كلية الحوسبة والمعلوماتية |
College of Computing and Informatics
الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

Second Semester – 2021/2022

Course Code	DS660
Course Name	Deep learning
Assignment type	Project
Module	All Modules

Student ID	G200007615
Student Name	Abdulaziz M. Alqumayzi
CRN	21604

Student ID	G200002614
Student Name	Enad S. Alotaibi
CRN	21604

Deep Fakes - Generative Adversarial Network (GAN)

Enad Saud Alotaibi
G200002614@seu.edu.sa,
Saudi Electronic University,
Riyadh, Saudi Arabia

Abdulaziz Mohammed Alqumayzi
G200007615@seu.edu.sa,
Saudi Electronic University,
Riyadh, Saudi Arabia

ABSTRACT

Machine Learning (ML) become one of the hot topics of 21st century and has progressed a lot. Deep Learning is of ML technique that concerned with algorithm called as Neural Network (NN), which mimics the working of human brain in processing data, has plenty of applications in real world i.e., fraudulent transactions in banking industry to detection of diseases in medical industry. Along the bright side there's dark side as well, adding noise to data can fool the algorithms and results misclassification. Generative Adversarial Network (GAN) is one of the special kinds of NN. GAN is composed of two neural networks called the generator and discriminator, which keep pitting against each other, generator by name gives an idea to generate synthetic data and discriminator try to differentiate and label the image either it's synthetic data or real one. The process of generating the synthetic data is also called as deep fakes, machine learning technique in which the characteristics of one character such as face, and voice is tried to replicate on other character. Which recently have got a lot of attention as being the feature it's also the have dark impact on society. In this project we developed a GAN to generate fake images of celebrities and explored latent space, performed arithmetic operations between points which have meaningful effects on the generated data. Along this we also have used the pretrained model to predict the ages of the faces which are generated by the generator model.

Keywords: GAN; Deep Fakes; Neural Networks; Discriminator; Generator

1. Introduction

Artificial Intelligence is the branch of computer science which concerned to build machine smarts that capable of performing task that require human intelligence. Machine learning is the subfield of artificial intelligence which refers to a system's ability to acquire, and integrate knowledge through large-scale observations, and to improve, and extend itself by learning new knowledge rather than by being programmed with that knowledge [1]. Deep learning is Machine Learning technique concerned with algorithms inspired by function of human brain called artificial neural network. Which is made up of neurons and those neurons

communicate with each other to pass the message across the body. Similarly, neural network is composed of neurons which are inter-connected with each other and pass the data to other neurons of the layers. The error rate is observed, and which is being tried to minimize in each iteration known as back propagation.

Neural Networks sound like a weird combination of biology and math with a little Computer Science sprinkled in, but it has been influential innovation in the field of Deep learning and computer vision. an image for humans it's simple image and easy to understand but computers see images differently in the form of pixels. Each of the numbers in the image range is between 0 to 255 with 3 color channels RGB.

The neural network work is similar to the brain after being analyzing the picture we can easily decide on the picture of an object, but a computer cannot understand it with that much ease, it detects patterns among image and based upon those patterns it decides to which object it belongs.

Generative Adversarial Network is the type of neural network. Which introduced the concept of adversarial learning, is a machine learning technique which aims to fool machine learning model by providing deceptive data. Which enable researchers to create completely realistic computer-generated images of an object which is known as Deep Fakes.

GANs are generative models with implicit density estimation, part of unsupervised learning and are using two neural networks generator and discriminator. These two neural networks try to fool each other. The generator generates an image while the discriminator tries to distinguish between real and fake images. The generator can be interpreted as artist where discriminator as an art critic. During the training phase the generator progressively become better at creating an image that look like real and the discriminator become better at differentiating between real and fake. The discriminator employs a feedback mechanism to help the generator in generating more realistic images. This process reaches to an equilibrium when the discriminator can no longer differentiate between real and fake. This mechanism has a lot of benefits to generate images that can be useful in increasing the low amount of data. Training machine learning models on those images

can help in better training and less venerable to spoofing, helping in creating artwork and colorization of images.

2. Body section

Literature review

Goodfellow, et al. [2] introduced a new framework for estimating generative models using an adversarial process in which two models are simultaneously trained: a generative model G that captures the data distribution and a discriminative model D that estimates the likelihood that a sample came from the training data rather than G. The generative model is fought against an opponent in the proposed adversarial nets framework: a discriminative model that learns to distinguish whether a sample comes from the model or data distribution. The generative model can be compared to a group of counterfeiters attempting to create false currency it without being detected, the discriminative model is comparable to a police officer attempting to discover counterfeit money. In this game, the competition forces both sides to develop their strategies until the fakes are indistinguishable from the real data. Their model is being trained on datasets including MNIST, Toronto Face Database (TFD) and CIFAR-10. In comparison to earlier modelling frameworks, the new framework has both advantages and downsides. The disadvantages are that there is no explicit representation of the data z, and that D and G will be well synchronized during training (in particular, G must not be trained too much without updating D, in order to avoid "the Helvetica scenario," in which G collapses too many variables' values to the same value). must have enough diversity to model Page | 7 of x. Adversarial networks have the virtue of being able to express exceedingly sharp, even degenerate distributions. Adding c as an input to both G and D yields a conditional generative model, according to this approach. Also, by training an auxiliary network to predict z given x, learnt approximation inference may be achieved. [2]

According to Kwak and Zhang [3, picture formation remains a basic difficulty in artificial intelligence, despite numerous proposed strategies to solve it. The author of this research introduces a model known as a

composite generative adversarial network (CGAN) that detangles intricate components of pictures with several generators, each of which creates a portion of the imageCGAN is a GAN variant that use numerous generators in conjunction with a recurrent neural network (RNN). CGAN generators vary from GAN generators in that they have more alpha channels in the output. After that, the photos are blended progressively using alpha blending to create the final image. By calculating the common components of pictures and constructing realistic examples, CGANS assigns duties to each generator. The CelebA and Oxfold 102 Flowers datasets, as well as the pororo comic film, are being used to train CGANs. Ant aliased resizing of all pictures to 64 × 64 pixels. CGANS successfully created picture parts by part using three generators, resulting in final images with backgrounds, faces, and hair sections, respectively. The third generator initially failed to produce meaningful images, but after adding alpha loss, the problem is alleviated, and the images become less hazy. [3]

According to Radford, et al. [4], supervised learning using convolutional networks (CNNs) is more popular than unsupervised learning with CNNs, which gets less attention. However, they seek to help bridge the gap by creating a type of CNNs dubbed DCNN. They apply the trained discriminator for picture classification tasks, exhibiting competitive performance with other unsupervised algorithms, then experimentally depict it to demonstrate the filters inside this study [4Previous research has shown that supervised CNN training on big picture datasets produces highly effective learning. Object detectors are also learned by supervised CNNs trained on scene categorization. However, according to the findings, an unsupervised DCGAN trained on a big picture dataset may also learn a hierarchy of useful characteristics. Using backpropagation with a guide [4]

By replacing MLP with Convolutional neural networks (CNN) and deleting pooling layers, Zhang et al. [5] research explores how to increase the quality of produced face pictures with GAN. The original GAN network for face image creation is unstable in terms of image quality created by generators during training and is unable to get high-quality generators. The fact that generators and discriminators utilise the same back-

propagation network is the root of the problem. This research presents strategies for modifying the original GAN design to overcome this issue. They use DCGAN architecture to train the model. First, instead of using the deterministic space pooling function, the entire CNN employs stride convolution. They apply the approach of network learning its own spatial down-sampling in the generating network, allowing the discriminating network to learn its own spatial up-sampling. The fully linked layer is then removed. The studies carried out using the LFW and CelebA face datasets and indicate that their strategy is successful. [5]

2.1 Data

CelebFaces Attributes Dataset (CelebA) is a large-scale face attributes dataset with more than 200K celebrity images, each image is 178 pixel in width and 218 pixel in height each with 40 attribute annotations for what appears in given photos, such as glasses, face shape, hats, hair type, etc. The images in this dataset cover large pose variations and background clutter. the authors provide a version of each photo centered on the face and cropped to the portrait with varying sizes around 150 pixels wide and 200 pixels tall CelebA has large diversities, large quantities, and rich annotations, including

- 10,177 number of identities,
- 202,599 number of face images, and
- 5 landmark locations, 40 binary attributes annotations per image.

The dataset can be employed as the training and test sets for the following computer vision tasks: face attribute recognition, face recognition, face detection, landmark (or facial part) localization, and face editing & synthesis.

2.2 Methods

The images of the dataset are loaded by using the pillow library as an array of pixels. The images are around 200k in a dataset, so we loaded around 50k for training our model.

As we are only interested among the faces in an image. So, we adopted an approach to crop the image and only obtain the images of faces. After cropping the images are then resized into 64x64 in size. which is then converted into numpy array and used those arrays of images as our training examples to train the model. The training data is size 60000 out of 200000 images.

Generative Adversarial network (GAN) is the combination of two neural networks generative and discriminative networks. The generative network work as an artist which generate arts and discriminative model work as art critic which distinguish between real and fake art generated by the generative and real training image.

The discriminator in a GAN is a classifier. It tries to distinguish real data from the data created by the generator. The data for discriminator come from two sources real data as a training example from dataset and fake data generated by the generator. We defined a discriminator which simply uses convolution layers and for every layer of the network, we perform a convolution, then we perform batch normalization to make network faster and more accurate and leakyRelu is performed. Discriminator model is compiled using binary cross-entropy as a loss function as it just distinguish among two images and Adam as an optimizer is used.

The defined discriminative model is taking input a color image of 64x64 and outputs a binary prediction as to whether the image is real (class=1) or fake (class=0). Using the best practices of GAN design it's implemented as a modest convolution neural network, by using LeakyRelu as an activation function with slope of 0.2 and 2x2 stride to down-sample the data. Final output layer of the discriminator model is using sigmoid as an activation function as its prediction is binary in nature 0 or 1. So by keeping that in mind the model is compiled by using binary crossentropy loss and Adam as an optimizer with learning rate of 0.0002 and momentum as 0.5. It tries to distinguish among both the images which one is real and which one is fake.

The architecture of our discriminator model is:

Layers	Output	Param
conv2d (Conv2D)	(None, 64, 64, 128)	9728

leaky_re_lu (LeakyReLU)	(None, 64, 64, 128)	0
conv2d_1 (Conv2D)	(None, 32, 32, 128)	409728
leaky_re_lu_1 (LeakyReLU)	(None, 32, 32, 128)	0
conv2d_2 (Conv2D)	(None, 16, 16, 128)	409728
leaky_re_lu_2 (LeakyReLU)	(None, 16, 16, 128)	0
conv2d_3 (Conv2D)	(None, 8, 8, 128)	409728
leaky_re_lu_3 (LeakyReLU)	(None, 8, 8, 128)	0
conv2d_4 (Conv2D)	(None, 4, 4, 128)	409728
leaky_re_lu_4 (LeakyReLU)	(None, 4, 4, 128)	0
flatten (Flatten)	(None, 2048)	0
dropout (Dropout)	(None, 2048)	0
dense (Dense)	(None, 1)	2049

The generative model tries to generate images from noise data and keep improving itself through the feedback of discriminator. So, the generator model is taking an input a point from latent space and outputs an image of 64x64 in size. The output layer of our generator model is consisting of the sigmoid as an activation function the other, we've tried was Tanh, and from the result, the sigmoid is used as an activation function of the final as the result of sigmoid function has much clarity in the images.

Our generative model is developed by using a fully connected layers to interpret a point from latent space. The output image of generator in up sampled by the discriminator model by using transpose convolution layers.

The architecture of our generator model is:

Layers	Output	Param
dense_1 (Dense)	(None, 8192)	499712
reshape (Reshape)	(None, 4, 4, 512)	0
conv2d_transpose (Conv2DTranspose)	(None, 8, 8, 256)	2097408
leaky_re_lu_5 (LeakyReLU)	(None, 8, 8, 256)	0
batch_normalization (BatchNormalization)	(None, 8, 8, 256)	1024

conv2d_transpose_1 (Conv2DTranspose)	(None, 16, 16, 128)	524416
leaky_re_lu_6 (LeakyReLU)	(None, 16, 16, 128)	0
batch_normalization_1 (BatchNormalization)	(None, 16, 16, 128)	512
conv2d_transpose_2 (Conv2DTranspose)	(None, 32, 32, 64)	131136
leaky_re_lu_7 (LeakyReLU)	(None, 32, 32, 64)	0
batch_normalization_2 (BatchNormalization)	(None, 32, 32, 64)	256
conv2d_transpose_3 (Conv2DTranspose)	(None, 64, 64, 3)	3075

Once we defined both the components of GAN, the GAN model combines both the generator and discriminator model in to one larger model. This larger model will be used to adjust the generator model weights, using the output and error calculated by the discriminator model. The discriminator model training is set to False and ensured that only the weights of generator are updated.

This larger GAN model takes as input a point in the latent space, uses the generator model to generate an image, which is fed as input to the discriminator model, then output or classified as real or fake.

The architecture of complete GAN model is:

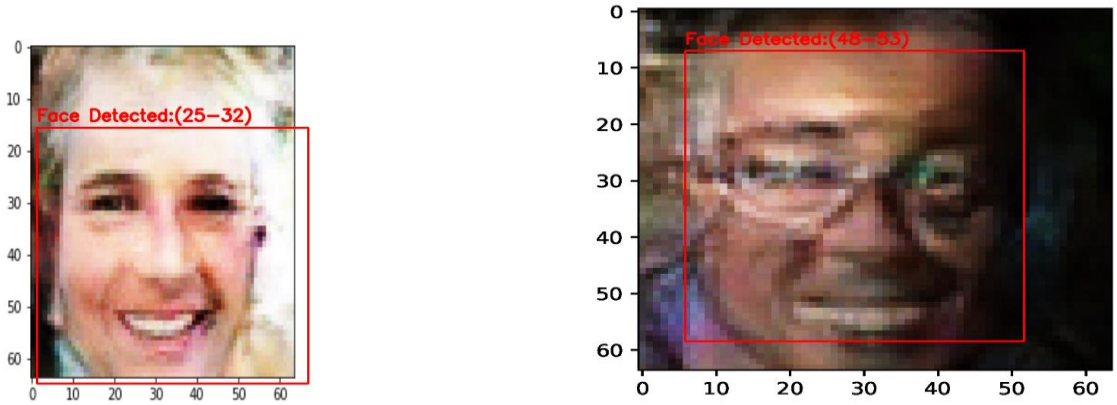
Layers	Output	Param
sequential_1 (Sequential)	(None, 64, 64, 3)	3257539
sequential (Sequential)	(None, 1)	1650689

Finally, after defining GAN, it's time to train the model over the training data. The training of GAN is difficult since GAN contain two separate network and its convergence is hard to identify.

During the training of the GAN model, we keep on saving the models in h5 file so we can use later on. These trained models are later utilized to generate images and explore the latent space for generated images. We created an interpolation path between two points in the latent space and generate faces along this path. We defined a function which provides the series of linearly interpolated vectors between two points in latent space. We can then generate two points in the latent space, perform the interpolation, then generate an image for each interpolated vector. The below image demonstrates the results.

Performed some vector arithmetic to explore the latent space on the generated images. Large number of faces are generated and their corresponding latent spaces. The latent spaces of the images can be accessed by using the index point for manipulation, 36 random faces are generated and plotted through which we selected the smiling women faces by using the index, similarly, for neutral women and neutral men from those images. Then average point for each of the categories are calculated and use those average point to perform vector arithmetic in latent space and after vector arithmetic we got a face of a men as end result.

We also used the pretrained network to check whether the model can predict the ages on those generated faces or it fails. The pretrained model of caffe model called age net is used and the above arithmetic operations generated image and some images are provided to the model and model successfully predicted the ages of those persons. Below attached images are the outputs of the pretrained network predictions.



2.3 Analysis

The data we have for training consists of images of more than 200K celebrity images, each image is 178 pixels in width and 218 pixels in height. Due to hardware limitations, we must reduce the size of the images, so our system doesn't crash during the training. The Images are reduced to size 64 pixels in width and 64 pixels in height and color frame of 3 RGB. After that the images are normalized between range 0 to 1 by dividing on the max pixel 255.

Most of the images contained a lot of background of as well so we cropped the images by manually defining the coordinates of the faces. The other technique such as pretrained models can also be used for face coordinates detection like MTCNN.

The generator generates an image from the noise which keep on improving itself from the feedback of the discriminator. The initial noise of 100 was defined but the images generated were blurry, so we keep reducing the noise and find out that 60 was good noise where model generated better images than any other one.

2.4 Results

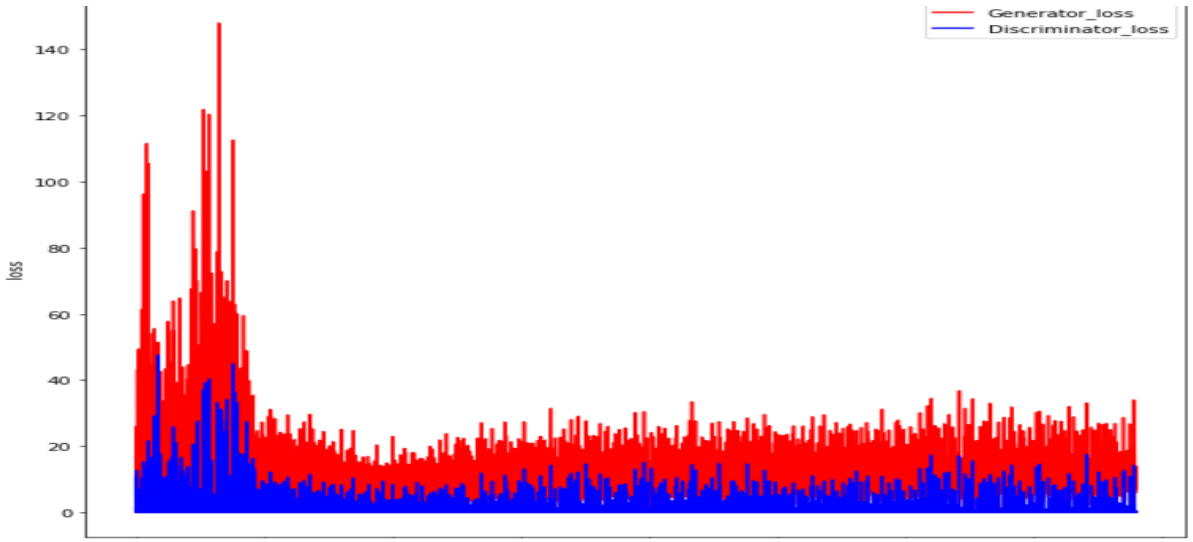
The GAN model is the combination of Discriminator and Generator. During training we monitored the loss of both the models. The loss function is below, this loss is tried to be minimized by the generator and maximized by the discriminator.

$$\min_G \max_D V(D,G)$$

$$V(D,G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z))]$$

D(x) is the discriminator estimate that data real data is real. D(G(z)) is the discriminator estimate that fake data is real.

The below graph represents the loss of the model. It shows that in start the model make large fluctuations but after that it keep on moving in a same range.



3. Conclusion

To sum up, we built a GAN to generate deepfakes of the CelebA dataset and performed human test how well they can detect fake and real images and find out that its real distinguishable by human. Which can be further improved by tunning model and changing hyperparameters. Adding more layers, making deeper model can also help in generating more realistic images but will take more and more time for training with deepening the model. The model doesn't know when to stop training so proper evaluation metric can help during training of model and let him know when to stop training further, that's one of the reasons of not having good GAN model. Furthermore, providing more data to the model will help in generating more realistic images.

References

[1] B.P. Woolf. (2009). Building Intelligent Interactive Tutors. 2nd ed.

[2] I. Goodfellow et al., "Generative adversarial nets," Advances in neural information processing systems, vol. 27, 2014.

[3] H. Kwak and B.-T. Zhang, "Generating images part by part with composite generative adversarial networks," arXiv preprint arXiv:1607.05387, 2016.

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[5] Z. Zhang, X. Pan, S. Jiang, and P. Zhao, "High-quality face image generation based on generative adversarial networks," Journal of Visual Communication and Image Representation, vol. 71, p. 102719, 2020.

[6] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks" arXiv preprint arXiv:1604.02878, 2016

[7] Sorin, V., Barash, Y., Konen, E., & Klang, E. (2020). Creating artificial images for radiology applications using generative adversarial networks (GANs)–a systematic review. *Academic radiology*, *27*(8), 1175-1185.

[8] Striuk, O. L. E. K. S. A. N. D. R., & Kondratenko, Y. U. R. I. Y. (2021). Generative Adversarial Neural Networks and Deep Learning: Successful Cases and Advanced Approaches. *Int. J. Comput*, *20*, 339-349.

[9] Shen, T., Liu, R., Bai, J., & Li, Z. (2018). "deep fakes" using generative adversarial networks (gan).

[10] Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, *1*(1), 100004.