Creating Lifelike Human Faces with Deep Convolutional GANs

Dr. David Raj Micheal
Division of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
davidraj.micheal@vit.ac.in

Anusha Puppala
Division of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
puppala.anusha2023@vitstudent.ac.in

***Abstract*—This project explores the use of Deep Convolutional Generative Adversarial Networks (DCGANs), an advanced deep learning architecture designed to generate realistic human faces from random noise. The GANs' adversarial training framework involves two neural networks, the generator and the discriminator, competing to create images that mimic real human faces. As training progresses, the generator improves its ability to create highly realistic images, while the discriminator becomes better at identifying subtle discrepancies between real and generated images. The CelebA dataset, consisting of over 200,000 celebrity images with rich annotations, is used as the training ground. The DCGAN architecture utilizes convolutional layers for feature extraction, capturing intricate patterns,textures, and structures inherent in human faces. The generative process is guided by adversarial loss functions, which evaluate the quality of the images produced. With a focus on human face synthesis, the research seeks to demonstrate the strength and adaptability of GANs in the creation of synthetic images. This talent has consequences for numerous sectors, such as identity protection, gaming, and entertainment. The study also investigates the subtleties of training GANs, such as minimizing mode collapse, improving image quality, and stabilizing the training procedure.**

***Index Terms*—DCGAN, GAN, Image generation, Human face synthesis.**

I. Introduction

Significant progress has been made in the field of generative modelling, especially since the introduction of Generative Adversarial Networks (GANs). Deep Convolutional Generative Adversarial Networks (DCGANs), in particular, have shown to be incredibly successful in picture generation tasks, especially when it comes to creating human faces. The goal of this research is to create realistic human faces out of random noise by employing DCGANs. A generator and a discriminator are two competing neural networks that are a part of the DCGAN design. The discriminator compares the pictures produced by the generator, which imitates real human faces, to actual photos in order to establish the veracity of the images. Both networks iteratively develop through an adversarial process, allowing the generator to generate human faces with excellent quality and photorealistic appearance. Virtual avatars, privacy-preserving technologies, and synthetic dataset generation can all benefit from this capacity to produce realistic faces.

II. Problem Definition

Because human features are so complex, it is challenging to train a model that can generate realistic, varied, and high-quality human faces from random noise. A key aspect of this task is ensuring that the generated faces are different while still having the structure, texture, and traits of real human faces.

III. Objective

The primary goal of this research is to create realistic, high-quality human face photographs by designing and implementing a Deep Convolutional Generative Adversarial Network (DCGAN). The objectives are specifically to:

• Provide a generator that uses random noise to produce aesthetically pleasing face images.

• Use a discriminator to tell the difference between faces that are real and ones that are made up.

• To increase the realism of the generated faces, train both models in an adversarial fashion.

• Measure the loss during training and visually analyze the output images to validate and assess the DCGAN's performance.

• Examine the possible uses of face generation, such as the development of virtual avatars, artificial datasets, and privacy-preserving technology.

IV. Dataset

We use the CelebA dataset, which contains over 200,000 celebrity images with rich annotations. This dataset provides a diverse set of human faces, which is ideal for training generative models to learn realistic human face structures.

V. Literature Review

Kapalavai et al. [1] focuses on generating realistic high-resolution human face images using deep learning models. It combines two types of Generative Adversarial Networks (GANs): DCGAN (Deep Convolutional GAN)and ESRGAN(Enhanced Super-Resolution GAN). DCGAN generates images from random noise by training a generator to produce convincing images and a discriminator to identify whether an image is real or fake. The CelebA dataset is used to train the DCGAN model, and the Structural Similarity Index (SSIM) is employed to assess image quality. ESRGAN is then applied to enhance the resolution of the generated images. The results show that combining DCGAN with ESRGAN yields high-quality, detailed human face images.

Shariff et al. [2] explores how GANs (Generative Adversarial Networks) are utilized to generate artificial human faces. The study implemented a DCGAN (Deep Convolutional GAN) to train a model using the CelebA dataset, which contains celebrity face images. The GAN architecture comprises a generator network that creates fake images and a discriminator network that classifies them as real or fake. The generated images are evaluated using the Structural Similarity Index (SSIM) to measure their quality against real images. The results show that while the generated images are not perfect, they progressively improve in quality as the model trains, with a maximum SSIM score of 0.34. The study concludes that this method could be used for creating high-quality images in various applications and suggests future work in generating higher-resolution images using more advanced GAN architectures.

Nekmaiche et al. [3] presents a study on using Deep Convolutional Generative Adversarial Networks (DCGANs) for generating images of human faces. It discusses the challenges of evaluating the performance of GAN models and introduces a new hybrid evaluation metric called Measuring the Quality of the Features of an Image (MEQFI). The authors demonstrate that DCGANs can produce high-quality face images when trained on well-curated datasets and highlight the importance of human evaluation in assessing generated images.The performance of GANs is often evaluated using various metrics such as Inception Scores and Frechet Inception Distance, but these can be complemented with human evaluations for better accuracy.The document reviews existing GAN variants aimed at enhancing image generation quality through architectural and loss function improvements.The experiments suggested that using around 100 epochs was preferable for training, as increasing the number of epochs beyond this did not yield significant improvements in image quality.Experimental results indicate that the generated images improve over epochs, with higher quality observed in images created from the CelebA-HQ dataset, underscoring the importance of dataset quality in model training. [4] It explores the optimization of a loss function for generating realistic human faces through deep convolutional generative adversarial networks (DCGANs). By training the model with the CelebA dataset, the authors focus on improving the distinct features of human faces as well as the ability to identify partially visible images. The study outlines a systematic approach involving preprocessing of images, training of generator and discriminator models, and iterative loss function optimization, ultimately demonstrating a reduction in generator loss and failure in discriminator recognition, resulting in the generation of high-fidelity synthetic facial im ages.A significant challenge addressed is achieving differentiation between unique human facial features, particularly in cases of identical twins.

Generating Human Face with DCGAN and GAN explores the use of Generative Adversarial Networks (GANs) and Deep Convolutional GANs (DCGANs) for human face generation [5]. GANs, known for their ability to generate data through adversarial training, consist of a generator and a discriminator network that work together to create realistic images. DC GANs build on this by using convolutional layers, improving the model's ability to capture intricate facial features and produce higher-quality images. The study uses the CelebA dataset and evaluates the models using metrics like Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) to assess image quality.Key topics include pre-processing techniques (noise removal, image normalization, resizing), training methods, and challenges such as mode collapse and training instability. DCGANs show superior performance compared to traditional GANs, offering better stability and more realistic face generation. The paper also addresses future advancements in GANs, including tackling mode collapse and exploring multi-modal generation techniques, with an emphasis on practical applications like virtual avatars, facial recognition, and digital art.

Yi et al. [6] introduces a novel approach for transforming face photos into artistic portrait drawings using a deep learning model called APDrawingGAN. The authors highlight the unique challenges associated with artistic portrait drawing, which is highly abstract and consists of sparse, continuous graphical elements like lines. Existing methods fail to generate high-quality artistic portraits due to their inability to capture these nuances.The paper introduces a Hierarchical GAN model with global and local networks, allowing the model to learn different drawing strategies for various facial regions. The authors propose a novel loss function based on distance transforms, which helps the model generate strokes that align more closely with the abstract style of artistic portrait drawings. A dataset of 140 high-resolution face photos paired with professional artistic drawings was created for training and testing the model.The results show that APDrawingGANoutperforms existing state-of-the-art methods like Cycle GANand Pix2Pix in generating high-quality, expressive artistic portraits.A user study confirms the superiority of the proposed method, with APDrawingGAN ranking highest in 71.39

Wanget al. [7] explores the problem of generating facial images and videos based on given attribute labels using deep learning, specifically conditional generative adversarial networks (DCGANs). This topic is of significant in terest for applications such as law enforcement and entertainment, where realistic face generation from descriptions is valuable. The paper proposes two models: a 2D model, which generates face images based on attribute labels like glasses, gender, hair color, smile, and age, and a 3D model, which generates facial videos, particularly focusing on smiling videos, us ing the UvA-NEMO dataset. Both models incorporate attribute labels in the generation process, with the 2D model concatenating labels with a noise vector and passing to both the generator and discriminator. The models can generate realistic images and videos with good attribute con sistency, and future work could explore matching generated faces with real faces in existing databases for improved realism and diversity.

VI. Methodology

The DCGAN architecture is based on Generative Adversarial Networks (GANs) and incorporates deep convolutional layers to expand on the basic GAN structure. Because of this, the model can learn hierarchical feature representations, which makes it particularly suitable for jobs involving the creation of images.

A. Dataset and Preprocessing

- **Dataset**: We use the CelebA dataset, a large-scale dataset with 200,000+ images of celebrity faces, which serves as an excellent source of varied human facial structures, expressions, and attributes.
- **Preprocessing Steps**:
  - **Resize Images**: Images are resized to a fixed size, typically 64x64 pixels, which reduces computational load while maintaining sufficient resolution for training.
  - **Normalization**: Image pixel values are scaled to a range of $[-1, 1]$ to improve training stability, as the generator's output layer (usually a Tanh function) is also configured to produce values within this range.

B. Model Architecture

The DCGAN model consists of two primary components:

- **Generator (G)**: This network takes a random noise vector z as input and generates synthetic images that resemble real human faces. It uses transposed convolution layers to upsample the noise vector into an image, gradually adding detail.
- **Discriminator (D)**: The discriminator is a convolutional neural network that classifies input images as "real" or "fake." Its goal is to learn to detect generated images and provide feedback to the generator.

C. Weight Initialization

The weights of both networks are initialized with a normal distribution to prevent issues like dead neurons. DCGAN weight initialization typically follows the approach:

$$w \sim \mathcal{N}(0, 0.02)$$

where weights are drawn from a normal distribution with a mean of 0 and a standard deviation of 0.02.

D. Hyperparameter Settings

- **Learning Rate**: 0.0002 for both generator and discriminator.
- **Batch Size**: 128.
- **Noise Vector Size (nz)**: 100.
- **Optimizer**: Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.999$.
- **Number of Epochs**: 50 or more, depending on computational resources.

E. Training Process

The adversarial training process involves two loss functions and alternating updates:

- **Discriminator Loss (Binary Cross-Entropy Loss)**:
  - For real images:
$$\text{loss}_{D_{\text{real}}} = -\mathbb{E}_{x \sim p_{\text{data}}}[\log D(x)]$$
  - For fake images:
$$\text{loss}_{D_{\text{fake}}} = -\mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$

  The total discriminator loss:
$$\text{loss}_D = \text{loss}_{D_{\text{real}}} + \text{loss}_{D_{\text{fake}}}$$
- **Generator Loss**: The generator is optimized to maximize $\log(D(G(z)))$ (equivalent to minimizing $\log(1 - D(G(z)))$):
$$\text{loss}_G = -\mathbb{E}_{z \sim p_z}[\log D(G(z))]$$

F. Function to Display Generated Images

The function for displaying generated images allows visualization of the generator's output during and after training, providing insight into its progress toward realism.

G. Post-Training Evaluation Metrics

- **Inception Score (IS)**: Measures the diversity and realism of generated images.
- **Frechet Inception Distance (FID)**: Compares the statistics of generated images with real images, where a lower FID indicates a higher similarity to the real distribution.

VII. Results

After training, the generator produces high-quality images of human faces. Evaluation using Inception Score and Frechet Inception Distance shows the model generates human-like faces effectively.

VIII. Applications and Future Work

DCGANs have vast applications, including:

- **Entertainment & Gaming**: Synthetic avatars and background characters.
- **Privacy**: Generating faces that look real but are not associated with specific individuals.
- **Art & Design**: Artists can create unique characters for digital art or animation.

Future work may involve improved training stability with Wasserstein GANs and exploring conditional GANs to generate faces conditioned on specific attributes.

IX. Conclusion

In this project, we trained a DCGAN to generate realistic human faces using the CelebA dataset. DCGANs demonstrate great potential in generating photorealistic images, with applications across multiple domains.

REFERENCES

[1] HarshaVardhan Kapalavai, Sourav Mondal, et al. Generating new human faces and improving the quality of images using generative adversarial networks (gan). In *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, pages 1647–1652. IEEE, 2023.

[2] Daanish Mohammed Shariff, H Abhishek, D Akash, et al. Artificial (or) fake human face generator using generative adversarial network (gan) machine learning model. In *2021 fourth international conference on electrical, computer and communication technologies (ICECCT)*, pages 1–5. IEEE, 2021.

[3] Noha Nekamiche, Chahnez Zakaria, Sarra Bouchareb, and Kamel Smaïli. A deep convolution generative adversarial network for the production of images of human faces. In *Asian conference on intelligent information and database systems*, pages 313–326. Springer, 2022.

[4] Vivek Bhalerao, Sandeep Kumar Panda, and Ajay Kumar Jena. Optimization of loss function on human faces using generative adversarial networks. *Machine Learning Approaches for Urban Computing*, pages 189–208, 2021.

[5] M Manimegala, V Gokulraj, K Karisni, and S Manisha. Generating human face with dcgan and gan. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2(05):1348–1354, 2024.

[6] Ran Yi, Yong-Jin Liu, Yu-Kun Lai, and Paul L Rosin. Apdrawinggan: Generating artistic portrait drawings from face photos with hierarchical gans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10743–10752, 2019.

[7] Yaohui Wang, Antitza Dantcheva, and Francois Bremond. From attribute-labels to faces: face generation using a conditional generative adversarial network. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.