Machine Learning for Soil and Crop Management
Professor Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 36
UAV and ML Applications in Agriculture

Welcome, friends, to this NPTEL Online Certification Course of Machine Learning for Soil and Crop Management. We are going to start week 8, and the topic of this week will be UAV and Machine Learning Applications in Agriculture.

(Refer Slide Time: 00:39)

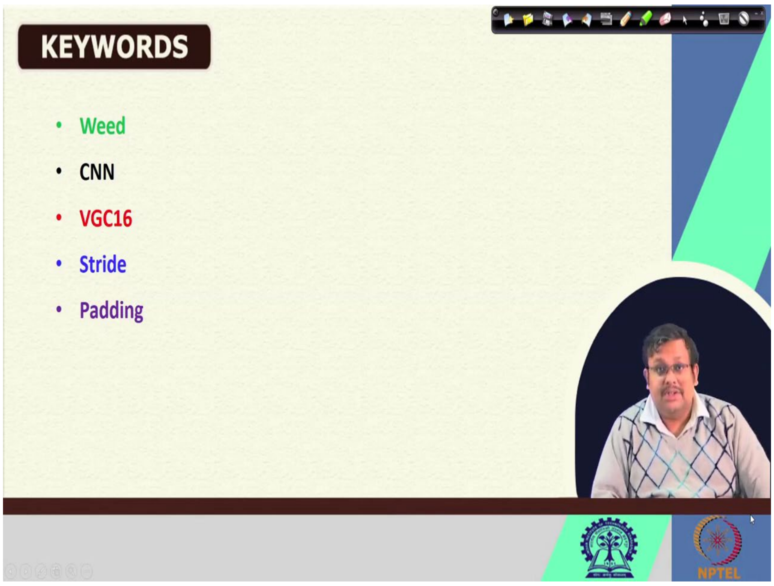
So, this is lecture number 36, and in this week we are going to discuss in details about the UAVs, that is unmanned aerial vehicle, and machine learning and deep learning applications in agriculture, specifically for soil and crop image processing. Now, in this first lecture, before we actually go to UAV, it is important to understand some more application of image based soil and crop characterization.

And in this week, I am going to, in this lecture I am going to talk about one soil image application. Also I am going to talk about another application where crop images actually were used for identification of weeds in the crop field by using deep learning architecture of Convolutional Neural Network. So, let us start.

(Refer Slide Time: 02:24)

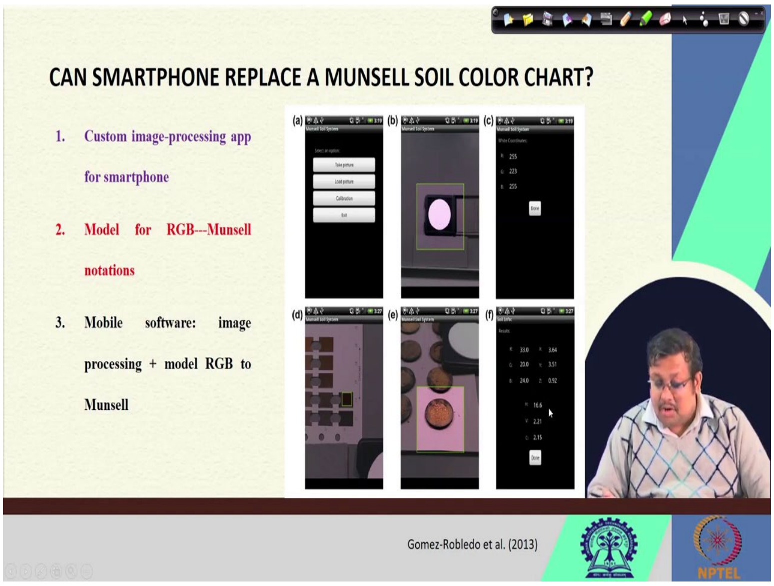
So, here in this lecture, I am going to focus, as I have mentioned, I am going to focus on these two concepts. One is the image based soil property prediction, and the second one is crop image plus Convolutional Neural Network for weed identification.

(Refer Slide Time: 02:43)



And these are some of the keywords which we are going to discuss in this lecture. First of all, the weed, then Convolutional Neural Network, then VGC16, then Stride, Padding, these things we are going to discuss. We have already discussed the structure of Convolutional Neural Network. So, some of the important terminologies, we are going to define, which will be necessary for understanding the application.

(Refer Slide Time: 03:20)



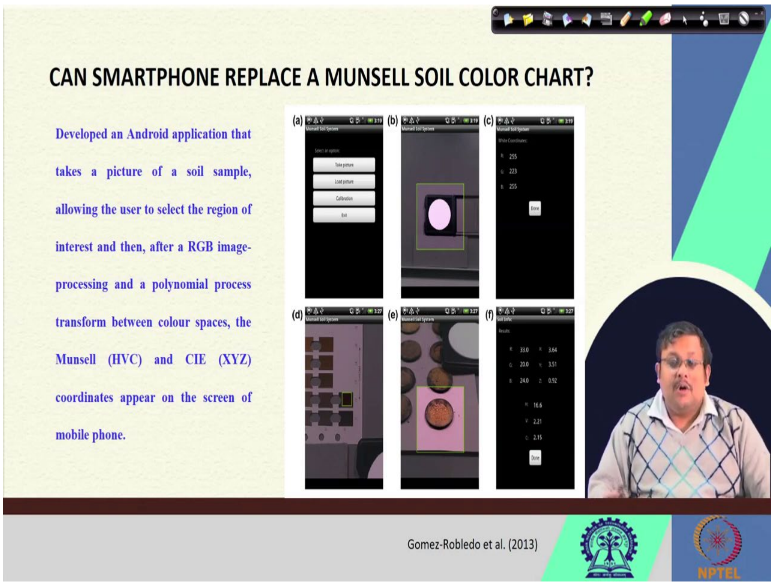
So, let us start with the first point, that is can, smartphone replace a Munsell soil color chart. As I have already mentioned that soil color is traditionally qualitatively measured in the field or in the lab using Munsell soil color chart, by the scientists. And there are different parameters of Munsell soil color chart like hue, value chroma, we have discussed. And there are certain other apps or certain optical sensors like Nix which are being tested by the scientists nowadays to replace the soil color chart.

But, given the widespread availability of smartphone based images and android app, this scientist group, Gomes, Robledo et al, in 2013, they have tried to seek the answer of this question that can smartphone replace a Munsell soil color chart, because if it is possible, then the whole concept of qualitative description of soil color will change with the quantitative description and subsequent modeling.

So, what they have done, they have developed a custom image processing app for smartphone. Secondly, they try to model for converting these RGB values which you can extract from an image to Munsell notation. And thirdly, they try to develop a mobile software which can combine both image processing as well as they can model this RGB to Munsell. And you can see here, the different steps of this app.

For example, they are first, you have to take the image of a white standard and then calibrate the RGB values. Once you calibrate the RGB values, in the next step, you can take the image of the chip and you can take the image of the soil, and then you can produce the values of different soil crop, different types of color models. So, you can see here RGB, then XYZ, and HVC and So, on.

(Refer Slide Time: 05:58)



So, this group has developed an android application that takes a picture of a soil sample, allowing user to select the region of interest, and then after an RGB image processing and a polynomial process transformation between the color space. So, once you capture the image, you extract this RGB information, you can use some polynomial process transform to convert that RGB color space model to the other color space model.

We have discussed the color space model like CMYK, HVC, LCH, LAV in our previous lectures. So, these color space models can be transformed from each other using a set of defined equations. So, this app can convert this RGB image to the other colors crop spaces, the Munsell HVC values as well as CIE XYZ coordinates. So, ultimately you can see from these RGB values, this XYZ and HVC, the results are appearing in the output.

So, that shows that, yes, using this machine learning, using different types of algorithms as well as the smartphone captured images, it is now possible to convert or to replace the traditional qualitative description of soil color using the smartphone. I have already showed you that similar way, the same concept was utilized for Nix sensor for using a set of color extracted values.

And then converting them into other color space model features and then using them together in combination with Artificial Neural Network or other deep learning method or other machine learning method to replace the traditional Munsell soil color chart for assigning the soil color based classification. So, that shows that the application of soil color using the smart, using the smartphone images along with some data mining tools are hugely important for future advancement of quantitative soil modeling.

(Refer Slide Time: 08:45)

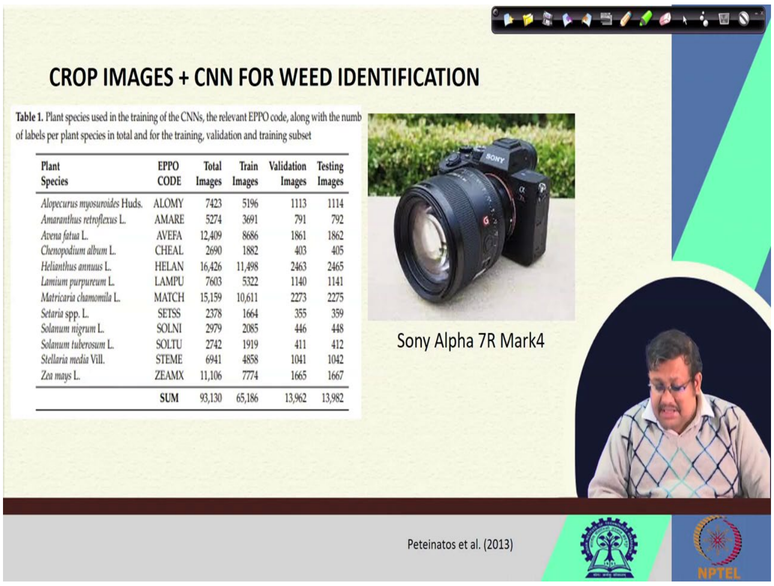


Table 1. Plant species used in the training of the CNNs, the relevant EPPO code, along with the numb of labels per plant species in total and for the training, validation and training subset

Plant Species	EPPO CODE	Total Images	Train Images	Validation Images	Testing Images
Alopecurus myosuroides Huds.	ALOMY	7423	5196	1113	1114
Amaranthus retroflexus L.	AMARE	5274	3691	791	792
Avena fatua L.	AVEFA	12,409	8686	1861	1862
Chenopodium album L.	CHEAL	2690	1882	403	405
Helianthus annuus L.	HELAN	16,426	11,498	2463	2465
Lamium purpureum L.	LAMPU	7603	5322	1140	1141
Matricaria chamomila L.	MATCH	15,159	10,611	2273	2275
Setaria spp. L.	SETSS	2378	1664	355	359
Solanum nigrum L.	SOLNI	2979	2085	446	448
Solanum tuberosum L.	SOLTU	2742	1919	411	412
Stellaria media Vill.	STEME	6941	4858	1041	1042
Zea mays L.	ZEAMX	11,106	7774	1665	1667
	SUM	93,130	65,186	13,962	13,982

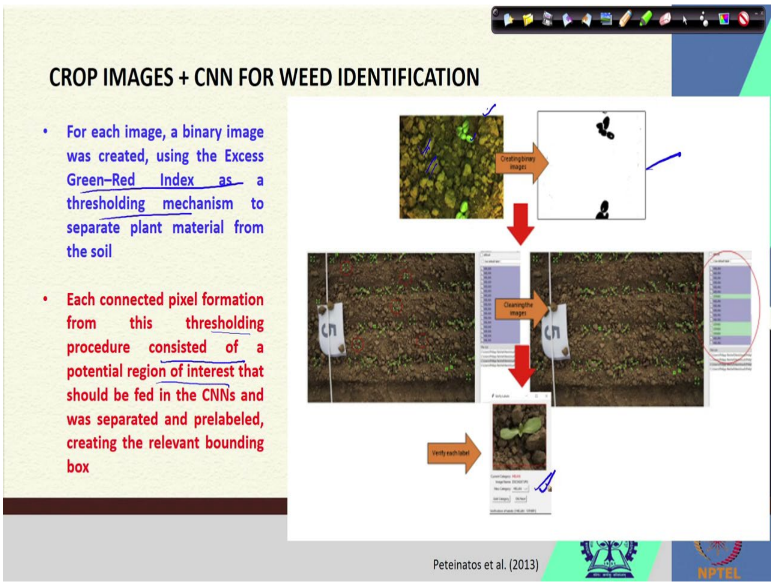
Now, I would like to also discuss another very important application using the digital crop images and Convolutional Neural Network for weed identification. The reason I am telling, explaining this before we go to UAV, because in case of UAV also they are utilizing different types of RGB camera. UAV can utilize RGB camera to capture the high resolution photographs of the crop.

And similar type of operation, we can mimic there for weed identification or for crop health identification or crop health status identification. We are anyway going to discuss this, but before we go to the actual UAV, it is important to at least understand one application where crop images were utilized in combination with Convolutional Neural Network for weed identification.

So, this research was done by this Peteinatos et al, in 2013, where they have combined the crop images from 1, 2, 3, for different, 12 different crops. They have grown those crops in the field and they have taken the image using the Sony Alpha 7R Mark4 camera. High resolution images from, at different growth stages.

So, there is a total of 93,130 images they have collected, and for training, they have used 65,000. Then, for validation of the images they have used 13,962, and for testing, they have used these 13,982 images. So, you can see here the major crops they had grown were Zea Mays, which is maize, Solanum Tuberosum, which was, which is potato, and then the Helianthus Annuus, which is sunflower. And rest of the crops, as you can see, these are basically the weeds.

(Refer Slide Time: 11:14)



So, these, all these 12 crops were grown in the field, and their images were taken. So, for each image, a binary image was created. I told you that a binary image is represented by two digits, that is 0 and 1. So, a binary image was created using the excess Green-Red index. So, excess Green-Red index they have calculated as a thresholding mechanism to separate.

Because once you take the photograph using your camera, you will see both the cuff as well as the surrounding background which is showing the soil, but you do not need to incorporate the soil information in your machine learning model, because soil is not important. Because suppose your ultimate goal is to classify the winds, or identify the weeds in the plant, in the field.

But here, soil is immaterial. So, you have to remove this soil. So, for this, they have created a binary image. As you can see here, they have used the binary image using the excess Green-Red index as a thresholding mechanism. So, here, in the binary image you can see they have only kept, they have segregated the crop from the background.

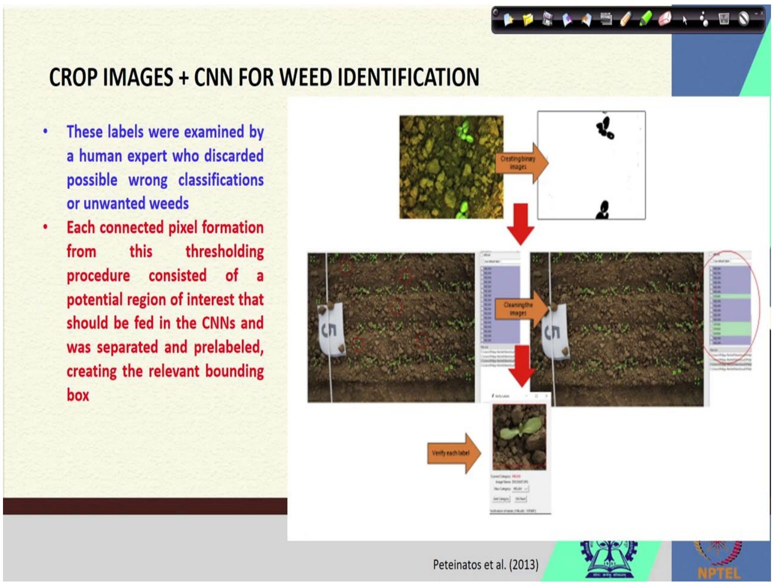
So, after they have done this, what they did, they have can they have basically connected pixel formation that using the, this thresholding procedure, they have done, using procedure consistent, using a potential region of interest. So, here you can see these red circles are showing the region of interest or ROI that should be fed.

So, these ROIs were created by connecting pixel method from this thresholding procedure, and then fed into the subsequent Convolutional Neural Network. And then, they were separated and prelabeled creating the relevant bounding box. So, these images of the crop were separated from the background information, and then they were fed into the subsequent machine learning models or Convolutional Neural Network model.

And then, they were prelabeled. As you can see here, they were clearly labeled, here, and then labeled were further verified by an expert. So, this is very important part of image processing before they execute this Convolutional Neural Network. Again, in a nutshell, they have taken the image, after taking the image they converted into the binary image.

And then using this thresholding formation they used this, connected pixels and they identified this region of interest which can only capture the images of the crop and then subsequently, they fed it into the Convolutional Neural Network after the labeling.

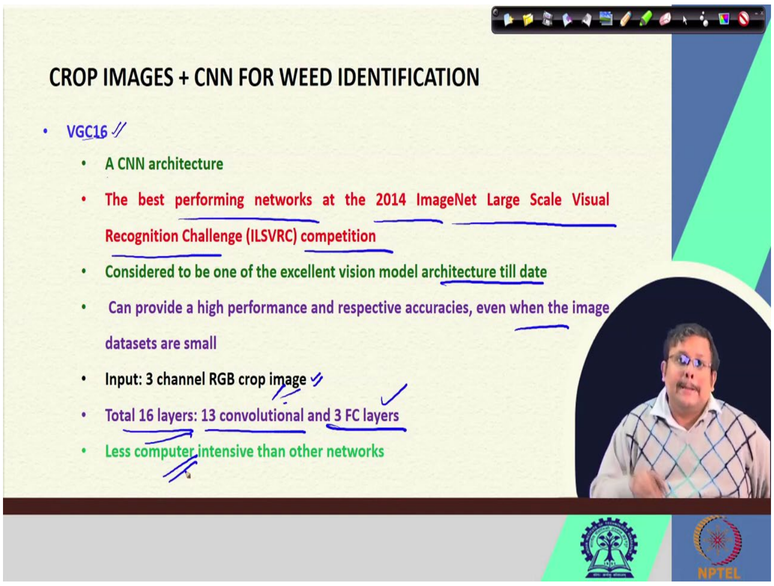
(Refer Slide Time: 14:43)



So, these labels were examined by a human expert who discarded possible wrong classification or unwanted weeds. Remember, once you identify this region of interest, it is very, very important that you correctly label them, because if you are labeling them wrongly then your classification will suffer. So, they were further validated by a human expert, and they discarded this possible wrong classification of unwanted, discarded these possible problems from this miss classification.

Now, each connected pixel formation from this thresholding procedure consists of a potential region of interest, and then they went to the next, this Convolutional Neural Network. And then, in the Convolutional Neural Network they have tried three different types of CNN models.

(Refer Slide Time: 15:37)



So, the first one, I am going to discuss this in details, So, the first one they have tried is called the VGC16 model. So, this is a basically and Convolutional Neural Network architecture. So, this model was first introduced in 2014, because it was appeared as the best performing network at this 2014 ImageNet Large Scale Visual Recognition Challenge competition.

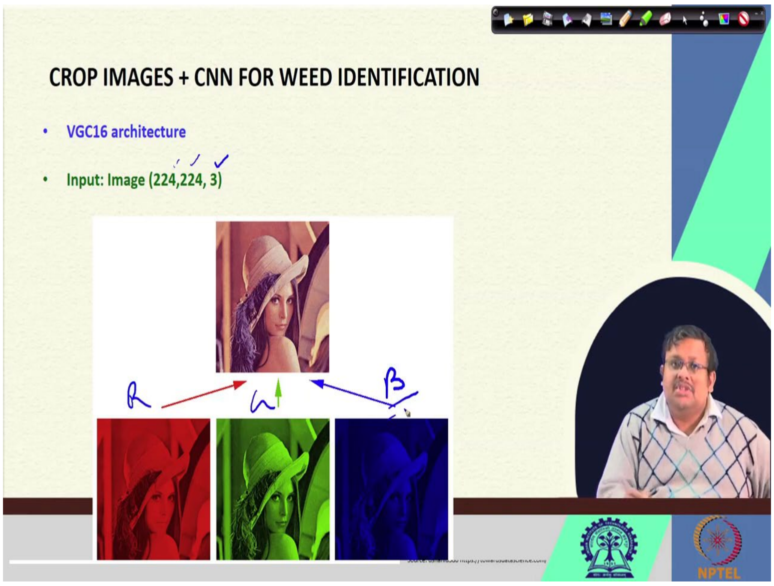
So, this ImageNet is a big data set where they, this VGC16 has performed best as a CNN architecture. It is nothing but a, it is also CNN model, but it is having a specialized architecture. So, we are going to discuss the specialized architecture. So, it is considered to be one of the excellent vision model architecture till date. So, it is an excellent vision model architecture till date, because it can provide high performance and respective accuracies even when the image data sets are small.

So, this is one benefit of using this VGC16 Convolutional Neural Network. So, what is the input in this VGC16? So, in this VGC16, the input is a 3 channel RGB crop image and there are total, because the three channels means one image or color image can be represented by three channels R, G and B, I am going to show you in the next slide. So, this RGB crop image appeared in this research as the input and then this CNN architecture of VGC16 was executed.

Now, in this VGC16, as the name suggests, it has total 16 layers and out of them 13 are convolutional layers and 3 fully connected layers. So, we have already discussed what are the convolutional layers and what are the fully connected layers, So, I am not going to discuss this again. But remember, one of the major challenges of this VGC16 model, is it is you it is less computer intensive than other networks.

So, you can see clearly that after you take the image and do the thresholding to remove the unwanted material from the from the background, you clearly label them, and then you feed the image directly into the Convolutional Neural Network for subsequent image based classification.

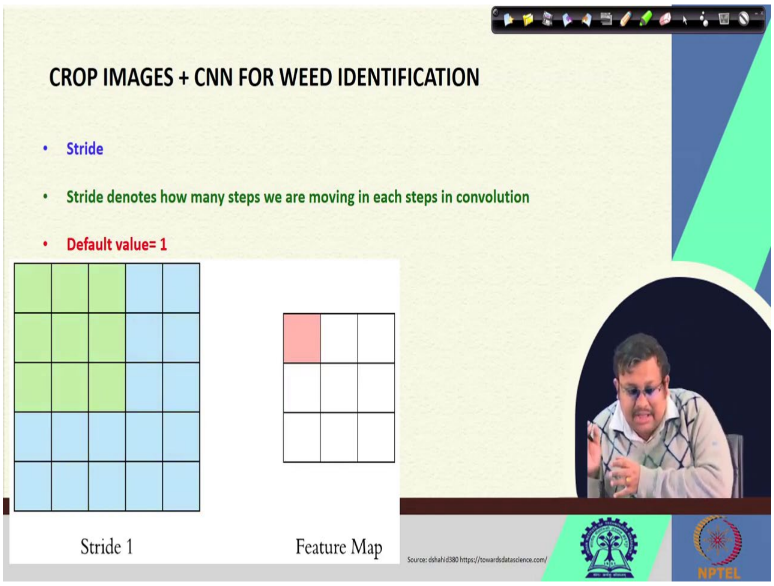
(Refer Slide Time: 18:38)

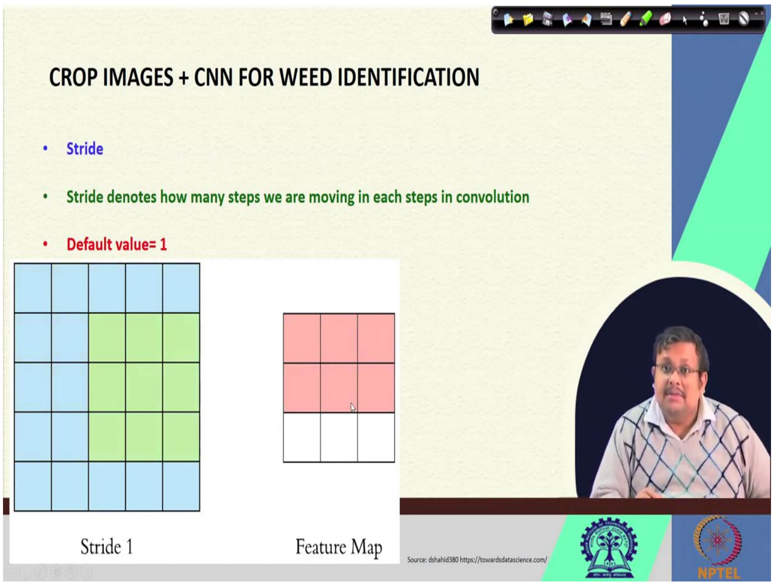


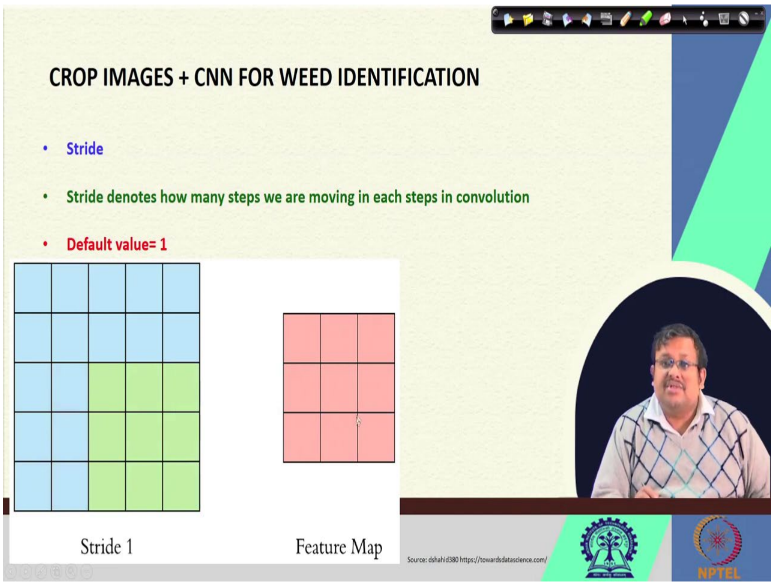
So, let us see the input image. So, you know, that any image can be decomposed into, any color image can be decomposed into 3 channels of RGB. So, as you can see this color image can be decomposed into these R, G and B channel image. And when they combine them, they will be producing this final color image. So, this VGC16 architecture is basically having the inputs from these color images.

In this research it was crop images. Of course, there will be 224 by 224 pixel resolution, and there will be 3 layers or 3 channels, one is for R, another is for G, another is for B. So, this is the input in this VGC16.

(Refer Slide Time: 19:32)





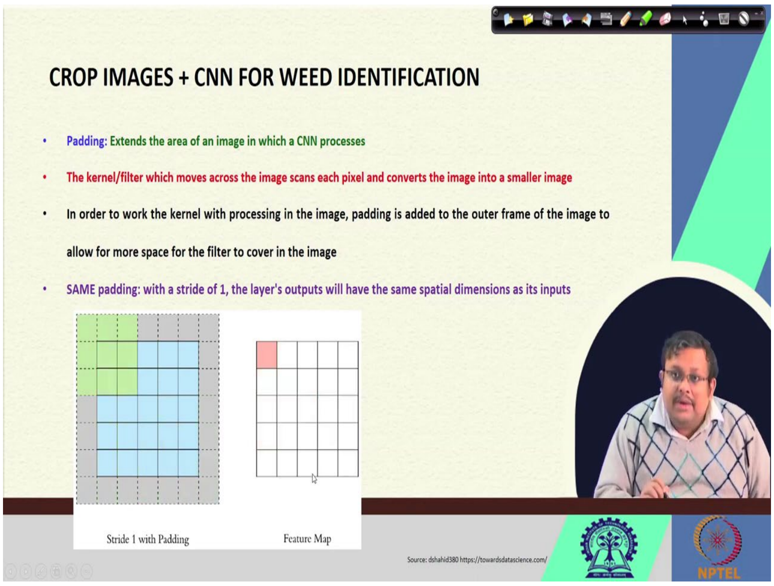


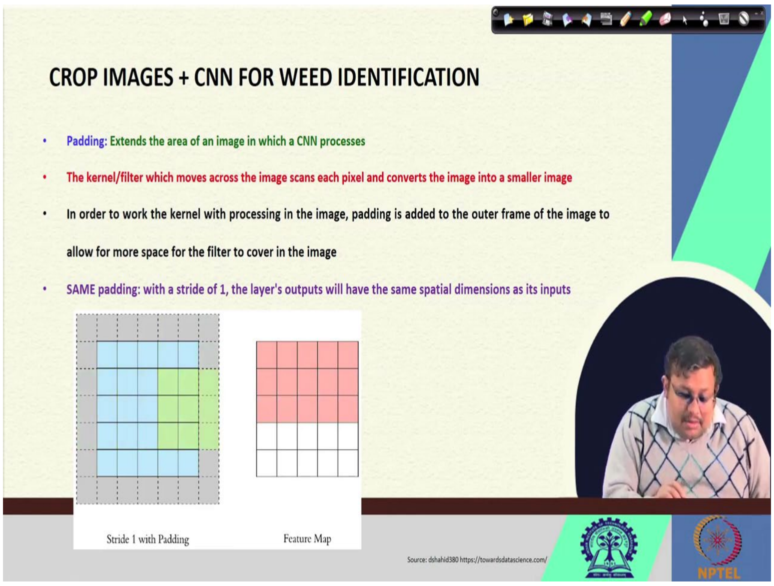
Now, before we, before, let me just go back quickly. So, before we go and discuss in details about the about the structure of the VGC16, it is also important to understand two important definition. One is called Stride, another is Padding. So, the, first comes the question, what is a Stride?

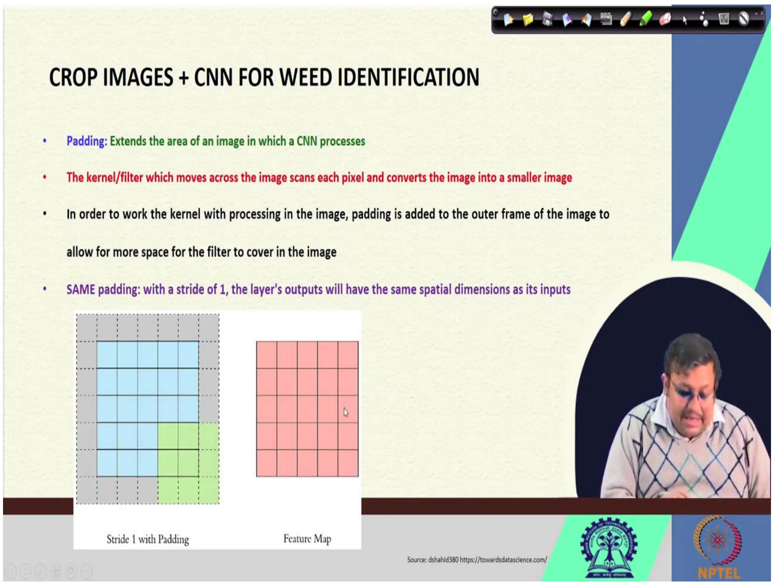
Stride denotes how many steps we are moving in each step in convolution. So, let me just go back for a quick help. So, if you can see, this is called stride. So, you can see we are moving one step at a time, when this is a stride one, the default value is always one. So, that means when the kernel is moving through the image, how many steps we are moving in this total convolution.

So, when, in other words, when the kernel is moving during the convolutional process, in each step, how many we are actually making the progress. So, here you can see one at a time, So, here, this is called the stride equal to 1 so, the default value is 1.

(Refer Slide Time: 21:10)







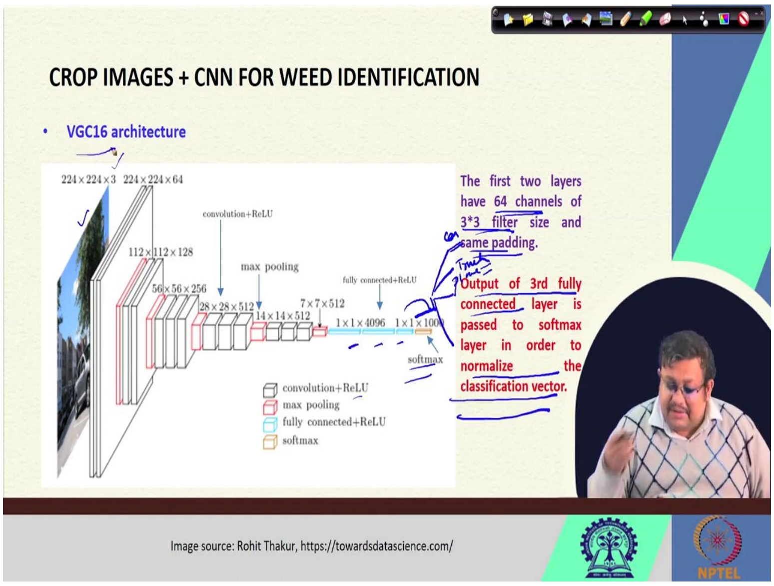
Now, another important feature is called the Padding. So, Padding, what is Padding? Padding is basically, it is an operation which extends the area of an image in an Convolutional Neural Network process, because you can see as the kernel is moving, so, this blue area is actually the image. So, to feed this kernel to cover all these individual pixels of an image, it is necessary to extend this area of this image at this edge so, that this kernel can fit perfectly to this, all these individual pixels.

So, the kernel or filter which moves across this image scan each pixel, and then they convert the image into smaller image as you can see from here. So, in order to work with the kernel with processing in the image, padding is added to the outer frame of the image to allow for more space for the filter to cover the, in the image. This external shaded area is given just to ensure that by moving this filter or kernel, it is possible to cover the whole image.

So, there are different types of padding. One padding is known as the SAME padding, that is, with a variable stride of 1, I have already discussed what is stride. So, with a variable padding of stride 1, the layers output will have the same spatial dimension as this input. So, here, as you can see, here, since it is a stride 1, that means we are moving one step per convolution. So, that means, when we are, so, this is the default value, so, when this condition is met, you can see that the output image will have the same spatial dimension as the input image.

So, here you can see it is 1, 2, 3, 4, 5 by 5. So, 5 by 5 pixel, and you can see the output image is also 5 by 5 pixel. So, when the stride is 1, same padding will show the same special dimension of the output as the input. So, I hope now, these two terms are clear to you. Now, what is the importance for these, in this whole VGC16 architecture?

(Refer Slide Time: 23:50)



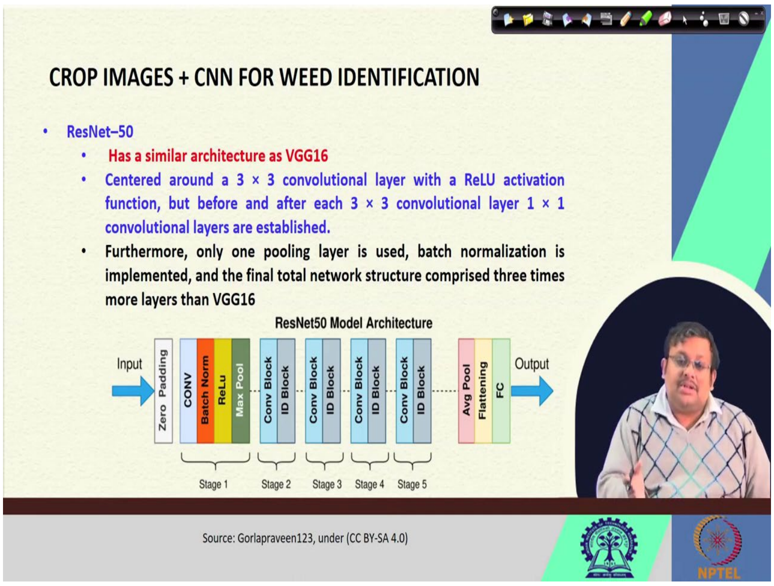
So, this is the total VGC16 architecture, and if you can focus on it, you can see this is the input image of 224 by 224 pixel, with 3 layers, and the first two layers, so, these are the, these white are the convolutional layers, plus these ReLU are rectified linear image, and then you can see here, these red layers are max pooling layers, and these are the fully connected layer. There are three fully connected layers as we have discussed, and of course this is softmax criteria.

Now, the first two layers, so, here you can see the first two layers have 64 channels of 3 by 3 filter size and SAME padding. So, we have already discussed what is SAME padding, and these two layers, so, these individual layers have 3 by 3 filter and total 64 channels, each of these. So, similarly, the layer constructions are given here, step by step, and you can see the output of the third fully connected layer, So, this thirdly third fully connected layer is passed to this softmax layer in order to normalize the classification vetcor.

So, ultimately, the final end product is suppose we are using different features or objects in this picture, can we classify these objects suppose, car or truck, or suppose plane, so, suppose these are our objects in an image. Ultimately, using this architecture, can we classify this features, or classify these objects in the image. So, this is our final target or objective. So, I hope now it is clear to you. Of course, if you want to have more and more information you should go and read some literature regarding this VGC16.

But our time is limited so, we are not, we cannot go further. So, here, they have tried, in this research, they have tried this VGC16 architecture.

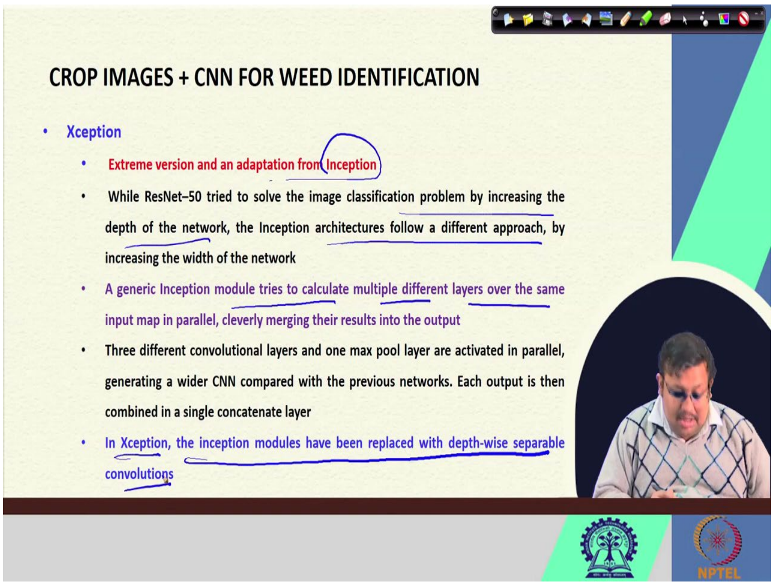
(Refer Slide Time: 26:25)



Similarly, they have also tried this ResNest-50 architecture, which has a similar architecture as VGC16, and it is centered around 3 by 3 convolutional layers with a ReLU activation function. But before and after each 3 by 3 convolutional layer, 1 by 1 convolutional layers are established. Further, only one pooling layer is used, and batch normalization is implemented and the final total network structure comprises 3 times more layers than VGC16.

This is, this ResNet-50 is another very widely used CNN structure. The only variation in these models are their structures in the CNN. So, you can see the number of layers are varying from one structure to another structure, and also the number of pooling layers, number of features and all these things. So, only in the construction, based on the constructional variation, we can, there are different types of Convolutional Neural Network.

(Refer Slide Time: 27:36)

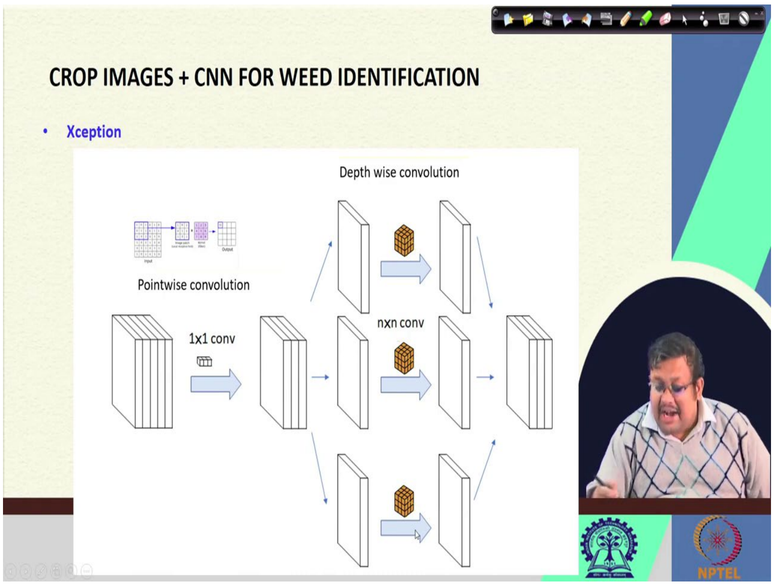


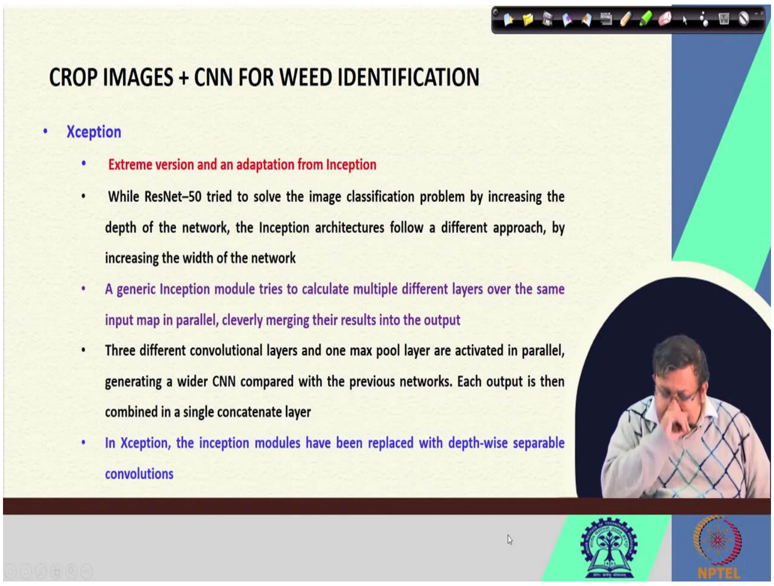
Another convolutional network, they have tried, it is called Xception. So, this Xception is also known as the extreme version of an adaptation from an Inception, which is another architecture. So, while ResNet-50 tried to solve the image classification problem by increasing the depth of the network, because you know, the major difference between the ResNet-50 and VGC16 is ResNet-50 has more number of layers, and they try to get the accuracy of the image classification based of a more number of layers or increasing the depth of the network.

The inception architecture follows a different approach by increasing the width of the network. So, a generic inception model tries to calculate multiple different layers over the same input in parallel. I will show you in the next slide. So, it will try to calculate the multiple different layers over the same input map in parallel fashion, clearly merging their results in the output. So, 3 different convolutional layers, and 1 maxpool layers are activated in parallel. So, they are run in parallel, generating a wider CNN compared with the previous networks.

And each output is then combined in a single concatenate layer. So, in Xception, these are all about the Inception. So, in Xception, the Inception modules have been replaced with a depth-wise separable convolutions.

(Refer Slide Time: 29:15)

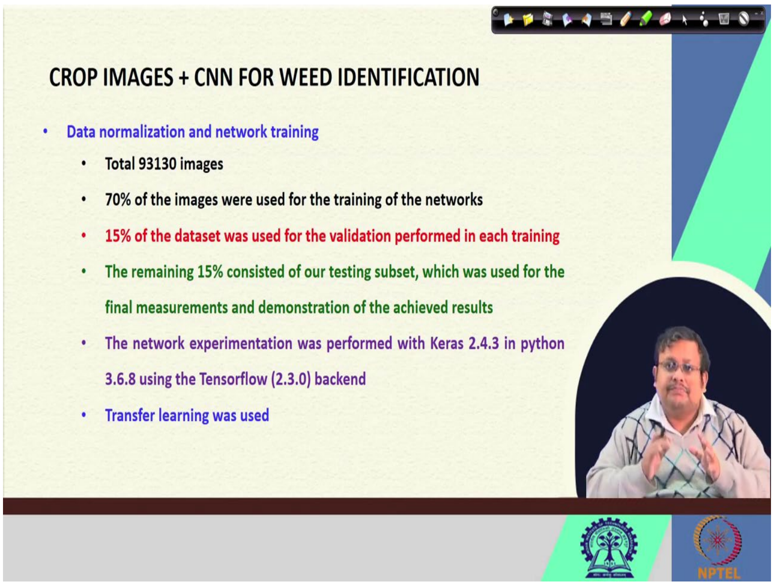




So, you can see, this is the architecture of the Xception. As I have mentioned that in Xception, the Inception modules have been replaced with a depth-wise separable convolution. So, you can see depth-wise separable convolutions are maintained, and ultimately the results are combined and concatenated together to get the final classification.

So, it starts with the point-wise convolution, and then we get this depth-wise convolution. Ultimately, we are getting the results together to finally get this final prediction or classification.

(Refer Slide Time: 29:50)

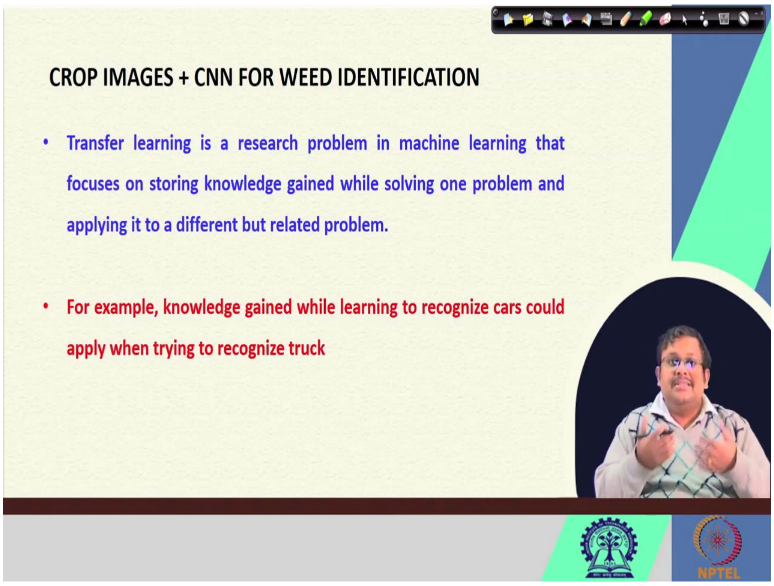


So, what they did in this research, they used these three different types of Convolutional Neural Network, using this total 93,130 images, they tried with the 70 percent of the images for training, the 15 percent of the data set used for the validation purpose in each training, and the remaining 15 percent, they used for testing the subset which was used for the final measurement and demonstration of the achieved results.

And the network experimentation was performed with this Keras 2.4.3 in python using the Tensorflow backend. You know Tensorflow is a system which is a package which is developed by Google brand team to execute different types of advanced machine learning algorithms. And then this is very, very favorite nowadays in the machine learning domain.

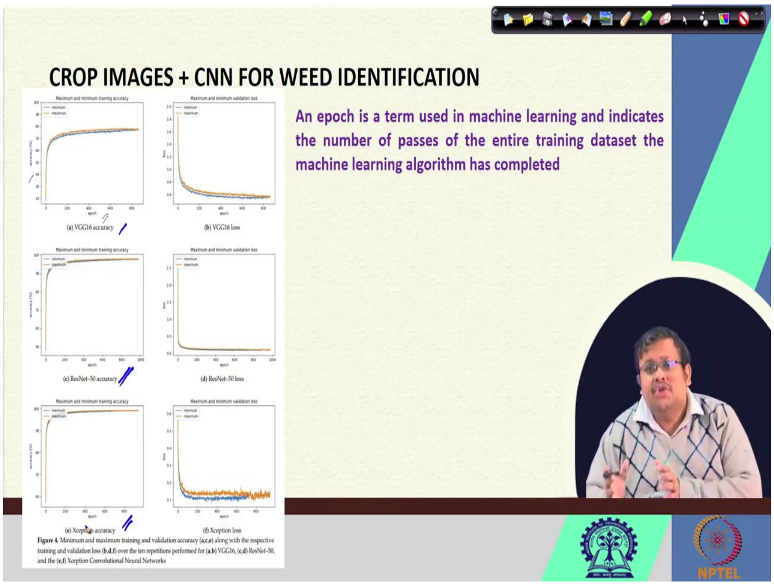
So, if you want to have more information about this Tensorflow backend, you please go and read some literature regarding this Tensorflow. And of course, in this total exercise, this transfer learning was used.

(Refer Slide Time: 31:12)



Now, what is a transfer learning? Transfer learning is a research problem in machine learning that focuses on storing the knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize car could be applied when trying to recognize trucks. So, this is a kind of a transfer learning.

(Refer Slide Time: 31:36)

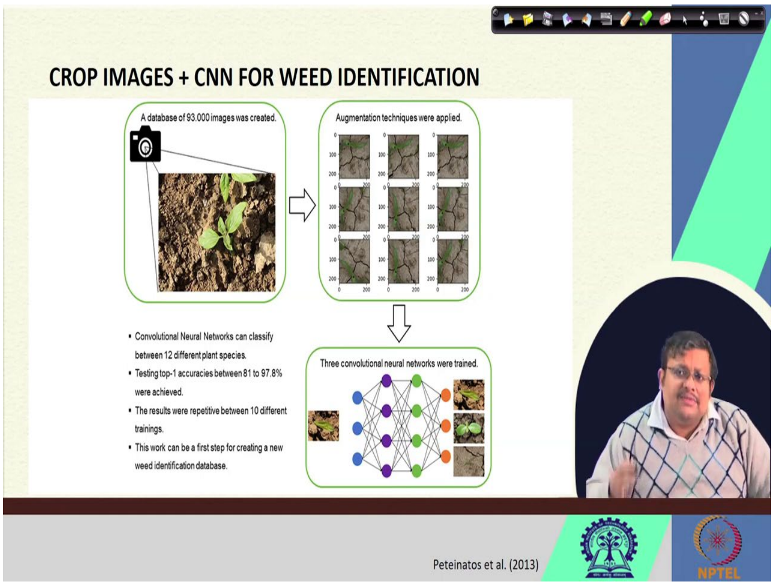


So, in this research, they have tried the transfer learning also and finally, this is the result they have got. They have found that while using these different types of Convolutional Neural Network, it is possible, so, they have tried to see the accuracy versus the epochs. Epochs is basically a term which is used in the machine learning, and indicates the number of passes of the entire training dataset the machine learning algorithm has completed.

So, you can see that as this number of epochs are increasing the accuracy of both the training set and testing set is increasing up to a certain point, and then they are reaching a plateau. And also if we plot the loss function with the number of epochs, we can see it will reduce to a certain point and then it will reach a plateau.

So, for all these VGC16 and then ResNet-50 and then Xception, they have tried to plot these accuracy versus the epochs, and they have find the optimal number of epochs for which they are getting the highest accuracy, and then they have found that this type of architectures can be used to identify or classify the weeds in the in the field using the crop images and deep learning methods.

(Refer Slide Time: 32:58)



Now, in a nutshell, what they have used, a database of 93,000 images were created and augmentation trainings were applied, and then they have tried this Convolutional Neural Network for predicting or classifying the weeds in the field. So, Convolutional Neural Network can classify between 12 different plant species and testing this top-1 accuracies between 81 to 97.8 percent was achieved.

These results were repeated between 10 different trainings. And this work was shown that it was the first step for creating a new weed identification database. So, ultimately this can be used for future with identification in the field also.

(Refer Slide Time: 33:45)

So, these are the references, guys. I hope that you have learnt something new from this information, you have gained some understanding of how these crop images can be extended or combined with different image processing and Convolutional Neural Network to finally predict the, or classify the weeds.

And we have also learned the structural variation of three different types of Convolutional Neural Networks, VGC16, ResNet-50 and Xception, and we have seen the differences. And basically when they are using different types of structures, they are having, they are offering some advantages over each other, and then they are producing more accurate results.

So, of course, these results are encouraging and showing the applicability of crop images with machine learning for better identification, classification of crops, as well as their properties. So, guys, let us wrap up our lecture here. These two are the references which I used.

I would request you to please go and read these two papers to gain more and more comprehensive information from this, for this research. And, so, let us wrap up our lecture here. And in the next lecture, we will be discussing about the UAVs and their agricultural operations. Thank you.