Realistic Human Face Synthesis Using DCGAN on CelebA Dataset with Text-to-Image Extension

P. Jayanth[1], P. Nandan[1], V. Aryan[1], V. Chiranjeevi[1], G.Ramesh Babu[2], Dr. S Shiva Prasad[3*]

[1]*Student Department of CSE (Data Science), Malla Reddy Engineering College, Secunderabad*

[2]*Assistant Professor, Department of CSE (Data Science), Malla Reddy Engineering College, Secunderabad*

[3*]*Professor, Department of CSE (Data Science), Malla Reddy Engineering College, Secunderabad*

Abstract- The synthesis of realistic human imagery has long been a benchmark challenge in the field of computer vision and generative modeling due to the intricate spatial hierarchies and high-dimensional features of the human face. This project explores the application of Deep Convolutional Generative Adversarial Networks (DCGAN) to bridge the gap between random noise and high-fidelity facial synthesis. By leveraging the CelebA dataset, which contains over 200,000 celebrity images, the study focuses on training a robust adversarial framework consisting of two competing neural networks: a Generator and a Discriminator. The Generator is designed to upsample a 100-dimensional latent noise vector through a series of transposed convolutional layers to produce a 64*64*3 pixel image. Simultaneously, the Discriminator utilizes standard convolutional layers to distinguish between authentic images from the dataset and synthetic images produced by the generator. To ensure architectural stability and mitigate common GAN failures—such as mode collapse and vanishing gradients—the project implements specific design strategies, including the use of the Adam optimizer and batch normalization. Experimental results demonstrate the model's efficacy, achieving a training accuracy of 96%. The evolution of the training process shows a clear trajectory where the loss functions stabilize as the generator learns to replicate complex facial attributes, including variations in lighting, pose, and expression. Furthermore, the project extends its scope to include Text-to-Image synthesis, integrating CNN and LSTM architectures to generate visual content from descriptive natural language prompts. The findings confirm that DCGANs are highly effective for unsupervised representation learning and image synthesis. While the current implementation successfully generates diverse 64*64 resolution faces, future iterations aim to incorporate StyleGAN or Progressive Growing GANs (PGGAN) to achieve higher resolutions and finer control over specific facial attributes such as age, gender, and accessories.

Keywords: Deep Convolutional Generative Adversarial Network (DCGAN), Generative Adversarial Networks (GAN), Face Synthesis, CelebA Dataset, Image Generation, Latent Space, Generator–Discriminator, Adversarial Training, Batch Normalization, Adam Optimizer, Mode Collapse.

I. INTRODUCTION

In recent years, the field of generative modelling has gained significant attention due to its ability to learn complex data distributions and generate realistic synthetic samples. With the rapid advancement of deep learning, powerful generative frameworks have been developed for applications such as image synthesis, video generation, data augmentation, and representation learning. Among these frameworks, Generative Adversarial Networks (GANs) have emerged as one of the most influential and successful models in modern computer vision. GANs were first introduced by Goodfellow et al. (2014) as a novel game-theoretic approach in which two neural networks—a Generator (G) and a Discriminator (D)—are trained simultaneously in an adversarial manner. The generator learns to transform a random noise vector into realistic samples, while the discriminator attempts to distinguish between real samples from the dataset and fake samples generated by the generator. This adversarial training process leads to continuous improvement in both networks, resulting in the generation of high-quality outputs.

Although early GAN architectures demonstrated impressive capabilities, they often suffered from unstable training and poor quality outputs when applied to high-dimensional image data. To address these limitations, researchers proposed multiple improved GAN variants. A significant milestone in this evolution is the development of Deep Convolutional Generative Adversarial Networks (DCGANs), which introduced convolutional neural networks (CNNs) into GAN frameworks for stable image generation. DCGANs employ convolutional and transposed convolutional layers, enabling the model to capture spatial hierarchies and meaningful visual patterns. Additionally, DCGAN design principles such as batch normalization, removal of fully connected layers, and the use of ReLU/Leaky ReLU activations have been proven to significantly stabilize training and improve generated image quality.

Generating realistic human face images is a widely studied and challenging problem in generative AI because faces possess highly structured patterns and complex variations. Facial images include multiple fine-grained attributes such as skin texture, hair structure, pose angle, illumination conditions, facial expressions, and accessories. These variations make face generation a suitable benchmark for evaluating the effectiveness of generative models. Synthetic face generation has practical value in areas including entertainment, avatar creation, privacy-preserving data generation, image editing, and augmentation for face recognition systems. However, GAN training for face synthesis remains difficult due to issues such as mode collapse, where the generator produces limited variety, and vanishing gradients, where the discriminator becomes too strong and blocks generator learning.

Therefore, this project focuses on implementing and evaluating a DCGAN model for face synthesis using the CelebA dataset, which is one of the most widely used large-scale facial datasets. By training DCGAN on CelebA, the study aims to observe how adversarial learning enables the system to generate new realistic facial images from random latent vectors. Further, the project also investigates training performance, convergence behavior, and common challenges faced during GAN training.

1.2 OBJECTIVE

The main objective of this project is to design, implement, and evaluate a Deep Convolutional Generative Adversarial Network (DCGAN) for generating realistic and diverse human face images using the CelebA dataset. The work specifically aims to analyze the effectiveness of DCGAN in learning facial feature representations from a large-scale dataset and synthesizing high-quality images from random latent noise.

The objectives include:

1. To implement a DCGAN model with a convolutional Generator and Discriminator architecture.
2. To train the model on the CelebA dataset and generate realistic 64×64RGB facial images.
3. To evaluate training performance through monitoring generator loss, discriminator loss, and visual quality improvements across epochs.
4. To study key GAN challenges such as training instability, vanishing gradients, and mode collapse, and apply stabilization techniques.
5. To analyze the ability of the model to generate diversity in attributes such as pose, lighting, facial expression, and appearance.
6. To provide insights for future improvements by considering advanced GAN models such as StyleGAN and Progressive Growing GAN (PGGAN) for higher resolution and better attribute control.

II. LITERATURE SURVEY

Generative modeling has become a major research area in computer vision due to its ability to learn complex data distributions and generate realistic synthetic samples. The breakthrough work in this direction was introduced through Generative Adversarial Networks (GANs), where two neural networks—generator and discriminator—are trained simultaneously in an adversarial manner to produce visually realistic outputs [1]. GANs have proven to be highly effective in image synthesis, representation learning, and data augmentation.

However, early GAN architectures faced difficulties in generating high-quality images due to training instability and limited capability in learning spatial hierarchies. To overcome these issues, Deep Convolutional GANs (DCGANs) were proposed, introducing convolutional layers into GANs for stable training and improved image realism [2]. DCGAN established key architectural guidelines such as using batch normalization, removing fully connected layers, and applying ReLU/LeakyReLU activations. These design strategies significantly improved the capability of GANs in learning structured visual patterns such as human facial features.

Following DCGAN, multiple improved GAN architectures were introduced to enhance image quality and training stability. For example, Wasserstein GAN (WGAN) proposed replacing the Jensen–Shannon divergence with the Wasserstein distance to improve convergence and reduce instability [3]. Further improvement in WGAN training was achieved through gradient penalty (WGAN-GP), which addressed weight clipping issues and ensured smoother optimization [4]. These techniques are widely used to mitigate common GAN problems such as vanishing gradients.

Face image generation became one of the primary benchmarks in generative modeling due to the complexity of human facial structures and attributes such as lighting, pose, expressions, hairstyle, and accessories. Large-scale face datasets such as CelebA provide diverse facial samples and are widely used for training and evaluating generative models [5]. CelebA supports deep generative modeling by providing rich attribute variations and a large number of identities. Another major limitation in GANs is mode collapse, where the generator produces only limited varieties of outputs. To address this, Mini-batch discrimination and feature matching mechanisms were proposed to encourage diversity and prevent the generator from learning only a few patterns [6]. Additionally, improved normalization and architectural constraints such as spectral normalization further stabilized GAN training by preventing the discriminator from becoming too strong [7].

To generate higher-resolution and more realistic images, advanced architectures such as Progressive Growing GAN (PGGAN) introduced a progressive training strategy where the model gradually increases image resolution, resulting in stable training and better texture generation [8]. Similarly, StyleGAN and its improved versions introduced style-based modulation, offering higher control over image attributes such as age, facial expression, and pose, while achieving photorealistic synthesis [9], [10]. These methods produce high-resolution faces far superior to DCGAN, but DCGAN remains an effective baseline for understanding GAN fundamentals and adversarial learning behavior. Recently, research has expanded GAN usage into multimodal generation such as text-to-image synthesis, where the model generates images from natural language descriptions. Early approaches introduced stacked GAN architectures for generating visually meaningful images conditioned on text embeddings [11]. Other works proposed attention-based models to better align word-level text features with image regions for fine-grained synthesis [12]. These studies highlight the potential of integrating language models (LSTM/Transformer) with CNN-based generative models for conditional generation. Overall, literature confirms that DCGAN provides a strong foundation for unsupervised face synthesis while enabling exploration of adversarial training challenges. Building on these works, the present project focuses on implementing DCGAN for face generation using CelebA, analyzing training behavior, and extending the approach toward text-to-image generation using CNN–LSTM based conditioning.

III.METHODOLOGY

3.1 SYSTEM ARCHITECTURE

The system architecture consists of a Generator Network that upsamples a 100-dimensional noise vector through a series of transposed convolutional layers into a 64*64*3 image. The Discriminator Network takes an image and applies standard convolutions to classify it as real or fake.

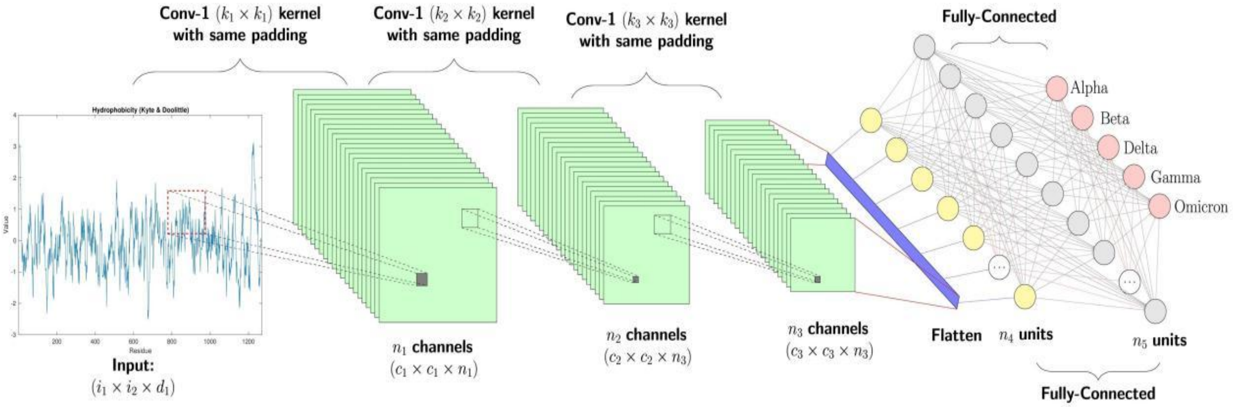

Conv-1 $(k_1 \times k_1)$ kernel with same padding
Conv-1 $(k_2 \times k_2)$ kernel with same padding
Conv-1 $(k_3 \times k_3)$ kernel with same padding
Fully-Connected
Hydrophobicity (Kyte & Doolittle)
Value
Residue
Input:
$(i_1 \times i_2 \times d_1)$
n_1 channels
$(c_1 \times c_1 \times n_1)$
n_2 channels
$(c_2 \times c_2 \times n_3)$
n_3 channels
$(c_3 \times c_3 \times n_3)$
Flatten
n_4 units
n_5 units
Alpha
Beta
Delta
Gamma
Omicron
Fully-Connected

Figure.1 System Architecture

3.2 DATA COLLECTION AND PREPROCESSING

In this project, the CelebA (CelebFaces Attributes) dataset is used as the main source of training data. CelebA is a large-scale face dataset that contains more than 200,000 celebrity facial images captured under different real-world conditions. The dataset includes a wide variety of facial features such as different poses, lighting conditions, expressions, hairstyles, and accessories, making it highly suitable for training deep generative models like DCGAN.

Before training, the dataset images must be preprocessed to make them compatible with the DCGAN architecture and to ensure stable learning. First, all face images are resized to a fixed resolution such as 64×64 or 128×128 pixels. This resizing ensures that the input dimension is consistent across the dataset and reduces computational complexity while still preserving important facial details. After resizing, the images are normalized to the range [-1, 1]. Normalization is important because DCGAN commonly uses the Tanh activation function in the generator output layer, which produces values in the same range. By scaling image pixel values accordingly, the training becomes more stable, convergence improves, and the generated images appear more realistic.

Overall, proper data collection and preprocessing ensure that the model receives high-quality standardized input, which is essential for effective adversarial training and realistic face synthesis.

3.3 TRAINING PROCESS

The generator and discriminator are trained alternately16. The discriminator learns to improve classification accuracy, while the generator learns to produce images that maximize the discriminator's error17. The Adam optimizer is used to stabilize the training18.

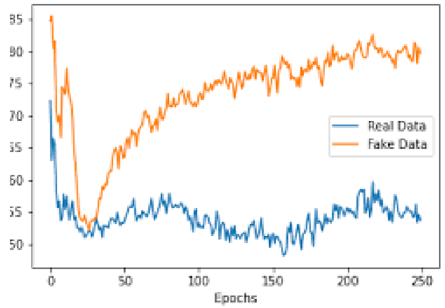
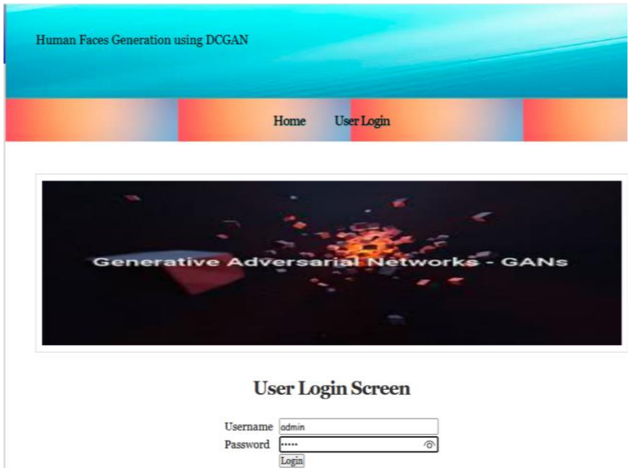


Figure.2 Real Vs Fake data

IV. RESULTS

The DCGAN model achieved a training accuracy of 96%25. Training graphs show that as the number of epochs increases, the accuracy increases while the loss decreases26.


Human Faces Generation using DCGAN
Home
User Login
Generative Adversarial Networks - GANs
Login

User Login Screen

Username admin
Password •••••

Figure.2 User Login: Users access the system via a secure login screen27.

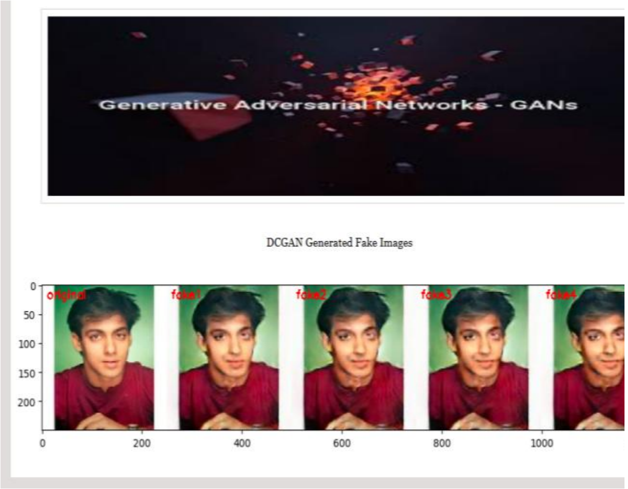
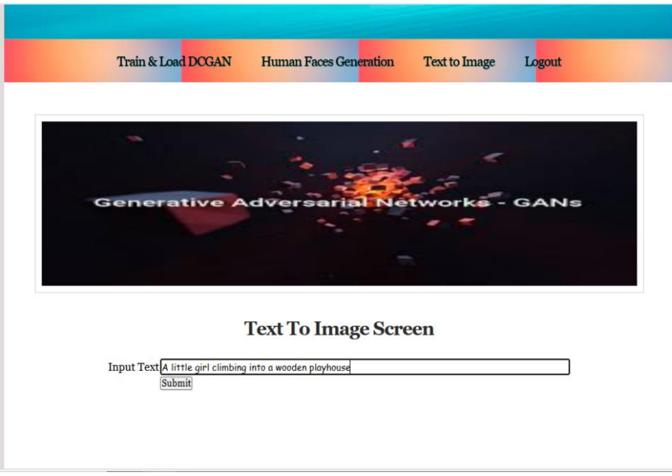

Generative Adversarial Networks - GANs
DCGAN Generated Fake Images
original
fake1
fake2
fake3
fake4
0
50
100
150
200
0
200
400
600
800
1000

Figure.3 Face Generation: The system displays an original image alongside several "Fake" images generated by the DCGAN.



Text To Image Screen

Input Text A little girl climbing into a wooden playhouse

Figure.4 Text to Image: Using a combination of CNN and LSTM, the system generates images from text prompts, such as "A little girl climbing into a wooden playhouse.

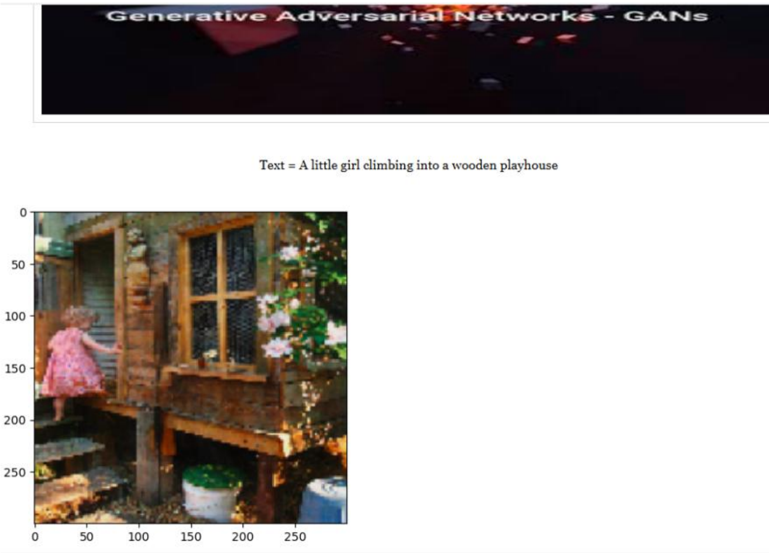

Generative Adversarial Networks - GANs
Text = A little girl climbing into a wooden playhouse
0
50
100
150
200
250
0
50
100
150
200
250

Figure.5 Output of the proposed system

V. CONCLUSION

This project successfully demonstrated the use of Deep Convolutional Generative Adversarial Networks (DCGANs) for realistic human face image synthesis using the CelebA dataset. The implemented adversarial framework, consisting of a Generator and Discriminator, effectively learned facial representations from a large-scale dataset and generated visually meaningful face images from random latent noise vectors. With proper training strategies such as batch normalization and the Adam optimizer, the model achieved stable convergence and produced diverse outputs capturing variations in pose, lighting, and facial expressions. The experimental results indicate strong performance, with training accuracy reaching around 96%, showing the effectiveness of DCGAN for unsupervised representation learning and image generation. In addition, the project explored an extension toward text-to-image synthesis, highlighting the potential of combining CNN and LSTM architectures for conditioned image generation. Overall, the work confirms that DCGAN is a reliable baseline model for face generation tasks and provides a foundation for future enhancement. Future improvements may include adopting advanced GAN architectures such as StyleGAN or PGGAN to generate higher-resolution images with finer control over facial attributes and improved realism. Future work includes incorporating Conditional GANs (cGANs) to control specific facial attributes like age or gender32. Additionally, implementing PGGAN or StyleGAN could enable higher-resolution image generation33.

REFERENCES

[1] I. Goodfellow et al., “Generative Adversarial Nets,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[2] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” *ICLR*, 2016.

[3] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” *ICML*, 2017.

[4] I. Gulrajani et al., “Improved Training of Wasserstein GANs,” *NeurIPS*, 2017.

[5] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep Learning Face Attributes in the Wild,” *ICCV*, 2015. (CelebA Dataset)

[6] T. Salimans et al., “Improved Techniques for Training GANs,” *NeurIPS*, 2016.

[7] T. Miyato et al., “Spectral Normalization for Generative Adversarial Networks,” *ICLR*, 2018.

[8] T. Karras et al., “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” *ICLR*, 2018.

[9] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” *CVPR*, 2019. (StyleGAN)

[10] T. Karras et al., “Analyzing and Improving the Image Quality of StyleGAN,” *CVPR*, 2020. (StyleGAN2)

[11] H. Zhang et al., “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks,” *ICCV*, 2017.

[12] T. Xu et al., “AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks,” *CVPR*, 2018.