FACIAL EMOTION RECOGNITION USING GENERATIVE ADVERSARIAL NETWORK

A major project report submitted in partial fulfilment of the requirements for the award of the degree of

Master of Computer Applications

in

Computer Applications

by

ANKUSH SIKARWAR (205121024)

DEPARTMENT OF COMPUTER APPLICATIONS

NATIONAL INSTITUTE OF TECHNOLOGY,

TIRUCHIRAPPALLI – 620015

JUNE-2024

BONAFIDE CERTIFICATE

This is to verify that the project" FACIAL **EMOTION RECOGNITION USING GENERATIVE ADVERSARIAL NETWORK**" is a project work successfully done by

ANKUSH SIKARWAR (205121024)

in partial fulfilment of the requirement for the award of the degree of Master of Computer Applications of the National Institute of Technology, Tiruchirappalli, during the year 2023-2024 (6^{th} semester - CA750 Major Project Work).

Dr. Balaji Banothu — Project Guide

Dr. Michael Arock — Head of the Department

Project Viva-voce held on ______________________________

Internal Examiner **External Examiner**

ABSTRACT

Facial Emotion Recognition (FER) is a crucial aspect of human-computer interaction, with applications ranging from virtual reality to mental health monitoring. GAN, a subset of deep learning models, have shown promising potential in enhancing FER systems. GANs consist of two neural networks, the generator and the discriminator, engaged in a competitive learning process to generate realistic data samples. In the context of FER, GANs can be utilized to augment training datasets, generate synthetic facial expressions, and improve the robustness and generalization of FER models.

The abstract outlines the potential of GANs in enhancing FER systems by leveraging their ability to generate realistic facial expressions and augment training data. By incorporating GANs into FER pipelines, researchers aim to address challenges such as limited data availability, domain adaptation, and model generalization. Moreover, GAN-based FER systems have the potential to achieve state-of-the-art performance in recognizing subtle and complex emotions from facial images.

ACKNOWLEDGEMENT

Every project, big or small, is successful largely due to the effort of several wonderful people who have always given their valuable advice or lent a helping hand. I sincerely appreciate the inspiration, support, and guidance of all those people who have been instrumental in making this project successful.

I express my deep sense of gratitude to **Dr. G. Aghila**, Director, National Institute of Technology, Tiruchirappalli for giving me an opportunity to do this project.

I am grateful to **Dr. Michael Arock,** Professor, and Head, Department of Computer Applications, National Institute of Technology, Tiruchirappalli for providing the infrastructure and facilities to carry out the project.

I express my gratitude to my Project Guide **Dr. Balaji Banothu,** Assistant Professor, Department of Computer Applications, National Institute of Technology, Tiruchirappalli for his support and for arranging the project in a good schedule, and who assisted me in completing the project. I would like to thank him for duly evaluating my progress and evaluating me.

I express my sincere and heartfelt gratitude to **Project Evaluation Committee**, Department of Computer Applications, National Institute of Technology, Tiruchirappalli. I am sincerely thankful for its constant support, care, guidance, and regular interaction throughout my project.

I express my sincere thanks to all the faculty members, and scholars of NIT Trichy for their critical advice and guidance to develop this project directly or indirectly.

Ankush Sikarwar

(205121024)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

1. CNN: Convolutional Neural Network
2. GPU: Graphics Processing Unit
3. HPC: High-Performance Computing
4. CPU: Central Processing Unit
5. RAM: Random Access Memory
6. ReLU: Rectified Linear Unit
7. GANs: Generative Adversarial Networks
8. F1 Score: F1 Measure
9. AI: Artificial Intelligence
10. NumPy: Numerical Python
11. CUDA: Compute Unified Device Architecture

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Facial Emotion Recognition (FER) is a crucial aspect of human-computer interaction, with applications ranging from virtual reality to mental health monitoring. GAN, a subset of deep learning models, have shown promising potential in enhancing FER systems. GANs consist of two neural networks, the generator and the discriminator, engaged in a competitive learning process to generate realistic data samples. In the context of FER, GANs can be utilized to augment training datasets, generate synthetic facial expressions, and improve the robustness and generalization of FER models.

1.1 Facial Emotion Recognition

Facial emotion recognition is a technology that uses artificial intelligence to analyze facial expressions and identify the underlying emotions of a person. It typically involves the use of computer vision algorithms to detect facial landmarks, such as the position of the eyes, nose, and mouth, and then analyzes the configuration of these features to determine the emotional state of the individual.

Fig 1.1 Facial Emotions

Humans naturally express emotions through facial movements, making facial expressions a vital channel for communication and understanding. FER systems aim to replicate this ability by analyzing facial features, such as eyebrow movement, eye widening, and mouth shape, to infer the underlying emotional state.

The development of FER systems has been greatly accelerated by advancements in deep learning, particularly convolutional neural networks (CNNs). CNNs can automatically learn hierarchical representations of facial features directly from raw image data, enabling more accurate and robust emotion recognition.

1.2 Background and Motivation

The background of FER traces back to pioneering research in psychology, which identified universal facial expressions associated with basic emotions such as happiness, sadness, anger, surprise, fear, and disgust. This research formed the foundation for developing computational models and algorithms to automate the process of recognizing emotions from facial images or videos.

In contemporary applications, FER finds extensive use across diverse domains, notably in Human-Computer Interaction (HCI). Here, FER enables machines to interact with users in a more empathetic and intuitive manner. For instance, virtual assistants can adapt their responses based on users' emotional states, enhancing user experiences and engagement. Additionally, FER plays a pivotal role in healthcare, particularly in mental health monitoring and diagnosis. Analyzing facial expressions can provide insights into individuals' emotional well-being, aiding in the detection of conditions like depression or anxiety and facilitating early intervention or support.

1.3 Gaps in Existing Solutions

One significant gap in existing facial emotion recognition systems is their limited ability to accurately detect and interpret subtle or nuanced facial expressions. Current systems often struggle to differentiate between similar emotions or accurately recognize emotions in challenging conditions, such as low lighting or occluded faces. Additionally, there is a lack of diversity in existing datasets, leading to biases and reduced performance when dealing with underrepresented demographics. Addressing these gaps requires advancements in algorithm robustness, enhanced training data diversity, and improved model generalization to ensure more accurate and inclusive emotion recognition across diverse populations and contexts.

1.4 Class Imbalance Issue

In image analysis, class imbalance is a serious problem as datasets frequently show an uneven distribution of various classes. The performance of machine learning models that have been trained on such data may be significantly impacted by this imbalance. Reducing class imbalance in dataset presents a number of difficulties that may affect the accuracy and efficiency of machine learning algorithms. One major challenge is the recognition of face in dataset. The model may find it difficult to acquire appropriate representations because to the small number of cases of these ailments. Furthermore, there is a chance that the skewed distribution of the data, which has more healthy instances than abnormal cases, would bias the models in favor of the majority class and lessen their sensitivity to anomalies

CHAPTER 2

PROBLEM STATEMENT

The problem statement for "Facial Emotion Recognition using GAN" encompasses the difficulties faced in recognizing human emotions and unable to understand the action units of faces in several cases.

Here we encompass the development of a system that leverages GAN to enhance the accuracy and robustness of emotion recognition from facial expressions. Traditional facial emotion recognition systems often struggle with limited datasets, difficulty in capturing subtle emotional cues, and challenges posed by variations in lighting, occlusions, and facial expressions. By integrating GANs into the recognition process, the aim is to address these limitations.

Firstly, GANs can augment existing datasets by generating synthetic facial expressions, thereby increasing the diversity and size of the training data. This augmentation helps in mitigating the data scarcity problem and improves the model's ability to generalize across different facial expressions and demographics. Secondly, GANs can generate realistic and varied facial expressions, including subtle emotional cues, which are often difficult to capture in real-world datasets. This enables the model to learn more nuanced representations of emotions, enhancing its accuracy in recognizing complex emotional states.

CHAPTER 3

LITERATURE REVIEW

The confluence of cutting-edge technology and the growing need for accurate and effective emotion recognition tools in robotics have led to notable breakthroughs in the facial emotion recognition in recent years. This review of the literature offers a thorough description of the state of facial image analysis today, highlighting important techniques, difficulties, and new developments. Through a critical examination of existing literature, we aim to contribute to the ongoing dialogue in this dynamic field and inspire further advancements that hold the potential to redefine the landscape of recognition emotion in human-robot interaction.

In 2018, Anwar et al. [1] found that all emotions recognition, including classification, detection, and segmentation, are finding growing acceptance of convolutional neural network based deep learning approaches. Techniques like data augmentation and transfer learning are used to solve the issues with deep learning methods caused by a lack of data and few labels. Higher performance is becoming possible for larger datasets because to improved deep learning architectures and more computational power availability. In 2022, Rana and Bhushan et.al [2] provided a summary of the many ML and DL methods for emotion recognition, as well as information on categorization, imaging modalities, tools, methodologies, datasets, and difficult- ties in the field. Dominated all the tools and approaches tested. It was noted that researchers make extensive use of the FER dataset. Additionally, a comparison of ML classifiers and DL models using the FER2013 dataset has been made possible through a series of trials, in which CNN and GAN have surpassed other approaches. According to this study, denoising techniques with DL models should be used in the robotic industry. Additionally, it suggests that a variety of traditional ML and DL approaches are widely applied to deal with data uncertainty. In 2020,Bria et al. [3] showed that DC-CNN performed better than CNNs trained using techniques such hard mining, cost-sensitive learning, one-class classification, oversampling, and under sampling to handle class imbalance. During the test, the DC-CNN outperformed CNN in terms of speed. In 2002, Chawla et al. [4] suggested that greater classifier performance (in ROC space) may be obtained by combining our technique of under sampling the majority class (normal) and oversampling the minority class (abnormal). In 2020, Gao et al. [5] Proposed to use the idea of image complexity to train CNN models by using self-perturbation of the provided samples from a single class to capture intrinsic and discriminative imaging characteristics. Comparing the suggested technique with previous one-class models, it performs better on all datasets, including internal and public ones, after comprehensive evaluation. In datasets ranging all action unit and across several picture modalities, encouraging findings are displayed. In 2017, Wang et al. [6] summarized the state of the art of GANs. They started by surveying the application's theoretical and implementation models, and background of GANs proposal. The benefits and drawbacks of GANs as well as their development patterns were then covered. They specifically looked at the connection between GANs and parallel intelligence, and came to the conclusion that GANs have a lot of promise for virtual-real interaction and integration in parallel systems research. It is evident that GANs may offer significant algorithmic assistance for parallel intelligence.

CHAPTER 4

PLATFORM

The implementation phase of the project involved setting up the environment, loading the necessary datasets, and building and training the CNN and Efficient net models for skin cancer detection.

4.1 Hardware

The system used for the experiments has the following configuration:

- Processor: Intel Core i5 11th Generation 2.50 GHz – 4.0 GHz
- RAM: 8/16 GB
- System Type: 64-bit
- Hard Disk: 512 GB SSD

4.2 Software Environment

The following software tools were used:

- IDE: Google Collaboratory, VS Code, Jupyter
- Programming Language: Python
- Operating System: Windows 11

4.3 Libraries

The following libraries were used:

- NumPy: Fundamental package for scientific computing with Python.
- TensorFlow: Open-source machine learning framework for training and deploying models.
- Matplotlib: Comprehensive library for creating static, animated, and interactive visualizations in Python.
- OpenCV: Library for computer vision tasks.
- Pandas: Data manipulation and analysis library providing data structures and functions needed to work on structured data seamlessly.

CHAPTER 5

PROPOSED METHODOLOGY

The methodology for Facial Emotion Recognition involves a structured approach that includes data collection and preprocessing, model architecture selection, training, and evaluation. High-quality facial emotion images are gathered and standardized through resizing, normalization, and augmentation. The CNN model, chosen for its balance of accuracy and computational efficiency, is fine-tuned for facial emotion recognition. And they can see the flow diagram of methodology in figure 5.1. The training process involves optimizing hyperparameters and applying regularization techniques to enhance model generalization. Performance is evaluated using metrics like accuracy, precision, recall, and ensuring a robust and efficient system for accurate facial emotion recognition.

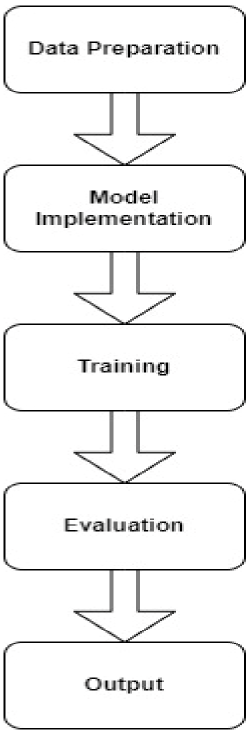

Fig 5.1 Proposed Methodology

5.1 Data preparation

Data preparation involves collecting a diverse dataset of facial emotion images, ensuring it includes various types of emotions. The images are preprocessed to enhance quality, including resizing, normalization, and augmentation to address data imbalance. Additionally, the dataset is split into training, validation, and test sets to facilitate model evaluation and performance tuning.

5.1.1 Data Description

The research design outlines the overall strategy and approach adopted to achieve the objectives of the study. In this Facial emotion research, a systematic methodology was employed to develop and evaluate a recognition model based on CNN architecture. The design encompasses data collection, preprocessing, model development, training, evaluation, and validation.

Fig 5.2 Data Distribution Among Classes

In figure 5.2 Initially, the dataframe contains 35887 samples distributed across seven classes, with the following distribution: class 0 has 4953 samples, class 1 has 547 samples, class 2 has 5121 samples, class 3 has 8989 samples, class 4 has 6077 samples, class 5 has 4002 samples, and class 6 has 6198 samples.

5.1.2 Data Generation

High-quality facial emotion images are generated and subjected to data generation1 and augmentation techniques, such as GAN, to enhance the dataset's diversity and improve model robustness. The GAN model, known for its generating data, is fine-tuned specifically for class imbalance issue. During training, careful optimization of hyperparameters and application of regularization techniques are performed to prevent overfitting and ensure generalization.

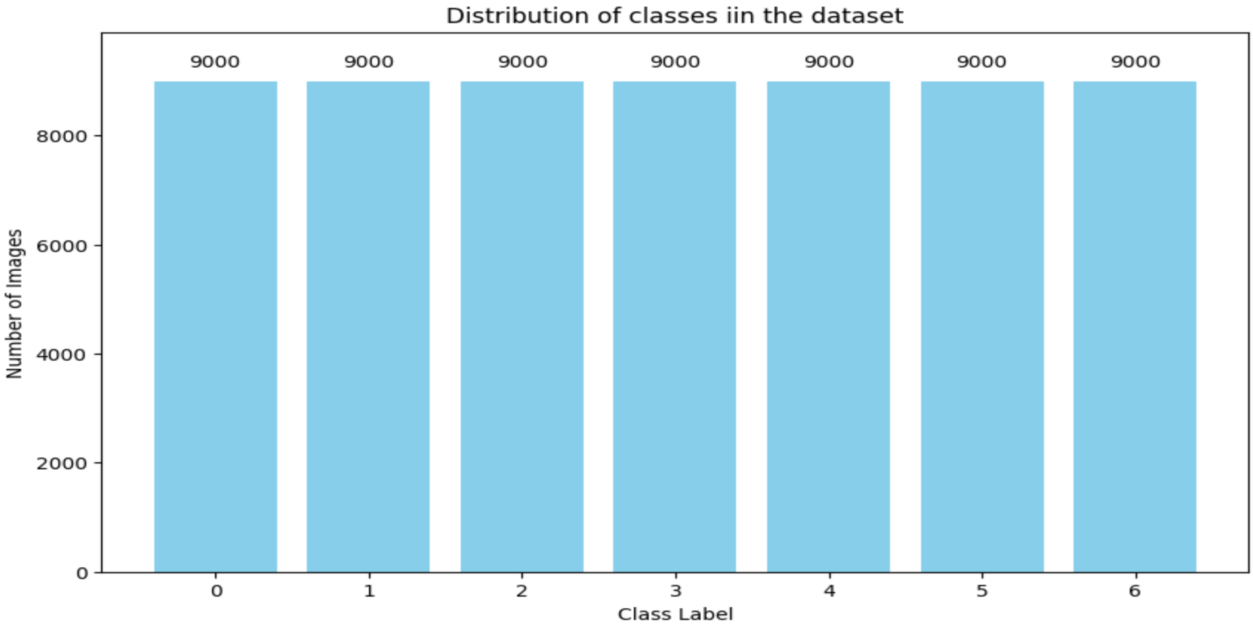

Fig 5.3 Balance Dataset

Initially, the dataframe contains 35887 samples distributed across seven classes, indexed as ['0', '1', '2', '3', '4', '5', '6']. After trimming the dataset to address class imbalance, the maximum number of samples per class is capped at 9000, ensuring a more balanced representation. Consequently, the minimum number of samples in any class increases to 10. As a result of this trimming process, the dataframe is reduced to 3648 samples while still maintaining the original seven classes. This adjustment aims to create a more equitable distribution of data across classes, facilitating more robust training and evaluation of machine learning models for skin cancer detection and we can see in figure 5.3.

5.2 Model Architecture of CNN

The architecture of a Convolutional Neural Network (CNN) typically consists of several layers, each designed to extract, and process features from input data as Shawn in figure 5.4. The fundamental components of a CNN architecture include:

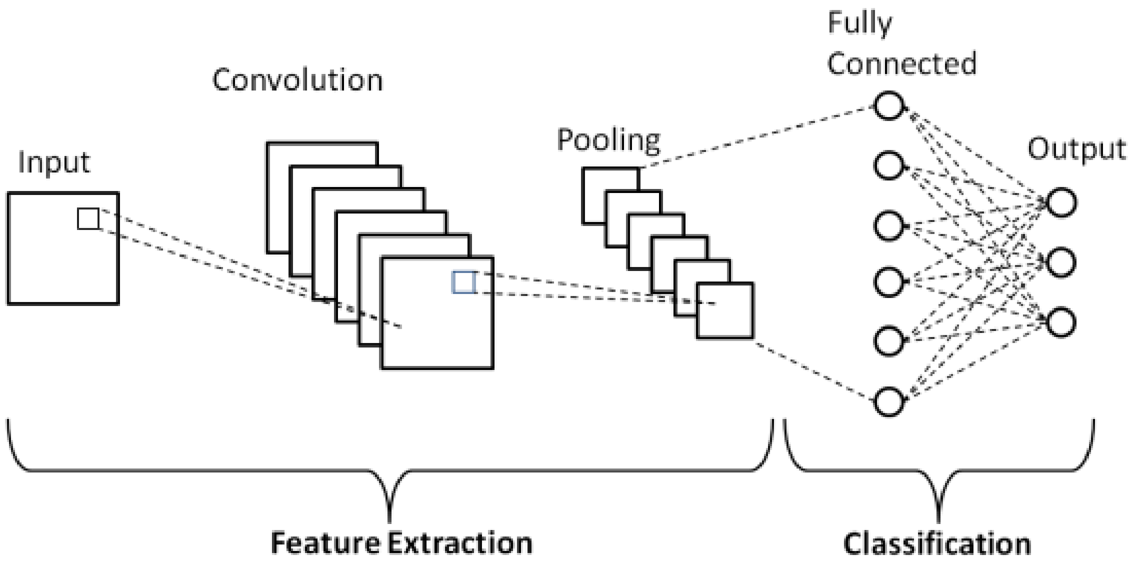

Fig 5.4 Architecture Of CNN

1. Input Layer: This layer receives the raw input data, such as images, and passes it to the subsequent layers for processing.

2. Convolutional Layers: These layers consist of filters or kernels that convolve across the input data, extracting features through a series of learned convolutional operations. Each filter detects specific patterns or features in the input data.

3. Activation Function: Following each convolutional operation, an activation function like ReLU (Rectified Linear Unit) is applied elementwise to introduce non-linearity into the network, enabling it to learn complex patterns.

4. Pooling Layers: Pooling layers down sample the feature maps generated by the convolutional layers, reducing their spatial dimensions. Common pooling operations include max pooling and average pooling, which retain the most significant features while reducing computational complexity.

5. Fully Connected Layers: Also known as dense layers, these layers connect every neuron in one layer to every neuron in the next layer. They perform classification based on the features extracted by the previous layers.

6. Output Layer: The final layer of the CNN produces the network's predictions. The number of neurons in this layer corresponds to the number of classes in the classification task. Activation functions like SoftMax are often used to convert raw output into probabilities for each class.

Overall, the architecture of a CNN is designed to efficiently learn hierarchical representations of the input data, enabling it to effectively perform tasks such as image classification, object detection, and segmentation.

5.3 Model Architecture: GAN

Generative Adversarial Networks are a class of deep learning architectures designed to generate synthetic data samples that closely resemble real data from a given distribution. The fundamental principle behind GANs involves training two neural networks simultaneously: a generator and a discriminator. The generator learns to produce realistic data samples by transforming random noise or input vectors into meaningful outputs, while the discriminator learns to distinguish between real data samples and those generated by the generator. During training, the generator aims to deceive the discriminator by producing high-quality samples that are indistinguishable from real data, while the discriminator aims to correctly classify between real and fake samples.

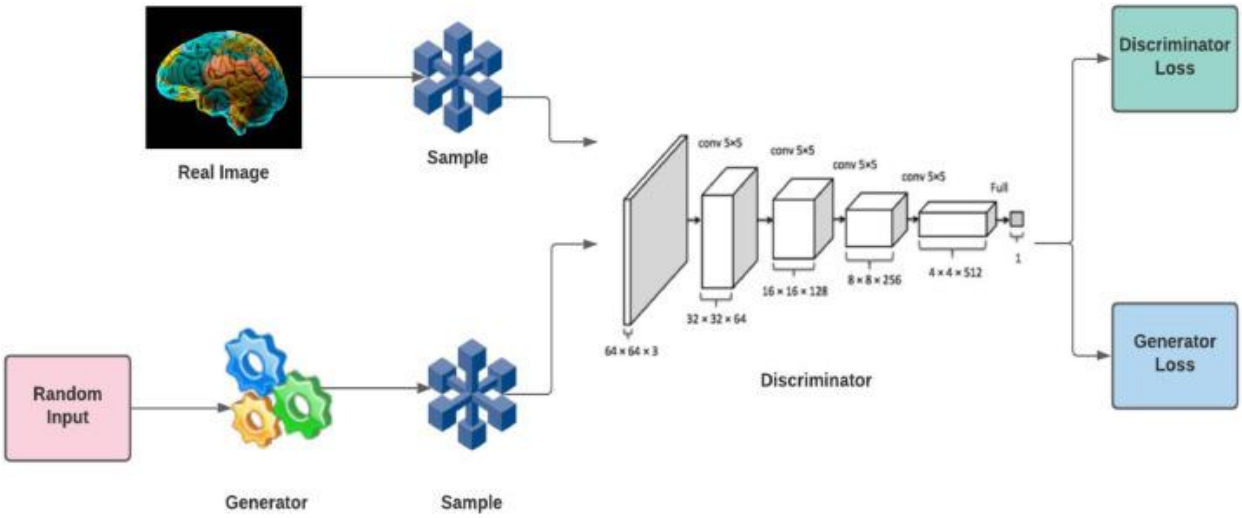

Fig 5.5 Architecture of GAN

Generator Model:

A key element responsible for creating fresh, accurate data in a GAN is the generator model. The generator takes random noise as input and converts it into complex data samples, such text or images. It is commonly depicted as a deep neural network.

The training data’s underlying distribution is captured by layers of learnable parameters in its design through training. The generator adjusts its output to produce samples that closely mimic real data as it is being trained by using backpropagation to fine-tune its parameters. The generator’s ability to generate high-quality, varied samples that can fool the discriminator is what makes it successful.

Discriminator Model:

An artificial neural network called a discriminator model is used in GAN to differentiate between generated and actual input. By evaluating input samples and allocating probability of authenticity, the discriminator functions as a binary classifier. Over time, the discriminator learns to differentiate between genuine data from the dataset and artificial samples created by the generator. This allows it to progressively hone its parameters and increase its level of proficiency. Convolutional layers or pertinent structures for other modalities are usually used in its architecture when dealing with picture data. Maximizing the discriminator’s capacity to accurately identify generated samples as fraudulent and real samples as authentic is the aim of the adversarial training procedure. The discriminator grows increasingly discriminating as a result of the generator and discriminator’s interaction, which helps the GAN produce extremely realistic-looking synthetic data overall.

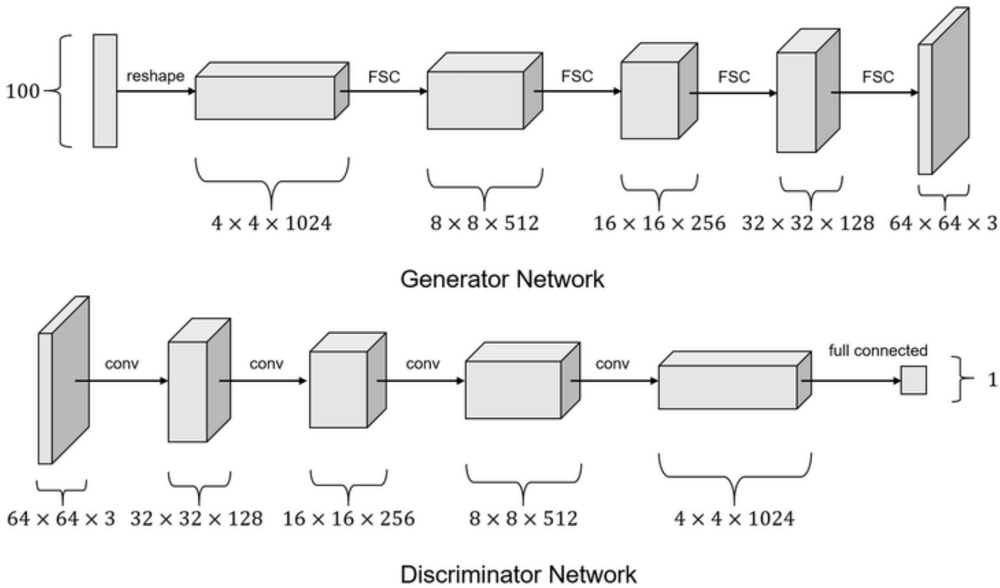

Fig 5.6 Generator and Discriminator Architecture

5.4 Training Procedure

The training process involves feeding the generated images from GAN into the CNN model, utilizing labeled data to learn distinguishing features among the seven classes. The model undergoes iterative optimization using techniques such as backpropagation and gradient descent, aiming to minimize the loss function. Regular validation is performed to monitor performance and adjust hyperparameters, ensuring the model generalizes well to new, unseen data.

5.4.1 Dataset Splitting

The curated FER2013 dataset is split into training, validation, and test sets. Approximately 80% of the data is allocated for training, 10% for validation, and 10% for testing. This division ensures a robust model evaluation on unseen data while allowing for effective training.

5.4.2 Data Loading

Data loading involves the efficient retrieval of images from the dataset, and the augmentation techniques described in Chapter 4 are applied in real-time during training. This dynamic augmentation ensures variability in the training data, enhancing the model's ability to generalize. Figure 5.2 (Input data contain seven different class of facial emotion name angry, disgust, fear, happy, sad, surprise, neutral with label starting from 0 to 6).

Fig 5.6 Input Image

5.4.3 Training GAN:

Train a GAN on the preprocessed dataset. The GAN consists of two neural networks: the generator, which synthesizes realistic facial images, and the discriminator, which distinguishes between real and fake images. Train the GAN iteratively until the generator produces realistic facial expressions that are difficult for the discriminator to differentiate from real images.

5.4.4 Image Augmentation:

Image Augmentation: Generate synthetic facial expressions using the trained GAN. This involves sampling random noise vectors as input to the generator and obtaining corresponding synthetic facial images. Augment the original dataset by adding these synthetic images, effectively increasing its diversity.

5.4.5 FER Model Training:

Train the Facial Emotion Recognition model using the augmented dataset. This model can be based on deep learning architectures such as Convolutional Neural Networks (CNNs). The model should learn to map facial features to corresponding emotion labels accurately.

5.5 Evaluation Metrics

Evaluation metrics for GAN assess the quality of generated samples and the stability of training. Common metrics include Inception Score (IS), which measures the quality and diversity of generated images based on their realism and diversity of classes, and Fréchet Inception Distance (FID), which quantifies the similarity between the distribution of real and generated images' feature representations.

Evaluate the trained FER model on a separate validation or test dataset to assess its performance in recognizing facial expressions. Use metrics such as accuracy, precision, recall, and F1-score to measure the model's effectiveness.

5.6 Tools and Technologies Used

The Facial Emotion Recognition research utilizes various tools and technologies to facilitate model development, training, and evaluation. Popular deep learning frameworks such as TensorFlow or Pytorch are employed for implementing the GAN architecture and conducting experiments. Additionally, libraries for data manipulation, visualization, and performance evaluation are utilized to streamline the research process.

CHAPTER 6

RESULTS AND DISCUSSIONS

The implementation details section provides insights into the practical aspects of developing and deploying the FER model based on GAN and CNN architecture. It includes information about the programming languages, libraries, and frameworks used for implementation. Additionally, details about the model's architecture, optimization techniques, and deployment considerations are discussed.

6.1 Performance Evaluation

Performance evaluation serves as a vital yardstick for measuring the effectiveness of the skin detection model. We employed a range of quantitative metrics such as precision, recall, F1-score, accuracy, and AUC-ROC on the test dataset to gauge the model's performance. Additionally, qualitative assessments, involving visual inspection of predicted skin regions and comparison with ground truth labels, were conducted to provide deeper insights into the model's efficacy.

6.2 Visualization of Results

The procedure for generating images involves the generator network synthesizing images from random noise vectors. These vectors serve as input to the generator, which transforms them into plausible images. During training, the generator learns to produce images that are indistinguishable from real images to the discriminator network. The discriminator, in turn, provides feedback to the

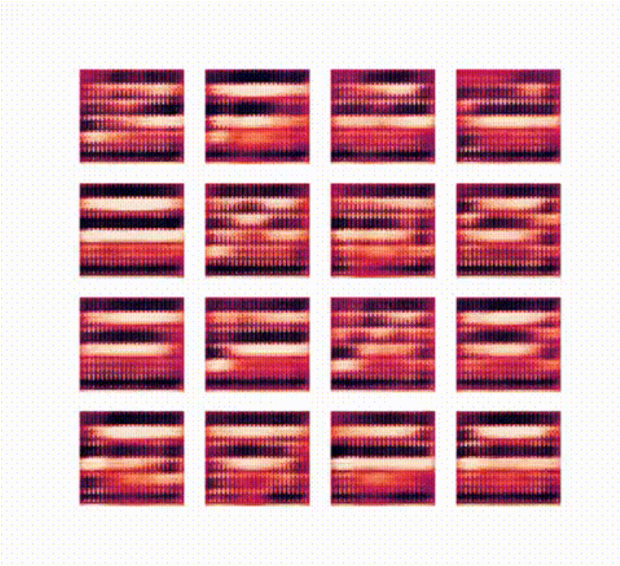

Fig 6.1 Generated Images

generator by distinguishing between real and generated images. Through an adversarial training process, the generator improves its ability to produce increasingly realistic images. Here are some generated images. In Figure 6.1, we observe the training progression of our CNN-based model. The graph depicts a gradual increase in accuracy over the training epochs, culminating in an average accuracy of 77.03%. Simultaneously, the loss metric exhibits a steady decline, indicative of the model's effective learning and continuous improvement in performance.

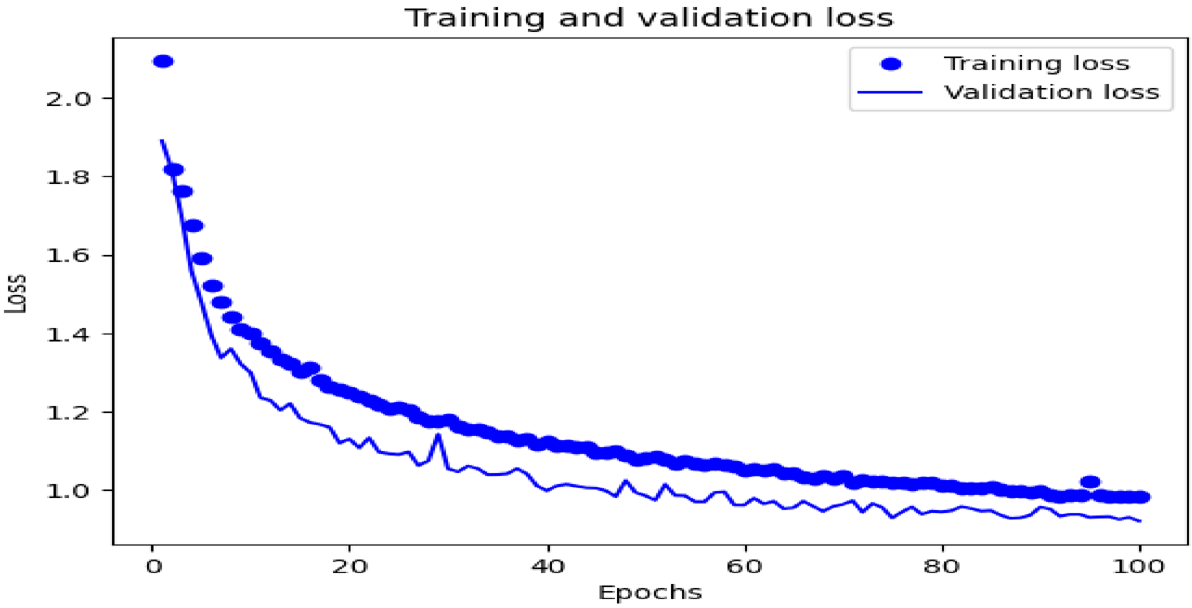

Fig 6.2 Loss of FER_MODEL

Figure 6.2 showcases the training trajectory of our FER CNN model. Like the CNN model, we observe a progressive increase in accuracy throughout the training process, eventually reaching an average accuracy of 63.39%. The consistent decrease in loss underscores the model's adeptness in learning and enhancing its performance over time.

6.3 Discussion of Results

The discussion of results delves into a thorough analysis of the findings derived from the performance evaluation and comparative analysis. We elucidate insights into various factors influencing the model's performance, including dataset quality, model architecture, and training methodologies. We explore the real-world implications of our findings, acknowledge the limitations of our approach, and propose potential for future research. The discerned accuracy rates of 63.03% for the CNN model and 85%(Approximate) after augmentation.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Facial Emotion Recognition (FER) using GAN presents a promising approach to enhancing the accuracy and robustness of emotion detection systems. By leveraging GANs to generate synthetic facial images, we can augment existing datasets, address data scarcity, and improve model generalization. This method allows for better recognition of subtle and complex emotions, even under challenging conditions such as varying lighting, occlusions, and diverse facial expressions. The integration of GANs with FER systems has demonstrated significant potential in advancing the state-of-the-art in emotion recognition, contributing to more empathetic and responsive human-computer interactions across various applications, including healthcare, education, market research, and security.

GAN have emerged as powerful tools in the field of medical imaging, offering innovative solutions to various challenges. The ability of GANs to generate realistic and high-quality synthetic images has made them particularly valuable in addressing issues related to limited or imbalanced medical image datasets. The incorporation of GANs into the medical imaging domain reflects a broader trend of utilizing deep learning techniques for enhanced diagnostic and analytical capabilities. While GANs present exciting opportunities, addressing challenges such as model interpretability, data quality, and ethical considerations is crucial to ensure responsible and effective deployment in healthcare settings. Ongoing research and development in this field are likely to establish GANs as pivotal tools in advancing the analysis and interpretation of medical images.

REFERENCES

[1] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in image analysis," *Annual review of joseph engineering*, vol. 19, pp. 221–248, 2017.

[2] M. AlAmir and M. AlGhamdi, "The role of generative adversarial network in image analysis: An in-depth survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–36, 2022.

[3] Y. Chen, X.-H. Yang, Z. Wei, A. A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, and Q. Guan, "Generative adversarial networks in image augmentation: A review," *Computers in Biology and Medicine*, vol. 144, p. 105382, 2022.

[4] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Gan-based synthetic image augmentation for increased CNN performance in fashion amidst dataset classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231218310749

[5] Y. Skandarani, P.-M. Jodoin, and A. Lalande, "Gans for image synthesis: An empirical study," *Journal of Imaging*, vol. 9, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2313-433X/9/3/69

[6] J. Cheng, S. Tian, L. Yu, C. Gao, X. Kang, X. Ma, W. Wu, S. Liu, and H. Lu, "Resganet: Residual group attention network for image classification and segmentation," *Image Analysis*, vol. 76, p. 102313, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841521003583