**Course Name: Machine Learning and Deep learning - Fundamentals and Applications**

**Professor Name: Prof. M. K. Bhuyan**

**Department Name: Electronics and Electrical Engineering**

**Institute Name: Indian Institute of Technology, Guwahati**

**Week-12**

**Lecture-41**

Welcome to NPTEL online course on machine learning and deep learning fundamentals and applications. In my last week of this course, I explained the concept of the deep learning and also I explained the concept of the convolutional neural network that is the CNN. After this I discussed the concept of some popular CNN architectures like the Alex net, VGG net, Google net. So, all these concepts I have explained. In this week, mainly I will consider some popular deep architectures and also the applications of these deep architectures. So, it is not possible to discuss all the deep architectures, popular architectures in this course.

So, mainly I will focus some very important deep architectures and also the applications of these architectures. So, today I am going to discuss the concept of one very popular architecture that is the generative adversarial networks. So, in case of the generative adversarial network that is the GAN, we have two models model 1 and model 2. Model 1 and model 2 they compete with each other and because of this competition, the performance of the models will improve.

So, that is the fundamental concept of the GAN. So, this concept I am going to explain today and also some of the applications of the GAN and particularly the biomedical image analysis. So, how the GAN can be employed for the application? The application is the biomedical image analysis, but there are many other applications. So, I will show all the applications of the GAN. So, let us begin this class, the generative adversarial network that is a GAN and the applications.

So, this is the generative adversarial network, the concept the fundamental concepts and the applications of the generative adversarial networks. So, why actually we need generative models? So, we know what is the discriminative models. So, in case of the discriminative models, suppose given an image x, so an image is given and from the image I have to predict the label y. That means I have to estimate the probability of y given x. So, y is nothing but the class and x is the input image that is the objective of the discriminative models.

So, discriminative models have several limitations. So, one limitation is it cannot model P(x) that is the probability of x that is the probability of seeing a certain image. That means with the help of the discriminative model, I cannot generate new images. So, that means I cannot determine probability of x. So, it is not possible to generate new images.

So, that is why the generative models are important. So, with the help of the generative models, I can model probability of x and I can generate new images. So, let us discuss about these generative models. So, generative adversarial network, the concept is the generative models. And in the figure you can see I am showing the samples belonging to two classes and you can see the samples one is the dog another one is the cat in the figure and you can see the decision boundary between the classes.

So, generative modeling aims to model the distribution that a given set of data came from. So, in this case you can see I am showing the samples and these samples belonging to different classes. So, in this case I am showing two classes and also I am showing the decision boundary between the classes. So, that is the concept of the generative learning. So, from where the sample is coming, so that we can determine.

So, what is the corresponding class that also I can determine. So, this is the background of the generative learning. So, now let us see what is actually the GAN, the generative adversarial network. So, already I told you in case of the GAN there are two networks that means there are two models and in this figure I have shown this GAN. So, one model is the generator that is the one model the model number one and another model is the discriminator.

So, that means in this case there are two networks. So, input to the generator that is actually the random noise and the generator is generating fake samples and the input to the discriminator is the real data samples and the fake samples. So, discriminator identify which one is the real data sample and which one is the fake sample. So, if you see here, so we have a discriminator network that takes samples from the true and the generated data. So, here you can see this discriminator is taking the inputs from the generator output and also the real data samples.

So, that is one is the fake samples another one is the real data samples and tries to classify them as much as possible. So, discriminator tries to classify which one is the real samples and which are the fake samples that is nothing but a binary classification problem. And what is the objective of the generator? So, a generator network that is trained to fool the discriminator by generating real looking images. So, that means the generator the objective of the generator is to generate fake samples. So that it can fool the discriminator by generating real looking images.

So, you can see the objective of these two networks or the two models are directly opposite. So, one is the discriminator, it tries to distinguish the real data samples and the fake data samples and generator it is generating the fake samples. So, what is the actually the adversarial networks? So both networks try to beat each other and doing so, they are both becoming better and better. So, that means both are competing discriminator and generator they are competing with each other and because of this competition, the performance of the models will improve. So, at each iteration of the training process, the weights of the generative networks are updated.

So, during the training we have to update the weights of the generative network and update it in order to increase the classification error. So, what is the objective of the generator? The objective of the generator is to increase the classification error. Whereas the weights of the discriminative network are updated to decrease the error. So, in case of the generator, the objective is to increase the classification error that is the classification between these two classes one is the real data sample and another one is the fake samples. So, that is the objective of the discriminator.

So, from this discussion, you can understand the objective of these two networks or the two models are directly opposite. The generator is increasing the classification error and the discriminator is reducing the classification error. So, that is the fundamental concept of the generative adversarial network that is the GAN. So, I am giving one example here. So, I am showing one model 1 and model 2.

Model 1 is suppose one criminal and model 2 is the police. So, model 1 he is making the fake notes the fake currency and model 2 police he is detecting the fake notes. So, this police can detect whether the notes or the currency is the real one or the fake one, but this model 1 that is the criminal he is making the fake currency or the fake notes. So, both are competing with each other because of this competition the performance of these models will improve with each and every iteration. So, this already I have explained here you can see.

So, this criminal he is making the notes the fake notes and that was caught by the police and these are the real notes from the bank. So, this police can identify which one is the real note and which one is the fake note. You can see this statement. So, these notes above that I got from the banks are real and these notes below are the counterfeit notes. So, that means the police can identify which one is the real note and which one is the fake note.

After this you can see what actually the criminal is thinking the police is getting better at catching my fake notes. So, I should work hard in making better fake notes. So, now you can see he has to make better fake notes and you can see the performance of the model 1 that is the criminal he is improving the performance is improving. So, he is creating better fake notes now. After this you can see what the police is doing now.

So, just got this new batch of the fake notes I have to say these guys are getting better in making those and I should get even better to identify fake notes. Now the police is also improving because the problem is now difficult. So, this better fake notes and the real notes from the bank he has to distinguish. So, which one is the real note and which one is the fake note. So, you can see during this process both the model 1 and model 2 they are improving the performance is improving because of this competition.

So, police is getting better and that criminal is also he is doing better. So, that is the concept of the generative adversarial network. So, both are competing with each other and because of this competition the performance will improve. So, the GAN are designed based on the idea of the generative adversarial learning. So, that already I have explained it learns a generative model that can sample from a distribution without explicitly estimating it.

So, it was actually proposed in the year 2014 by Goodfellow. So, this GAN is a very popular technique mainly the applications like the synthesis of audio, text, image, video and the 3D model in the machine learning community. So, there are many applications of the GAN. So, I will discuss about all these applications briefly. So, these are some interesting applications of the GAN.

One is the image generation using the deep convolution GAN. Another one is the generation of the anime characters using the adversarial network and also the another one is the sketching to color photograph generation using the GAN. After this another application is the unpaired image to image translation using the cycle GAN. So, I can convert one image to another image by considering this GAN that is a cycle GAN. Text to image synthesis by stack GAN.

So, there is one popular GAN that is a stack GAN. Generation of new human poses by GANs. So, we can generate new human poses by GAN and also the single image super resolution that means we can improve the resolution of an image by GAN and the GAN based inpainting of the photographs. So, these are very interesting and the popular applications of GAN. I will be explaining some of these applications.

So, you can see the magic of GAN. So, it is very difficult to identify which one is computer generated. So, one is the real image another one is the computer generated image. So, it is really very difficult to identify which one is the real and which one is the computer generated. In this case you can see which one is real and which one is the fake. So, it is very difficult to identify which one is real and which one is the fake image.

In this case which one is the real and which one is the fake image. So, just try to identify. So, I can give the answer. So, this is the real face and the second one is the fake image the fake photo. Similarly, this is the real image and the second one is the fake image.

So, like this the GAN can generate number of images it can generate images. So, for this actually the this paper was published in 2020. So, new image can be generated by GAN the Adversal Networks. So, you can see in 2014 and this was the development in 2018 and this is you can see the generated image by the deep architectures. So, in this case the GAN was used for generating the new images and also the image generation like we can generate animation characters like this if you see the figure we can generate the animation characters and this GAN is called the style GAN and this was also published in the year 2020.

So, you can see the paper. So, the proposal is the style GAN for the generation of the animation characters and these are the examples like you can see the first one is the label to street scene. So, the street scenes are labelled. So, labeling to street scene, labels to fake it. So, you can see that the labeling can be done the black and white to color image conversion. So, we can convert the black and white into color image and in this you can see this is the input image and we are getting this output image that means mainly we are getting the roads.

So, for this type of applications we can employ GAN and here you can see that this is the input image and output image is the this that is the output of the GAN and again another application is the input is this image and output is this image. So, we can convert this or we can do this conversions by GAN the Adversal Network and also one application is that is the important application is the conversion of the low resolution image into high resolution image. So, in a figure you can see I am showing the low resolution image and that is converted into high resolution image by GAN. So, the first image is the low resolution image and second image is the high resolution image that is the single image super resolution by GAN and also the generation of the new human poses.

So, different poses can be generated by GAN. So, you can see this paper the paper was published in 2017 and how to apply the GAN for the generation of the new human poses. This is one important and interesting application. So, we can generate new human poses and also you can see the one popular application is the Deepfake that is very interesting application. So, privacy at stake. So, you can see the first image is the original image and second one is the fake image that is the deep fake image and that is also generated by the GAN the Adversal Network.

So, there are many such applications you can see one important application is the image to image translation and that paper you can see that was published in 2017. So, you can see from zebra to horse and from horse to zebra you can see this application. So, image to image translation from summer to winter and from winter to summer this application. So, like this we can do the translation that is the image to image translation I can do with the help of the GAN. So, we can generate a number of images this type of images we can generate and this was actually proposed in 2015.

So, unsupervised representation learning with deep convolutional generative adversarial networks. So, this was proposed in 2015 and you can see we can generate number of images and one important application is the biomedical application. So, we can do the segmentation that is the image segmentation by GAN. So, with the help of the Adversal Network we can do the segmentation the image segmentation and particularly for the biomedical applications like in this case I am showing the eye image that is the retina image and we are doing the segmentation.

So, this paper was published in 2022. So, you can see all these papers for understanding these concepts. So, now briefly I will explain the concept of the GAN. So, already I told you in this case of the Adversal Network there are two networks or the two models one is the generator network and another one is the discriminator network and both are competing with each other. So, the input to the generator network is the random noise. So, generator is generating fake samples and what are the inputs to the discriminator the inputs to the discriminator network are the fake samples and the real samples.

The fake samples confuse the discriminator and also the input to the discriminator is the real sample. So, real samples are inputted to the discriminator and also the fake samples are inputted to the discriminator. So, fake samples are generated by the generator network. The input to the generator network is the random noise. So, the discriminator tries to identify which one is the fake sample and which one is the real sample and the generator is continuously making the fake samples better and better fake samples.

The discriminator is identifying which one is the fake sample and which one is the real sample and because of this competition the both are improving. So, this is about the generative Adversal Network. So, this concept already I have explained. Now mathematically how to consider this problem.

So, here I have shown the generator and the discriminator. So, this is generator and this is the discriminator. So, input to the generator that is the random noise. So, we are considering the Z that is the code vector we are considering. So, the generator network Z which tries to generate realistic samples. So, it takes an input the input is the Z and computes X is equal to G(Z).

So, I can say it is the fake sample. So, the discriminator network is the D which is a binary classification network because the problem is the binary classification problem which tries to classify the real versus fake samples. So, input to the discriminator you can see the real world image that is the real images and the fake images these are the fake images generated by the generator. So, you can see it takes an input X it is the real and the generated. So, X is the real image and also the generated image and it computes D(X) that is the difference between this real and the fake and the probability is assigned to X being the real.

So, the probability it assigns to X being real. So, I am repeating this the input to the discriminator is the real and the generated and it computes D(X). So, that measure we are considering the D(X) that is the I can say the difference between the real and the fake. So, for the training of the discriminator and the generator we have to consider some loss functions. So, the discriminator is trained just like a logistic regression classifier. So, that concept I have explained in my class of the logistic regression.

So, we can consider the cost function the cost function may be the cross entropy we can consider for the classification of the real versus fake. So, this $J_D$ is the loss function and that is the cross entropy function. So, X~D denotes sampling from the training set if the discriminator has low cross entropy that means it can easily distinguish real from the fake. So, that means I am repeating this for the discriminator low cross entropy is better because it means it can easily distinguish real from the fake. So, for the discriminator I have to consider low cross entropy that means it can distinguish which one is the real and which one is the fake.

And for the generator we have to consider high cross entropy that means it cannot because both are competing with each other. So, the objective will be directly opposite. So, in case of the discriminator we have to consider low cross entropy and in case of the generator I have to consider high cross entropy. So, now I am considering the loss function that is the cross entropy function that I can write in this form. So, the previous expression I have shown that is the cross entropy that I can write in this form.

Because $J_G = - J_D$, because the objective of these two models are directly opposite.

That means this is equivalent to making the generators cross function the negative cross entropy. So, the most straightforward criterion for the generator is to maximize the discriminator cross entropy and this is equivalent to making the generators cross function the negative cross entropy because the objective or the goal of these two models are directly opposite and from the previous expression I am getting this expression that is the expression for the cross entropy and that is actually the loss function. So, if you see the previous expression that I have shown in the previous slide it has two terms the first term is the first term and is the second term, but this equation I have modified here. So, first term is the constant and we are only considering the second term because the generator has no control over the first term.

So, that is why we are considering it as constant. So, the previous loss function I can consider like this. So, the constant term is because the generator has no control over the first term. So, that is why I can consider it as a constant and we are considering the second term only. So, the overall cross function can be considered as a min max formulation.

So, we are considering this cross function $J_D$. So, it should be minimum for the discriminator, but it should be maximum for the generator. So, that means I can consider this formulation min max formulation. So, it is a perfectly it is a perfectly competitive game or I can say zero sum game since the generator and the discriminator have exactly opposite objectives. Now in this formulation there is a problem here.

So, the problem is the problem of the saturation. So, suppose in generator loss function when $D(G(Z)) = 0$ the meaning is the discriminator is doing well. So, you can see the curve here. So, if I consider this point. So, $D(G(Z)) = 0$ and corresponding to this point if you see here.

So, if I take the gradient of the loss function then it will be close to 0. So, that is actually the saturation the problem of the saturation. So, if I take the gradient of the loss function then it will be close to 0 during the back propagation. So, that means the generator cannot make a large correction update. So, that is why I am changing the definition of the loss function.

So, the loss function was like this, but it is modified like this. So, corresponding to this modification I am getting this curve that is the modified cost representation and the saturation problem will be eliminated because of this representation. So, how to do the training of the GAN? So, actually how to do the training? So, first I will show how to do the training for the discriminator and after this I will show how to do the training for the generator. So, for this training we consider the back propagation training. So, in the figure you can see I am showing the training of the discriminator that is in the first figure. So, you can see here update the discriminator weights using back propagation on the classification objective.

So, the first I am doing the training of the discriminator and in case of the generator you can see in the second figure we can update the generator weights using the back propagation technique. So, how to do the training? I have shown in the flow chart, initialize weights of the generator and the discriminator. So, randomly I am initializing weights of the generator and the discriminator. So, first the weights of the generator are frozen that means it is fixed and we are learning the weights of the discriminator. So, corresponding to this you can see we have the ground truth that is the input to the discriminator and it can identify the real one which one is the real it can identify and the input to the generator is the noise and we are having the generator output that is inputted to the discriminator and it can identify the fake samples.

So, for this we are considering the discriminator loss function we are considering. So, that is for the training of the discriminator. After this we have to consider the training of the generator. The first one is the training of the discriminator the second one is the training of the generator. For the training of the generator weights of the discriminators are frozen.

So, it is fixed and weights of the generators are learned. So, how to learn the weights of the generator that I will show you. So, input to the generator is the noise and we are getting the generator output and that is inputted to the discriminator and discriminator can identify this is the real one. So, for this we are considering the generator loss function. So, this process I have to do repeatedly it is an iterative algorithm until the termination criteria is not obtained.

So, we have some conditions. So, until these conditions are not satisfied I have to do the training of the generator and the discriminator. So, during the training of the discriminator the generator weights are frozen. So, it is fixed and during the training of the generator the discriminator weights are frozen. So, that is the concept of the training of the discriminator and the generator.

So, in this figure also I have shown the concept. So, input to the generator is the latent random variable that is the random noise and you can see it is generating G(z) that is the fake samples it is generating. The input to the discriminator one is the real sample that is obtained from the real world images. So, x is the real world images and G(z) that is nothing but the fake samples. So, discriminator can identify which one is the real and which one is the fake.

So, corresponding to the real this is a green and corresponding to the fake it is a red. So, it is a two class classification problem and corresponding to this we can determine the loss. The loss is nothing but the cross entropy function we are considering. So, this is the concept of the GAN architecture.

Now how to do the training. So, already I told you so if I want to train the discriminator. So, the generator weights are frozen that is they are fixed they are locked and we are doing the training for the discriminator. So, we are considering the back propagation training algorithm. After this we have to do the training for the generator. So, in this case the weights of the discriminator are frozen that is they are fixed they are locked and we are doing the training of the generator by the back propagation training.

So, we have to we have to update the weights of the generator. The input to the generator is nothing but the latent random variable that is the random noise. So, we can do the training like this. So, first the discriminator and after this the generator. So, now I will explain the application some of the applications.

So, in my last week I explained the concept of the convolutional neural network that is the CNN. So, same thing here I am showing here. So, input image classification problem you can see the input is the image and after this we are considering the convolution operation. So, for this we are considering the kernel that is the filter we are considering and the ReLU activation function is considered.

So, convolution plus the ReLU convolution plus ReLU. So, like this we can do and also the pooling. So, pooling already I have explained. So, to get the information of the local neighborhood statistics I can consider the pooling and with the help of the pooling I can reduce the size of the feature map. So, in a particular convolutional neural network I may have many convolution layers and I may have many pooling layers. So, after this you can see finally, the feature map is converted into 1D vector and that is nothing but the flattening and after this we are considering the fully connected layers for the classification and we are considering the softmax activation function.

So, that I am getting the probability values corresponding to each and every classes. So, here you can see 0.2 for the horse, 0.7 for the zebra, 0.1 for the dog. So, these are the probability values and based on this I can recognize. In case of the convolutional neural network the one important point is I can reduce the number of parameters as compared to MLP that is the multi-layer perceptron because we are considering the concept of the convolution pooling and also the concept of the stride. So, because of this we can reduce the number of parameters. So, if I consider a particular image suppose $200 \times 200 \times 3$ that is the RGB image. So, if I consider simple connection then ultimately it will be 1 2 0 0 0 0 weights.

So, that is the case of the fully connectivity. But in case of the convolutional neural network we consider pacer extraction by the process of convolution. So, that advantage already I have explained in my previous class. So, we can first extract the features with the help of the convolution and the pooling and finally, we can do the classification and finally, we can determine a particular class based on the softmax activation function that is the output layer.

So, one GAN is very popular that is the deep convolutional GAN. So, you can see the structure of the GAN. So, input is the code and after this we are doing repeatedly deconvolutions to generate the new images. So, new images can be generated by this technique that is the deep convolutional GAN. In this case we are considering number of deconvolutions. So, this DCGANs are able to generate high quality images when trained on restricted domain of images such as images of the bedrooms and this GAN that is the

DCGANs were the first GAN model to learn to generate high resolution images in a single shot.

So, it can generate the high resolution images. So, there are many applications of the DCGAN, but here briefly I have shown the structure of the DCGAN and with the help of the DCGAN you can generate these type of images all the images you can generate. So, for more details you can read the paper that was published in 2015. So, these are some applications of the DCGAN. In this case I am giving one example of the DCGAN. So, DCGANs demonstrated that GANs can learn a distributed representation that disentangles the concept of gender from the concept of the wearing of glass.

So, if you see the input we are considering the man and the lady and another one is the wearing of the glass. So, this the DCGAN can disentangles the concept of the gender from the concept of the wearing of the glass. After this another type of GAN is the cycle GAN. So, we can do the transformation of the images with the help of the cycle GAN. So, what are the applications of the cycle GAN? So, we can generate some images based on some target style. So, target style is given and based on this we can generate some images to preserve the structure of the original image.

So, we have to preserve the structure of the original image but the matching the target style. So, this concept you can read from the paper. So, it was published in 2017. So, I can show the applications of this cycle GAN . So, like this in this application I have shown the generation of the images from one image to another image. So, in this example I have shown one input image that is the horse image that is the real horse image and we are considering the generator G1.

So, first one is a generator G1 and this generator G1 is generating the generated samples that is the fake samples and we are getting the zebra. So, the this generator G1 is generating the fake sample that is the zebra from the horse and if you see the discriminator. So, this is the discriminator the input to the discriminator is the real zebra image and the fake zebra image. As already I have explained the discriminator and the generator they are competing with each other and because of the competition they are improving.

So, generator G1 can generate the real like zebra image. So, input to the discriminator is the real zebra image and the fake zebra image. So, you can see how we can do the transformation that is the one image can be transformed to another image and again we are considering another generator that is a generator 2. So, you can see we can do the reconstruction of the original image. So, G2 is generating the horse image and this is the reconstructed image and the first one is the original image.

So, we can determine the loss. So, one loss is the discriminator loss and another one is the reconstruction loss. So, this paper was published in 2017. So, you can see this paper. So, that is the image to image translation.

So, image to image generation. So, one is the target image and one is the input image. So, in this example the input image is the horse image and we are getting the zebra image from the horse image and after this again we are reconstructing the original image that is the horse image. Now, I will consider one application of the GAN that is for the image segmentation. So, we are considering the biomedical image segmentation problem and this application is this glaucoma detection that glaucoma is the eye condition that damage the optic nerve. This damage is caused by abnormal high pressure in the eye. So, glaucoma is one of the leading cause of the blindness for the people over the age of 60 and it can occur at any age, but is more common in older adults.

So, for detection of the glaucoma. So, one important parameter is the optic disk. So, if you see here this is the optic disk. If you see this image that is the optic disk. So, that can be considered as a feature in the retinal fundus image. So, this is the retinal fundus image and you can see the optic disk and this optic disk can be considered for the detection of the glaucoma.

So, we can consider these are the features like the sense of the shape, the color, depth of the optic disk. So, these are the features we can consider for detection of the glaucoma. So, mainly we are considering the size of the optic disk and based on this I can estimate or I can determine whether the glaucoma is present or not. So, that is the objective of the segmentation. For the segmentation of the optical disk this model can be proposed. So, you can see one is the input fundus image that is the eye image and we can see the optic disk, we can do the pre-processing and the first image is the real image, this is the real image and second image is the generated image.

So, that is generated by the generator. So, discriminator can determine which one is the real and which one is the fake. So, based on this we can segment out the optical disk from the rest of the image. So, for this the generator network may be considered like this unit based network. So, generator network may be unet based network may be considered. So, what is unet that I will be explaining in my next class and this is the structure of the generator and for the discriminator we can consider simple convolutional neural network that is the CNN we can consider.

So, here you can see this is the convolutional neural network. So, we have number of convolutional layers and also the pooling the max pooling is considered. So, this type of structure we can consider for discriminator and for the training of these networks we have to define the loss function. So, maybe we can consider the binary cross entropy function BCE we can consider along with the dice loss. So, what is the dice loss I can show you, but this is the formula for the BCE that is the binary cross entropy we can consider.

So, ŷ that is nothing but the predicted value of y. The dice loss function gives the measure of similarity between the true value and the predicted value. So, you can see this is the formula for the dice one. So, one is the true value the true value is the y and ŷ is the predicted value. So, the loss function we can consider like this one is the BCE that is the binary cross entropy and another one is the dice loss we can consider.

So, based on this loss function we can train the GAN the adversarial network. So, the binary cross entropy loss function is used for the discriminator. So, this BCE we can consider for the training of the discriminator this ŷ actually the predicted value by the prediction model. So, y is the real value and ŷ is the predicted value. So, here I have shown the results of the segmentation you can see the fundus image and we are considering the ground tooth mask this is the predicted mask and the grayscale image with the disk contour.

So, this is one example that is the biomedical application example. In this class I briefly explained the concept of the GAN and also I have shown some applications of the GAN. In the GAN actually there are two networks one is the generator networks and another one is the discriminator networks. So, both are competing with each other and because of this competition they are improving. The goal of these two networks are directly opposite and they are competing with each other.

And after this I discussed the concept of the training of the generator and the discriminator. So, maybe I can consider the cross entropy function as a loss function for the training of the generator and the discriminator. So, this is the concept of the generative adversarial network and it has many applications and it is very popular network. So, let me stop here today. Thank you.