# Image To Image Translation Using Generative Adversarial Network

**Boddu Manoj, Boda Bhagya Rishiroop**

**Abstract:** With the increasing potential of Deep fakes in the field of computer vision has made many toilsome tasks effortless. In this paper, we will be discussing one such task. We will demonstrate how we can generate a real like images that don't even exist in the real world. We will be implementing this with the DCGAN (Deep Convolution GAN) algorithm which is an extended network of GAN (Generative Adversarial Network). Although there are other algorithms available such as encoder and decoder DCGAN has demonstrated to be an incredible accomplishment in generating better quality images. Also, we have talked about the conceptual parts of GAN and examined our technique to make a DCGAN model. For training purposes, we will be using the CelebA dataset which consists of more than 200k faces of celebrities.

**Index Terms**: computer vision, DCGAN, deep learning, Deepfakes, generative adversarial networks, GAN.

## 1 INTRODUCTION

Image-to-Image altering such as superimposing a new face on an existing image, altering the body parts such as eyes, hair, ears, etc. Creating such images especially with the help of the manual method is an onerous task in obtaining better results. One can easily detect the modified image effortlessly with his naked eye. There are many advantages of creating such images in the field of healthcare, film industry, educational media, games, fashion technology, etc. The film industry has advantages of creating a forgery of actors who lost their lives in between shoots, Gaming industry can create their players face, recreating the photos of the deceased, etc. To achieve these manually is a toilsome work. With the growing popularity in artificial Intelligence, deep fakes have emerged which have been very popular among recent years giving rise to completely new technology in the field of computer vision. There are many categories in the deep fake such as superimposing a person face on another person, creating a whole new realistic image, video, and audio which doesn't even exist in reality, predicting the older age, etc. In the year 2017, a user named Deepfake posted a video on Reddit which led to an increase in popularity of the deep fake. There have been many software and android applications developed to create a deep fake by superimposing their favorite celebrity such apps were Zao, Faceapp, etc. These apps were built on deep fake algorithms. Such deep fake algorithms are autoencoder and decoder algorithm and GAN (Generative Adversarial Network). In this paper, we are demonstrating the extended network of GAN which is DCGAN (Deep Convolution GAN). Our work in this paper is Generating a realistic image from the Celeba dataset and identifying the loss in generator and discriminator.

## 2 RELATED WORK

The common approach in creating a fake image (which doesn't even exist) of people, pets, cartoons, objects are with the Generative Adversarial Network (GAN) only. This Generative Adversarial Network was first introduced in 2014 by Ian Goodfellow. GAN works on image-to-image translation. GAN consists of 2 models that compete with each other. The two models are generative model G and a discriminative model D the former one captures the data distribution while the later one estimates the probability of sample data. The results show by the GAN in image generation, transfer of images, and some other related tasks are quite impressive. Many variants of GAN have emerged in recent years such as DCGAN, StyleGAN, StyleGAN2 BigGAN, ProGAN, StarGAN, CycleGAN, GauGAN, etc.

## 3 METHODOLOGY

The framework we used in this paper is DCGAN.

### 3.1 Generative Adversarial Network

The abbreviation of GAN is Generative Adversarial Network, which is a combination of 2 neural networks. The word 'Generative' means capable of creating it's own whereas 'Adversarial' refers to two things or sides that oppose each other. So as per the definition, GAN's are responsible for generating different data and tries to compete with each other. GANs consists of two networks, a Generator, and a Discriminator. The Two neural networks play an Adversarial game where the Generator strives to ruse the discriminator by inducing data resembling those given in the training set. The role of the Discriminator is to discern counterfeit data from the real data. The generator model produces data such as audio, video, images from a given arbitrary noise and then assimilate how to generate realistic data. In our case, we will be given training data of images to generate real-like looking images but in fact, they don't exist.

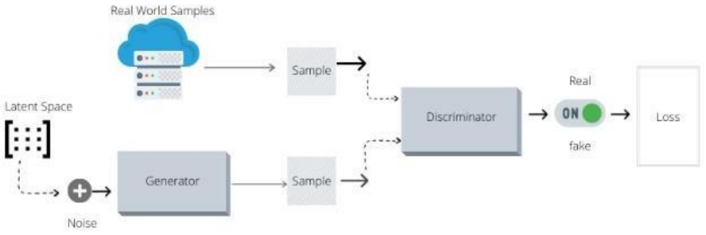

***Figure 1*** *Architecture of GAN*

Random noise is fed into the generator which generates a counterfeit image based on the given input features. Later on the both the real and generated images are fed into the

- *Boddu Manoj, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.*
- *Boda Bhagya Rishiroop, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.*

discriminator which differentiates counterfeit images from the real images. The Discriminator usually has a probability function D(x) to identify how much the generated images are close to the real images. The values lie between 0 to 1 where 1 means close to real images whereas 0 means converse of that.

Let's dig a wee deeper and apprehend how it functions numerically. The Generator and Discriminator have a formula for Minimax function as shown below:

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

### 3.1 Deep Convolution GAN (DCGAN)

DCGAN is an extention to GAN. It is an updated and expanded network compared to GAN. DCGAN uses layers like convolutional and convolutional-transpose in discriminator and generator, respectively. The discriminator is again comprised of various layers like 1) strided convolution layers, 2) batch norm layers and 3) LeakyReLU activations.

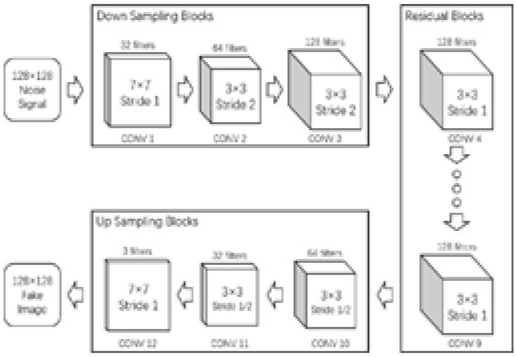

*Figure 2 block diagram of a Generator*

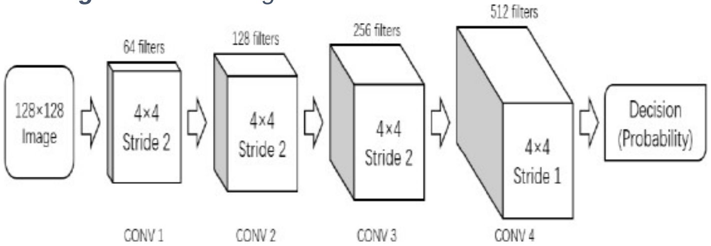

*FIGURE 3 BLOCK DIAGRAM OF A DISCRIMINATOR*

## 4 IMPLEMENTATION

For implemention of this work we used python PyTorch and the data set is collected from CelebA which contains more than 200 thousands of celebrity faces. We used a group size of 128 for this and 64x64 for spatial size of images. For training purpose, we used a dataset contains of 202599 images. The implementation starts with the initialization of weights with normal distribution method of mean=0 and a standard deviation = 0.02. We will be starting with generator and Discriminator implementation. A Discriminator is a binary classified network that takes image as input and then outputs the respective scalar probability that tells weather the images are real or fake with a decimal score of 0 to 1. The Discriminator process images through series of different layers namely 1) Conv2d 2) BatcNorm2d, and 3) LeakyReLU. These gives outputs for final probability through sigmoid activation function.

Some of the terms mentioned in the formula are as follows:

•Loss_D – Loss in Discriminator calculated as the sum of losses for the all real and fake images (log(D(x))+log(D(G)))).

•Loss_G – Loss in Generator calculated as log(D(z)))

•D(x) – The mean Discriminator output for all real images.

• D(G(z)) - The mean Generator outputs for all fake images.

After completing all the epochs our model will be trained and fake images will be displayed. The results will be better on the number of epochs.

5 SYSTEM SPECIFICATIONS

- Operating System – Windows 10
- System – Predator PH315-51
- Processor – Inter® i7-8750H
- Ram – 16.0 GB
- System type – 64bit operating system
- GPU – NVIDIA GeForce GTX 1060

## 6 RESULTS

Total Training Time – 1 hour 15 Mins

```
Starting Training Loop...
[illegible]
```

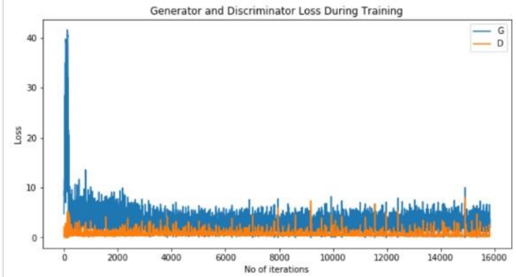

*Figure 4 Loss values generated in between the iteration*

For better understanding, we have displayed the loss values of each function in between the training of the model.

*Figure 5 Graph of Generator and Discriminator loss during training*

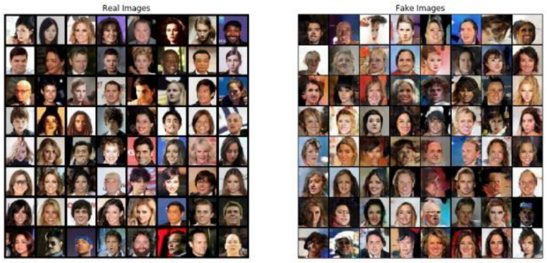

*__Figure 6__ Display of real images and fake images*

The result gets better and better with an increase in training time i.e increase in epochs and number of iterations.

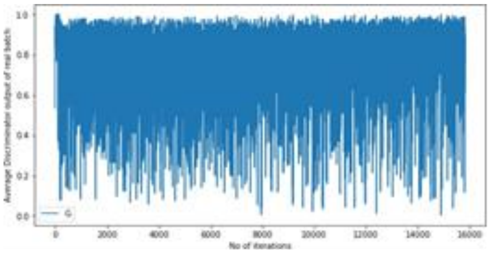

*__Figure 7__ Average Discriminator output of real images*

The value should be close to 1 then theoretically it should converge to 0.5 when G gets better.

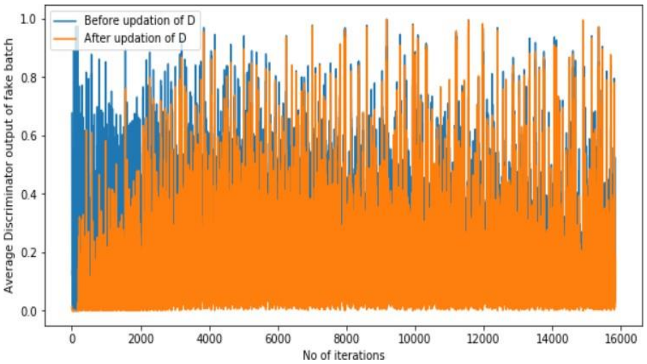

*__Figure 8__ Average Discriminator output of fake images*

These values should start near 0 and converge to 0.5 as G gets better.

## 7 FUTURE WORK

Some of the improvements further we can do is improving the quality of the generated images. And some major drawbacks are it consumes too much time to train the model even with GPU. We can further optimize the model to decrease the training time. Latterly, GANs have been displayed numerous usage in videos as well furthermore they are expanding its usage to make 3D model objects using images. In the near future, there will be an expanding number of GANs in many streams.

## 8 CONCLUSION

We demonstrated the creation of fake human images using a celebrity dataset. By integrating a clean celebrity dataset with GAN, we are able to build a model that can give rise to realistic human images that don't even exist in the real world. Similarly, with the images, GAN's can also be implemented with videos, music, etc. By following the same approach which we have created with images. There's a lot we can create with GAN's. Proper research in GAN's can be helpful in creating more efficient ways to generate the desired outputs. However, with growing technology in creating deepfakes has lead impossible to differentiate between the real and fake images. Proper detection methods need to be implemented for safe practices of deep fakes.

We desire that our work would prod more research in Deep Convolution GAN in generating realistic celebrity images and eventually it may help both neophyte and experts to create a new realistic fake image of humans. Furthermore, the implementation can be extended to other items such as animals, objects, nature scenarios, cartoons, etc

## ACKNOWLEDGMENT

This Project provides us an opportunity to have a greater understanding of the subject and explore beyond. This work was supported in part by Vellore Institute of Technology.

## REFERENCES

[1] Carlini, Nicholas & Farid, Hany. (2020). Evading Deepfake-Image Detectors with White- and Black-Box Attacks.

[2] Li, Muyang & Lin, Ji & Ding, Yaoyao & Liu, Zhijian & Zhu, Jun-Yan & Han, Song. (2020). GAN Compression: Efficient Architectures for Interactive Conditional GANs.

[3] Frank, Joel & Eisenhofer, Thorsten & Schönherr, Lea & Fischer, Asja & Kolossa, Dorothea & Holz, Thorsten. (2020). Leveraging Frequency Analysis for Deep Fake Image Recognition.

[4] Lyu, Siwei. (2020). DeepFake Detection: Current Challenges and Next Steps.

[5] Nguyen, Thanh & Nguyen, Cuong & Nguyen, Tien & Nguyen, Duc & Nahavandi, Saeid. (2019). Deep Learning for Deepfakes Creation and Detection.

[6] Goodfellow, Ian & Pouget-Abadie, Jean & Mirza, Mehdi & Xu, Bing & Warde- Farley, David & Ozair, Sherjil & Courville, Aaron & Bengio, Y.. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems. 3.Radford, Alec & Metz, Luke & Chintala, Soumith. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.

[7] Mobini, Majid & Ghaderi, Foad. (2020). StarGAN Based Facial Expression Transfer for Anime Characters. 1-5.
10.1109/CSICC49403.2020.9050061.

[8] Maksutov, Artem & Morozov, Viacheslav & Lavrenov, Aleksander & Smirnov, Alexander. (2020). Methods of Deepfake Detection Based on Machine Learning. 408-411.
10.1109/EIConRus49466.2020.9039057.

[9] Yang, Xin & Li, Yuezun & Lyu, Siwei. (2019). Exposing Deep Fakes Using Inconsistent Head Poses. 8261-8265. 10.1109/ICASSP.2019.8683164.

[10] Li, Yuezun & Chang, Ming-Ching & Lyu, Siwei. (2018). In

Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. 1-7.
10.1109/WIFS.2018.8630787.

[11] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2015). Deep Residual Learning for Image Recognition. 7.

[12] Matern, Falko & Riess, Christian & Stamminger, Marc. (2019). Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations. 83-92.
10.1109/WACVW.2019.00020.

[13] Li, Yuezun & Lyu, Siwei. (2018). Exposing DeepFake Videos By Detecting Face Warping Artifacts.

[14] Guera, David & Delp, Edward. (2018). Deepfake Video Detection Using Recurrent Neural Networks. 1-6.
10.1109/AVSS.2018.8639163.

[15] Guera, David & Delp, Edward. (2018). Deepfake Video Detection Using Recurrent Neural Networks. 1-6.
10.1109/AVSS.2018.8639163.

[16] Jin, Yanghua & Zhang, Jiakai & Li, Minjun & Tian, Yingtao & Zhu, Huachun & Fang, Zhihao. (2017). Towards the Automatic Anime Characters Creation with Generative Adversarial Networks.

[17] Hamada, Koichi & Tachibana, Kentaro & Li, Tianqi & Honda, Hiroto & Uchida, Yusuke. (2018). Full-body High-resolution Anime Generation with Progressive Structure-conditional Generative Adversarial Networks.

[18] Pantserev, Konstantin. (2020). The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability. 10.1007/978-3-030-35746-7_3.

[19] Zhang, Zhifei & Song, Yang & Qi, Hairong. (2017). Age Progression/Regression by Conditional Adversarial Autoencoder.

[20] Chapagain, Ashutosh. (2019). DCGAN--Image Generation. 10.13140/RG.2.2.23087.79523.

[21] Westerlund, Mika. (2019). The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review. 9. 39-52. 10.22215/timreview/1282.

[22] Koopman, Marissa & Macarulla Rodriguez, Andrea & Geradts, Zeno. (2018). Detection of Deepfake Video Manipulation.

[23] Korshunov, Pavel & Marcel, Sébastien. (2018). DeepFakes: a New Threat to Face Recognition? Assessment and Detection.

[24] Neekhara, Paarth & Hussain, Shehzeen & Jere, Malhar & Koushanfar, Farinaz & Mcauley, Julian. (2020). Adversarial Deepfakes: Evaluating Vulnerability of Deepfake Detectors to Adversarial Examples. 10.13140/RG.2.2.26227.48168.

[25] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision, pages 694–711. Springer, 2016.

[26] Xiaolong Wang and Abhinav Gupta. Generative image modeling using style and structure adversarial networks. In European Conference on Computer Vision, pages 318–335. Springer, 2016.

[27] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593, 2017.

[28] Rafael C Gonzalez and Richard E Woods. Image processing. Digital image processing, 2, 2007.

[29] Guo, Y., Jiao, L., Wang, S., Wang, S., and Liu, F. (2017). Fuzzy sparse autoencoder framework for single image per person face recognition. IEEE Transactions on Cybernetics, 48(8), 2402- 2415.

[30] Yang, W., Hui, C., Chen, Z., Xue, J. H., and Liao, Q. (2019). FV-GAN: Finger vein representation using generative adversarial networks. IEEE Transactions on Information Forensics and Security, 14(9), 2512-2524.

[31] Liu, F., Jiao, L., and Tang, X. (2019). Task-oriented GAN for PolSAR image classi_cation and clustering. IEEE transactions on Neural Networks and Learning Systems, 30(9), 2707-2719.

[32] Cao, J., Hu, Y., Yu, B., He, R., and Sun, Z. (2019). 3D aided duet GANs for multi-view face image synthesis. IEEE Transactions on Information Forensics and Security, 14(8), 2028-2042.

[33] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. N. (2019). StackGAN++: Realistic image synthesis with stacked generative adversarial networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(8), 1947-1962.

[34] Chesney, R., and Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign A_airs, 98, 147.