of this, but it also has a long depth which indicates that the same thing is captured at multiple at or at a large number of; at a large number of wavelengths which means that instead of having 3 or 4 channels this has like several hundred channels and so on.

Now, let us come to another work remote sensing of image fusion with deep convolutional neural network. So, this is also along the same lines. Remote sensing images with different spatial and spectral resolution such as panchromatic images and multispectral MS images can be captured by many earth observing satellites. Normally PAN images possess high spatial resolution, but low spectral resolution similar to the hyperspectral image which we talked about sorry, the multispectral image we which we just talked about.

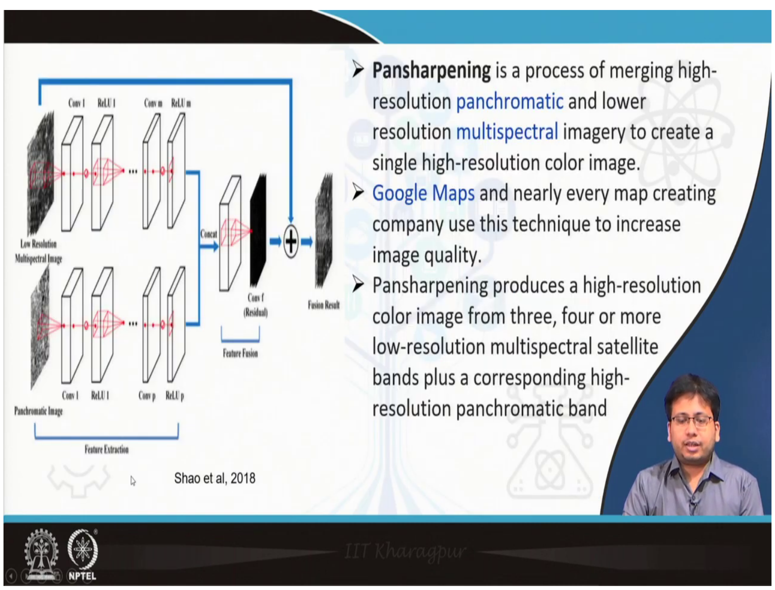
Now, the MS images they have high spectral resolution with low spatial resolution. So, like here you can see there is a hierarchy of these images in terms of spatial and spectral resolution when compared to hyperspectral images, multispectral images have higher spatial, but lower spectral resolution, but compared to the panchromatic images like even multispectral images have low spatial resolution and high spectral resolution.

That is a hyperspectral image is captured at hundreds of wavelengths a multispectral image is captured at maybe 5 or 6 wavelengths, and panchromatic image is captured at a single wavelength. But, like so, at the spectral radius is in the order hyperspectral maximum followed by multispectral and then finally, panchromatic; in case of spatial resolution, it is in the other way.

So, the previous work was about fusion of hyperspectral and multispectral, this one is about fusion of panchromatic and multispectral. So, like in this case also in this study based on deep convolutional neural network, a remote sensing image fusion method that can adequately extract spectral and spatial features from source image is proposed. The major innovation of this study is that the proposed fusion method includes two branches network with the deeper structure which can capture salient features of both MS and PAN images.

Besides, the residual learning is adopted in our network to thoroughly study the relationship between high and low-resolution multispectral images. The proposed method mainly consists of two procedures first the spatial and spectral features are respectively extracted from the MS and PAN images by convolution neural networks at different depths. Second, the feature fusion procedure utilizes the extracted features from the former step to yield fused images.

(Refer Slide Time: 19:46)



And, so, basically as you can see it is the same paradigm that is being followed. In this case, like there is this is the multispectral image, this is the panchromatic image as you can see this is multiple channel, but this is at a single channel. So, like here also there is a series of convolution layers which produces some kind of a intermediate representation.

Here also there is a parallel, but different convolutional neural network which produces an intermediate representation. They are concatenated and residual is computed and then so, there is this skip correction. So, the residual is then associated with the or like with the input low-resolution image and finally, we get the output.

So, the main way or the main innovation in the in this particular architecture is this kind of skip connection which was not there in the previous model. So, this is basically what is happening in this case we can say is pansharpening; that means, it is a process of merging a high-resolution panchromatic with a low resolution imagery to create a single high resolution color image.

And, so this kind of technique is used in Google Maps and various other map creating applications.

(Refer Slide Time: 21:09)

Hyperspectral Image Super-Resolution with Optimized RGB Guidance

Ying Fu[1] Tao Zhang[1] Yinqiang Zheng[2] Debing Zhang[3] Hua Huang[1]
[1]Beijing Institute of Technology [2]National Institute of Informatics [3]DeepGlint
{fuying,tzhang,huahuang}@bit.edu.cn yqzheng@nii.ac.jp debingzhang@deepglint.com

Abstract

To overcome the limitations of existing hyperspectral cameras on spatial/temporal resolution, fusing a low resolution hyperspectral image (HSI) with a high resolution RGB (or multispectral) image into a high resolution HSI has been prevalent. Previous methods for this fusion task usually employ hand-crafted priors to model the underlying structure of the latent high resolution HSI, and the effect of the camera spectral response (CSR) of the RGB camera on super-resolution accuracy has rarely been investigated. In this paper, we first present a simple and efficient convolutional neural network (CNN) based method for HSI super-resolution in an unsupervised way, without any prior training. Later, we append a CSR optimization layer onto the HSI super-resolution network, either to automatically select the best CSR in a given CSR dataset, or to design the optimal CSR under some physical restrictions. Experimental results show our method outperforms the state-of-the-arts, and the CSR optimization can further boost the accuracy of HSI super-resolution.

hybrid camera system [1, 2, 3, 12, 14, 24, 28, 30, 31, 45] employ various hand-crafted priors to model the underlying structure of the latent high resolution HSI. Nevertheless, to hammer out proper priors for a specific scene remains to be an art.

Recent alternative approaches [13, 35] leverage on deep learning to alleviate the dependence on hand-crafted priors, and show that the CNN scheme can effectively exploit the intrinsic characteristics of HSIs. Nevertheless, these methods either use the CNN scheme to refine the initialized results in a supervised way [3], or resort to step-by-step alternating optimization [35]. In this work, we present a simple and efficient CNN-based end-to-end method for HSI super-resolution with RGB guidance, which can effectively approximate the spectral nonlinear mapping between the RGB and the spectral space, and utilize the spatial consistency. Neither delicate hand-crafted priors nor training data are needed in our method. This allows our method to handle various scenes more easily.

In addition, all these methods mainly focus on RGB-guided HSI super-resolution under a given CSR function of the RGB camera. Recent researches on HSI super-

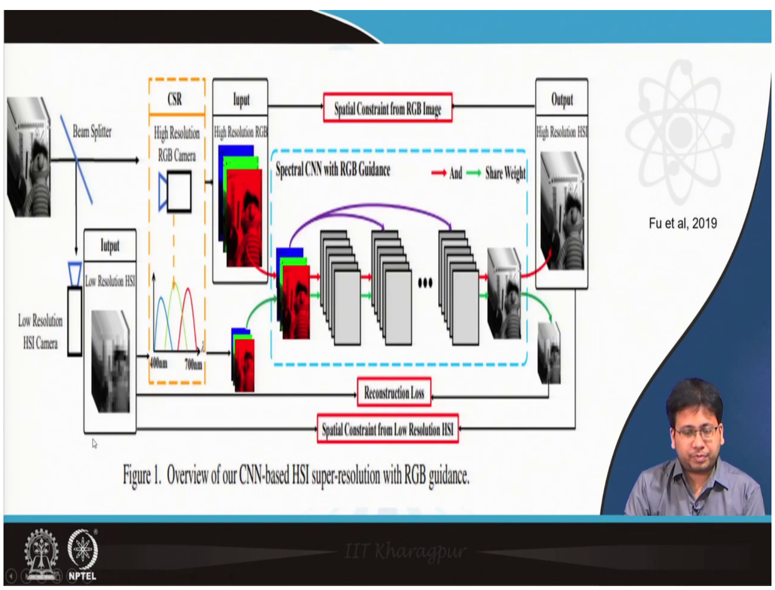
So, and the next is one more in the same family. So, this is about hyperspectral image super resolution with optimized RGB guidance. So, now, we are talking about hyperspectral image and RGB images. So, what is super resolution? Super-resolution basically means increasing the spatial resolution of an image.

To like you may remember that in one of our earlier lectures we had discussed about super resolution when we are talking about downscaling of rainfall. So, we had talked about these models like SRCNN etcetera which are you like which are used for taking a low resolution map of precipitation and mapping and then projecting it to a high resolution map of the same quantity.

So, in this case also we are doing something along those same lines only to overcome the limitations of existing hyperspectral cameras on spatial or temporal resolution fusing a low resolution hyperspectral image with a high resolution or RGB image into a high resolution HSI has been prevalent. Previous methods for this fusion task usually employ handcraft priors to model the underlying structure of the latent high resolution model HSI and the effect of camera spectral response of the RGB camera on super resolution accuracy has rarely been investigated.

In this paper, we first present a simple and efficient CNN based method for HSI super resolution in an unsupervised way without any prior. Later we append a CSR optimization layer into onto the HSI super resolution network either to automatically select the best CSR or to design the optimal CSR under some physical restrictions.

(Refer Slide Time: 23:16)

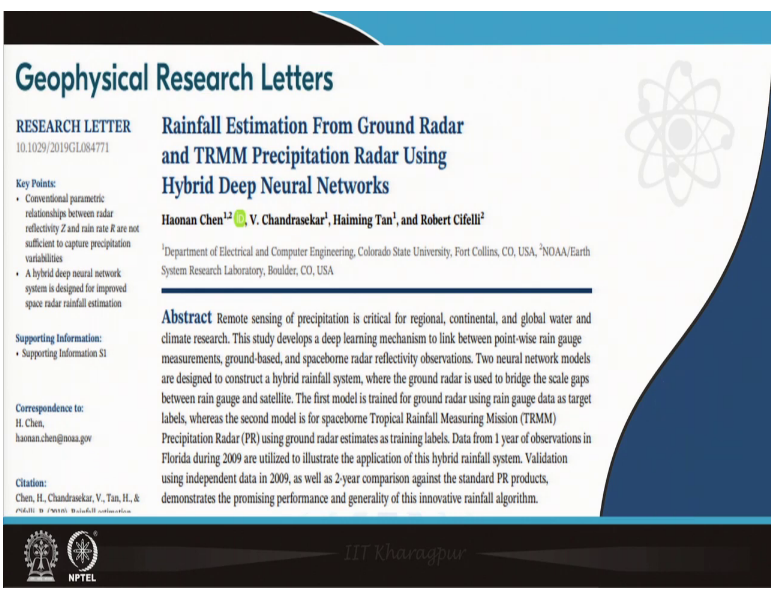


So, this will be like roughly this is what we have. So, this is the original like let us say this is the original image which we are getting for the HSI the hyperspectral imagery camera. So, this is at a low resolution I mean low resolution compared to an RGB image. So, this is like likes and then there is the RGB image. So, it is basically. So, in this case also we have the same components as we saw in the previous two cases.

So, there is this reconstruction loss and there is this special constraint also. So, like. So, basically it is all about reconstruction just like the previous case. So, there is a series of these convolutional neural networks and when the convolutional neural networks are provided with an input, the high single high resolution RGB image that is also provided as an input to the CNN and like as you can see the there is also weight sharing between the layers to reduce the number of parameters.

And, then there are also skip connections across the different layers and so the output is once again like we can say upscaled version of the same of the input. So, this is at low resolution this will be at a high resolution, but this is now this is have this will have to be compared with the RGB for spatial constraints and it will also have to be compared with the original low resolution HSI image that is also at a like for the spatial constraints. So, like basically once again these are the reconstruction losses that we earlier talked about.

(Refer Slide Time: 25:10)



Now, finally we talked about a different problem based on similar ideas. So, let us say this is related to rainfall estimation. So, for rainfall estimation we have primarily three sources – one is the automatic weather stations or the ground-based in situ measurements which like which are I mean located at specific points on the surface of the earth.

Now, they have very good accuracy because they are using specialized sensors to measure rainfall. But, they can they these are essentially in situ measurements that is the measurements they acquire is specifically for that region that location, but not for a larger range. But, then there are ground-based radars and there are space based radars or satellites.

Now, these radars are it is like their strengths are complimentary in the sense that their accuracy is not great because they are not directly measuring rainfall they are simply taking images at different wavelengths. So, the I am these images are basically obtained is basically some kind of measure of the reflectance. So, they will be sending out some signals. Those signals will reflect in the on the rain drops and come back to and the radars will capture the that reflected signal.

And, on the basis of that or on the strengths of that reflection like we will somehow have to like estimate how much rainfall there is. So, that kind of mapping from the radar imagery which is basically the strengths of the reflected signals to the actual quantity of rainfall that kind of relationship which is a very which might be a very complicated relationships that will have to somehow be represented either as a mathematical formula or as a machine learning model preferably a neural network.

But, but on the other hand, the these radars they have the advantage that they can cover a large spatial resolution. So, what is lost in terms of accuracy is gained or is compensated for in terms of the spatial resolution. Similarly, and these radars or satellites they can I mean the radars can be ground-based as well as space based and then from the space we also have satellite imagery.

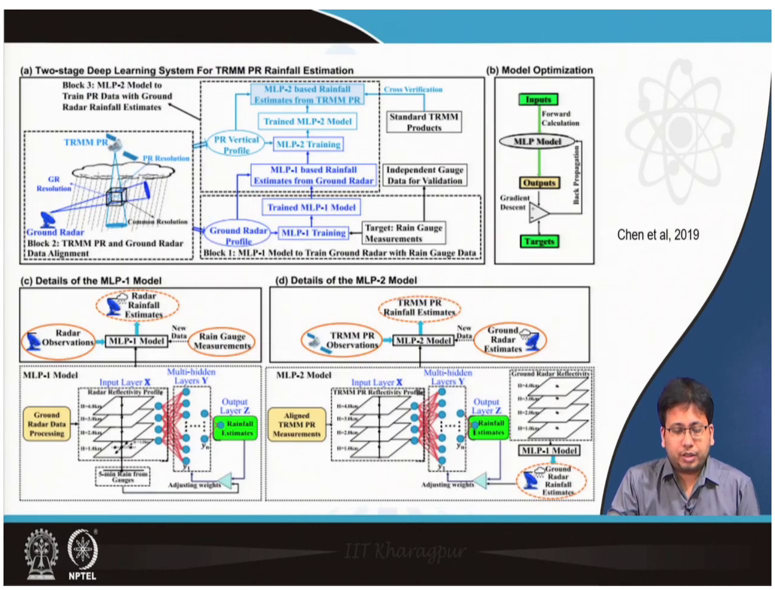
So, now the question is can like just like in the previous cases we were like taking the best of both worlds by that is we one source had high spatial resolution, the other source had high spectral resolution in this case also can be in instead of spectral resolution can we talk about accuracy that is can be one source has high accuracy other sources have high spatial resolution. Now, can we take the best of both worlds that is the idea of this paper.

Remote sensing of precipitation is critical for regional, continental and global water and climate research. This study develops a deep learning mechanism to link between point-wise rain gauge measurements, ground-based and spaceborne radar reflectivity observations. Two neural network models are designed to construct a hybrid rainfall system where the ground radar is used to bridge the scale gaps between rain gauge and satellites.

By scale gaps we talking about the spatial resolution etcetera which we just mentioned. The first model is trained for ground radar using rain gauge data as target levels whereas, the second model is for spaceborne tropical rainfall measuring mission or TRMM precipitation radar using ground radar estimates as training levels.

Data from one year of observations in Florida during 2009 are utilized to illustrate the application of this hybrid rainfall system. Validation using independent data in 2009, as well as 2-year comparison against the standard PR products, demonstrates the promising performance and generality of this innovative rainfall algorithm.

(Refer Slide Time: 29:36)



So, in this case there is a difference from the previous models in the sense that there is no intermediate or fused like intermediate representation from the different sources which are used to make a fused representation something like that, instead, there is direct mapping. So, the also the model in this case is not a CNN, it is a multi-layer perceptron. So, we can consider it as a like a fully connected neural network.

So, like what is happening is here is that you have the ground like if you focus on this block 2, here is the ground radar and there is a like precipitate the TRMM precipitation. So, like. So, this is taking it is imagery of the region where the rainfall is happening and this one is also taking you can say it is a taking a profile view of the rainfall.

So, now like what they are doing is they are developing two neural networks, two-multi layer perceptrons MLP-1 and MLP-2. In case of MLP-1 the input here is the ground radar profile. So, like it is basically a sequence of images. So, as you can understand it is the ground radar is taking the images like across is taking a profile view. So, it will be taking the images that or it will be getting the measurements at different altitudes.

And, so, the input is the profile obtained from the ground radar and it is to be compared against the rain gauge measurements over the like from some rain gauges which might be located in this particular region. Now, what if there are not enough rain gauges located here? So, the idea is that we will, so, now, not all locations on the world have rain gauges, there are some places where the infrastructure is good while in other places the infrastructure is poor, but these radar images can be obtained in many places, especially the spaceborne radar.

So, that way also it is what they train to do is they plan to train these neural networks in those regions where like both things are present, so that they learn to learn some kind of a mapping about how the these radars can the radar observations can be mapped to the rainfall image to the rainfall quantities. And, then once they have learned that relation that same relation can be deployed elsewhere also where they have radar based measurements, but no ground based measurements.

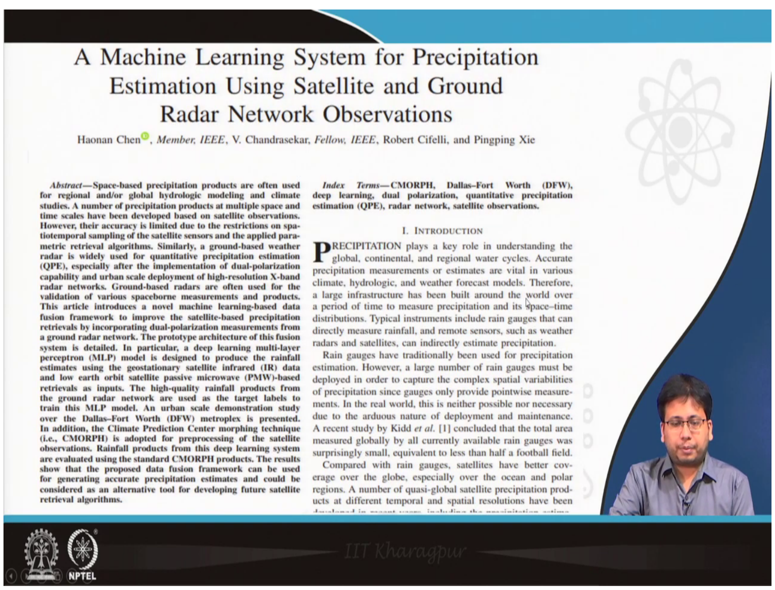
So, the MLP-1 that basically is used to calibrate the ground radars observations with the actual rain gauge measurements, so that it basically the MLP-1 what it does is it represents the mathematical relation between with the imagery captured by a ground radar to the actual quantity of rainfall. Now, once that is done this is the next step is to calibrate the spaceborne radar.

So, here the we have the MLP-2. What MLP-2 does is it receives it is input is the like the vertical profile of the rainfall as captured by the spaceborne satellite in this case TRMM which is a very well known satellite which measures precipitation. And it is calibrated against the data obtained from the ground-based radars. So, the input is the so, so we can say the input is the this thing the spaceborne radar and the output is what is the what is obtained from MLP 1.

Now, MLP 1 as we know is like has already been calibrated with the ground based observation. So, these two MLPs MLP 1 and MLP 2 their combined effect is the space borne observations are actually calibrated with the ground radar of the with the ground radar observations. And, so, then like we can of course, validate it using cross validation techniques against standard TRMM products and so on.

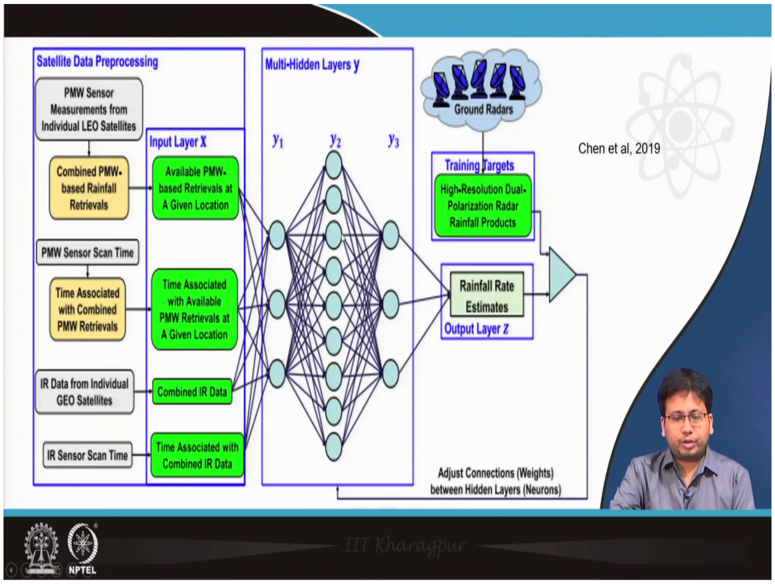
Now, if you look consider these MLP 1 or MLP 2 models so, there these are like these are simple like fully connected neural networks having several multiple hidden layers, these are not convolutional though it might be interesting to go for CNNs also in this case and see how it plays out, but so, this is the paper.

(Refer Slide Time: 34:10)



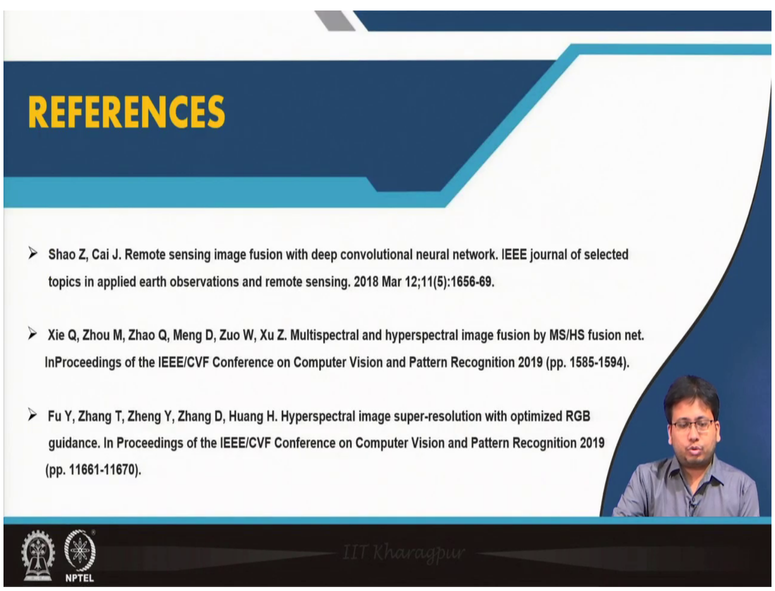
And, so, similarly like this can be extended to like satellite and radar ground radar fusion also will for that are similar neural net MLP based neural network has been developed by the same research group which is based on University of Colorado at Boulder.

(Refer Slide Time: 34:19)



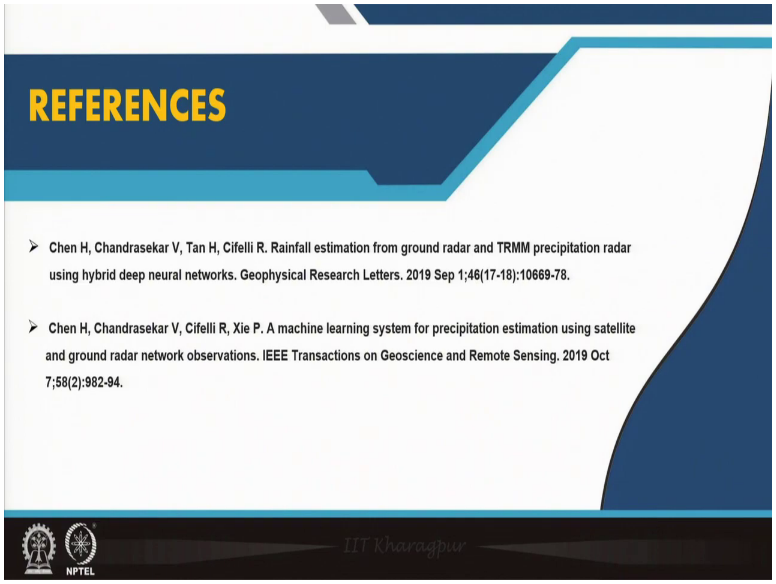
So, here like there is the instead of the instead of ground radar and space radar here on the input are the measurements from the satellites and the outputs are the measurements from the ground radars and the model remains the same which is the multilayer perceptron.

(Refer Slide Time: 34:52)



So, there so, these are the references.

(Refer Slide Time: 34:55)



So, that brings us to the end of this lecture where we discussed different ways in which the different sources of imagery or of remote sensing imagery can be fused with a neural network because different sources of imagery have complementary strengths. Some have high resolution high spatial resolution, some have high spectral resolution, some can cover a wider area, some have more accuracy etcetera.

So, these approaches or these papers which we discussed today basically what it does is they propose some way to get the best of all these worlds. So, that brings us to the next slide in to the to that brings us to the end of this presentation. In the next lecture, we will discuss some other applications of remote sensing in which machine learning and especially deep learning can be useful.

So, till then bye.