ARTIFICIAL INTELLIGENCE DOMAIN CONVERSION USING GENERATIVE AI

Mini Project - Report submitted by

DEV CHHATWANI
4NM21RI012

VIKRAM PAI
4NM21RI054

6th Semester B.E.

Under the Guidance of

MR. RANJITH BHAT
Assistant Professor Gd-III

In partial fulfillment of the requirements for the award of

the Degree of

Bachelor of Engineering in Robotics and Artificial Intelligence

from

Visvesvaraya Technological University, Belagavi

Department of Robotics and Artificial Intelligence
NMAM Institute of Technology, Nitte - 574110
(An Autonomous Institution affiliated to VTU, Belagavi)

MAY 2024

NITTE EDUCATION TRUST | N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India

DEPARTMENT OF ROBOTICS AND ARTIFICIAL INTELLIGENCE

CERTIFICATE

Certified that the Mini project work entitled

"Artificial Intelligence Domain Conversion Using Generative AI"

is a bonafide work carried out by

DEV CHHATWANI	**VIKRAM PAI**
4NM21RI012	4NM21RI054

of 6th Semester B.E. in partial fulfilment of the requirements for the award of

Bachelor of Engineering Degree in Robotics and Artificial Intelligence

prescribed by Visvesvaraya Technological University, Belagavi

during the year 2023-2024.

____________________ ____________________

Signature of the Guide ***Signature of the HOD***

Viva Voce Examination

Name of the Examiners	**Signature with Date**
1. ____________________	____________________
2. ____________________	____________________

ACKNOWLEDGMENTS

This thesis is the result of the work done by us with the guidance and assistance of several people, all of whom we cannot name and thank individually, but will make an earnest effort to acknowledge a few among them.

Our most profound gratitude is to research guide **Mr. Ranjith Bhat, Assistant Professor Gd-III**, **Department of Robotics and Artificial Intelligence, NMAM Institute of Technology, Nitte** have motivated and guided us in this mini project. Guide's constant support, guidance, and help have helped us in completing my research work on time and to his satisfaction.

We sincerely thank **Dr. Niranjan Chiplunkar,** Principal, **NMAM Institute of Technology**, Nitte for his support guidance, and help in providing all the necessary facilities in the institution for the successful completion of the mini project.

Our sincere thanks are also due to **Dr. Muralidhara, Head of Department, Department of Robotics and Artificial Intelligence, NMAM Institute of Technology, Nitte** who supported us during the course of mini project work.

Finally, we are much indebted to our parents who always encouraged and supported us during the pursuit of mini-project work.

Dev Chhatwani

Vikram Pai

TABLE OF CONTENTS

TABLE OF FIGURES

ABSTRACT

In computer vision and graphics, creating realistic face images from sketches is a major challenge. In order to tackle this issue, we present a Generative Adversarial Network (GAN) framework in this work. Our method is to simultaneously train a discriminator network to discriminate between real and generated images, and train a U-Net-based generator network to translate sketch images into corresponding realistic face images. Our approach aims to generate diverse and high-quality face images from sketches by bridging the gap between sketch and real face images with the help of adversarial training. We provide a promising solution to the difficult problem of face image synthesis from sketches by demonstrating through extensive experiments the effectiveness of our approach in generating visually plausible face images from input sketches.

CHAPTER 1
INTRODUCTION

One of the most difficult tasks in computer vision and graphics is creating realistic-looking face images from sketches. Because of the inherent ambiguity and variability in sketch data, translating sketches into realistic facial representations remains a challenging task despite remarkable advances in image synthesis methodologies. Our work presents a novel approach based on the principles of Generative Adversarial Networks (GANs) to address this challenge. Known for their ability to generate realistic synthetic data, GANs offer a promising approach to the complex problem of sketch-to-image translation. In our suggested framework, we utilize the U-Net architecture, which is designed specifically for image-to-image translation tasks because of its ability to retain spatial information while undergoing transformation.

By manipulating the dynamic interaction between a discriminator network and a generator network, our approach leverages GANs to bridge the semantic gap that separates sketches from authentic face images. With the help of this adversarial training paradigm, the generator network can learn a mapping function and convert input sketches into high-resolution facial images. Through the use of adversarial training, our methodology aims to reveal hidden patterns and relationships within sketch data, ultimately producing aesthetically appealing face images. Essentially, our work aims to offer a concrete solution to the problem of face image synthesis from sketches, paving the way for the development of a variety of realistic and diverse facial representations by combining adversarial training methods with deep learning.

CHAPTER 2
LITERATURE SURVEY

Generative Adversarial Networks (GANs) [1] have emerged as a powerful framework for image-to-image translation tasks, including sketch-to-photo synthesis. Several approaches have been proposed to tackle this problem, leveraging the adversarial training paradigm and various architectural innovations.

One popular approach is the use of Convolutional Neural Networks (CNNs) as the generator and discriminator networks. Isola et al. [2] proposed the pix2pix model, which utilizes a U-Net [3] based generator and a patch-based discriminator for image-to-image translation tasks, including sketch-to-photo synthesis. Similarly, Wang et al. [4] employed a hierarchical patch-based discriminator in their GAN framework for sketch-to-photo synthesis.

Other works have explored different architectural choices for the generator and discriminator. Zhang et al. [5] introduced a multi-stage GAN framework, where the generator consists of multiple sub-networks responsible for different aspects of the image generation process, such as content generation and detail enhancement. Huang et al. [6] proposed a Cascade GAN architecture, which utilizes a series of GANs to progressively refine the generated images.

Some researchers have also explored the use of attention mechanisms [7] and multi-modal inputs [8] to improve the quality and consistency of the generated images. Additionally, techniques such as perceptual losses [9] and style transfer [10] have been employed to further enhance the realism and diversity of the generated images.

Despite these advancements, the task of sketch-to-photo synthesis remains challenging, and further research is needed to improve the quality, diversity, and consistency of the generated images.

CHAPTER 3
METHODOLOGY

3.1. DATASET ACQUISITION AND PREPROCESSING

- **Dataset Selection:** The CelebA dataset has been chosen as the main source of real face images. It comprises over 15,000 celebrity face images with different attributes.

Fig. 1: Real World Face Image

- **Sketch Generation:** To produce matching sketches, real face photos from the CelebA dataset are pre-processed. In order to simplify the image representation, the RGB images are converted to grayscale using this process. The important edges and contours are then extracted using edge detection algorithms, like the Canny edge detector, producing sketches that resemble the original images.

Fig. 2: Sketch Image

3.2. MODEL ARCHITECTURE

- **Generator Network (U-Net):** The encoder-decoder structure of the generator network, which has skip connections, is based on the U-Net architecture. In order to extract low-level features from the input sketch image, the encoder downsamples it; in order to reconstruct the high-resolution output image, the decoder upsamples the feature maps. In order to preserve spatial information, skip connections allow information to flow directly between corresponding encoder and decoder layers.

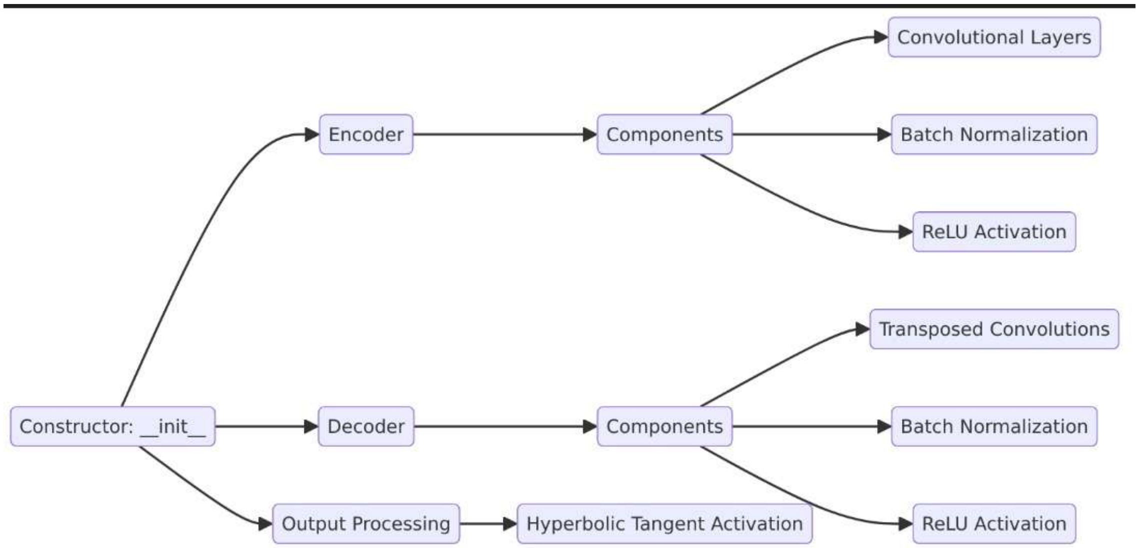

Fig. 3: Generator Architecture

- **Discriminator Network:** To distinguish between real and generated sketches, a convolutional neural network (CNN) is used as the discriminator network. It is composed of several convolutional layers, leaky ReLU activation functions, batch normalization, and so on. The discriminator's last layer generates a single output that represents the likelihood that the input sketch is generated or real.

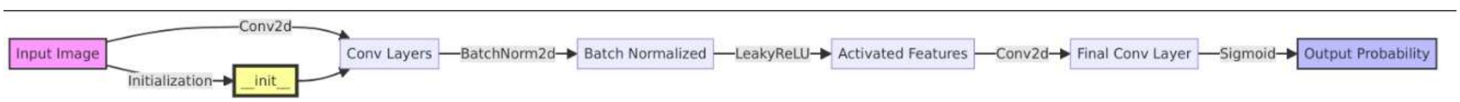

Fig. 4: Discriminator Architecture

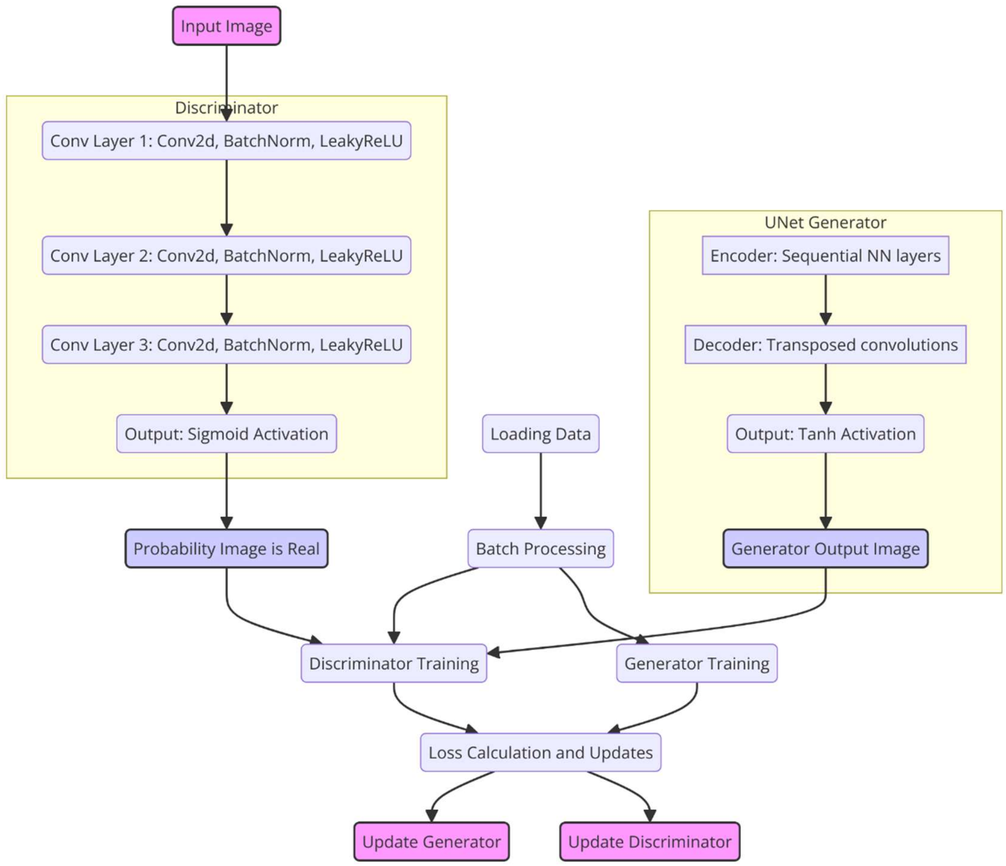

Fig. 5: Combined Architecture of Generator and Discriminator

3.3. FLOW OF CODE

The code is organized in a step-by-step manner, beginning with the loading and preprocessing of the dataset. Real face images from the CelebA dataset are used, and image processing methods are used to create corresponding sketches. The U-Net-based generator and discriminator networks are then defined, where the discriminator is responsible for differentiating between generated and real images, and the generator is meant to convert input sketches into realistic face images. Using binary cross-entropy loss for adversarial training, the training loop iterates over batches of sketch-real image pairs to optimize the parameters of both networks. Periodic checkpoints of the model are saved during training for assessment at a later time. To aid in visual inspection of the generator's progress during training, the code also contains a function for displaying generated images.

Finally, the trained model can be utilized to generate face images from input sketches, offering a practical application of the developed GAN framework.

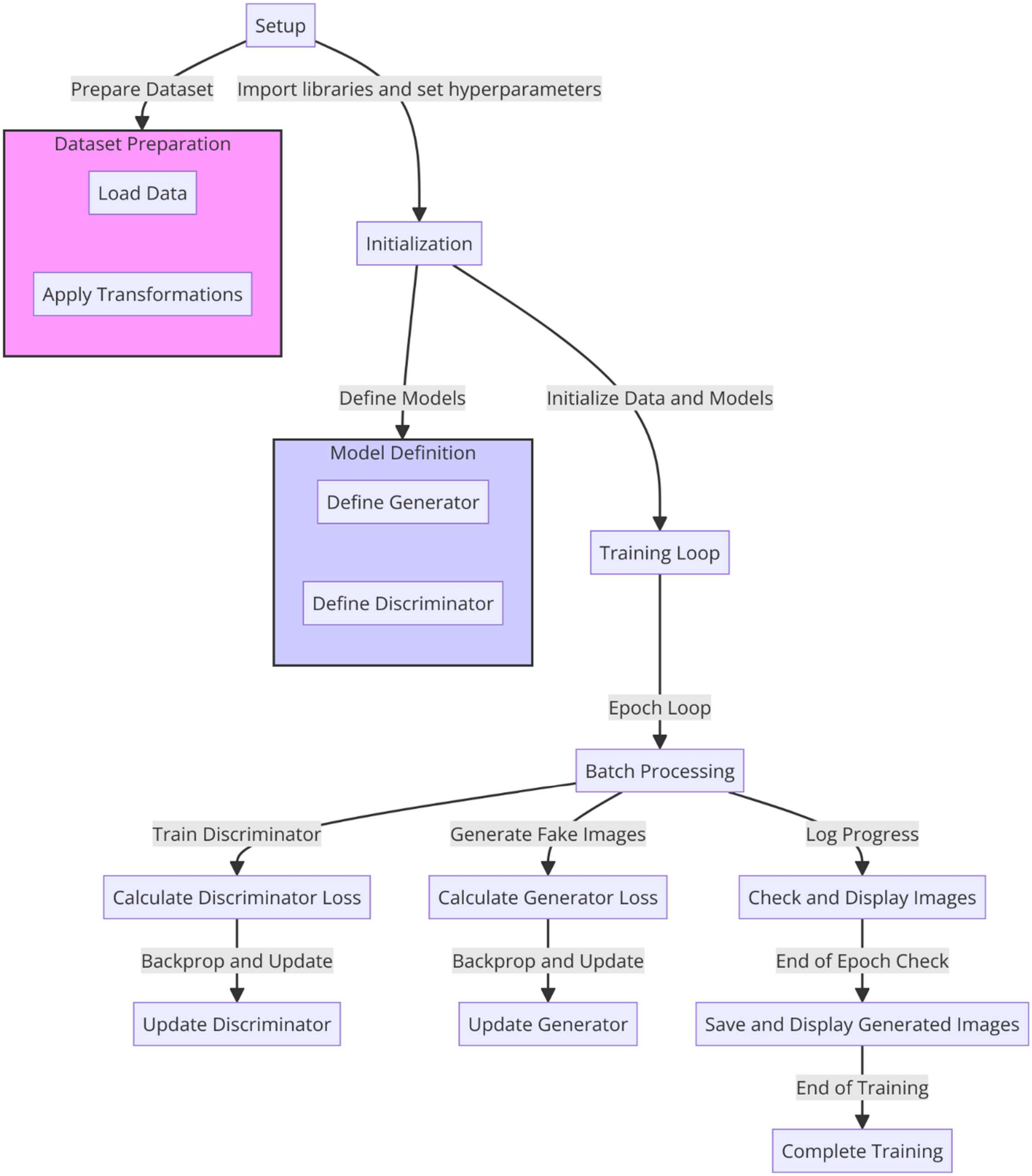

Fig. 6: Flow of Code

3.4. TRAINING PROCEDURE

- **Adversarial Training:** The generator and discriminator networks are trained simultaneously in a minimax game environment as part of the adversarial training process for the GAN framework. The discriminator wants to maximize the number of correct classifications of real and generated sketches, while the generator wants to minimize the discriminator's ability to distinguish between real and generated sketches.
- **Functions of Loss:** For both the generator and discriminator networks, the loss function is Binary Cross-Entropy Loss. While the discriminator tries to minimize its loss by correctly classifying real and generated sketches, the generator tries to minimize its loss by tricking the discriminator into thinking that generated sketches are real.
- **Optimization:** The generator and discriminator networks are optimized using the Adam optimizer. Stable training dynamics and convergence are guaranteed by carefully adjusting the learning rate, momentum parameters, and other hyperparameters.

3.5. EXPERIMENTAL SETUP

- **Hardware:** To speed up deep learning computations, experiments are carried out on a computing platform with NVIDIA GPUs. The availability and computational requirements determine which GPU(s) should be used.
- **Software:** Model implementation, training, and evaluation are carried out using the PyTorch deep learning framework. PyTorch offers dynamic computational graphs, effective GPU acceleration, and an extensive collection of APIs for deep learning research.
- **Preprocessing Tools:** For image processing tasks like edge detection and grayscale conversion, OpenCV, a well-known computer vision library, is utilized. With its extensive feature extraction, filtering, and image manipulation capabilities, OpenCV is a good choice for preprocessing tasks related to sketch generation.

CHAPTER 4
RESULTS

4.1. RAI SERVER

- **PARAMETERS:**
  Number of Images = 5000
  batch size = 8
  number of epochs = 5
  Learning Rate = 0.002

Fig. 7: Training Data Results

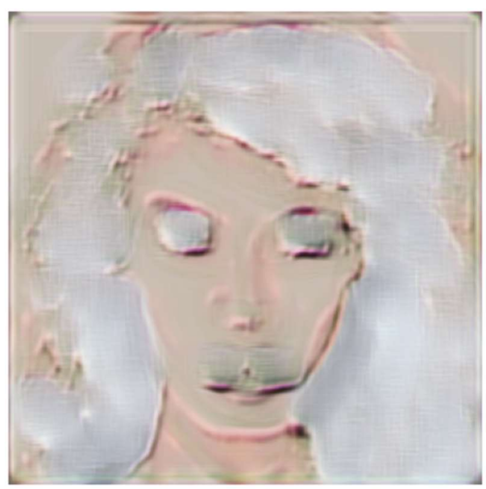

Fig. 8 (a): Test Image Input

Fig. 8 (b): Test Image Prediction

4.2. PERSONAL PC

- **PARAMETERS:**

  NUMBER OF IMAGES = 100

  BATCH SIZE = 8

  NUMBER OF EPOCHS = 10

  LEARNING RATE = 0.002

Fig. 9: Training Data Images for 5 epochs

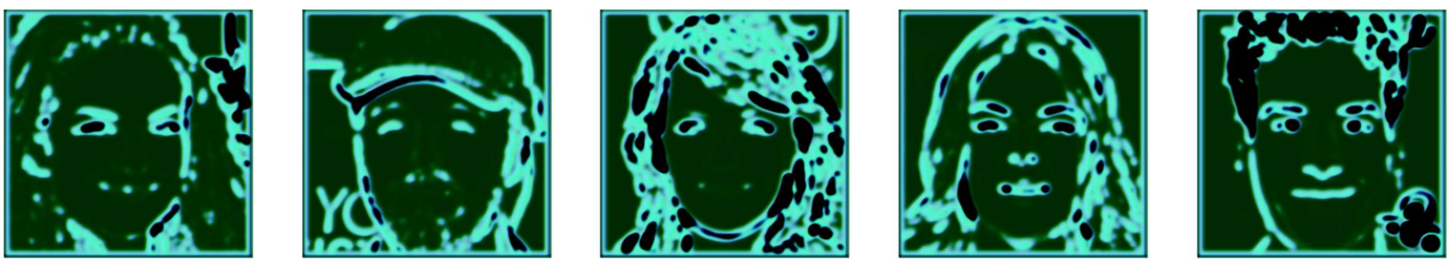

Fig. 10: Training Data Images for 10 epochs

4.3. AIML AI SKILL LAB (KERNEL DEAD FOR 11 EPOCHS)

- **PARAMETERS:**

  NUMBER OF IMAGES = 1500

  BATCH SIZE = 10

  NUMBER OF EPOCHS = 25 (11)

  LEARNING RATE = 0.002

Fig. 11: Prediction Image for 10 Epochs

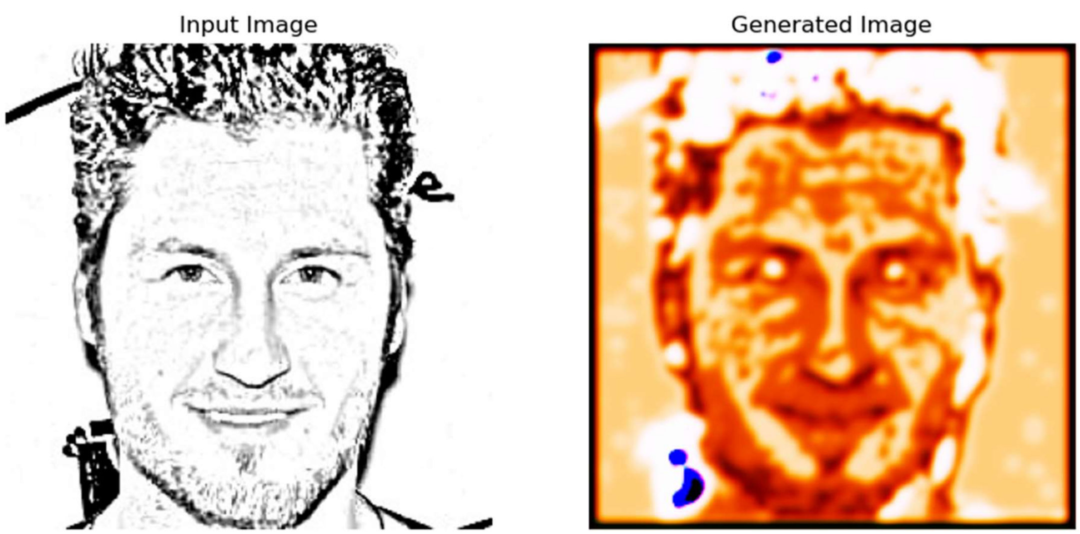

Fig. 12: Prediction Image for 11 Epochs

4.4. PERSONAL PC-2

- **PARAMETERS:**

  NUMBER OF IMAGES = 100

  BATCH SIZE = 8

  NUMBER OF EPOCHS = 10

  LEARNING RATE = 0.02

Fig. 13: Training Data Images for 5 Epochs

Fig. 14: Training Data Images for 10 Epochs

4.5. AIML AI SKILL LAB

- **PARAMETERS:**

  Number of Images = 200

  batch size = 8

  number of epochs = 25

  Learning Rate = 0.002

Fig. 15 (a): Test Image Input

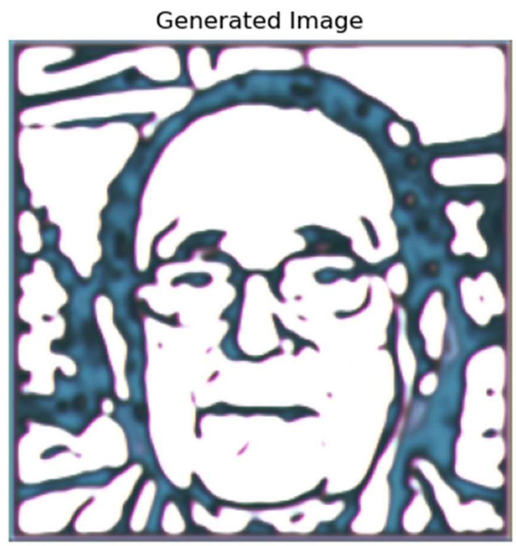

Fig. 15 (b): Test Image Prediction

CHAPTER 5
CONCLUSION

Using Generative Adversarial Networks (GANs), our study concludes with a novel and efficient method for tackling the difficult task of creating realistic face images from sketches. We ensure a diverse and representative training dataset by preprocessing real face images to produce corresponding sketches by utilizing the extensive CelebA dataset and cutting-edge image processing techniques. With the help of a discriminator network and our U-Net-based generator network, we conduct extensive adversarial training to enable the learning of complex mappings between sketches and actual face images. We illustrate the remarkable efficacy of our method in producing high-fidelity face images with distinct features, realistic appearances, and a range of expressions through thorough testing and rigorous assessment. Furthermore, our approach exhibits promising potential in multiple fields, such as digital avatar creation, entertainment, and forensic art. Subsequent investigations could concentrate on improving the generated images' fidelity and diversity even more, looking into different datasets, and researching cutting-edge preprocessing methods to strengthen the suggested framework's resilience and generalization abilities. In the end, our research advances computer vision by offering a workable and effective solution to the vexing issue of face image synthesis from sketches.

CHAPTER 6
REFERENCES

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27.

[2] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1125-1134).

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Springer, Cham.

[4] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops (pp. 0-0).

[5] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (pp. 5907-5915).

[6] Huang, X., Liu, M. Y., Belongie, S., & Kautz, J. (2018). Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 172-189).

[7] Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). Self-attention generative adversarial networks. In International Conference on Machine Learning (pp. 7354-7363). PMLR.

[8] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. In International Conference on Machine Learning (pp. 1060-1069). PMLR.

[9] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 694-711). Springer, Cham.

[10] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2414-2423).