Conversion of Iris Image from Off-Angle to Frontal Using Hybrid Attention and Multi-Scale-Discriminator GAN

Hamba Abebe Shere

A Thesis Submitted to the Department of Computer Science and Engineering,
College of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

June, 2025

Adama, Ethiopia

Conversion of Iris Image from Off-Angle to Frontal Using Hybrid Attention and Multi-Scale-Discriminator GAN

Hamba Abebe Shere

Advisor: Bahiru Shifaw Yimar (Ph.D)

A Thesis Submitted to the Department of Computer Science and Engineering, College of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

June, 2025

Adama, Ethiopia

DECLARATION

I declare that this Master's Thesis, titled "**Conversion of Iris Image from Off-Angle to Frontal Using Hybrid Attention and Multi-Scale-Discriminator GAN,**" is my original work and has not been submitted to any academic institution for similar purposes. All the sources of information used in this thesis have been properly acknowledged through citations.

_____	_____	_____
Name of student	Signature	Date

RECOMMENDATION OF ADVISORS

I, the advisor of this, hereby certify that I have read the revised copy of the thesis titled "**Conversion of Iris Image from Off-Angle to Frontal Using Hybrid Attention and Multi-Scale-Discriminator GAN"** prepared under my guidance by **Hamba Abebe Shere** submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science and Engineering. Hence, I recommend the submission of the revised version of the thesis to the department following the applicable procedures.

Major Advisor	Signature	Date

APPROVAL SHEET

I, being the thesis advisor of "**Conversion of Iris Image from Off-Angle to Frontal Using Hybrid Attention and Multi-Scale-Discriminator GAN**" by **Hamba Abebe Shere**, hereby confirm that the suggestions and comments of the board of examiners are appropriately incorporated in the final version of the thesis.

_____	_____	_____
Major Advisor	Signature	Date

APPROVAL OF THE BOARD OF REVIEWERS

We, the undersigned individuals whose signatures appear below, being members of the Board of Reviewers for the open defense of the thesis of **Hamba Abebe Shere**, have thoroughly read and evaluated the thesis titled "**Conversion of Iris Image from Off-Angle to Frontal Using Hybrid Attention and Multi-Scale-Discriminator GAN**". Thus, we hereby certify that the thesis is acceptable as partial fulfillment of the requirements for the degree of Master of Science in Computer Science and Engineering.

Chairperson	Signature	Date
Internal Examiner	Signature	Date
External Examiner	Signature	Date

Final approval and acceptance of the thesis is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPDS) through the Department Graduate Council (DGC) and College Graduate Committee (CGC).

Department Head	Signature	Date
College Dean	Signature	Date
Office of Postgraduate Studies, Dean	Signature	Date

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to the Almighty God for blessing me with the strength, wisdom, determination, and persistence necessary to pursue this study.

I want to extend my warmest gratitude to my advisor, Bahiru Shifaw Yimar (Ph.D), for his expert guidance, help, and mentorship throughout this study. His expertise and positive criticism have greatly contributed to the successful completion of this study.

I want to express my heartfelt thanks to my family for their unconditional love, encouragement, and belief in my abilities. Their continuous support and understanding have facilitated my academic journey. I am grateful to my classmates for their encouragement, motivation, and support throughout the study process. Their presence has been a great source of inspiration, making this process more enjoyable.

Lastly, I would like to express my gratitude to all the individuals, institutions, and organizations that have been associated with this study directly or indirectly. Their assistant has played a crucial role in the successful completion of this study.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

LIST OF SAMPLE CODE SNIPPETS ix

LIST OF ACRONYMS x

ABSTRACT xi

CHAPTER ONE 1

1. INTRODUCTION 1

1.1 Background of the Study 1

1.2 Motivation of the Study 3

1.3 Problem Statement of the Study 3

1.4 Research Questions 4

1.5 Objective of the Study 5

1.5.1 General Objective 5

1.5.2 Specific Objectives 5

1.6 Significance of the Study 5

1.7 Scopes and Limitations of the Study 6

1.7.1 Scopes of the Study 6

1.7.2 Limitations of the Study 6

1.8 Contributions of the Study 6

1.9 Organizations of the Study 7

CHAPTER TWO 8

2. LITERATURE REVIEW AND RELATED WORKS 8

2.1 Introduction 8

2.3 Challenges in Off-Angle Iris Recognition 9

2.4 Methods for Off-Angle to Frontal Iris Conversion 9

2.4.1 Geometric Transformation-Based Methods 10
2.4.2 Image Transformation Models 10
2.4.3 Deep Learning Based Approaches 10
2.5 Generative Adversarial Network (GAN) 12
2.5.1 Conditional GAN 13
2.5.2 CycleGAN 14
2.5.3 Deep Convolutional GAN (DCGAN) 14
2.5.4 Pixel-to-Pixel GAN (Pix2PixGAN) 15
2.6 Hybrid Attention Mechanism 16
2.6.1 Self-Attention Mechanism 17
2.6.2 Channel attention Mechanism 17
2.7 Multi-Scale Discriminator 18
2.8 U-net 19
2.9 Related Works 20
2.10 Summary of Related Works 22
CHAPTER THREE 24
3. RESEARCH METHODOLOGY 24
3.1 Overview 24
3.2 Dataset Acquisition 24
3.2.1 University of Beira Interior Iris Recognition Version 2 (UBIRIS.v2) Dataset 25
3.2.2 MOBIUS (Mobile Ocular Biometrics in Unconstrained Settings) Dataset 26
3.3.1 Data Augmentation 27
3.3.2 Normalization 28
3.3.3 Resize 28
3.4 Development Tools 28
3.4.1 Hardware Tools 28

3.4.2 Software Tools 28
3.5 Baseline Works 29
3.6 Feature Extraction Network 30
3.7 Performance Evaluation Metrics 31
3.7.1 PSNR (Peak Signal to Noise Ratio) 31
3.7.2 SSIM (Structural Similarity Index Measure) 31
3.7.3 Hamming Distance 32
CHAPTER FOUR 33
4. PROPOSED ARCHITECTURE 33
4.1 Overview of Chapter 33
4.2 Model Architecture 33
4.3 Generator Architecture 35
4.4 Discriminator Network 38
4.5 Model Learning Functions 40
4.5.1. Adversarial loss 40
4.5.2. Identity Loss 41
4.5.3. Perceptual loss 41
4.6 Training Pseudo Code 42
CHAPTER FIVE 43
5. IMPLEMENTATION OF PROPOSED MODEL 43
5.1 Chapter Overview 43
5.2 Working Environments 43
5.3 Environmental Setup 44
5.4 Training Procedure 45
5.5 Dataset Preparation 45
5.5.1 Data Splitting and Loading 45

5.5.2 Dataset Augmentation Implementation 46
5.6 Implementation of Proposed Model 46
5.6.1 Implementation of Self-attention 49
5.6.2 Implementation of Channel Attention 50
5.6.3 Implementation of Multi-Scale Discriminator 51
5.6.4 Implementation of Adversarial Loss Function 52
5.6.5 Implementation of Identity Loss Function 53
5.6.6 Implementation of Perceptual Loss Function 53
5.7 Hyperparameter Configuration 54
5.8 Experimental Class 55
CHAPTER SIX 58
6. RESULTS AND DISCUSSION 58
6.1 Overview of Chapter 58
6.2 Experimental Results 58
6.3. Results based on Evaluation Metrics 58
6.3.1 Evaluation of Experiments on MOBIUS Dataset 59
6.3.2 Evaluation of Experiments on UBIRIS.V2 Dataset 59
6.4 Converted Off-angle Iris Image to Frontal View: Results of Experiments 61
6.4.1 Converted off-angle iris image Results of Experiments for UBIRIS.v2 dataset 61
6.4.2 Converted off-angle iris image Results of Experiments for MOBIUS dataset 62
6.4.3 Pix2Pix-GAN 63
6.4.4 Pix2Pix-GAN+SA+CA 63
6.4.5 Pix2Pix-GAN+SA+CA+MSD 64
6.4.6 Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss} 64
6.5 Sample Training Log for HAM-MSD-GAN 65
6.6 Results Discussion 68

6.7 Research Question and Answer Discussion 69
CHAPTER SEVEN 71
7. CONCLUSION AND FUTURE WORKS 71
7.1 Conclusion 71
7.2 Future Works 73
REFERENCES 75
Appendix A: MOBIUS Dataset Test Result i
Appendix B: UBIRIS.v2 Dataset Test Result ii
Appendix C: Sample Code iii
Appendix D: Model Training Complexity Analysis iv

LIST OF TABLES

Table 2.1: Summary of Related Works 22
Table 3.1: Hardware Requirements 28
Table 5. 1: Content of encoder block in generator architecture 47
Table 5. 2: Content of bottleneck layer 47
Table 5. 3: Content of decoder block in generator architecture 48
Table 5. 4: Content of multi-scale discriminator architecture 49
Table 5.5: Hyperparameter Configuration 55
Table 5.6: Experiment class 56
Table 6.1: PSNR, SSIM, and Hamming Distance evaluation on MOBIUS Dataset 59
Table 6.2: PSNR evaluation on UBIRIS.V2 Dataset 59
Table 6.3: SSIM evaluation on UBIRIS.V2 Dataset 60
Table 6.4: Hamming Distance evaluation on UBIRIS.V2 Dataset 60

LIST OF FIGURES

Figure 2. 1: Off-Angle to Frontal Iris Image Conversion Process 9
Figure 2. 2: General Architecture of GAN 12
Figure 2. 3: CGAN architecture 13
Figure 2. 4: CycleGAN Architecture 14
Figure 2. 5: DCGAN Architecture 15
Figure 2. 6: Pix2PixGAN Architecture 16
Figure 2.7: Self-attention GAN module 17
Figure 2.8: Channel attention module 18
Figure 2. 9: Network architecture of the multi-scale discriminator 19
Figure 2. 10: U-Net Architecture 20
Figure 3.1: Overall process 24
Figure 3. 2: UBIRIS.v2 (Padole & Proenca, 2012) dataset samples 26
Figure 3. 3: MOBIUS (Vitek et al., 2020) dataset samples 27
Figure 3. 4 Mobile Net V2 architecture 30
Figure 4.1: Overall architecture of the proposed model 34
Figure 4. 2: Generator architecture 36
Figure 4. 5 Architecture of Multi-scale discriminators 39
Figure 6.3: Result of Pix2Pix-GAN 63
Figure 6.4: Result of Pix2Pix-GAN+SA+CA 63
Figure 6.5: Result of Pix2Pix-GAN+SA+CA+MSD 64
Figure 6.6: Result of Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss} 64
Figure 6.8: Discriminator loss during training 66
Figure 6.9: Adversarial loss during training 67

LIST OF SAMPLE CODE SNIPPETS

Sample code 5.1: Dataset Loading and Splitting implementation 46
Sample code 5.2: Dataset augmentation implementation 46
Sample code 5.3: Self-attention Implementation 50
Sample code 5.4: Channel Attention Implementation 51
Sample code 5.5: Multi-scale Discriminator implementation 52
Sample code 5.7: Identity Loss implementation 53

LIST OF ACRONYMS

2D	Two-Dimensional
3D	Three-Dimensional
CA	Channel Attention
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HAM-MSD	Hybrid Attention and Multi-Scale Discriminator
MOBIUS	Mobile Ocular Biometrics in Unconstrained Settings
PSNR	Peak Signal-to-Noise Ratio
Pix2Pix-GAN	Pixel-to-Pixel Generative Adversarial Network
PatchGAN	Patch Generative Adversarial Network
RAM	Random Access Memory
ReLU	Rectified Linear Unit
SA	Self-Attention
SSIM	Structural Similarity Index Measure
UBIRIS	University of Beira Interior Iris Recognition Version 2

ABSTRACT

Iris recognition has extensive applications in border control, healthcare, mobile authentication, and security systems. In real-world situations, the captured iris images often suffer from distortions, occlusions (by eyelids or blinking), and off-angle captures due to acquisition angle and the eye's 3D structure, leading to information loss. These off-angle images significantly reduce recognition accuracy, making frontal view reconstruction a critical preprocessing step. Although GANs have shown promise in this area, they struggle to handle extreme gaze angles (greater than 30˚), occlusions caused by eyelids or blinking, and distortions arising from the 3D structure of the eye. To address these challenges, this study proposed a GAN-based approach that integrates attention mechanisms and a multi-scale discriminator to convert off-angle iris images to high-quality frontal views. The generator is specifically based on a U-Net architecture improved by introducing channel attention in the encoder and self-attention in the decoder. Meanwhile, the discriminator utilizes a multi-scale approach to capture local and global features more effectively. The proposed method is tested on the MOBIUS and UBIRIS.v2 datasets and compared with baseline models. Quantitative measurement of the quality of the converted iris images was done using metrics such as Peak Signal-to-Noise Ratio, Structural Similarity Index Measure, and Hamming Distance. Our proposed model significantly outperforms the baseline Pix2Pix-GAN across these metrics. For PSNR, it improves from 28.00 to 30.34 (−30°), 28.01 to 30.31 (30°), 22.97 to 27.99 (−45°), and 23.97 to 28.98 (45°). SSIM improves from 0.69 to 0.79, 0.70 to 0.80, 0.61 to 0.76, and 0.60 to 0.75, respectively. Hamming Distance is reduced from 0.47 to 0.34, 0.44 to 0.36, 0.49 to 0.42, and 0.50 to 0.44. Additionally, our model achieves overall improvements with PSNR rising from 18.75 to 28.94, SSIM from 0.61 to 0.82, and achieves a lower Hamming Distance of 0.44 from 0.97 for converting occluded iris images caused by eyelids, partially visible iris images, and distorted iris images caused by its 3D structure of the eye or due to a large angle to frontal iris images. These results indicate that our model is more effective at handling large-angle up to 45^0, partially occluded, and structurally distorted iris images, making it a valuable advancement for iris recognition and surveillance systems that require accurate frontal iris representations.

Keywords: Conversion of Iris Image from Off-Angle to Frontal, Generative Adversarial Network, Hybrid Attention, Multi-Scale discriminator

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the Study

Iris recognition is a robust biometric identification technique that is characterized by utilizing unique patterns of the iris for allowing accurate identification in sensitive fields such as border control and medicine, mobile technologies, and security systems. The majority of systems in use today require frontal views, which might cause issues when participants turn their eye's away, resulting in low-quality photographs and the need for drawn-out recaptures. This emphasizes the necessity of an iris recognition model that is more flexible. Technological developments in this field would significantly improve dynamic biometric systems, enabling quicker identification and improved tracking of people, especially potential criminals.

Traditional iris recognition algorithms perform effectively under circumstances where high-resolution frontal images are accessible within controlled settings. Nevertheless, in suboptimal conditions, such as variations in pupil size, iris occlusion, and gaze angle, their performance decreases (Wang and Kumar, 2019). However, in more flexible environments, the introduction of standoff biometric systems enables the identification of cooperative and non-cooperative types (Nguyen et al., 2017). These systems provide users with increased flexibility during data collection, permitting photographs to be taken from a distance, even when the subjects are in motion. The resultant images are often less than ideal and may be captured from angles that are not frontal because individuals are not compelled to look directly at the camera. As a result, traditional iris recognition systems can sometimes misinterpret their similarity scores and mistakenly identify them as images of different people when compared to frontal photographs of the same individual.

These traditional frameworks tend to employ a methods similar to Daugman's original model (Daugman, 2007). Iris recognition generally consists of four fundamental steps: first, segmenting the inner and outer iris boundaries; second, transforming the segmented image to be dimensionless rectangular polar coordinates; thirdly, the representation of the binary attributes obtained from the standardized images in order to generate iris codes; and lastly, comparing the iris codes with enrolled subjects by estimating the Hamming distances. However, these systems are heavily reliant

on image quality and tend to be quite rigid in adapting to variations in iris texture; frequently, some parameters need to be adjusted for different datasets. Additionally, off-axis iris images pose new challenges for biometric recognition: limbus obstructions from freckles, eyebrows, or eyelashes affect the iris's distinctive features, blur results from a loss of depth of field, and three-dimensional iris textures are created by the cornea's refraction of light (Karakaya, 2016). As a result, off-angle iris biometrics has become a prominent research area focused on addressing these challenges.

Early attempts at improving the recognition accuracy in conventional iris recognition systems have employed techniques such as elliptical unwrapping and perspective transformation (Schuckers et al.,2007). Although these techniques aim to address perspective distortions, their effectiveness significantly declines when they are applied to images where the gaze angles exceed 30 degrees. Present algorithms struggle to confirm that two images depict the same individual, highlighting the need to confront various degradation factors related to off-angle images while creating robust standoff biometric recognition systems. Motivated by the accomplishments of the convolutional neural network in computer vision, researchers have employed deep learning techniques to tackle intricate challenges in fields like computer vision and speech recognition. In the area of iris biometrics, substantial research has been conducted on iris segmentation (Hofbauer et al., 2019), spoof detection (Menotti et al. , 2015), and enhancement of recognition efficacy (Gangwar and Joshi, 2016). The aforementioned deep learning techniques have led to advancements over traditional methods in feature extraction from iris textures but have not specifically tackled challenges that are distinct to off-angle images and lack extensive results.

Recent studies on iris recognition present significant potential, particularly the application of Generative Adversarial Networks (GANs), in converting off-angle iris images into frontal views. Modified Pix2Pix-based models combined with specific loss functions have emphasized the importance of maintaining fine-grained iris textures and thus effectively addressing problems such as corneal refraction and pupil dilation (Kota and Karakaya, 2023a). Nonetheless, obstacles persist in capturing fine-grained and multi-scale features from a partially visible iris image because of occlusions from eyelids or blinks, as well as a distorted iris image resulting from large gaze angle or due to 3D structure of eye. Therefore, we introduced a novel hybrid attention and multi-scale discriminator-based GAN for improved accuracy, enhanced detail in iris images generated from

off-angle views, and occluded iris images resulting from blinks, partially visible, or distorted iris images.

1.2 Motivation of the Study

Iris recognition technology has shown good performance with frontal iris images, but its efficiency significantly reduces when applied to non-frontal images and off-angle conditions. These are difficulties that arise in real-world instances where iris images are subject to factors such as partial occlusion by eyelids, blinks, the complex 3D structure of the eye, and distortions from very off-angles, difficulties encountered routinely due to environmental factors and constraints on camera positioning. Because of these difficulties, several approaches have been proposed to convert off-angle iris images into frontal views, which are crucial for accurate iris recognition in security and identity verification applications.

Despite the recent progress, most notably through the application of Generative Adversarial Networks (GANs) for improving image quality, current methods continue to fall short of producing high-quality, highly detailed frontal iris images, particularly when dealing with large-angle, occluded, or distorted inputs. This limitation has become more critical due to the increasing need for reliable contactless, robust biometric systems, which is being driven by the need for enhanced surveillance technologies, increasing global security concerns, and the expanding integration of iris recognition in consumer devices.

Hence, new methods that can successfully address the challenges posed by off-angle and off-angle situations, like partially occluded and distorted iris images, are in critical demand. Progress in this area is urgently required to help guarantee the reliability and accuracy of modern iris recognition systems for applicable, real-world scenarios.

1.3 Problem Statement of the Study

In most computer vision applications, such as iris recognition, border control, healthcare, mobile technology, and security systems, a frontal view of an iris image is crucial. However, environmental factors and limitations in camera positioning often result in iris images being captured at off-angles. This can lead to the circular shape of the iris appearing elliptical when observed from an angle, complicating feature extraction. Parts of the iris may be covered by eyelids, eyelashes, and reflections, which lowers the quality of the image that is taken. This issue

has emerged as a notable challenge for iris recognition systems. Unlike frontal images, where iris features are distinctly visible, off-angle images, particularly those with large angles and occluded or distorted, frequently lose critical iris features, making them more difficult to use in applications such as recognition, border control, health, mobile devices, and security systems. As a result, achieving reliable iris recognition with mostly off-angle data continues to be a persistent challenge in these domains. Therefore, converting off-angle iris images to realistic frontal iris images is crucial for producing high-quality and consistent iris images from off-angle input before using them in such applications.

Prior studies, including those conducted by Jalilian et al. (2020), Diab and Karakaya (2022), and Jalilian et al. (2022), have introduced various techniques to address the challenges associated with off-angle iris image conversion. Despite these advancements, significant limitations remain, particularly in generating high-quality, detail-preserving frontal images from off-angle iris image inputs captured at extreme angles (greater than 30˚), or under challenging conditions such as partial occlusions from eyelids or blinks, and distortions resulting from the eye's 3D structure. More recently, Kota and Karakaya(2023a) applied the Pix2Pix-GAN framework to improve off-angle to frontal iris transformation, showing improvements in handling typical features such as corneal refraction and pupil dilation. However, their approach still struggled with images captured at angles exceeding 30˚, as well as with occluded, partially visible, or heavily distorted iris images. To overcome these challenges, this study proposes a novel GAN-based method incorporating hybrid attention mechanisms and a multi-scale discriminator. This architecture was specifically created to enhance the conversion of high-fidelity, consistent frontal iris images, even in the presence of severe distortion, large gaze angles, and partial occlusions.

1.4 Research Questions

This study seeks to examine the following research questions:

RQ1: What are the impacts of incorporating an attention mechanism for transforming off-angle iris images?

RQ2: What are the impacts of using a multi-scale discriminator concerning the quality and transformation of off-angle iris images?

RQ3: What effect would Adversarial loss, identity loss and perceptual loss have on the quality and consistency of generated frontal iris images?

1.5 Objective of the Study

1.5.1 General Objective

The general objective of this study is to develop a hybrid attention and multi-scale discriminator-based generative adversarial network focused on achieving realistic frontal view iris image conversion from off-angle iris images.

1.5.2 Specific Objectives

The specific objectives listed below were addressed to achieve the general objective.

- To determine a method for converting off-angle iris images to frontal view.
- To develop a generator and discriminator network focused on enhancing the quality and details in the converted iris images.
- To design a Generative Adversarial Network (GAN) architecture that incorporates attention mechanisms and a multi-scale discriminator.
- To train the model utilizing a dataset that consists of both off-angle and frontal iris images.
- To evaluate the model's performance using different evaluation metrics on the test dataset.

1.6 Significance of the Study

This research contributes significantly to computer vision and biometrics by tackling the issues that off-angle iris images pose in iris recognition systems. This study introduced a novel framework that integrates hybrid attention mechanisms with a multi-scale discriminator within the architecture of the Generative Adversarial Network. The hybrid attention mechanism enhances the capability of the model to focus on significant iris characteristics by improving both the quality and accuracy of the resulting images. Multi-scale discriminators, on the other hand, evaluate the realism across various levels of detail more effectively and thus ensure that transformed irises are not just a simulation of authentic irises but include complex structural differences. The paper serves as a comprehensive solution for producing high-fidelity iris images taken from different

angles, and in different scenarios, such as occluded by eyelid or distorted thereby advancing the efficiency of identity verification and access control security. This advancement encourages the development of more accurate and reliable iris recognition systems.

1.7 Scopes and Limitations of the Study

1.7.1 Scopes of the Study

The main goal of this study is to generate photo-realistic frontal-view iris images from off-angle input images with angles ranging from 30^0 to 45^0 and challenging cases such as partial visibility of the iris due to eyelid occlusion, blinking, or irregularities introduced by the 3D structure of the eye.

To achieve this, the following specific tasks were carried out:

- Frontal iris conversion from off-angle inputs, even under partial occlusion and severe distortion.
- Synthetically generated frontal iris image quality and consistency improvement with GAN
- Architecture and implementation of a GAN model with attention mechanisms and a multi-scale discriminator to improve the realism of iris image transformation.
- Evaluation of the model on a benchmark iris dataset, with comparisons to existing state-of-the-art methods.

1.7.2 Limitations of the Study

This study did not include the following due to time and resource limitations:

- Our study does not focus on end-to-end iris recognition; instead, the model is intended to convert off-angle iris images to frontal-view images.
- The conversion is limited to iris images taken at angles from 30 to 45 degrees.
- It does not involve the conversion of the iris when it is occluded by glasses, and the off-angle exceeds more than 45 degrees.

1.8 Contributions of the Study

Our main goal was to enhance the performance of the discriminator and generator networks in our proposed model.

- Channel attention mechanisms were incorporated into the encoder to emphasize important feature channels, and a self-attention mechanism in the decoder part of the U-net-based generator network to capture long-range dependencies.
- A multi-scale based discriminator was implemented within the discriminator module to evaluate fine-grained details and overall consistency.
- We used adversarial loss combined with identity loss (L1 loss) to maintain identity information and perceptual loss utilizing MobileNetV2 to preserve structural details.

1.9 Organizations of the Study

This study consists of seven chapters, which are summarized below.

Chapter One: It describes the importance of the study, the problems at hand, and the issues that inspired us to do this study.

Chapter Two: It details literature reviews, deep learning and machine learning algorithms, and previous works related to off-angle to frontal view conversion tasks.

Chapter Three: It describes the research methodologies, including dataset collection, data preprocessing, tools used for development, baseline work, and evaluation metrics.

Chapter Four: It provides a thorough explanation of the proposed model's architecture, including the generator and discriminator network designs, model learning functions, and pseudocode detailing the training process.

Chapter Five: It delves into the model implementation, presenting a code sample that illustrates the proposed solution, specifying the hyperparameters applied, and detailing the experiment class.

Chapter Six: It outlines the results of the proposed method and compares them with previous studies.

Chapter Seven: It summarizes the study findings and provides recommendations for future research.

CHAPTER TWO

2. LITERATURE REVIEW AND RELATED WORKS

2.1 Introduction

In this chapter, we reviewed the literature on off-angle iris image to frontal iris image conversion. Background information on off-angle iris recognition challenges is covered. A review of relevant studies that have concentrated on creating deep learning models for off-angle conversion is presented in this chapter.

2.2 Off-angle to Frontal Iris Image Conversion

Iris recognition is widely used in biometric authentication due to its high accuracy and long-term stability (Daugman, 2009). However, one of the key issues facing iris recognition systems relates to processing off-angle images, where the iris is captured at an oblique angle rather than a frontal view. Off-angle iris images often lead to distortions, occlusions, and reduced recognition accuracy, consequently affecting biometric performance significantly (Zhou & Kumar, 2017). To alleviate this problem, several techniques and approaches have been proposed for converting off-angle iris images into frontal ones to enhance feature extraction and recognition performance. Off-angle iris image conversion to a frontal view is a method for converting iris images taken at off-angle to a frontal view. This process is vital in biometric systems since frontal iris images are more reliable for identification. Images taken off-angle commonly suffer from various distortions caused by corneal refraction, occlusions, and perspective shifts that decrease accurate matching and recognition (Kota & Karakaya, 2023).

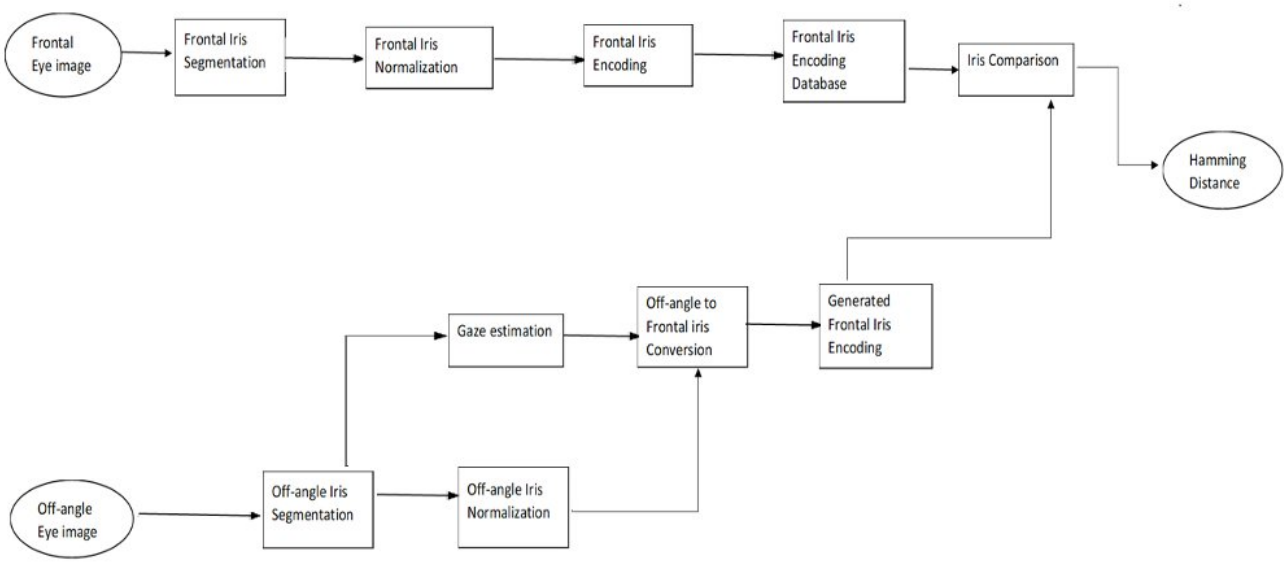


Figure 2. 1: Off-Angle to Frontal Iris Image Conversion Process
(Kota & Karakaya, 2023b)

2.3 Challenges in Off-Angle Iris Recognition

Iris images taken at non-perpendicular positions pose several challenges, most notably perspective distortion, where the iris morphology undergoes elliptical distortion due to off-axis viewing angles (Li et al., 2013). Partial occlusion is also a common issue since eyelids and eyelashes have the potential to cover parts of the iris, making feature extraction more challenging (Proença & Alexandre, 2007). Also, the feature misalignment is caused by the angular rotation of significant iris features, which results in decreased accuracy of template matching (Wildes, 1997). Furthermore, lighting conditions make the problem worse because variations in illumination can alter the iris texture's appearance, which impacts how well identification systems work (Zhang et al., 2018).

2.4 Methods for Off-Angle to Frontal Iris Conversion

Various approaches have been developed for dealing with such distortions of off-angle iris images conversion to frontal view. Most can be roughly grouped as geometric transformation-based methods, image transformation models and deep learning-based approaches.

2.4.1 Geometric Transformation-Based Methods

Geometric Transformation-based methods are designed to convert the distorted iris images into a frontal view by assuming the iris is a three-dimensional structure. One of the critical methods in this area is the mapping of the captured off-angle iris image on a standardized reference plane. One of the early contributions to this field was made by Abhyankar & Schuckers (2010), who proposed a projection-based method using image warping to composite for the geometric distortions in off-angle iris images. More sophisticated models use polar transformation techniques to map the iris image into a normalized polar coordinate system, thus efficiently compensating for the foreshortening effect. Among the many approaches, Daugman's rubber sheet model, as described (Daugman, 2004), is prominent in the way it can stretch the iris into a rectangular strip for easy feature extraction. However, there have been numerous modifications to address distortions caused by off-angles by improving the mapping functions. Despite such advancements, geometric normalization methods may become challenging when dealing with complex non-linear distortions in images taken at large angles.

2.4.2 Image Transformation Models

Image transformation approaches apply 2D and 3D models to simulate the real orientation of the iris, which leads to normalization; hence, the frontal view. In order to match iris images collected from an angle with those taken from a frontal position, (Karakaya, 2016) introduced 2D rectification approaches that perform affine and projective transformations. This was a straightforward linear relationship method and worked reasonably well for images taken at moderate angles, but less so for images at extremes. More recent attempts use 3D modeling, where the anatomy of the eye is factored in. For instance, (Frigerio & Marcon, 2014) created a 3D iris model, emulating corneal curvature, which corrects for off-axis distortions by remapping pixel locations based on estimated gaze direction. These 3D models, though computationally heavy, can give better reconstructions of highly distorted images. Effective for moderate angles, such methods usually entail an increase in computational demand and, at the same time, a drop in accuracy in the case of extreme angles.

2.4.3 Deep Learning Based Approaches

Deep learning-based approaches have become critical in overcoming the inherent challenges of iris image recognition taken from non-frontal angles. Traditional iris recognition methods, which

usually depend on frontal images, are faced with complications arising from distortions and variations when images are taken from different angles. Deep learning models, especially convolutional neural networks (CNNs), were used to automatically extract more reliable characteristics from iris scans in response to these difficulties, irrespective of the angle at which the images were taken. Trained on large and diverse image datasets, these neural networks can efficiently detect meaningful, discriminative features, including under challenging conditions. It is the chance of increasing segmentation and encoding of the iris pattern for the changing gaze angles that provides the benefit of accuracy rising in iris transformation and recognition (Karakaya, 2018).

These methods are frequently challenged by either their dependency on high-quality, annotated datasets or the considerable computational requirements for such models to be effective; this affects their generalization capability toward new, unseen situations. (Chavarro & Karakaya, 2023) proposed a multi-modal deep learning approach in which several shallow networks, like Squeeze Net, were trained to be the best at recognizing images captured under specific gaze directions. Part and parcel of this approach is the estimation of gaze angles, which is performed beforehand in sending the off-angle images to the respective network. This system has shown a promising result in compensating for gaze angle-induced distortions to improve the robustness of the recognition system. However, there are still a few significant challenges. A major limitation of the contemporary deep learning methodologies employed in off-angle iris recognition is that they mostly rely on the quality and diversity of the training datasets. Networks designed with a spectrum of gaze angles can show some resilience but remain fragile when confronted with large off-angle distortions, including occlusions caused by eyelid blinking. Moreover, the computational complexities involved in training and maintaining deep learning models capable of analyzing off-angle images form a substantial challenge, mainly for systems intended for real-time standoff biometric applications.

Jalian et al. (2022) have proposed an application of GANs in deep learning for the transformation of off-angle iris images, placing a strong emphasis on overcoming the challenges that arise in real-world applications due to non-ideal iris images. These off-angle representations can significantly degrade the recognition accuracy because of complications like iris distortion and suboptimal texture quality. GANs have promised to handle such issues by learning to generate frontal-view

iris images from the off-angle inputs to enhance the performance of the recognition systems. The deep learning techniques can make the key features of irises sharper, which in turn allows recognition in scenarios where control is less stringent, such as in standoff or dynamic environments. There are, however, some big challenges in applying GANs for the reconstruction of off-angle iris images. One of the primary concerns is that this method, when trained, needs massive datasets, as GANs need many varied data points to generalize well across various gaze angles and lighting conditions.

2.5 Generative Adversarial Network (GAN)

Goodfellow et. (2014) introduced Generative Adversarial Networks (GANs). A generative model learns a set of training samples to understand the underlying probability distribution by which they are governed. This, therefore, enables the generation of new, realistic images, similar to real ones, following the learned probability distribution. A GAN comprises two neural networks: A Generator(G) and a Discriminator (D), which are trained in a competitive setup, therefore creating a game-like scenario.

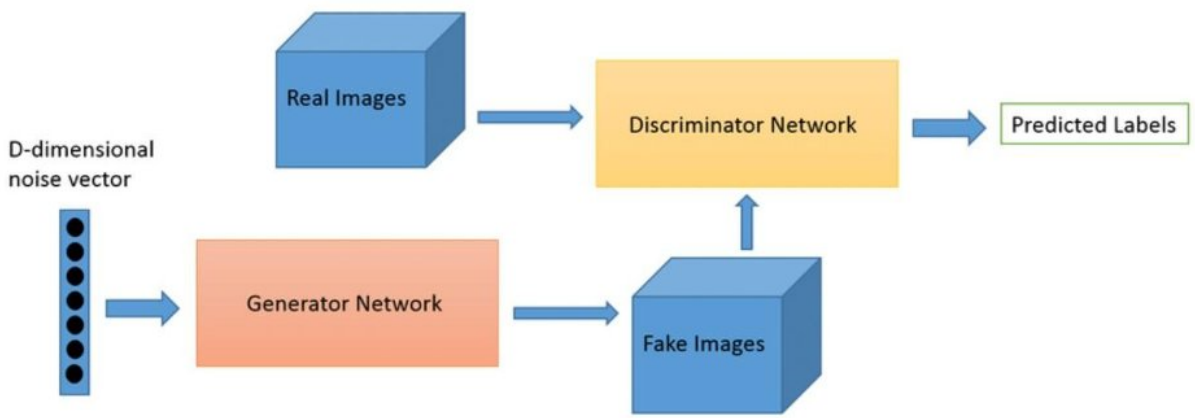


Figure 2. 2: General Architecture of GAN
(Alqahtani et al., 2021)

The discriminator network assesses how realistic a certain piece of data is, whereas the generator network creates new data. In order to trick the discriminator, the generator tries to generate data that looks like actual data, and the discriminator strives to get better at telling the difference between real and fake data. These two networks train in a competitive manner. The sustaining competition forces both networks to move forward and enhance their functionalities, to a point

where the generator becomes expert enough in data generation to succeed in fooling even a very good discriminator, despite such a discriminator being very hard to fool.

2.5.1 Conditional GAN

Generative models can learn the dataset's probability distribution and generate new samples from the likely distribution of the data on which they were trained. Nevertheless, the data being generated cannot be controlled. In a conditional generative adversarial network proposed by (Mirza & Osindero, n.d.), other information is utilized by the generator and discriminator networks, like class labels or other input features. The information is normally encoded either as a one-hot encoded vector or feature vector, which is then added to the internal data representation in a hidden space. The generated vector serves as the generator's input. Likewise, the discriminator is also fed the conditioning information as input, along with the generated or actual data sample. It uses this knowledge to differentiate between original and fake samples. This enables the cGAN to generate new data samples that are similar to the training data and share certain attributes. Many researchers used CGAN in research domains like translating one image to other (Isola et al., 2017; Chai et al., 2018), text-to-image generation, face generation (Gauthier, 2015) and video generation (Wang et al., 2019).

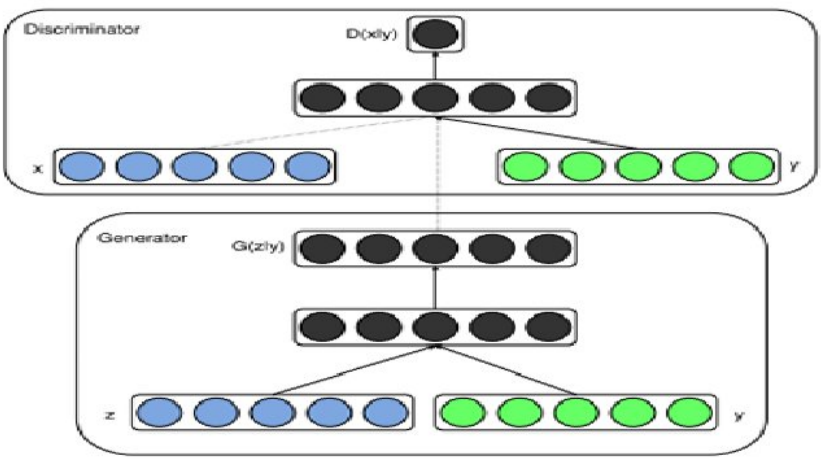


Figure 2. 3: CGAN architecture

(Mirza & Osindero, n.d.)

2.5.2 CycleGAN

The goal of image-to-image translation is to determine the relationship between an input image and the matching output image that is learned from a dataset of synchronized image pairs. But in most cases, the paired training data is not feasible. CycleGAN (Zhu et al., 2017) proposed a method of learning method of transferring an image from a source domain, like X, to a target domain, like Y, can be achieved without the need for paired samples. For unpaired photos, it is a type of generative adversarial network. CycleGAN (Zhu et al., 2017) creates a correspondence between two distinct spaces, X and Y, in order to achieve image-to-image translation. F: Y→X and G: X→Y. The attempt to portray these is what is new. Such mappings are inversely proportional to one another, and the two mappings are bijections. A cycle consistency loss function that encourages F(G(x)) ≈ x and G(F(y)) ≈ y accomplishes this. This loss, in addition to the losses on X and Y, provides the full goal for unpaired image-to-image translation.

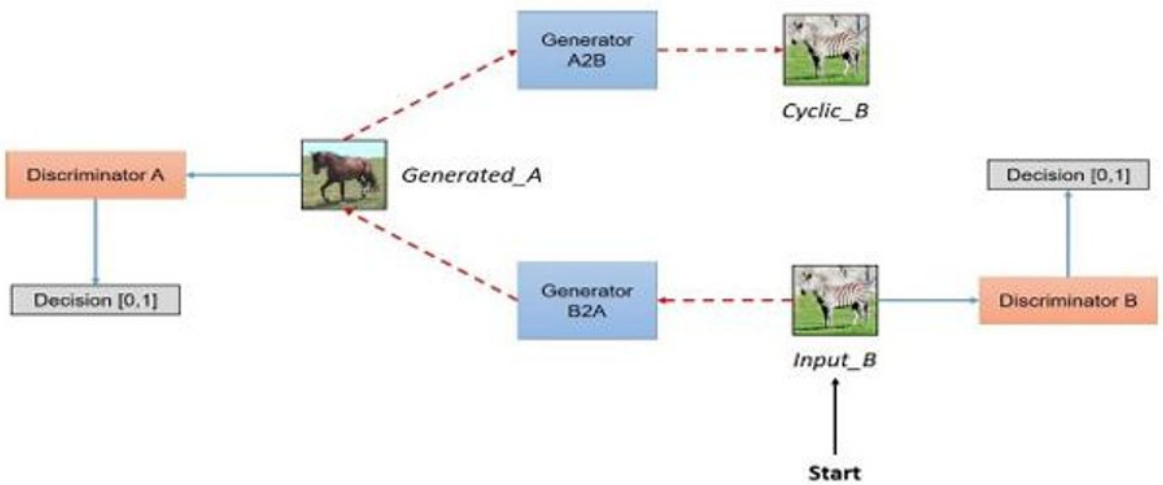


Figure 2. 4: CycleGAN Architecture

(Zhu et al., 2017)

2.5.3 Deep Convolutional GAN (DCGAN)

Significant advancements in the field of generative adversarial networks (GANs), especially in terms of stability and image quality, are shown by Deep Convolutional Generative Adversarial Networks (DCGANs) (Y. Zhang et al., 2018). Deep convolutional neural networks (CNNs), use both the discriminator and the generator which come from the fully connected networks used in

the previous GAN architecture. It used transposed convolutional layers (deconvolutional layers) to up sample the input noise vector to a full-resolution image in generator component and regular convolutional layers in discriminator component. DCGAN replaced stride convolutional pooling layers in the discriminator and used stride transposed convolutions in the generator to enable the networks to learn their own knowledge spatial down sampling or upscaling.

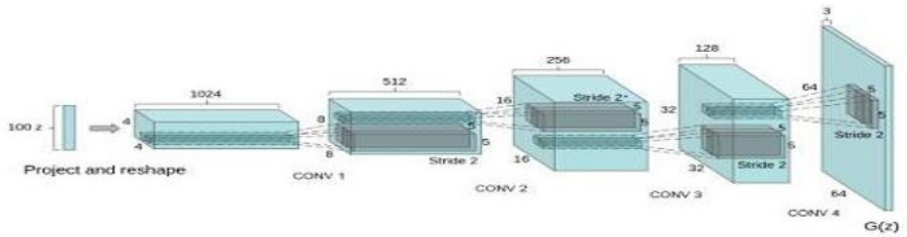


Figure 2. 5: DCGAN Architecture

(Y. Zhang et al., 2018)

2.5.4 Pixel-to-Pixel GAN (Pix2PixGAN)

Pix2PixGAN is a conditional Generative Adversarial Network (cGAN) for image-to-image translation tasks (Henry et al.,2021). Introduced as a general framework, Pix2PixGAN enables the generation of a target image from an input image through learning a mapping from paired datasets. The framework consists of a discriminator that evaluates the generated images' realism in comparison to the actual ones and a generator that attempts to produce realistic target images. It employs a U-Net-based generator architecture, which preserves spatial information through skip connections, and a PatchGAN discriminator, which decides the authenticity of patches, not the entire image. The patch-based approach promotes the network to focus on local details, enhancing the output sharpness and realism. Henry et al. (2021) state that Pix2PixGAN works well on a diverse range of tasks, including photo restoration, semantic label-to-image translation, and style transfer, thus being a useful tool for computer vision as a field. As noted by the research, its real-world applicability and ease of extension to various image modalities are accentuated regarding its capability to learn structured transformations between images from paired data.

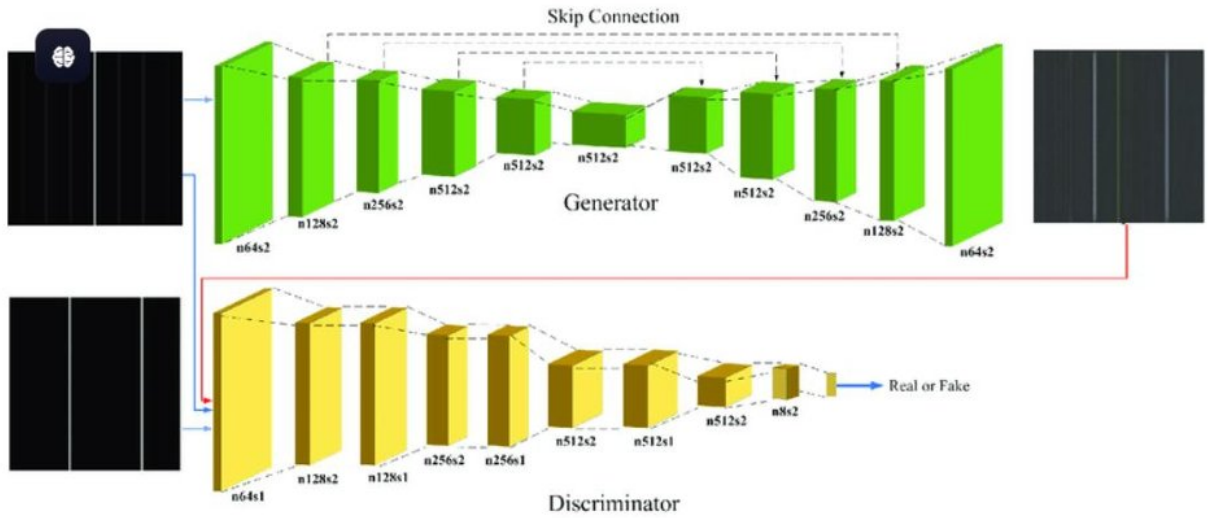


Figure 2. 6: Pix2PixGAN Architecture

(Huang et al., 2021)

2.6 Hybrid Attention Mechanism

Deep learning attention mechanisms are methods inspired from the cognitive processes of humans which enable models to highlight the most important aspects of the input disregarding irrelevant information This targeted strategy is maximized in impact by enabling the assignment of various weights to the different parts of the data to allow it to emphasize information most relevant to specific tasks. These operations allow the capacity of each element in a sequence to "attend to" every other component, enhancing contextual understanding. The attention mechanisms have been utilized in Generative Adversarial Networks (GANs) to allow models to concentrate on a specific subset area of an image; a feature strongly demanded for tasks requiring high spatial accuracy, for example, iris transformation (Habib et al., 2023). Hybrid attention, as introduced by (Jiang et al., 2022), combines various attention mechanisms under a single framework to render them stronger in computer vision and image processing tasks. It is particularly helpful in the iris reconstruction task that necessitates both global structural integrity and local detailed features to be preserved. In in this work, channel and self-attention hybrid attention mechanisms are employed in the building blocks of the U-Net architecture, namely the encoder and decoder parts, which were used as a generator to enable produce high-quality images.

2.6.1 Self-Attention Mechanism

Most of the existing image generation models based on GANs rely on convolutional layers. However, since convolution is performed over local regions, it is not computationally efficient to depend solely on convolutional layers for modeling long-range dependencies in images. To address this issue, (H. Zhang et al., 2019) proposed SAGAN, which introduced self-attention into the GAN framework so that the generator and discriminator can capture long-range dependencies effectively. The SAGAN also adopts spectral normalization to reduce the computational cost of training.

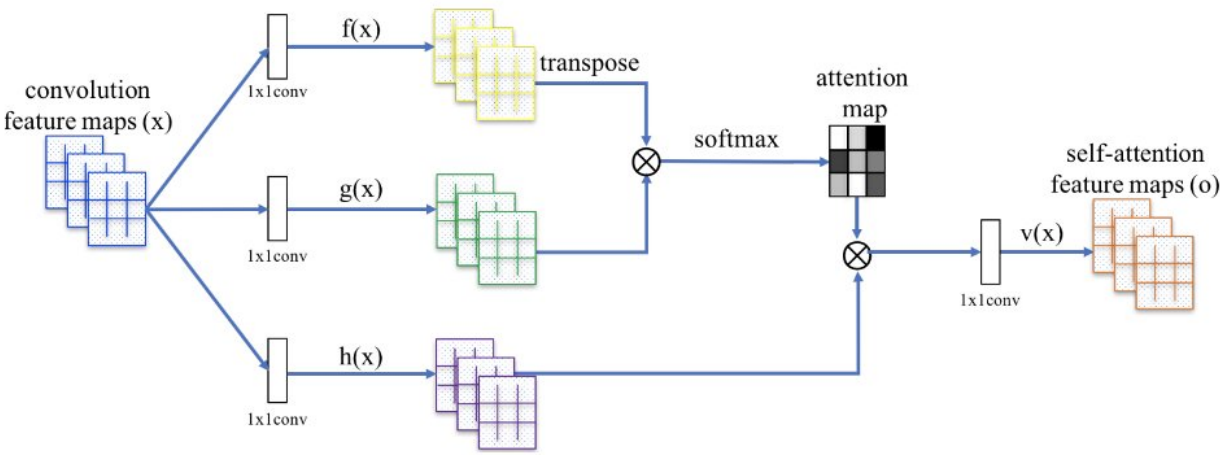


Figure 2.7: Self-attention GAN module

(H. Zhang et al., 2018)

2.6.2 Channel attention Mechanism

Channel attention was introduced by SENet (Hu et al., 2018) to estimate the importance of each channel on feature maps. This operation is used in deep CNNs to highlight the most important channels that appear within a feature map. It, therefore, improves the performance of the network by suppressing the noise and redundant information. A channel attention map is obtained by applying inter-channel relationships between features to emphasize the important elements related to a given input image by regarding each channel of feature map as a separate feature detector. To efficiently compute channel attention, the spatial dimensions of an input feature map are are down-scaled, and average pooling is utilized to summarize the spatial information (Woo et al., 2018).

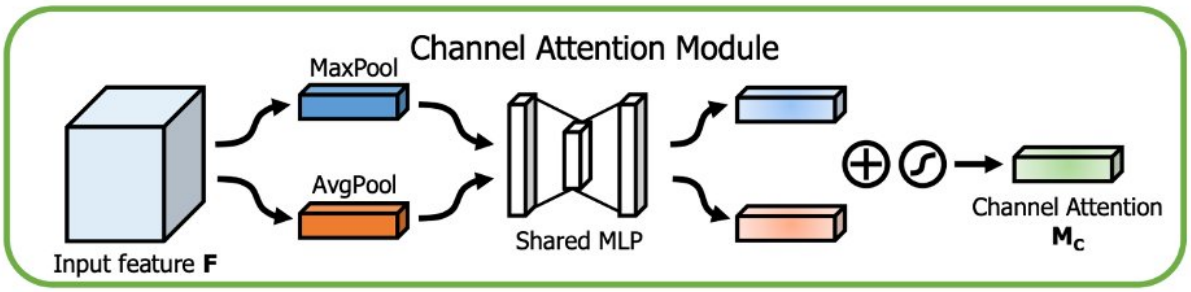


Figure 2.8: Channel attention module

(Woo et al., 2018)

2.7 Multi-Scale Discriminator

The design of a Multi-Scale Discriminator is common in Generative Adversarial Networks to make the discriminator effective at assessing images at multiple scales or resolutions. The key advantage of such an approach lies in bringing both global and local features into images, which in effect makes the discriminator sensitive to various spatial hierarchies in generated images. Consequently, this enhances the efficiency of training, especially for the tasks of generating high-resolution images. The multi-scale discriminator works by examining the outputs from a generator at different levels of granularity, which helps the model differentiate between real and fake images by fusing detailed and generalized features. This approach works especially well for complicated image generation tasks like high-resolution synthesis, where it's necessary to preserve fine details while keeping the overall structure (Wang et al., 2018). The multi-scale discriminator introduced in GANs means applying several discriminators that are operating at different resolutions. This idea allows the catching of big-picture and detailed features of the image, increasing the discriminator's capability to judge minute variations in the picture without compromising the quality of the context. Such kind of technique is useful when reconstructing images with which a GAN achieves a delicate balance between realistic textures and global structure (Lee et al., 2023).

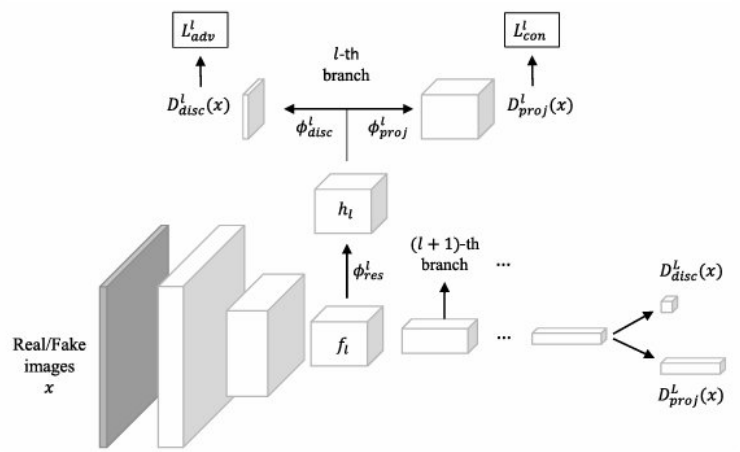


Figure 2. 9: Network architecture of the multi-scale discriminator (Lee et al., 2023)

2.8 U-net

The U-Net architecture, which was originally proposed by (Ronneberger et al., 2015), is a strong convolutional neural network that is mainly developed for semantic segmentation. Owing to its encoder-decoder architecture with skip connections, U-Net has been shown to work exceedingly well on several image-to-image translation and reconstruction tasks, thus making it apt for synthesizing high-quality images. For our hybrid attention and multi-scale based discriminator GAN, we utilize U-Net as the generator so that we can leverage its spatial information preserving property and fine-grained feature preserving property throughout the network. The addition of hybrid attention mechanisms within the U-Net framework also enhances its ability to focus on crucial features at different scales. This is especially helpful for difficult image transformations, e.g., mapping slanted iris images to realistic front views. Tiny texture and structure details have to be captured to obtain high-quality output.

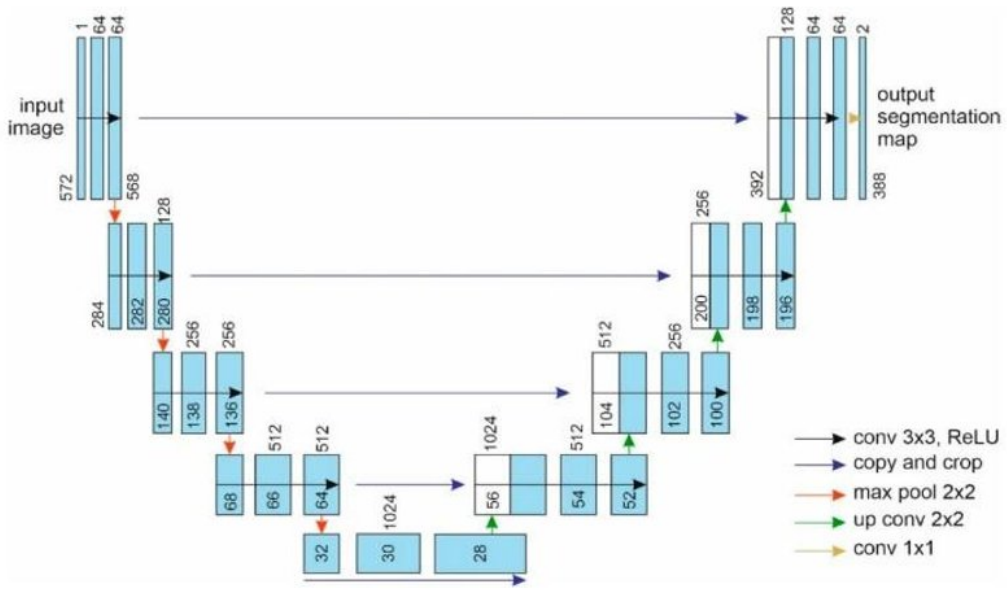


Figure 2. 10: U-Net Architecture

(Ronneberger et al., 2015)

2.9 Related Works

Kota & Karakaya (2023), discusses the challenges that biometric iris recognition systems face in dealing with off-angle images, which are generally afflicted by distortions introduced by factors like corneal refraction and the angle of gaze. The study proposes an adapted Pix2Pix-GAN framework for improving recognition accuracy by converting these off-angle images to frontal views using advanced loss functions to improve image quality. Experimental results demonstrate a strong performance improvement for features such as corneal refraction and pupil dilation, but the problem persists for larger angles exceeding 30^0 and occlusions, such as when eyelids cover most of the pixels, irregular limbus occlusion. Additionally, low contrast iris texture and the distortion due to the 3D structure of the iris images did not show improvement.

Jalilian et al. (2020) proposed a fully convolutional network (FCN) integrated with a gaze estimation module to process off-angle iris images. Although this method demonstrated potential in enhancing segmentation accuracy for such images, it encountered challenges with improving

the segmentation accuracy of extreme off-angle images and heterogeneous training setups for more generalization.

Jalilian et al. (2021) proposed new CNN-based algorithms, including a KNN-based gaze-angle estimator and morphological operations to enhance the accuracy of segmentation. The major finding is that training with well-aligned gaze angles and optimization in the process of segmentation significantly raises the rates of recognition at mostly extreme angles. The proposed methodologies are effective for extreme angles (±40° and ±50°). However, the challenge remains interpolation during corrections, introduced feature distortions, limiting its utility for near-frontal images, and a limited dataset, which necessitates additional refinement and wider validation.

Diab & Karakaya (2022) discusses the utilization of convolutional neural networks (CNNs) in estimating gaze angles to enhance the accuracy of iris recognition in challenging standoff conditions. It proposes a gaze estimation method to convert off-angle iris images into frontal representations using transfer learning with the AlexNet architecture to predict continuous gaze angles without the need for segmentation or additional hardware. The study uses a dataset of 108,600 images of irises and achieves an average gaze estimation error of 4° at more than 93% accuracy within the 10° margin. However, it outperforms existing approaches and demonstrates robustness in handling off-angle images; the model is limited in handling occlusions due to blinking and needs specific camera settings.

Jalilian et al. (2022) explore the application of the convolutional neural network CNNs) along with the triplet loss function for the task of off-angle iris recognition, which has become a challenge due to variations in gaze angles and environmental conditions while capturing iris images. The study noted limitations, such as decreased performance of the models when confronted with extreme angles or varying light conditions, which can obscure texture details.

2.10 Summary of Related Works

Table 2.1: Summary of Related Works

Related Paper	Methodology	Contribution	Limitation
(Jalilian et al., 2021)	- CNN - Gaze-Angle Estimation and Conversion Module - Use the geometric transformation method	- Gaze Angle Segmentation and Recognition Improvement - Gaze-Angle Estimation - Conversion Techniques and Interpolation Effects were identified	- Interpolation inaccuracies in gaze angle adjustments - Lack of generalization due to a limited dataset
(Jalilian et al., 2022)	- CNN with the triplet loss function	- Improvement on Gaze Angle Differences Recognition	- Minimal improvement for extreme angles - Lack of generalization - Performance deteriorates with large gaze angle differences
(Jalilian et al., 2020)	- RefineNet for Segmentation - Affine transformation method for Off-angle correction	- Geometric Correction via Parameterization - Morphological Segmentation Improvement	- The correction method showed limited effectiveness beyond 20^0 off-angle
(Kota & Karakaya, 2023b)	- Pix2Pix-GAN model	- Modified Pix2Pix GAN Architecture - Improved Recognition Performance - Asymmetric Angle Performance Improvement	- No improvement for iris image captured at angles exceeding 30^0 - Low-contrast images require improvement. - No improvement for distortion due to the 3D structure of the eye's
(Diab & Karakaya, 2022)	- CNN-based AlexNet using transfer learning	- Effective CNN for Gaze Estimation	- Limitation of handling occlusions due to blinking and requires specific camera settings.

As summarized in Table 2.1, prior studies have extensively explored Convolutional Neural Networks (CNNs) for processing off-angle iris images. For instance, Jalilian et al. (2021) utilized CNNs along with a geometric transformation method, but, as the table indicates, suffered from interpolation inaccuracy and poor generalization due to limited data. Similarly, Jalilian et al. (2022) employed a triplet loss function to enhance performance, yet their method showed minimal improvement with extreme gaze angles and again lacked generalization, especially when faced with large gaze-angle differences.

Refinement was also attempted by Jalilian et al. (2020) using RefineNet and affine transformation techniques, but as shown in the table, their correction approach was rendered ineffective beyond 20° off-angle deviation. Kota and Karakaya (2023a) explored the use of GANs, specifically a Pix2Pix model, for frontal iris image synthesis. While this GAN-based method showed some success for image-to-image translation, Table 2.1 shows that it could not handle the distortion of iris images due to large angles exceeding 30^0, low-contrast inputs, and could not correct 3D structure caused distortions from iris images.

In contrast to these previous methods, the proposed model addresses the limitation by incorporating a hybrid attention mechanism, which combines self-attention and channel-attention blocks within in Multi-Scale Discriminator GAN framework and utilizing a combination of adversarial, perceptual, and identity loss functions. This enables the model to enhance key iris features (e.g., texture, pupil border) extraction while suppressing irrelevant background noise, resulting in improved performance under difficult conditions like extreme gaze angles, occlusions, and 3D distortions. Furthermore, the multi-scale discriminator captures both global and fine-grained iris image information, resulting in more realistic synthesis. By including adversarial, perceptual, and identity loss functions, the model provides structural consistency, visual fidelity, and identity preservation, effectively addressing generalization, detail loss, and alignment difficulties reported in previous studies summarized in Table 2.1.

CHAPTER THREE

3. RESEARCH METHODOLOGY

3.1 Overview

This study implements a varied range of methodologies, strategies, instruments, and documentation supports to carry out research goals effectively. The key objective here was to design an advanced Hybrid Attention and Multi-Scale Discriminator-Based Generative Adversarial Network (GAN) framework aimed at transforming high-quality frontal iris images obtained from off-angles. To achieve this aim, this study adopts the application of different tools, implementation methodologies, literature analyses, dataset organization, and performance evaluation criteria.

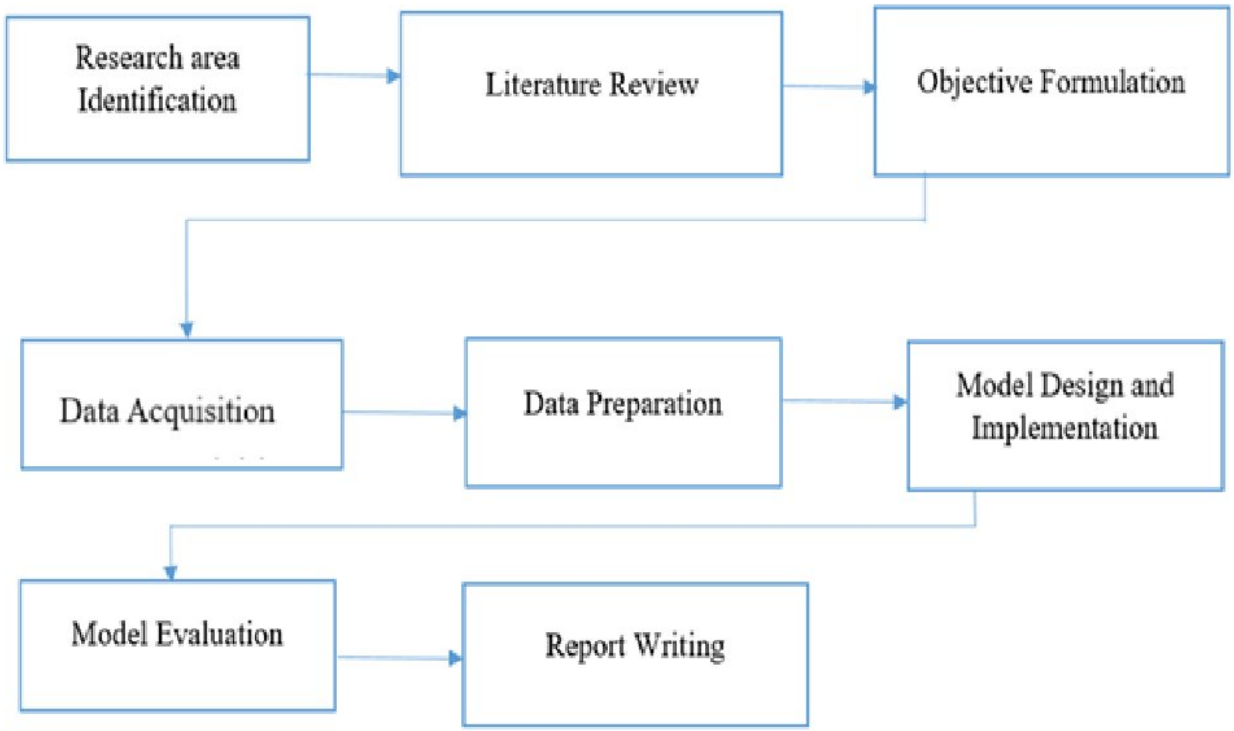


Figure 3.1: Overall process

3.2 Dataset Acquisition

This study addressed the issue of off-angle to frontal iris conversion with the power of deep learning approaches. Deep learning architectures mostly relying on data to learn. Hence, data has

played a significant part in research. We used publicly accessible iris datasets MOBIUS and UBIRIS.v2 for training and testing in order to perform off-angle to frontal iris conversion utilizing Generative Adversarial Networks, which combines attention and multi-scale-based discriminator. To assess the generalization capacity of our model that was trained on the MOBIUS dataset, we used the MOBIUS dataset for training and testing, and the UBIRIS.v2 dataset for testing. We selected those datasets to use in this study because they contain a variety of iris images with different illumination, occlusions, and angles.

3.2.1 University of Beira Interior Iris Recognition Version 2 (UBIRIS.v2) Dataset

It is a visible-spectrum (RGB) iris image dataset created by the University of Beira Interior in Portugal. It includes iris images that are intended to simulate real-life conditions. There are lots of adverse conditions like reflections, motion blur, off-angle imaging, and non-uniform lighting, making it highly valuable for research in robust iris recognition and image correction. Unlike typical databases derived from near-infrared images taken in controlled settings, UBIRIS.v2 includes over 11,000 images taken at a distance and while in motion from 260 subjects, and it was specifically created to simulate real-world biometric scenarios, making it ideal for research in robust iris recognition, image enhancement, and cross-spectral matching (Padole & Proenca, 2012). Due to its unconstrained nature and inclusion of partial and off-angle iris images, UBIRIS.v2 is also a suitable dataset for off-angle to frontal iris image conversion research. These kinds of research are aimed at removing perspective distortions from off-angle images and transforming them into an almost frontal view to improve the accuracy of recognition. The angle and illumination variability of the database provide a good test bed for training and testing algorithms that can handle unfavorable imaging conditions.

We used the UBIRIS.v2 dataset to test our model, which was trained on other dataset, to evaluate its generalization ability for converting off-angle iris images to frontal views, with a particular emphasis on angles greater than 30° in this work. We aimed to show that our model performs well in correcting extreme angular distortions, thus making iris biometrics more usable in unconstrained conditions.

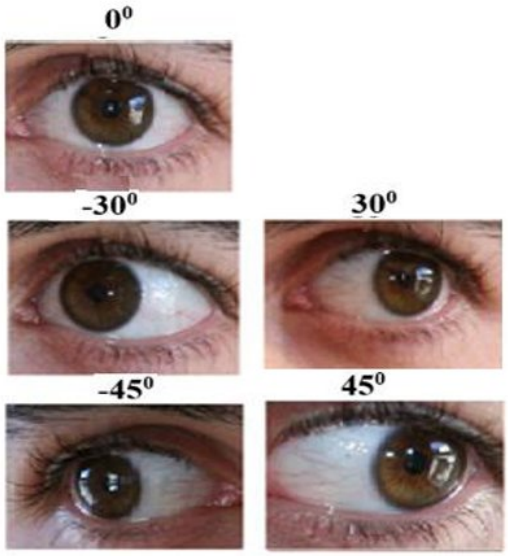


Figure 3. 2: UBIRIS.v2 (Padole & Proenca, 2012) dataset samples

3.2.2 MOBIUS (Mobile Ocular Biometrics in Unconstrained Settings) Dataset

It was created at the University of Ljubljana in 2019. It is a freely accessible dataset is mainly meant for studies on mobile ocular recognition research. However, it also includes iris images captured at various angles, including extreme off-angle views (large deviations from the frontal perspective), with occlusions (e.g., blinks, eyelids, reflections) (Vitek et al., 2020). This makes it an ideal dataset for testing and training our model to demonstrate how well it converts frontal iris photos from off-angle ones. There are 16,717 RGB ocular photos of 100 participants in the entire dataset. There are 3542 annotated RGB images from 35 patients (70 eyes) in the segmentation part of dataset. This part of the dataset is not used in this study. Images were captured in high resolution at four distinct gaze directions (straight, left, right, and up) and under three distinct acquisition conditions: good indoor lighting (Indoor), good outdoor lighting (Neutral), and poor indoor lighting (Poor) on three distinct mobile phones: The Xiaomi Pocophone, the Sony Xperia Z5 Compact, and the Apple iPhone 6s.

For this study, images occluded by eyelids or caused by blinks, partially visible iris image due to extreme angle taken at four different off-angle directions (straight, left, right, and upwards) were used in training and testing to illustrate our model's performance for converting partial visible iris image due to extreme off-angle and occluded iris image by eyelid or caused by blinks to frontal iris image. Images under bad lighting in the side (Poor) were not used.



Figure 3. 3: MOBIUS (Vitek et al., 2020) dataset samples

3.3 Pre-processing Techniques

3.3.1 Data Augmentation

Data augmentation techniques can be used to artificially expand the size and diversity of a training dataset. The augmented dataset can be utilized to train model in order to make it more robust and adaptable to different circumstances. It avoids overfitting. We incorporated additional training data through the assistance of a sequence of techniques to increase its variability. The enrichment procedure aimed to improve the model's generalization capability and perform better. In our study, we used the following data augmentation methods to increase our dataset.

- Horizontal flip
- Random rotation
- Random brightness, random contrast, and random saturation

3.3.2 Normalization

It is the process of transforming data into a uniform and comparable format. This typically includes normalization of data to a specific range or distribution of data according to some statistical property. This improves model performance, optimizes algorithms, and shortens training time. In this study, the dataset is normalized within the range of -1 to 1, which implies that the highest and lowest values for each attribute are constrained to -1 and 1, respectively.

$$x = \frac{y-\mu}{\delta} \qquad \text{eq. 3.1}$$

Where y denotes the input image normalized to a range between 0 and 1, μ represents the mean and δ refers to the variance. In this context, the mean is fixed at 0.5 and the variance at 0.5.

3.3.3 Resize

Image resizing is an important first step. It is the process of adjusting the images' sizes so they match the model's required input dimensions. As deep learning involves a huge amount of most situations, the model computes more to accommodate bigger images. Scaling to a smaller size can reduce the level of computational resources needed and significantly accelerate training. We resized the images to 128x128 to decrease the level of computation complexity.

3.4 Development Tools

This study used a wide range of tools and resources to be implemented successfully.

3.4.1 Hardware Tools

The following hardware tools were utilized to implement the.

Table 3.1: Hardware Requirements

No	Tools	Purpose
1	RAM	To accelerate the training process and improve the performance
2	Hard Disk	For storing data and models For storing data and models
3	GPU	To enhance computation and speed up the training process

3.4.2 Software Tools

The software tools help in preprocessing the data, developing the model, visualization, and deployment.

Anaconda: This popular open-source framework packages several programming languages for data analysis, machine learning, and scientific computation.

Jupyter Notebook: is a web-based, open-access platform that lets users create and share documents with executable code, definitions of mathematical terms, graphical representations, and explanations in prose.

TensorFlow: is open-source machine learning software that provides users with a comprehensive suite of libraries, tools, and community resources to help build and deploy their learning.

PyTorch: A powerful and flexible framework for deep learning research and development, loved for its user-friendliness and strong community backing.

3.5 Baseline Works

Our proposed model's performance was compared to the most advanced generative adversarial network (GAN) models designed specifically for off-angle iris image conversion. We chose the Pix2Pix-GAN model (Kota & Karakaya, 2023a) as our baseline model because it performs well in image-to-image translation tasks, which closely match our goal of generating frontal views from off-angle iris images. Pix2Pix-GAN is a conditional GAN framework that is very successful for structured image transformations since it learns a straight mapping from input to output images. Its usefulness and resilience in this particular field have also been confirmed by its successful application in earlier research addressing off-angle iris image conversion. For this reason, it serves as a solid and credible baseline to benchmark improvements introduced by our customized GAN architecture.

The Pix2Pix model (Kota & Karakaya, 2023a) is made up of a PatchGAN discriminator and a U-net-based generator. To keep spatial information in the input and output images, the generator employs an encoder-decoder structure with a skip connection. By reducing high-frequency detail, the PatchGAN discriminator, which analyzes local patches of an image instead of entire image, helps to ensure better realism in frontal iris images that are generated.

The original work's authors note that the process of conditional adversarial training used by Pix2Pix makes off-angle and frontal iris images more similar, hence making it an ideal choice for

comparison. Using this baseline, our objective is to compare how effective our model is at off-angle to frontal iris image conversion accuracy, structural consistency, and overall image quality.

3.6 Feature Extraction Network

We employed MobileNetV2 (Sandler et al., 2018) as a feature extractor to test the accuracy of our model. It is a CNN architecture designed to operate well on mobile devices. An inverted residual structure serves as the foundation for the residual connections among the bottleneck levels. The intermediate expansion layer employs lightweight depth-wise convolutions as a feature non-linearity filter. The MobileNetV2 architecture is composed of 19 residual bottleneck layers after the first fully convolutional layer with 32 filters. Because of its efficient depth-wise separable convolutions and inverted residual blocks, this lightweight architecture is highly favored for computer vision tasks.

These architectural features enable MobileNetV2 to provide high accuracy with the maintenance of computational efficiency (Sandler et al., 2018). In off-angle to frontal iris conversion, the ultimate goal is to convert an off-angle iris image of a person to a high-quality frontal iris view that accurately represents the subject's identity. In order to compare the synthetic frontal iris images with the ground truth frontal images, we used pixel-wise accuracy as a similarity metric. The success of the proposed model is ensured by measuring the proportion of pixels in the synthesized image that closely resemble matching pixels in the ground truth image.

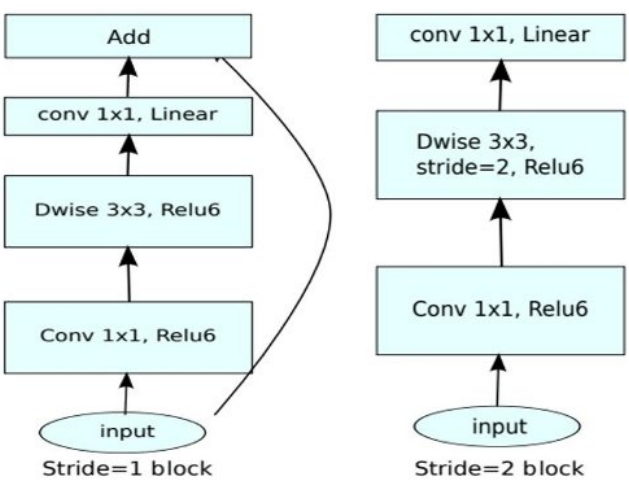


Figure 3. 4 Mobile Net V2 architecture

(Sandler et al., 2018)

3.7 Performance Evaluation Metrics

Evaluation metrics in image generation provide an essential way of assessing the quality and effectiveness of generative models. Such metrics help in determining whether the model, after training on a given dataset, generates realistic and diverse images characteristic of the training set. In this study, we used objective metrics (PSNR, SSIM, and Hamming Distance to assess the performance of the off-angle to frontal iris conversion results.

3.7.1 PSNR (Peak Signal to Noise Ratio)

Peak Signal-to-Noise Ratio, in decibels, measures the quality of an image reconstructed or signal to the original. This measure performs its analysis by figuring out the ratio between the signal's highest potential power and the power of the noise interfering it. Greater PSNR values point to higher fidelity in reconstruction and better quality in images (Isa et al., 2015).

$$PSNR = 20log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \qquad eq.\ 3.\ 2$$

Where:

- ✓ MAX_I is the maximum possible intensity value for a pixel in the image
- ✓ MSE is the average squared difference between the actual and the generated image. It is calculated as:

$$MSE = \frac{1}{M*N}\sum_{i=1\ j=1}^{i=M\ j=N} (f(ij) - g(ij))^2 \qquad eq.\ 3.\ 3$$

Where f and g are the actual and generated images of size M*N, respectively.

3.7.2 SSIM (Structural Similarity Index Measure)

It is a metric for measuring the similarity of two images and a quantitative assessment of the deterioration of image quality after performing a given process. It calculates scores between the original and the results in images by comparing three distinctive features: luminance, contrast, and structure.

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \qquad eq.\ 3.\ 4$$

Where $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are variables used to modify the three components, and l c and s are luminance, contrast, and structure, respectively.

3.7.3 Hamming Distance

It calculates the similarity of two iris codes by measuring the proportion of differing bits in iris recognition systems. A lower Hamming distance indicates higher similarity, suggesting that the iris images likely originate from the same eye, while a higher distance suggests dissimilarity. The Hamming distance between two iris codes C_1 and C_2 can be used to measure the dissimilarity of the two iris patterns:

Hamming Distance $(C_1, C_2) = \sum_{i=1}^{n} |C_{1i} - C_{2i}|$ *eq. 3.5*

Where:

- C_1 and C_2 are the iris codes being computed.
- n is the length of the iris code (typically 512 bits, but it can vary).
- The sum counts how many bits in the two iris codes differ.

CHAPTER FOUR

4. PROPOSED ARCHITECTURE

4.1 Overview of Chapter

In this chapter, the architecture of the proposed off-angle to frontal iris image conversion method is presented. It uses a multi-scale discriminator Generative Adversarial Network in conjunction with attention. It comprises mathematical and graphical representations. The chapter is structured into four parts: the initial section provides a background of the model's structure, followed by a description of the generator and discriminator architecture in the second part. The third part describes the learning functions of the model, and the last part describes the training pseudocode.

4.2 Model Architecture

The proposed approach relies on the principle of Generative Adversarial Network (Goodfellow et, 2020). The GAN network comprises two networks: Generator G and Discriminator D, which are trained competitively. Our proposed approach used U-Net (Ronneberger et al., 2015b) in the generator, which consists of an encoder-decoder structure with skip connections. A self-attention module (H. Zhang et al., 2018) was added in the decoder part of the generator to capture long-range contextual information. In the encoder part of the generator, we used a channel attention module (Hu et al., 2018) in the generator network to prioritize the most important iris features, which are essential for off-angle to frontal iris image conversion. We used a channel attention layer before using a skip connection, enabling the selective highlighting of relevant information before its transmission to the decoder half of the generator. The generator uses information from the input image to trick the discriminator into producing a fake image. To increase the efficiency of the discriminator network's effectiveness, we used a multi-scale discriminator (Lee et al., 2023). This is not just labeling something as true or false. Instead, it measures the extent to which a thing is real compared to something else that it is known to be fake. It is the discriminator's responsibility to differentiate between images produced by the generator (fake images) and a real image. The discriminator is constantly modifying itself to assess the images created by the generator, and the generator is constantly evolving to create images that cannot be distinguished from their real equivalents.

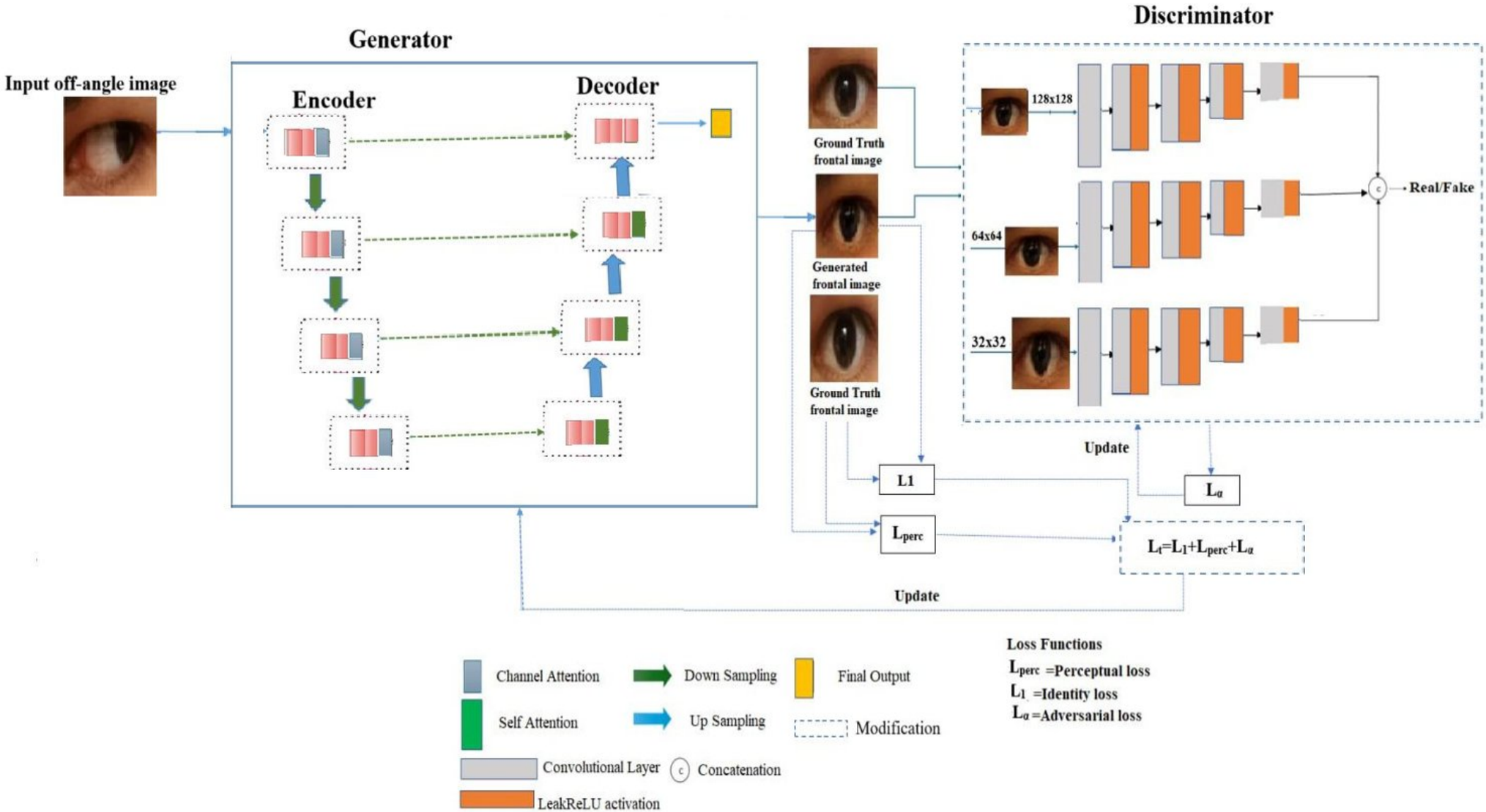


Figure 4.1: Overall architecture of the proposed model

4.3 Generator Architecture

Within the framework of Generative Adversarial Networks (GANs), the generator network is the network that produces simulated data that very closely resembles real data. This is typically expressed in the form of a neural network, usually with convolutional layers, which accepts random noise and outputs data samples that are intended to be indifferentaible to real data. Our model used U-Net architecture (Ronneberger et al., 2015b) that consists of an encoder-decoder structure combined with skip connections and hybrid attention (such as self-attention and channel-attention).

U-Net is a deep learning technique originally introduced by (Ronneberger et al. , 2015) and is characterized by two main pathways: contraction and expansion. The expansion pathway consists of decoder layers that decipher encoded information while also generating information based on the contraction pathway by skip connections to form a segmentation map. The contraction path is, in contrast, characterized by an architectural design of convolutional networks with encoder blocks capable of obtaining contextual information by downsampling spatial dimensions of the input (Ronneberger et al., 2015). Our encoder section of our generator contains Convolution and Batch normalization, and ReLU, while the decoder portion comprises Transposed convolution and Batch normalization and ReLU.

To get more abstract and compressed representations of the image, the bottleneck layer of the generator network compresses the input data and reduces the dimensionality of feature maps. Additionally, the generator part of the network uses skip connections to keep high-resolution feature details from the preceding layers, essential to preserving high-resolution iris detail features throughout the transformation process. The self-attention mechanism is used by the third and fourth decoder layers to allow the generator to capture both global and local dependencies in intermediate feature maps. This technique may be used to enhance the spatial complexities and details of the resultant frontal iris. Channel attention is applied in each of the encoder layers to enable the network to focus on the most crucial features.

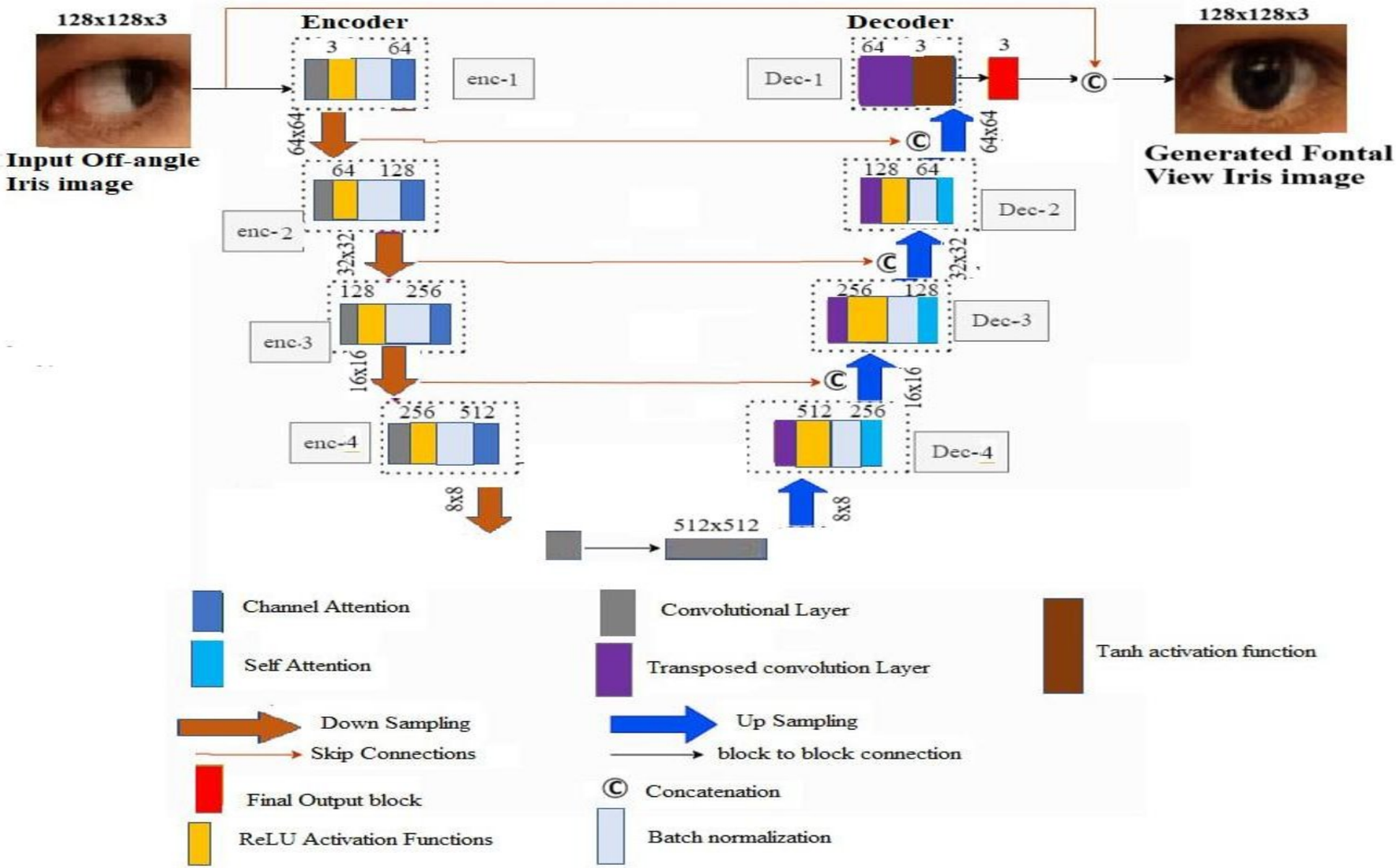


Figure 4. 2: Generator architecture

In the proposed approach, channel attention is used in the encoder block and self-attention in the U-Net-based generator decoder block in order to enhance its ability to produce a real-looking frontal iris image that can fool the discriminator. In the GANs framework, self-attention allows the generator and discriminator to learn long-range dependencies among image features more effectively (H. Zhang et al., 2018). In off-angle iris image to frontal representation conversion processes, preserving the interdependencies among significant iris features, such as edges, shape, and texture, is critical. This approach also preserves detailed iris features, i.e., crypts, furrows, freckles, and pigment spots, facilitating the generation of more realistic frontal iris images. Self-attention also removes typical GAN distortions and blur by introducing a way of refinement and enhancement of generated features according to their global context. This allows generation a cleaner and sharper frontal view iris image. To utilize these valuable features, we employed self-attention in the decoder part of the generator.

Channel attention is a technique used in deep convolutional neural networks (CNNs) to focus on the most important channel in a feature map. This improves network performance by minimizing the processing of noise and irrelevant information (Hu et al., 2018). It allows the model to focus on the more important regions of the iris, like edges, shape, and texture. This leads to better preservation of these elements in the generated frontal iris image. To make use of these important features to their best, we used channel attention on the encoder part of the generator. Applying channel attention in the encoder portion enables to establishment of a solid base for the subsequent layers of the model because the encoder improves key features by concentrating on the most significant features of the iris, which are important for precise conversion.

The encoder part in the proposed architecture acts as a down-sampling module, which consists of four convolutional layers (namely enc-1 to enc-4) and a channel attention mechanism after every convolutional layer. Every convolutional layer uses a stride of 2 and a kernel size of 4. Batch normalization and the ReLU activation function are also used after each convolutional layer. The decoder part, on the other hand, acts as an up-sampling module, which is made up of four Transposed Convolutional Layers or deconvolutional layers (namely Dec-1 to Dec-4). Self-attention block is added to the first, second, and third layers (Dec-4, Dec-3, Dec-2). A stride of 2 and a kernel size of 4 are used in all deconvolutional layers. ReLU activation function and Batch

normalization are used after each deconvolutional layer except the last layer (Dec-1). Tanh activation function is used in the last deconvolution layer (Dec-1).

4.4 Discriminator Network

The discriminator network serves a fundamental role in a Generative Adversarial Network (GAN), where its primary task is to differentiate between real and synthesized data samples. It is typically constructed as a neural network that commonly comprises convolutional layers, which takes input data samples both real and generated entries and outputs a probability score describing the probability of the input sample being real. The discriminator employed in this case is a multi-scale discriminator, following previous work, allowing it to generate high-quality frontal view iris images and improve the overall model performance. The generator consists of a network specifically crafted to produce a synthesized iris image, which is compared with the real iris image by the multi-scale discriminator. The discriminator sends gradient information to the generator, thereby allowing the generator to create images with higher quality. A multi-scale discriminator is a type of discriminator used in Generative Adversarial Networks. Unlike PatchGAN discriminators, which focus on classifying the authenticity of image patches rather than the whole image, multi-scale discriminators operate across multiple resolutions. This enables them to capture both global structural patterns and fine-grained details within an image. Given the enhanced ability of multi-scale discriminators to analyze visual content at various scales, we have adopted them in place of a PatchGAN discriminator of the baseline works. This choice significantly enhances the quality of the generated frontal iris images by ensuring consistency across spatial dimensions and improving the network's ability to learn complex textures and fine structural details.

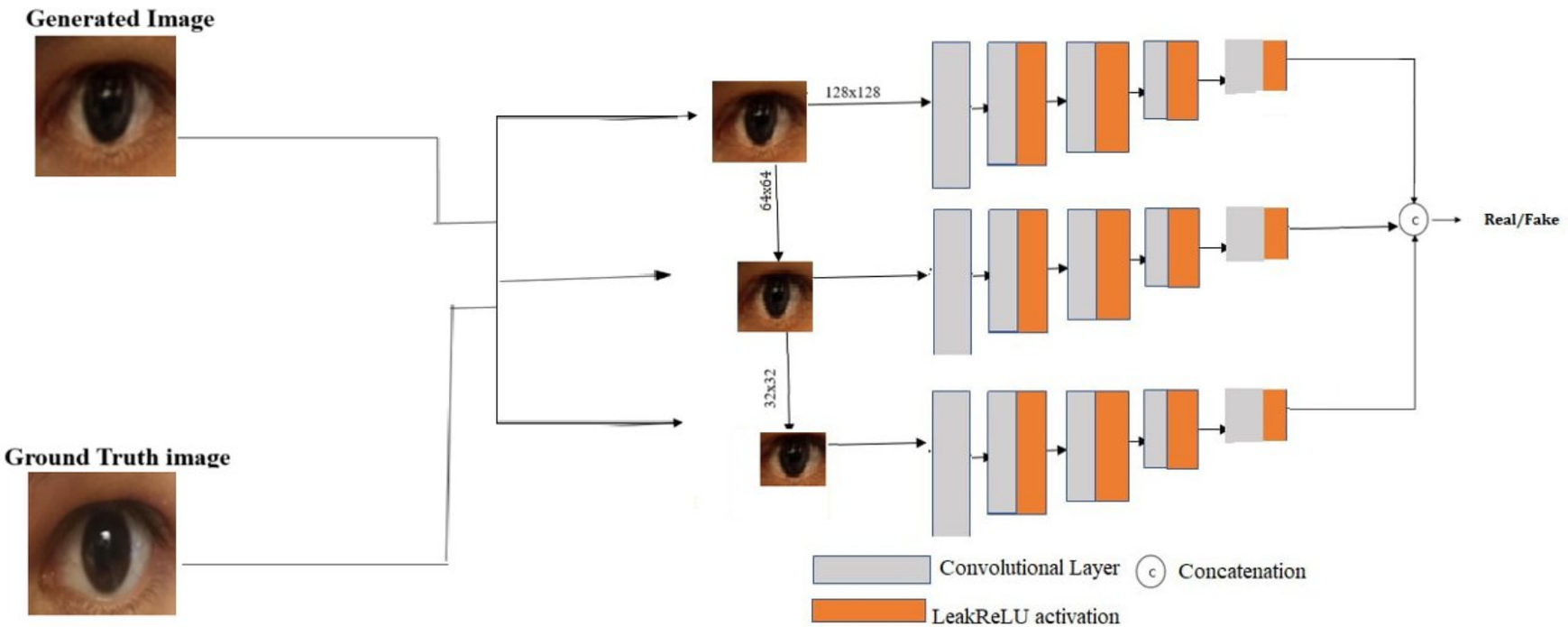


Figure 4. 5 Architecture of Multi-scale discriminators

4.5 Model Learning Functions

The primary objective of this study is to produce a high-quality image. To achieve this objective, various learning mechanisms like loss functions and attention mechanisms are introduced to the network. The loss functions employed in this work include adversarial loss $\mathcal{L}a$, identity loss ($\mathcal{L}1$), and perceptual loss ($\mathcal{L}_{perc}$).

$$\mathcal{L}_t = \mathcal{L}a + \mathcal{L}1 + \mathcal{L}_{perc} \qquad eq.\ 4.1$$

4.5.1. Adversarial loss

Adversarial loss is a loss function utilized in Generative Adversarial Networks (GANs. It calculates the difference between the predicted output of a generator and the real data distribution, forcing the generator to make its output look like real data as much as possible. The adversarial loss is a two-player game in which the discriminator attempts to correctly distinguish between actual and fake data while the generator attempts to fool the discriminator. The generator tries to minimize the adversarial loss, while the discriminator tries to maximize it.

The formula for adversarial loss can be defined as follows:

For the Discriminator

Discriminator is trained to increase the likelihood of correctly classifying the data as either real or fake data. The loss function for the discriminator is:

$$\mathcal{L}_D = -\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] - \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \qquad \text{eq. 4.2}$$

Where:

- x denotes real data samples from the actual data distribution
- G(z) represents fake data generated by the generator using random noise z.
- D(x) is the probability that the discriminator assigns to the real data being real.
- D(G(z)) is the probability that the discriminator assigns to the generated data being real.
- pdata(x) is the real data distribution, and pz(z) is the noise distribution used by the generator.

For the Generator G:

The generator is trained to minimize the adversarial loss, effectively forcing the discriminator to classify the generated data as real. The loss function for the generator is:

$$\mathcal{L}_G = -\mathbb{E}_{z\sim p_z(z)}[\log D(G(z))] \quad \text{eq. 4.3}$$

Where:

- $\mathcal{L}G$ is the generator's loss, representing how much the generator's output is classified as real by the discriminator.

4.5.2. Identity Loss

L1 is a popular loss function that lowers the absolute value of differences between estimated values and existing target values. The L1 loss function is more robust and, for the most part, is not affected by outliers. $\mathcal{L}1$ loss was used between the generated and the actual images (Isola et al.,2017).

$$\mathcal{L}1 = |It - I\tilde{t}| \quad \text{eq. 4.4}$$

Where It is the ground truth (real) image and $I\tilde{t}$ is the generated image

4.5.3. Perceptual loss

It calculates the difference in high-level features of images as seen by a pre-trained network (usually a convolutional neural network such as VGG, MobileNetv2) instead of directly comparing pixel values (Johnson et al., 2016).

$$\mathcal{L}_{\text{perceptual}}(\hat{I}, I) = \sum_{l\in L} \lambda_l \left\| \phi_l(\hat{I}) - \phi_l(I) \right\|_2^2 \quad \text{eq. 4.5}$$

Where:

- $\hat{I}$ is the generated image.
- I is the target image.
- ϕl represents the feature maps extracted at the l-th layer of the pretrained network (e.g., VGG).

- L is the set of layers from which features are extracted (e.g., layers like Conv1_2, Conv2_2, etc. in VGG).
- λl is a weighting factor that controls the importance of the feature map at each layer.

4.6 Training Pseudo Code

This section explains the overall training process of the hybrid attention and Multi-Scale Discriminator GAN model.

Algorithm 1: Overall training process of proposed model

1. **Initialize** *all generator G and discriminator D weight*
   Channel attention *at encoder and* ***self-attention*** *at decoder part of U-net*
   Multi-scale discriminator *with in discriminator part*
2. **Input***: batch of off-angle (I_o) and frontal iris images (I_f)*
3. ***Set*** *training parameters:*
   Batch size:16
   Optimizer: Adam
   Learning rate:0.0002 for Generator G and 0.0001 for discriminator D
   Number of epochs: 100

4. ***For*** *number of epochs* ***do***

5. ***For*** *number of batches in epoch* ***do***

6. ***Update*** *discriminator using adversarial loss*

7. ***Update*** *generator G using adversarial loss, identity loss and perceptual loss*

8. ***Back propagate*** *and* ***update*** *parameters*

9. ***Save*** *the generator and discriminator weights*

10. ***End for***

11. **Output**: *Converted frontal iris image*

Bold, indicates the updated components

CHAPTER FIVE

5. IMPLEMENTATION OF PROPOSED MODEL

5.1 Chapter Overview

This chapter provides a detailed description of the processes involved in carrying out the proposed hybrid attention and multi-scale discriminator-based Generative Adversarial Networks (GAN) design for off-angle to frontal iris image conversion. It consists of environment setup, data preprocessing, proposed model implementation, hyperparameter configuration, and experimental class undertaking. Furthermore, this section provides some code sample snippets to implement the proposed solution.

5.2 Working Environments

This chapter explains the environmental setups and details of the hardware tools used in model implementation. The primary characteristics of the proposed method are a multi-scale discriminator-based GAN and off-angle to frontal iris image conversion working settings with hybrid attention. It contains the hardware and software used in the study. For efficient data processing, the hardware consists of a PC with a powerful CPU and a GPU. The proposed method for off-angle to frontal iris image conversion employing a hybrid attention and multi-scale discriminator-based GAN may be developed, trained, and tested effectively in the following scenarios.

A laptop

- System software: Windows 11.
- Processor: 3.20 GHz, 3.19 GHz Intel® Core i5-8700 CPU
- System type: 64-bit operating system, x64-based processor;
- Installed memory (RAM): 8.00 GB

Desktop:

- System software: Windows 10
- Processor: 3.30GHz 3.19 GAHz Intel® Core i5-4590 CPU
- Installed memory (RAM): 4.00 GAB
- System type: x64-based processor, 64-bit operating system

Graphics processing unit (GPU):

- Graphics: Intel® HD Graphics 520/NVIDIA Corporation
- GeForce RTX 3060 with 12 GB RAM

5.3 Environmental Setup

This study was carried out using a range of software and Integrated Development Environments (IDEs) that support well-known machine learning frameworks like PyTorch.

Anaconda: Anaconda is a working environment that uses multi-scale discriminator-based GAN conversion and hybrid attention to process off-angle to frontal iris images. It is an open-source convenience utility used to manage packages with ease and is widely used to install Python and packages.

Anaconda version 1.9.7 is utilized in the execution of the proposed approach in order to retrieve and install all the required libraries and modules.

Jupyter Notebook: is a free environment used in this project to write and debug the hybrid attention and multi-scale discriminator-based GAN for off-angle to frontal iris image transformation. It combines code, results, text, and multimedia into a single interface and allows easy documentation, testing, and iteration.

Kaggle: is a cloud platform specializing in data science and machine learning, offering free GPU and TPU access and a vast array of datasets and code notebooks. Kaggle enables the writing and execution of Python code locally without setup, in addition to collaboration, learning, and sharing of projects, thus making it highly applicable for research and development of deep learning.

Colab: is a cloud-based environment for editing and running Python code. It eliminates the need for local installation, making it ideal for quick start-up and providing free access to GPUs for a limited time. These powerful graphics processing units can significantly accelerate the running of computationally intensive machine learning tasks, especially those involving deep learning models.

5.4 Training Procedure

The following steps were carried out to accomplish the objectives of this study.

- Data Collection: datasets of frontal and off-angle iris images were collected, e.g., from public Repositories.
- Data preprocessing: Image data was loaded, quality checked, and resized to standard sizes.
- Data Augmentation: Techniques involved in data augmentation, including geometric and photometric improvements and generative production techniques, were applied to expand the training data.
- GPU Configuration: NVIDIA GPU server was configured, which includes CUDA/PyTorch installation and driver installation.
- Implementing the proposed solution.

5.5 Dataset Preparation

5.5.1 Data Splitting and Loading

Data is randomly divided into three categories: train (70%), validation (15%), and test (15%). The test data is initially separated from the dataset, followed by the validation data from the training data to prevent data leakage. Resizing iris input images during training ensures consistent dimensions, improves computational efficiency, and also reduces the training time. Every input image is resized to a size of 128x128 pixels. The preprocessing step of rescaling by 1/255 normalizes the pixel values of the input images between 0 and 1. This process is called normalization or standardization. This leads to better numerical stability and gradient-based optimization, accelerates faster convergence, equates all the color channels equally, and may result in better model performance.

```
dataset = OffAngleToFrontalDataset("/kaggle/input/dataset/final MOBIUS/off-angle",
"/kaggle/input/dataset/final MOBIUS/frontal", transform=transform)
train_size = int(0.7 * len(dataset))
val_size = int(0.15 * len(dataset))
test_size = len(dataset) - train_size - val_size
train_dataset, val_dataset, test_dataset =
torch.utils.data.random_split(dataset, [train_size, val_size, test_size])
train_loader = DataLoader(train_dataset, batch_size=16, shuffle=True)
val_loader = DataLoader(val_dataset, batch_size=16)
test_loader = DataLoader(test_dataset, batch_size=16)
```

Sample code 5.1: Dataset Loading and Splitting implementation

5.5.2 Dataset Augmentation Implementation

As was previously discussed, in order to enhance the diversity of the training data and the model's performance, this study utilized a variety of data augmentation approaches to the training images. Many distinct versions of the images can be produced by these data augmentation approaches. Random flipping, random brightness, random saturation modification, random contrast modification, and brightness adjustment approach were among the various image augmentation methods employed.

```
#Data Augmentation
transform = transforms.Compose([
    transforms.RandomHorizontalFlip(),
    transforms.RandomRotation(15),  # Random rotation between -15 and 15 degrees
    transforms.ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.1),
    transforms.ToTensor(),
])
```

Sample code 5.2: Dataset augmentation implementation

5.6 Implementation of Proposed Model

The proposed model is made up of a single U-net-based generator and a multi-scale discriminator. The generator, made up of U-net, as described in section 4.3, consisted of two network blocks: an encoder block and a decoder block that contain hybrid attention mechanisms. Both the decoder block and the encoder block are four-layer neural networks with a bottleneck layer that consists of two convolutional layers followed by ReLU activations and a residual connection. Channel attention mechanism is applied in the encoder block, whereas the self-attention mechanism is employed in the decoder block. ReLU activation function and Batch normalization are employed in both blocks of the generator.

Table 5. 1: Content of encoder block in generator architecture

Architecture of Encoder block in Generator	
Layers in the Encoder	**Content of Each Block**
1	Conv2d (64,3, kernel size=4, stride=2, padding=1),BatchNorm2d(64),ReLU()
2	Conv2d (128,64, kernel size=4, stride=2, padding=1),BatchNorm2d(128),ReLU()
3	Conv2d (256, 128, kernel size=4, stride=2, padding=1), BatchNorm2d (256), ReLU ()
4	Conv2d (512, 256, kernel size=4, stride=2, padding=1), BatchNorm2d (512), ReLU ()

Table 5.1 describes about the contents utilized in the encoder block of the generator. Four convolutional layers used in encoder block with kernel size 4x4, stride two, and padding of one. We also used Batch normalization to stabilize the training of the generator so that the output samples have controlled distributions. ReLU activation function is also used in each layer to incorporate non-linearity in the model and prevent vanishing gradients issue.

Table 5. 2: Content of bottleneck layer

Bottleneck Layer	**Content of Bottleneck Layer**
1	Conv2d (512,512, kernel size=3, padding=1), ReLU()
2	Conv2d (512,512, kernel size=3, padding=1), ReLU()
3	Conv2d (512,512, kernel size=1)

The bottleneck layer in a U-Net structure has an important role in connecting the encoder and decoder paths while encapsulating and altering the most essential features extracted from the input data. As shown in table 5. 2, The bottleneck layer comprises two convolutional layers

followed by ReLU activations (layer 1 and layer 2) and a residual connection (Layer 3). This helps in learning more complex features while preserving important information through residual mapping. Both convolutional Layer blocks consist of kernel size 3x3, and padding one.

Table 5. 3: Content of decoder block in generator architecture

Architecture of Decoder block in Generator	
Layers in Decoder	**Content of each block**
1	ConvTransposed (256, 512, kernel size=4, stride=2, padding =1), BatchNorm2d (256), ReLU ()
2	ConvTransposed2d (128, 256, kernel size =4, stride=2, padding=1), BatchNorm2d (128), ReLU ()
3	ConvTranspose2d (64, 128, kernel size=4, stride=2, padding=1), BatchNorm2d (64), ReLu ()
4	ConbTranspose2d (3, 64, kernel size=4, stride=2, padding=1), Tanh()

Table 5.3, describe the contents utilized in the generator's decoder block. With a kernel size of 4x4, a stride of 2, and a padding of 1, the decoder block employs four deconvolutional layers (Transposed Convolutional Layers). In order to stabilize the generator's training and control the distribution of the generated samples, we additionally performed batch normalization to every layer save the output layer. ReLU activation function is also applied in all the layers except the last layer to avoid non-linearity into the model and avoid the vanishing gradients problem. Due to our image is normalized in the range of [-1,1], we have applied Than activation function in the last layer to obtain output in the range of [-1,1].

Table 5. 4: Content of multi-scale discriminator architecture

Layers in multi-scale discriminator	**Content of each block**
1	Conv2d (in chnnels, out channels, kernel size=4, stride=2, padding=1), LeakReLu (0.2)
2	Conv2d (out channels, 128, kernel size=4, stride=2, padding=1), LeakyReLU (0.2)
3	Conv2d (128, 256, kernel size=4, stride=2, padding=1), LeakyReLu (0.2)
4	Conv2d (256, 512, kernel size=4, stride=2, padding=1), LeakyReLU (0.2)
5	Conv2d (512, 1, kernel size=4, stride=1, padding=0)

As presented in table 5.4, the multi-Scale discriminator contains three parallel discriminators that work on the input image at multiple scales. The individual discriminators have five convolutional layers with kernel size 4×4. The first four convolutional layers take stride two and padding one, and the last convolutional layer takes stride one with no padding. LeakyReLU activation is used on all convolutional layers except the last one. To deal with several scales, the second and the third discriminators are given down sampled versions of the input by bilinear interpolation to resolutions of 64×64 and 128×128, respectively. The ultimate output is three sets of predictions at various scales, which enables the model to learn features at varying levels of details.

5.6.1 Implementation of Self-attention

According to section 4.3, self-attention enables the generator to pay attention to not just fine details (local dependencies) but also how various areas of the iris relate to one another across the entire image (global dependencies). This enables the generator to produce irises with much sharper, finer details and forms, making them look more realistic. The self-attention layer is utilized in the third

and fourth levels of the decoder. Before the application of self-attention, the input image is typically converted into a sequence of vectors. Each vector (being a pixel or pixels) is a query, a key, and a value, these arguments are computed employing the convolution over the input image with kernel size 1*1. Dot products are calculated for every query vector with all key vectors to indicate the relevance (attention) of every key regarding the query. Attention weights are therefore provided. To obtain attention weights, SoftMax normalizes these attention scores. The output of the self-attention mechanism for each question is then provided by the weighted sum of value vectors, which are weighted by attention scores. In order to preserve spatial information, the resulting vectors are often returned to their initial spatial arrangement configuration after the self-attention process has been applied to each layer in this instance.

```
class SelfAttention(nn.Module):
    """
    Self-Attention module for capturing long-range dependencies.
    """
    def __init__(self, in_channels):
        super(SelfAttention, self).__init__()
        self.query = nn.Conv2d(in_channels, in_channels // 8, kernel_size=1)
        self.key = nn.Conv2d(in_channels, in_channels // 8, kernel_size=1)
        self.value = nn.Conv2d(in_channels, in_channels, kernel_size=1)
        self.gamma = nn.Parameter(torch.zeros(1))  # Learnable parameter

    def forward(self, x):
        batch_size, C, H, W = x.size()
        query = self.query(x).view(batch_size, -1, H * W).permute(0, 2, 1)  # (B, H*W, C//8)
        key = self.key(x).view(batch_size, -1, H * W)  # (B, C//8, H*W)
        attention = torch.bmm(query, key)  # (B, H*W, H*W)
        attention = F.softmax(attention, dim=-1)  # Normalize attention scores
        value = self.value(x).view(batch_size, -1, H * W)  # (B, C, H*W)

        out = torch.bmm(value, attention.permute(0, 2, 1))  # (B, C, H*W)
        out = out.view(batch_size, C, H, W)
        out = self.gamma * out + x  # Add attention map to input
        return out
```

Sample code 5.3: Self-attention Implementation

5.6.2 Implementation of Channel Attention

Channel attention is utilized into the encoder block of the generator network to assist in focusing important iris features. The channel attention mechanism includes of two components: squeezing and excitation blocks. Global average pooling and global Max pooling are employed in squeezing block on the input feature maps to produce channel descriptors. In excitation block there are two full connected layers: dimensionality reduction and dimensionality restoration. The number of

channels is decreased from C to C/r by first feeding the channel descriptors of the squeeze block to a fully linked layer with reduction ratio r. The number of channels is then returned to C by passing the output of the first fully connected layer—that is, through another fully connected layer. Lastly, the channel attention weights are obtained by using a sigmoid activation function. Then, the α channel attention weights scale the weight of the importance of the early feature map X. The feature Maps are reweighted based on the respective attention weight given to each channel.

```
class ChannelAttention(nn.Module):
    """
    Channel Attention module for enhancing important feature channels.
    """
    def __init__(self, in_channels, reduction=16):
        super(ChannelAttention, self).__init__()
        self.fc1 = nn.Conv2d(in_channels, in_channels // reduction, kernel_size=1)
        self.fc2 = nn.Conv2d(in_channels // reduction, in_channels, kernel_size=1)
        self.relu = nn.ReLU()

    def forward(self, x):
        avg_pool = F.adaptive_avg_pool2d(x, 1)
        max_pool = F.adaptive_max_pool2d(x, 1)
        out = self.fc1(avg_pool) + self.fc1(max_pool)
        out = self.relu(out)
        out = self.fc2(out)
        return x * torch.sigmoid(out)
```

Sample code 5.4: Channel Attention Implementation

5.6.3 Implementation of Multi-Scale Discriminator

To improve the discriminator to differentiate real and generated images with different levels of detail, we employed a Multi-Scale Discriminator in the GAN model for off-angle to frontal iris image conversion. The Multi-Scale Discriminator works on three distinct scales, enabling closer inspection of features. At each level of analysis, convolutional layers are applied to the input image in order to capture characteristics that aim to capture features at different hierarchies. To ensure consistency between different scales, bilinear interpolation is utilized for down-sampling the input image for the second and third scales. Each scale outputs independently, and these outputs all contribute to the final adversarial loss, resulting in increased stability and robustness during training. By modeling features at several resolutions, the Multi-Scale Discriminator improves the model's capacity to capture fine details and enforce structural consistency, thereby generating more realistic converted frontal iris.

```
class MultiScaleDiscriminator(nn.Module):
    def __init__(self):
        super(MultiScaleDiscriminator, self).__init__()
        self.scale1 = self.build_discriminator(3, 64)
        self.scale2 = self.build_discriminator(3, 64)
        self.scale3 = self.build_discriminator(3, 64)

    def build_discriminator(self, in_channels, out_channels):
        return nn.Sequential(
            nn.Conv2d(in_channels, out_channels, kernel_size=4, stride=2, padding=1),
            nn.LeakyReLU(0.2),
            nn.Conv2d(out_channels, 128, kernel_size=4, stride=2, padding=1),
            nn.LeakyReLU(0.2),
            nn.Conv2d(128, 256, kernel_size=4, stride=2, padding=1),
            nn.LeakyReLU(0.2),
            nn.Conv2d(256, 512, kernel_size=4, stride=2, padding=1),
            nn.LeakyReLU(0.2),
            nn.Conv2d(512, 1, kernel_size=4, stride=1, padding=0)
        )
    def forward(self, x):
        # Generate outputs for different scales
        scale1_out = self.scale1(x)
        scale2_out = self.scale2(F.interpolate(x, size=(64, 64), mode='bilinear',
        align_corners=False))
        scale3_out = self.scale3(F.interpolate(x, size=(32, ), mode='bilinear',
        align_corners=False))
        return (scale1_out, scale2_out, scale3_out)
```

Sample code 5.5: Multi-scale Discriminator implementation

5.6.4 Implementation of Adversarial Loss Function

Adversarial loss is used in the off-angle iris image to the frontal view transformation process with hybrid attention and multi-scale discriminator-based GAN. In this framework, the adversarial loss drives the generator to create images that are photorealistic and identical to real frontal iris images, according to the judgment of the discriminator. Incorporating this loss function in the training process leads the model to synthesize high-quality iris textures along with identity-persistent features, generally lost in traditional methods that utilize mean square loss, matrix multiplication, or SSIM loss. This adversarial framework allows the generator to capture fine-grained structural and textural relationships, both resulting in visually and semantically improved frontal iris reconstructions from non-frontal inputs.

```
def adversarial_loss(logits, target_is_real):
    target = torch.ones_like(logits) if target_is_real else torch.zeros_like(logits)
    return criterion_adv(logits, target)
```

Sample code 5.6: Adversarial Loss implementation

5.6.5 Implementation of Identity Loss Function

To ensure the generated frontal iris image preserves the original input's distinctive identity characteristics, identity loss is used in this study. In this case, the off-angle iris images are geometrically distorted and occluded, which reduces the performance of recognition systems. By imposing identity loss during training, the GAN is persuaded to preserve the identity-salient features of the subject, which encourages the generator to generate frontal iris images that are not just visually indistinguishable but also semantically similar to the input given. This results in improved identity consistency and enhances the robustness of iris recognition at non-frontal orientations.

```
def identity_loss(real, fake):
    return F.l1_loss(fake, real)
```

Sample code 5.7: Identity Loss implementation

5.6.6 Implementation of Perceptual Loss Function

Traditional loss functions such as mean square error, matrix product-based similarity, or SSIM may not capture fully the high-level structural and perceptual features necessary for realistic iris synthesis. The incorporation of perceptual loss in the optimization process encourages the synthesis of frontal iris images that preserve complex texture and identity-relevant details, while at the same time adhering to human visual realism. This improvement enhances the realism of the synthesized iris, thus enabling more accurate and reliable biometric recognition.

```
def perceptual_loss(real, fake, feature_extractor):
    # Normalize images to the range expected by MobileNet ([0, 1])
    real = (real + 1) / 2
    fake = (fake + 1) / 2

    # Resize images to the expected input size for MobileNet (224x224)
    real = F.interpolate(real, size=(224, 224), mode='bilinear', align_corners=False)
    fake = F.interpolate(fake, size=(224, 224), mode='bilinear', align_corners=False)

    real_features = feature_extractor(real)
    fake_features = feature_extractor(fake)
    return F.l1_loss(fake_features, real_features)
```

Sample code 5.8: Perceptual Loss implementation

5.7 Hyperparameter Configuration

Hyperparameters are the parameters that are adjusted to manage the learning process of a machine learning model. They are responsible for controlling model training and can be optimized to achieve optimum performance. During our experimentation, we trained our model under different hyperparameter configurations such as 30 epochs, 50 epochs, batch size 8, and various learning rates to evaluate its performance across a range of setups. After iterative experiments, we selected the optimal hyperparameters based on the model's stability and ability to generate high-quality frontal iris images.

For our training pipeline, we employed the Adam optimizer to update the weights of the UNetGenerator and the Multi-Scale Discriminator. Adam is an adaptive optimization algorithm that leverages the advantages of AdaGrad and RMSProp (Kingma & Ba, 2014). The algorithm automatically computes the learning rates for different parameters, thereby providing better stability and convergence. The learning rate for the generator is 0.0002, and the discriminator is subjected to a lesser learning rate of 0.0001, which balances the networks effectively. The Adam momentum parameters are $\beta_1 = 0.5$ and $\beta_2 = 0.999$, which aid in minimizing oscillations and stabilizing training. Gradient clipping, with norm 1.0, is used to prevent the gradient explosion phenomenon.

A batch size of 16 is utilized during training, which continues for 100 epochs. Learning rate is modified with the help of a StepLR scheduler, defined by a step interval of 10 epochs and a decay factor of 0.5, thus allowing the model to gradually learn better during training. To improve perceptual quality, we merge a perceptual loss with a pre-trained Mobile Net V2 feature extractor on iris images and adversarial and identity loss for preserving fine details in off-angle to frontal iris image conversion. We utilize two steps of gradient accumulation to effectively increase the batch size in the memory-limited limitation. The model's performance is tracked during training, and loss values are recorded for the generator and discriminator at every epoch.

Table 5.5: Hyperparameter Configuration

Hyperparameter	**Values**
Batch size	16
Number of epochs	100
Optimizer	Adam
Learning Rate	0.0002

5.8 Experimental Class

To determine the effectiveness of hybrid attention mechanisms and multi-scale discriminators for enhancing iris image quality conversion, numerous experiments are carried out and well tested.

Four experiments have been conducted, as illustrated in Table 5.6. The baseline study is the first experiment. As discussed in Section 3.5, the Pix2Pix-GAN employed a generator based on the U-Net architecture and a PatchGAN discriminator to convert off-angle to frontal iris conversion with a combination of Mean Squared Error(MSE), Matrix Multiplication(MatMul) loss, and Structural Similarity Index (SSIM) loss functions to guide the training process.

The second experiment is the first test of the new model. We introduced a self-attention block into the decoder section of the U-Net-based generator and a channel attention block on the encoder part of the generator. In the third experiment, we used the multi-scale discriminator on the discriminator part of the GAN. The fourth experiment demonstrates the complete model proposed, which incorporates channel attention in the encoder blocks and employs self-attention mechanisms in the decoder block of the U-Net generator network. It employs a multi-scale discriminator in the discriminator network, with adversarial loss, perceptual loss, and identity loss instead of the mean square loss function, Matrix Multiplication, and SSIM loss function of the baseline work to generate sharper images that can capture the textural information of the original image better with high quality and consistency.

Table 5.6: Experiment class

Notation	**Experiment**	**Dataset Used**
Pix2Pix-GAN	Baseline	UBIRIS,v2 and MOBIUS Dataset
Pix2Pix-GAN+SA+CA	Using self-attention in the decoder part and channel attention in the encoder part of the U-net-based generator	UBIRIS,v2 and MOBIUS Dataset
Pix2Pix-GAN+SA+CA+MSD	Using multi-scale discriminator at discriminator parts of the GAN Network	UBIRIS,v2 and MOBIUS Dataset
Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}	Utilizing the combination of adversarial loss, perceptual loss, and identity loss instead of the mean square loss function, Matrix Multiplication, and SSIM loss	UBIRIS,v2 and MOBIUS Dataset

In Table 5.6, a summary of different experiments in this work is listed, which consist of off-angle to frontal iris image transformation using hybrid attention and multi-scale discriminator-based GAN.

In the first experiment, we used baseline work Pix2Pix GAN combined Mean Squared Error (MSE), Matrix Multiplication (MatMul) loss, and Structural Similarity Index (SSIM) loss with a PatchGAN discriminator and a U-Net-based generator.

The second experiment utilized a self-attention mechanism within the generator decoder and channel attention in the encoder to address issues of image blurriness and the challenge of maintaining detailed information, particularly during the transformation of off-angle iris images with significant off-angles exceeding 30° and off-angle scenario like occluded iris image, partial visible and distorted iris image due to large angle or 3D structure of eye. This self-attention process assists in enhancing the model's capacity to preserve long-range dependencies within the image and therefore, ensure better detail retention and reduced blurriness. The hybrid attention mechanism facilitates improved iris feature retrieval, including texture details, reflections, and occlusions. Channel attention helps to highlight the most significant feature channels, thereby enhancing the extraction of useful iris features.

A third experiment was carried out with a multi-scale discriminator to extend the adversarial learning model by evaluating generated images at different resolutions. The multi-scale approach maintains both the global structure and fine details of the iris, creating more realistic iris images after conversion.

The fourth experiment illustrates the complete proposed model, which includes channel attention in the encoder blocks and self-attention mechanisms in the decoder block of the U-net-based generator network. It uses the multi-scale discriminator in the discriminator network, with Adversarial loss, perceptual loss, and Identity loss.

CHAPTER SIX

6. RESULTS AND DISCUSSION

6.1 Overview of Chapter

This chapter presents the results of all the experiments, including the new approach we proposed, as well as the baseline work. Charts, graphs, and tables are used to present the results in an understandable way so that comparisons may be made.

6.2 Experimental Results

The network of the proposed model was trained through numerous iterations while taking into account a number of factors. To provide for a fair comparison and guarantee reliable results for this study, four experiments were carried out using the same datasets, hardware setups, and software settings. Two publicly accessible datasets the MOBIUS dataset (Vitek et al., 2020) and the UBIRIS.V2 dataset (Padole & Proenca, 2012) were used to test our model. As demonstrated by a comparison of our findings, the proposed model converts off-angle iris images and off-angle iris image scenarios such as occluded iris images caused by eyelids, partial visible iris images, and distorted iris images caused by 3D structure of eye's or due to large angle to frontal view and provides the best quantitative values based on quantitative evaluation metrics.

6.3. Results based on Evaluation Metrics

The performance of our proposed model was evaluated using three metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Hamming distance. PSNR quantifies the extent to which the generated frontal view iris images preserves the original frontal iris image's features and iris structure as a whole. And, rather than simply measuring noise levels, SSIM measures how accurately the generated front view precisely defines the spatial characteristics of the iris, such as the pupil, iris curves, and sclera boundaries. Thus, we obtained those measurements as crucial metrics to compare the proposed model's performance. Tables 6.1 illustrate the PSNR and SSIM, and Hamming Distance results of experiments on the MOBIUS dataset, respectively. Tables 6.2, Table 3 and Table 6.4 also present PSNR and SSIM, and the Hamming Distance experiment result on the UBIRIS.V2 dataset, respectively. The higher value of the PSNR, SSIM, and lower values of the Hamming Distance indicate that the generated frontal iris images are more similar to the original frontal iris image in both quality and consistency.

6.3.1 Evaluation of Experiments on MOBIUS Dataset

Table 6.1: PSNR, SSIM, and Hamming Distance evaluation on *MOBIUS* Dataset

S.No	Experiments	PSNR	SSIM	Hamming Distance
1	Pix2Pix-GAN	18.75	0.61	0.97
2	Pix2Pix-GAN+SA+CA	19.61	0.69	0.63
3	Pix2Pix-GAN+SA+CA+MSD	25.01	0.81	0.49
4	Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}	**28.94**	**0.82**	**0.44**

As shown in Table 6.1, the bold value shows the best results of our proposed model (Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}) in the experiments that outperforms all other approaches by improving PSNR values from **18.75** to **28.94** and SSIM values from **0.61** to **0.82** with the MOBIUS dataset image. It also achieves a lower Hamming Distance than other experiments by achieving **0.44** for converting occluded iris images caused by eyelids, partially visible iris images, and distorted iris images caused by its 3D structure of the eye or due to a large angle to frontal iris images. This implies that the model generates higher-quality image sequences than those found in previous experiments.

6.3.2 Evaluation of Experiments on UBIRIS.V2 Dataset

Table 6.2: PSNR evaluation on UBIRIS.V2 Dataset

S.No	Experiments	-30^0	30^0	-45^0	45^0
1	Pix2Pix-GAN	28.00	28.01	22.97	23.97
2	Pix2Pix-GAN+SA+CA	28.05	28.10	22.99	25.98
3	Pix2Pix-GAN+SA+CA+MSD	28.05	28.05	25.99	27.97
4	Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}	**30.34**	**30.31**	**27.99**	**28.98**

As shown in Table 6.2, the bold value shows the best results of our proposed model (Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}) in the experiments that outperforms all other approaches by improving PSNR values from 28.00 to 30.34 on -30^0,28.01 to 30.31 on 30^0,22.97 to 27.99 on -45^0, and 23.97 to 28.98 on 45^0 with the UBIRIS.V2 dataset image.

Table 6.3: SSIM evaluation on UBIRIS.V2 Dataset

S.No	Experiments	-30^0	30^0	-45^0	45^0
1	Pix2Pix-GAN	0.69	0.70	0.61	0.60
2	Pix2Pix-GAN+SA+CA	0.70	0.75	0.69	0.65
3	Pix2Pix-GAN+SA+CA+MSD	0.74	0.74	0.74	0.71
4	Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}	**0.79**	**0.80**	**0.76**	**0.75**

As shown in Table 6.3, the bold value shows the best results of our proposed model (Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}) in the experiments that outperforms all other approaches, achieving SSIM value from baseline work (Pix2Pix-GAN) results 0.69 to 0.79 on -30^0,0.70 to 0.80 on 30^0,0.61 to 0.76 0n -45^0, and 0.60 to 0.75 on 45^0, which indicates our proposed model generates frontal iris images that are more similar to the original frontal iris image in both quality and consistency.

Table 6.4: Hamming Distance evaluation on UBIRIS.V2 Dataset

S.No	Experiments	-30^0	30^0	-45^0	45^0
1	Pix2Pix-GAN	0.47	0.44	0.49	0.5
2	Pix2Pix-GAN+SA+CA	0.44	0.42	0.45	0.47
3	Pix2Pix-GAN+SA+CA+MSD	0.39	0.40	0.45	0.45
4	Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}	**0.34**	**0.36**	**0.42**	**0.44**

As shown in Table 6.4, the bold value shows the best results of our proposed model (Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}) in the experiments that outperforms all other approaches , achieving lower hamming Distance value from baseline work (Pix2Pix-GAN) results 0.47 to 0.34 on -30^0,0.44 to 0.36 on 30^0,0.49 to 0.42 on -45^0, and 0.5 to 0.44 on 45^0, with the UBIRIS.V2 dataset image which indicates our proposed model generates frontal iris images that are more similar to the original frontal iris image in both quality and consistency.

6.4 Converted Off-angle Iris Image to Frontal View: Results of Experiments

In this section, the converted off-angle iris images are presented for all experiments, including the baseline work. The first row indicates an input off-angle iris image with its angle, the next rows indicate the frontal results for the experiments, and the last row indicates the ground truth frontal iris image.

6.4.1 Converted off-angle iris image Results of Experiments for UBIRIS.v2 dataset

As shown on the figure 6.1, four experiments were conducted to convert off-angle iris images at -30^0, 30^0, -45^0 and 45^0 to frontal views. The first experiment (Row 2) used Pix2Pix-GAN with MSE, MatMul, and SSIM losses, but the results had prominent artifacts. In the second experiment (Row 3), Self-Attention (SA) and Channel Attention (CA) were added, which improved the structural integrity and details. The third experiment (Row 4) added a Multi-Scale Discriminator (MSD), which greatly improved the realism over the basic PatchGAN method. The last experiment (Row 5), titled *Ours*, integrated SA, CA, MSD, and a novel combination of losses adversarial, identity, and perceptual losses and generated outputs closest to the ground truth (Row 6). The figure illustrates incremental improvement seen with each experiments.

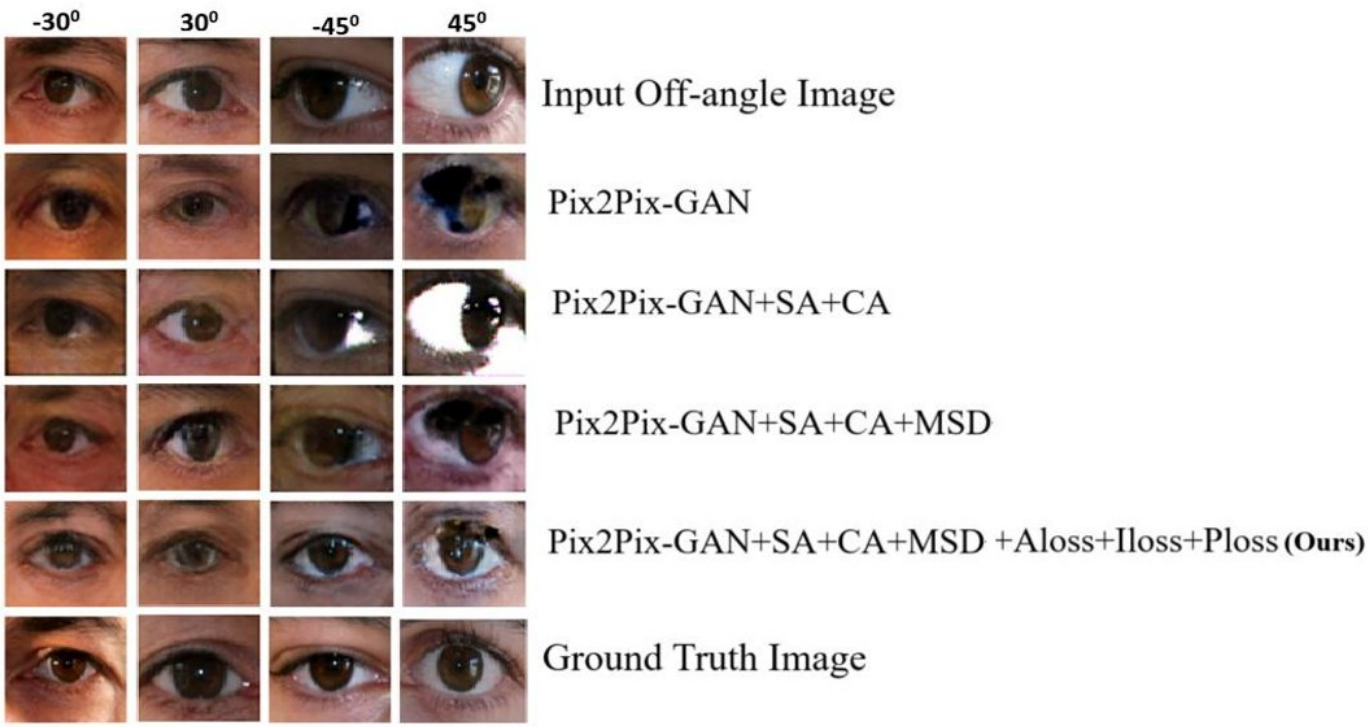


Figure 6.1: UBIRIS.v2 Converted off-angle iris image in different experiment classes.

6.4.2 Converted off-angle iris image Results of Experiments for MOBIUS dataset

As shown in the figure 6.1, four experiments were conducted to generate frontal iris images from a series of off-angle views challenges like partially visible iris images, distorted and occluded iris images. The first row illustrates input iris images that encounter off-angle issues such as partial visibility or occlusions. The second row displays the output generated by the baseline Pix2Pix-GAN, which encounters issues with artifacts and distortions in shape. The third row utilizes Pix2Pix-GAN with Self- Attention (SA) and Channel Attention (CA) for better image structure and sharpness. The fourth row uses a Multi-Scale Discriminator (MSD), further improving visual quality and realism of features. The final model, displayed in the fifth row and labeled as *Ours*, is the combination of SA, CA, MSD with adversarial, identity, and perceptual losses that generates the most accurate and consistent frontal iris images, with the least difference to the ground truth in the last row. This progression demonstrates the capacity of each component in handling complex gaze and occlusion conditions.

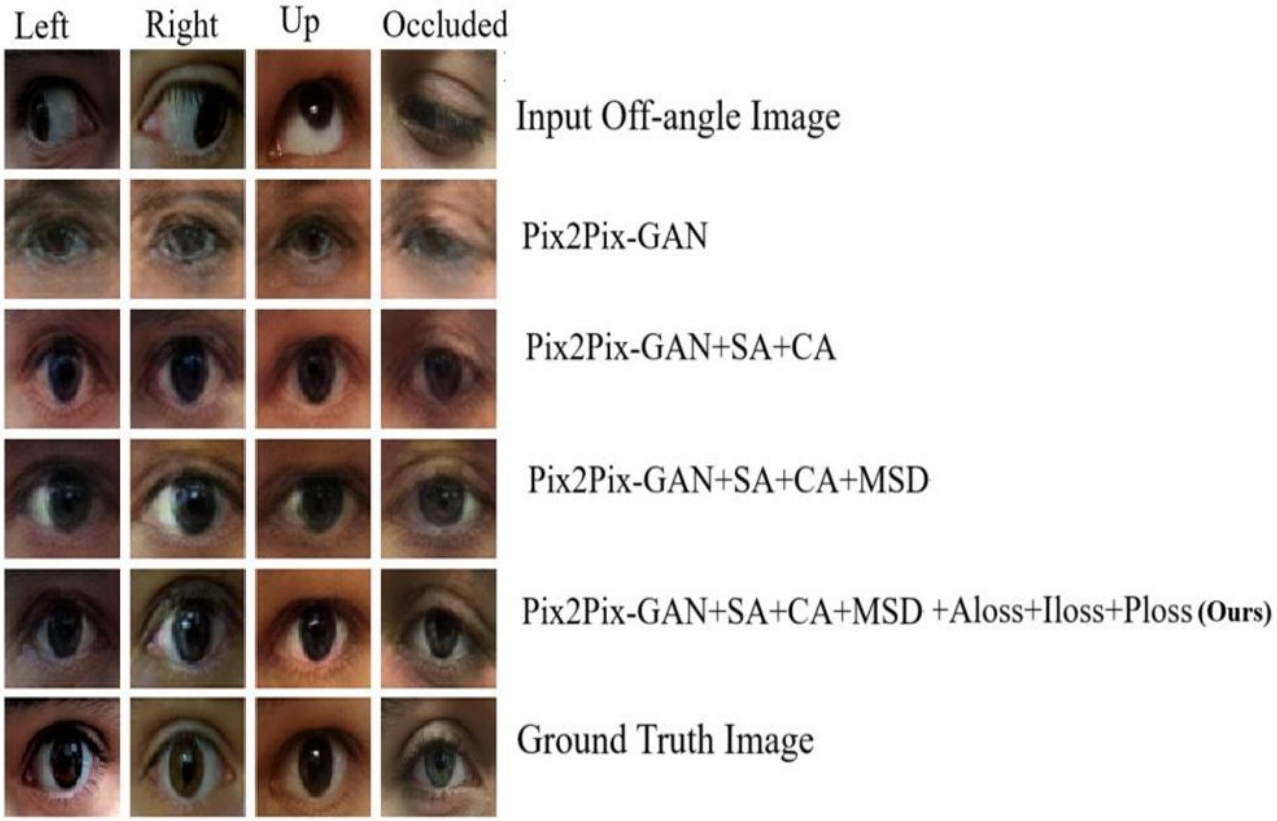


Figure 6.2: MOBIUS Converted off-angle iris image in different experiment classes.

6.4.3 Pix2Pix-GAN

In the first experiment, Pix2Pix-GAN baseline work was conducted using a mix of Mean Squared Error(MSE), Matrix Multiplication(MatMul) loss, and Structural Similarity Index (SSIM) loss functions to guide the training process.

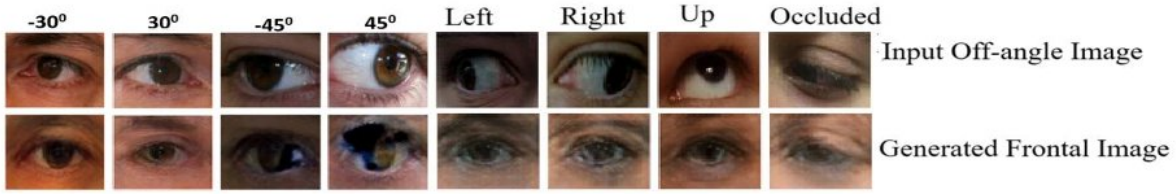


Figure 6.3: Result of Pix2Pix-GAN

In Figure 6.3, the first row indicates an input off-angle iris image, and the next row indicates the corresponding frontal iris image converted by the Pix2Pix-GAN baseline work experimental class. As shown in figure, the input off-angle iris images which are less angles are converted to its frontal view, but there is distortion on converted frontal image especially when the off-angle is large means greater than 30^0 and occluded iris image by eyelids, partially visible iris images and distorted iris images due to it is 3D structure of eye or large angle.

6.4.4 Pix2Pix-GAN+SA+CA

In the second experiment, we integrate attention mechanisms into the generator by utilizing a self-attention block in the decoder and a channel attention block in the encoder part of the generator network to address the challenge of preserving structural details.



Figure 6.4: Result of Pix2Pix-GAN+SA+CA

In Figure 6.4, the first row indicates an input off-angle iris image, and the next row indicates the corresponding frontal iris image converted by the Pix2Pix-GAN+SA+CA experimental class. As shown in figure 6.4, the images are converted to its frontal view, but there is some distortion on generated frontal image especially when the off-angle is large means greater than 30^0 or when iris image is occluded caused by eyelids, some part of iris is covered and distorted due to 3D structure of eye or large angle.

6.4.5 Pix2Pix-GAN+SA+CA+MSD

In the third experiment, we utilized a channel attention block in the encoder part and, self-attention block in the decoder part of the generator network. Instead of a PatchGAN discriminator of baseline work, we used a multi-scale discriminator in order to enhance the performance of the discriminator.



Figure 6.5: Result of Pix2Pix-GAN+SA+CA+MSD

Based on Figure 6.5, the model significantly reduces the distortion as compared to the previous model results, but it still does not convert an off-angle image to a frontal image with high quality and consistency.

6.4.6 Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}

In the fourth experiment, we utilized a channel attention block in the encoder part and, self-attention block in the decoder part of the generator network and we used a multi-scale discriminator instead of a PatchGAN discriminator with the combination of Adversarial loss, perceptual loss and identity Loss instead of Mean square loss function, Matrix Multiplication, and SSIM loss function of baseline work to generate sharper images that can capture the textural information of the original image better with high quality and consistency.



Figure 6.6: Result of Pix2Pix-GAN+SA+CA+MSD+A_{Loss}+I_{Loss}+P_{Loss}

In Figure 6.6, the first row indicates input off-angle iris images, and the next row indicates the corresponding frontal iris image converted by our proposed experiment.

Our model can convert off-angle iris images to their frontal view and reduce distortion on all off-angles and occluded iris images caused by eyelids, partially visible iris images, and distorted iris images due to 3D structure of eye or large angle.

6.5 Sample Training Log for HAM-MSD-GAN

We used graphs and charts (training log figures) to monitor the model's performance as it learned information. These figures tracked different aspects of the training process, such as the effectiveness of the model, minimized mistakes (loss), and gave valuable information about the model's capacity and training status for converting off-angle iris images.

The plot in figure 6.7, shows Generator Loss over Epochs. The curve's steep initial decline shows rapid improvement of generator's ability to produce realistic images that may fool the discriminator and meet the identity and perceptual objectives. The loss continues to decrease continuously, throughout the remaining epochs after this initial steep decline. The generator loss has significantly decreased by epoch 100, indicating that the generator is consistently enhancing the quality of the images it generates. The generator appears to have learned smoothly and efficiently during the training period, as indicated by the stable and effective decreasing trend.

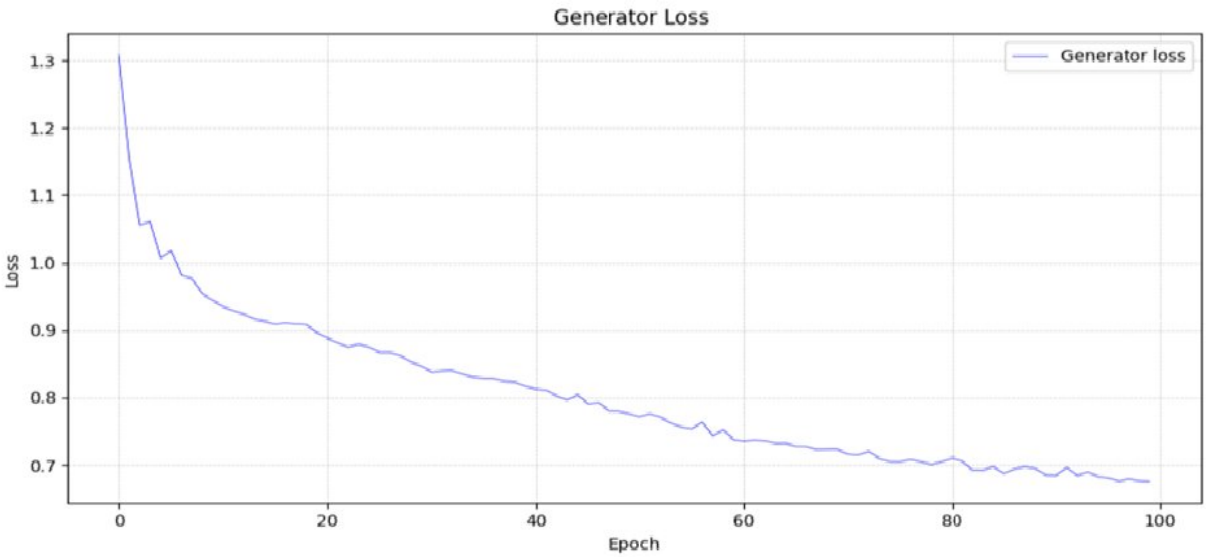


Figure 6.7: Generator loss during training

The graph in figure 6.8, illustrates discriminator loss over epochs during training phase. The discriminator rapidly learns to differentiate between real and fake images, as shown by the loss's sharp initial drop in the first few epochs. After this initial phase, for the remaining of the training,

the discriminator loss stabilizes and varies within a relatively small range, mostly between 0.52 and 0.54. This stable variation around 0.5 indicates that the discriminator is performing effectively, achieving a balance between not completely being fooled by the generator (which would indicate a dominant generator) and not accurately differentiating between real and fake images (which would indicate a struggling generator). The consistent performance of the discriminator indicates a healthy adversarial training process.

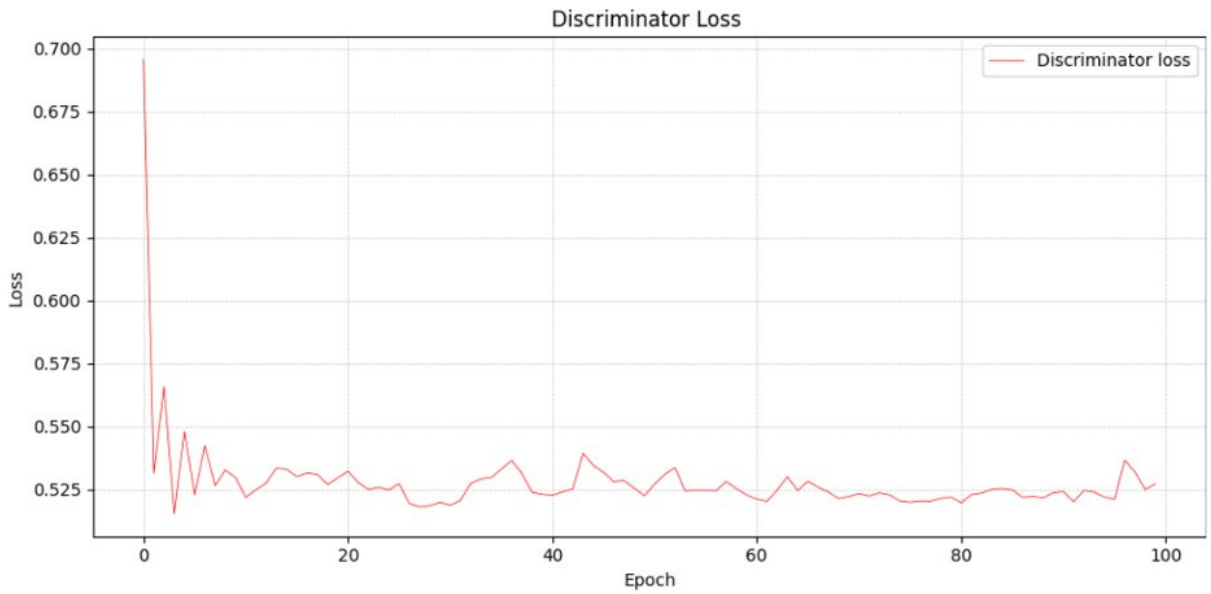


Figure 6.8: Discriminator loss during training

The graph in the figure 6.9 shows training and Validation Adversarial Loss during the training phase. Both the training and validation adversarial losses shows significant fluctuations and an overall decreasing trend, indicating the models (Generator and Discriminator) are learning actively. Beyond around 20-30 epochs, both loss curves reach a steady point between the value of 0.35 and 0.40, fluctuating within a quite small range. Adversarial loss during training and validation adversarial loss largely track each other, with minor deviations and random spikes (e.g., at epoch 55 and 80 for validation loss). The coherent action and similar magnitudes between the training and validation adversarial loss indicate that the GAN is keeping the generator and discriminator in balance, and is not significantly overfitting to the training data in terms of adversarial performance.

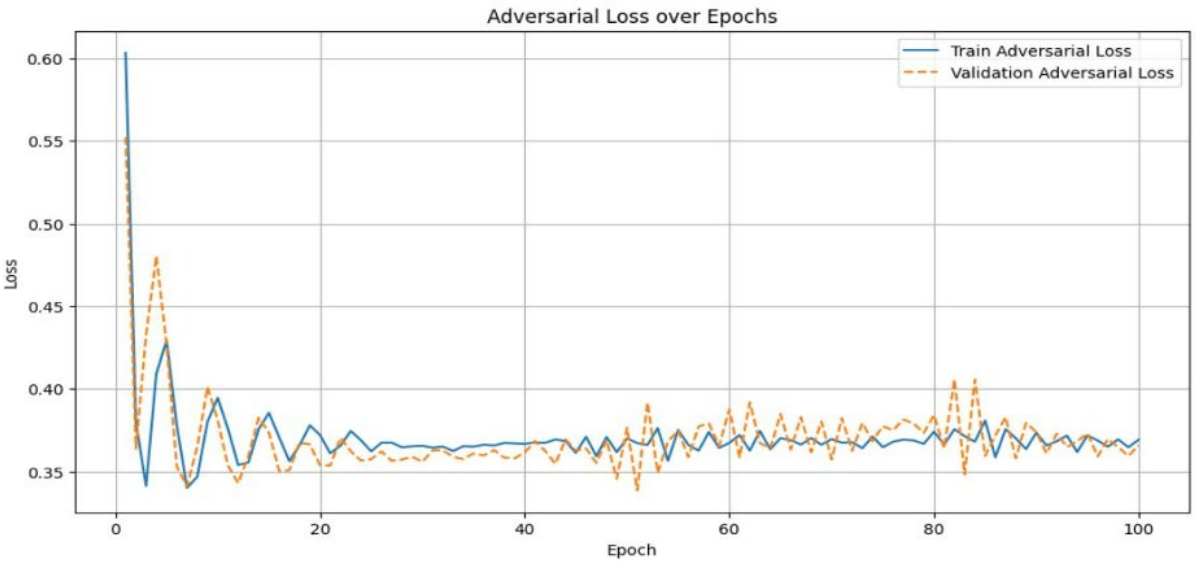


Figure 6.9: Adversarial loss during training

Figure 6.10, shows Identity Loss over Epochs during training phase. Identity losses for both training and validation show a sharp decrease initially, followed by a gradual decline and stabilization. Particularly in the latter epochs, the validation identity loss shows more oscillations and is consistently slightly higher than the training identity loss. This indicates that while the model is learning to preserve identity well on the training data, there's minor generalization gap on unseen validation data in terms of identity preservation, but the overall trend suggests to good learning.

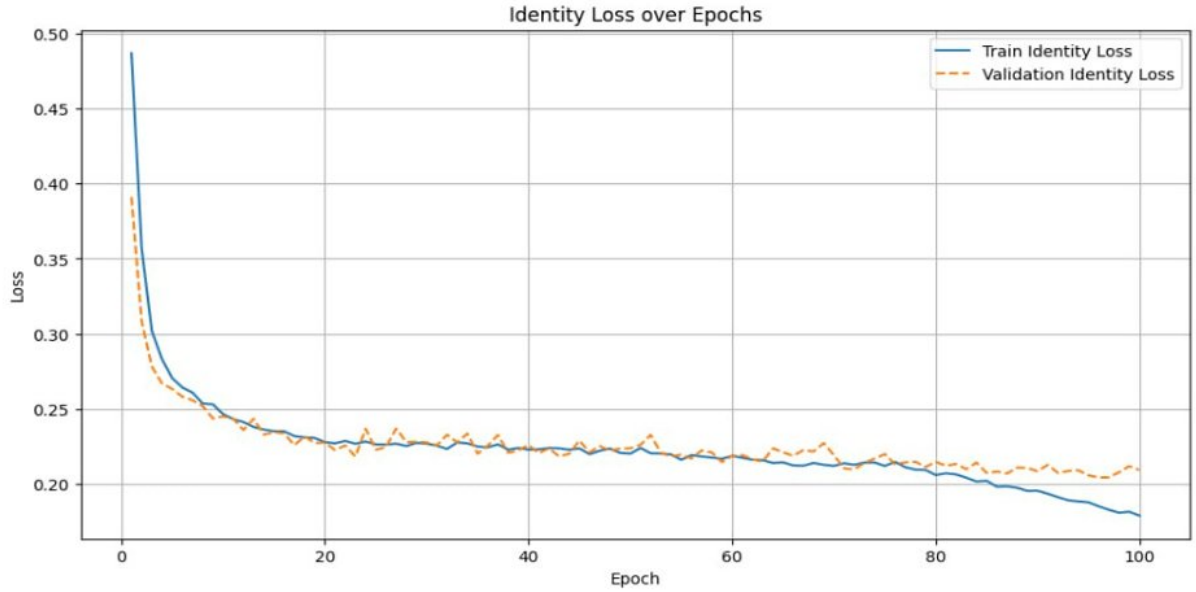


Figure 6.10: Identity loss during training

Figure 6.11 shows Perceptual Loss over Epochs during training phase. In the early epochs, training and validation perceptual losses both drop off quickly. After that, they continue to decline more slowly and show a similar general pattern. Compared to the training loss, the validation perceptual loss sometimes shows minor increases or more noticeable variations, especially after epoch 60, when it continuously stays slightly higher. This a slight and growing gap in performance on unseen validation data, indicating a minor degree of overfitting in terms of perceptual quality as training goes on, even while the model is learning to produce perceptually identical images on the training data, but the overall trend suggests to good learning.

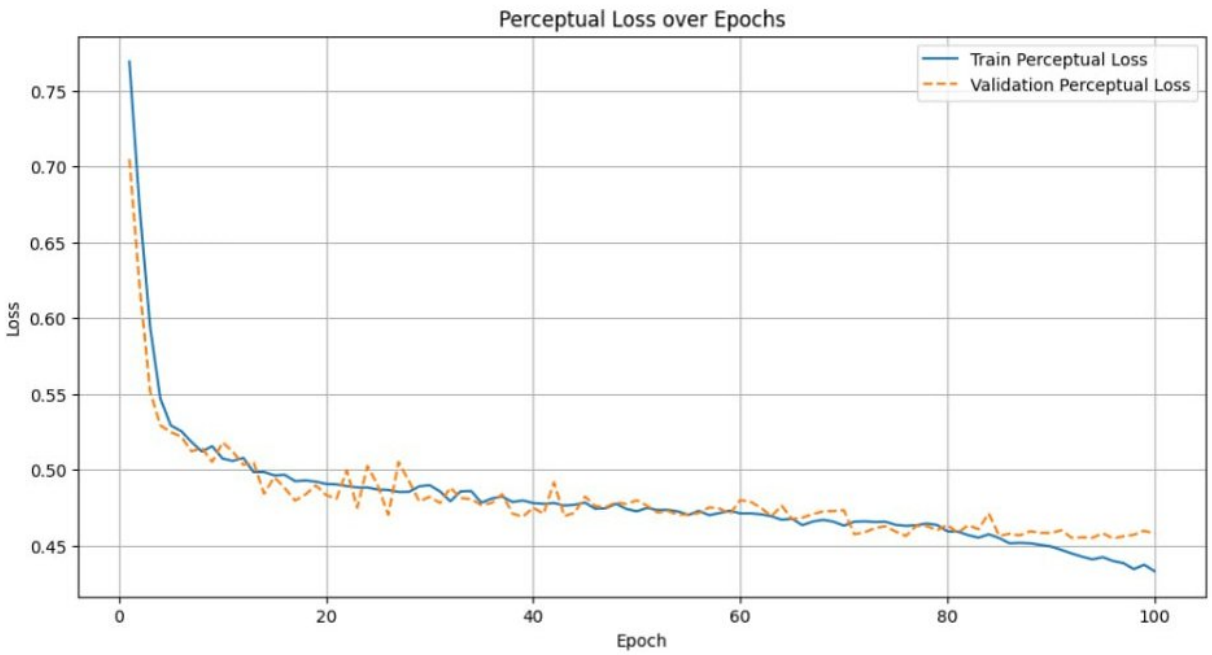


Figure 6.11: Perceptual loss during training

6.6 Results Discussion

In this study, it was shown that the proposed hybrid attention and multi-scale discriminator-based GAN (HAM-MSD-GAN) is successful in converting off-angle iris images to frontal view iris images. The proposed model incorporates hybrid attention mechanisms with self-attention and channel attention, along with a multi-scale discriminator design, to facilitate the feature synthesizing capacity of the generator and the perceptual sensitivity of the discriminator at various scales. Compared with previous methods, including Pix2Pix-GAN (Kota & Karakaya, 2023) and CNN(Jalilian et al., 2022), the proposed method achieves better results on various quantitative metrics. The hybrid attention mechanisms, inspired by the structures of self-attention (Zhang et

al., 2018) and channel attention (Hu et al., 2018), improve the model's capacity to capture fine-grained iris features. Meanwhile, the multi-scale discriminator gives a more detailed examination by assessing image realism at various resolutions. Consequently, the model effectively generates high-resolution frontal iris images that preserve identity-related features, thereby enhancing the overall quality and reliability of the conversion process.

6.7 Research Question and Answer Discussion

This study answered all the research questions: **1**. What are the impacts of incorporating an attention mechanism for transforming off-angle iris images? **2.** What are the impacts of using a multi-scale discriminator concerning the quality and transformation of off-angle iris images? **3.** What effect would Adversarial loss, identity loss and perceptual loss have on the quality and consistency of generated frontal iris images? The answers to each question are described in the following section.

Answer for Q1: This study incorporated a self-attention and channel attention mechanisms into a U-net-based generator part of the GAN network. Based on experimental results, there was improvement on the quality of converted off-angle iris images even when the angle is large and when the iris image is occluded, caused by eyelids, some part of the iris is covered and distorted. The PSNR, SSIM, and Hamming Distance results on the HAM-MSD-GAN experiment in table 6.1 for the MOBIUS dataset and tables 6.2, 6.3, and 6.4 for the UBIRIS, v2 dataset showed how well the self-attention and channel attention mechanism improved the quality of the converted off-angle iris images. So, this study concludes that incorporating self-attention and channel attention mechanisms in a U-net-based generator of GAN significantly improves the quality of converted iris images.

Answer for Q2: The efficiency of the GAN network in producing high-quality frontal-view iris images from off-angle inputs is demonstrated by the employment of a multi-scale discriminator at the discriminator portion of the network. Table 6.1 for the MOBIUS dataset and Tables 6.2, 6.3, and 6.4 for the UBIRIS.v2 dataset demonstrate enhanced performance in image quality metrics like PSNR, SSIM, and Hamming Distance in the HAM-MSD-GAN experimental results, confirming the benefit of the multi-scale discriminator in off-angle iris image conversion tasks. So, this study finds that by capturing features at multiple scales, a multi-scale discriminator, as opposed to a PatchGAN discriminator, at the discriminator part of the GAN Network of the

baseline work, greatly improves the quality of converted iris images and increases the GAN's capacity to convert realistic and detailed frontal iris images.

Answer for Q3: Adversarial, identity, and perceptual losses enhances quality and consistency of frontal iris image synthesis from off-angle inputs. As shown in Table 6.1 for the MOBIUS dataset and Tables 6.2, 6.3, and 6.4 for the UBIRIS.v2 dataset, these losses improve PSNR, SSIM, and Hamming Distance scores in the HAM-MSD-GAN experimental results. Adversarial loss enhances realism, identity loss maintains subject-specific details, and perceptual loss maintains structural integrity. Together, they lead to more precise and visually consistent frontal iris reconstructions.

CHAPTER SEVEN

7. CONCLUSION AND FUTURE WORKS

7.1 Conclusion

Computer vision tasks like iris recognition have significantly improved with recent advancements in Generative Adversarial Networks (GANs). In this work, we introduced a novel technique for converting off-angle iris images to a frontal view using a GAN model that combines a multi-scale based discriminator and a hybrid attention mechanism. Occlusion, texture distortion, and decreased feature visibility at large angles more than 30˚ make it extremely difficult to accurately create high-quality frontal iris images from off-angle iris images. Previous methods, such as Pix2Pix-GAN, have faced limitations when addressing off-angle iris images, often resulting in blurry or low-quality results at greater off-angles or when the iris images are covered by eyelids, partially visible, and distorted due to their three-dimensional structure or large angles. To address these issues, our proposed approach utilizes a U-Net-based generator equipped with hybrid attention mechanisms to enhance focus on iris-related areas across multiple scales. The conversion from off-angle to frontal iris image is adversarial trained through a multi-scale discriminator that assesses the shape of the iris in addition to the fine-scale texture of the iris, thus enhancing the realism and preservation in the converted iris outputs.

Enhancing the accuracy and realism of converting off-angle iris images to a frontal one is the primary aim of this study. To accomplish this, we introduced HAM-MSD-GAN, which incorporates channel attention in the encoder, self-attention in the decoder of a U-Net-based generator, and employs a multi-scale discriminator. This architectural framework facilitates the effective transformation of off-angle iris images into their frontal views even when they are significantly off-angle, occluded by eyelids, partially visible, or distorted due to their 3D nature or large angles. To improve the model's performance, we combine perceptual loss, which maintains texture and structural details, identity loss, which guarantees the distinctiveness and reliability of an individual iris pattern, and adversarial loss, to promote photorealistic results. The combination between attention mechanisms and these specialized loss functions enables the model to produce more precise and visually consistent transformed iris images. This methodology offers significant improvement in visual quality and identity preservation compared to baseline methods.

Several experiments were carried out to enable comparisons among them. The first experiment employed Pix2Pix-GAN, which was implemented a mix of Mean Squared Error (MSE), Matrix Multiplication (MatMul) loss, and Structural Similarity Index (SSIM) loss functions to steer the training process.

In the second experiment, Pix2Pix-GAN+ SA +CA was used, where channel attention is applied in the encoder and self-attention in the decoder section of the generator. This provides some improvements in the generator's performance compared to Pix2Pix-GAN.

The third experiment used a multi-scale discriminator in the discriminator part of the network. The performance of the discriminator is enhanced by applying a Multi-scale discriminator compared to the PatchGAN discriminator utilized in Pix2Pix-GAN.

In the fourth experiment, rather than relying on traditional loss functions like Mean Squared Error (MSE), Matrix Multiplication (MatMul) loss, or Structural Similarity Index (SSIM) loss, we combined perceptual loss, identity loss, and adversarial loss with our proposed HAM-MSD-GAN model to convert off-angle to frontal iris images. The primary reason for selecting this loss function is that pixel-level loss functions such as MSE and SSIM tend to produce smooth or blurry results and fail to reconstruct the intricate textures and structural variations essential for achieving high-fidelity iris synthesis. Perceptual loss addresses this issue by functioning at the feature level of a deep network, allowing the model to maintain a richer global context alongside detailed texture. Identity loss guarantees that the distinctive biometric attributes of the iris are retained after conversion, which is vital for identity-sensitive tasks like authentication and recognition. Adversarial loss, derived from the GAN framework, encourages the production of visually realistic and naturally appearing frontal iris images by compelling the distribution of the generated output to mimic real images. Together, these losses enable the model to produce high-quality, identity-preserving frontal iris images that are resilient to significant off-angle variations and more visually consistent than images generated using standard loss functions alone. HAM-MSD-GAN, along with the integration of perceptual loss, identity loss, and adversarial loss, proves to be an efficient approach for converting off-angle iris images with superior quality and finer features compared to all other experiments regarding PSNR, SSIM, and Hamming Distance Results.

Overall, to produce better results with high quality and consistency for off-angle to frontal iris image conversion, this proposed "Hybrid Attention and Multi-Scale based Generative Adversarial Network" offers additional enhancements to the generative adversarial network architecture. Furthermore, this study has improved the work's quality and performance at the technical level by integrating a number of procedures and constraints.

7.2 Future Works

This study, involves converting iris images that are specifically captured at angles between 30 and 45 degrees. Expanding the model's robustness to handle large off-angles, including those exceeding 45 degrees, and address the challenging problem of iris image conversion when occluded by glasses will be significant future directions. The current network depends on fully labelled training data and uses supervised learning. To increase the model's adaptability in scenarios with a lack of labeled data, utilization of unsupervised or self-supervised training process will be future work.

Special Acknowledgment

This research work was funded by grant number ASTU/SM-R/1122/25 from Adama Science and Technology University, **Adama, Ethiopia**.

REFERENCES

Abhyankar, A., & Schuckers, S. (2010). A novel biorthogonal wavelet network system for off-angle iris recognition. *Pattern Recognition*, *43*(3), 987–1007.

Alqahtani, H., Kavakli-Thorne, M., & Kumar, G. (2021). Applications of Generative Adversarial Networks (GANs): An Updated Review. *Archives of Computational Methods in Engineering*, *28*(2), 525–552. https://doi.org/10.1007/s11831-019-09388-y

Bi, X., Zhang, Y., & Goh, S. K. (2024). DSD-GAN: Double-Scale Discriminators GAN for Enhanced Chinese Brush Baimiao Painting Colorization. *2024 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN60899.2024.10651295

Chavarro, D., & Karakaya, M. (2023). Off-angle iris recognition using single and multiple deep learning frameworks. *Multimodal Image Exploitation and Learning 2023*, *12526*, 62–70.

Daugman, J. G. (2002). High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(11), 1148–1161.

Daugman, J. (2004). Recognising persons by their iris patterns. *Chinese Conference on Biometric Recognition*, 5–25.

Daugman, J. (2007). New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *37*(5), 1167–1175. https://doi.org/10.1109/TSMCB.2007.903540

Daugman, J. (2009). How iris recognition works. In *The essential guide to image processing* (pp. 715–739). Elsevier.

Diab, K., & Karakaya, M. (2022). Cnn-based gaze estimation for off-angle iris recognition. *SoutheastCon 2022*, 736–742.

Frigerio, E., & Marcon, M. (2014). Iris image correction method from unconstrained images. *Computational Modeling of Objects Presented in Images: Fundamentals, Methods and Applications*, 201–224.

Gangwar, A., & Joshi, A. (2016). DeepIrisNet: Deep iris representation with applications in iris

recognition and cross-sensor iris recognition. *2016 IEEE International Conference on Image Processing (ICIP)*, 2301–2305

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

Habib, M. A., Wadud, M. A. H., Pinky, L. Y., Talukder, M. H., Rahman, M. M., Mridha, M. F., Okuyama, Y., & Shin, J. (2023). GACnet-Text-to-Image Synthesis With Generative Models Using Attention Mechanisms With Contrastive Learning. *IEEE Access*, *12*, 9572–9585.

Henry, J., Natalie, T., & Madsen, D. (2021). Pix2pix gan for image-to-image translation. *Research Gate Publication*, 1–5.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Advances in Neural Information Processing Systems*, *30*.

Hofbauer, H., Jalilian, E., & Uhl, A. (2019). Exploiting superior CNN-based iris segmentation for better recognition accuracy. *Pattern Recognition Letters*, *120*, 17–23.

Huang, C.-M., Wijanto, E., & Cheng, H.-C. (2021). Applying a Pix2Pix generative adversarial network to a Fourier-domain optical coherence tomography system for artifact elimination. *IEEE Access*, *9*, 103311–103324.

Huang, Y., Ahmad, S., Han, L., Wang, S., Wu, Z., Lin, W., Li, G., Wang, L., & Yap, P.-T. (2022). *Longitudinal Prediction of Postnatal Brain Magnetic Resonance Images via a Metamorphic Generative Adversarial Network*. http://arxiv.org/abs/2208.04825

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

Jalilian, E., Karakaya, M., & Uhl, A. (2020). End-to-end Off-angle Iris Recognition Using CNN Based Iris Segmentation. *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1–7.

Jalilian, E., Karakaya, M., & Uhl, A. (2021). CNN-based off-angle iris segmentation and recognition. *IET Biometrics*, *10*(5), 518–535.

Jalilian, E., Wimmer, G., Uhl, A., & Karakaya, M. (2022). Deep learning based off-angle iris

recognition. *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4048–4052.

Jiang, W., Li, Q., Zhan, K., Fang, Y., & Shen, F. (2022). Hybrid attention network for image captioning. *Displays*, *73*, 102238.

Jalilian, E., Wimmer, G., Uhl, A., & Karakaya, M. (2022). Deep learning based off-angle iris recognition. *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4048–4052.

Jiang, J.-T., Wu, C.-P., Deng, H.-F., Lu, M.-Y., Wu, J., Zhang, H.-Y., Sun, W.-H., & Ji, M. (2004). Serum level of TSGF, CA242 and CA19-9 in pancreatic cancer. *World Journal of Gastroenterology: WJG*, *10*(11), 1675.

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. Computer Vision--ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, 694–711.

Karakaya, M. (2016). A study of how gaze angle affects the performance of iris recognition. *Pattern Recognition Letters*, *82*, 132–143.

Karakaya, M. (2018). Deep learning frameworks for off-angle iris recognition. *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 1–8.

Karnewar, A., & Wang, O. (2020). Msg-gan: Multi-scale gradients for generative adversarial networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7799–7808.

Kota, J. S., & Karakaya, M. (2023). Challenges in Off-angle to Frontal Iris Image Conversion using Pix2Pix Generative Adversarial Networks. *2023 IEEE International Joint Conference on Biometrics, IJCB 2023*, 1–8. https://doi.org/10.1109/IJCB57857.2023.10448976

Lee, M. B., Kang, J. K., Yoon, H. S., & Park, K. R. (2021). Enhanced iris recognition method by generative adversarial network-based image reconstruction. *IEEE Access*, *9*, 10120–10135.

Lee, H., Lee, S.-G., Park, J., & Shim, J. (2023). Improving Complex Scene Generation by Enhancing Multi-Scale Representations of GAN Discriminators. *IEEE Access*, *11*, 43067–43079.

Li, X., Wang, L., Sun, Z., & Tan, T. (2013). A feature-level solution to off-angle iris recognition. *2013 International Conference on Biometrics (ICB)*, 1–6.

Menotti, D., Chiachia, G., Pinto, A., Schwartz, W. R., Pedrini, H., Falcao, A. X., & Rocha, A. (2015). Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, *10*(4), 864–879.

Mirza, M., & Osindero, S. (n.d.). Conditional Generative Adversarial Nets.

Nguyen, K., Fookes, C., Jillela, R., Sridharan, S., & Ross, A. (2017). Long range iris recognition: A survey. *Pattern Recognition*, *72*, 123–143.

Pan, J., Wang, C., Jia, X., Shao, J., Sheng, L., Yan, J., & Wang, X. (2019). Video generation from single semantic label map. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3733–3742.

Padole, C. N., & Proenca, H. (2012). Periocular recognition: Analysis of performance degradation factors. 2012 5th IAPR International Conference on Biometrics (ICB), 439–445.

Proença, H., & Alexandre, L. A. (2007). Toward noncooperative iris recognition: A classification approach using multiple signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(4), 607–612.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241.

Schuckers, S. A. C., Schmid, N. A., Abhyankar, A., Dorairaj, V., Boyce, C. K., & Hornak, L. A. (2007). On techniques for angle compensation in nonideal iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *37*(5), 1176–1190.

Vitek, M., Das, A., Pourcenoux, Y., Missler, A., Paumier, C., Das, S., De Ghosh, I., Lucio, D. R., Zanlorensi Jr., L. A., Menotti, D., Boutros, F., Damer, N., Grebe, J. H., Kuijper, A., Hu, J., He, Y., Wang, C., Liu, H., Wang, Y., … Štruc, V. (2020). {SSBC} 2020: Sclera Segmentation Benchmarking Competition in the Mobile Environment. IEEE International Joint Conference on Biometrics (IJCB), 1–10. https://doi.org/10.1109/IJCB48548.2020.9304881

Wang, K., & Kumar, A. (2019). Toward more accurate iris recognition using dilated residual

features. *IEEE Transactions on Information Forensics and Security*, *14*(12), 3233–3245.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8798–8807.

Wildes, R. P. (1997). Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, *85*(9), 1348–1363. https://doi.org/10.1109/5.628669

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11211 LNCS*, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

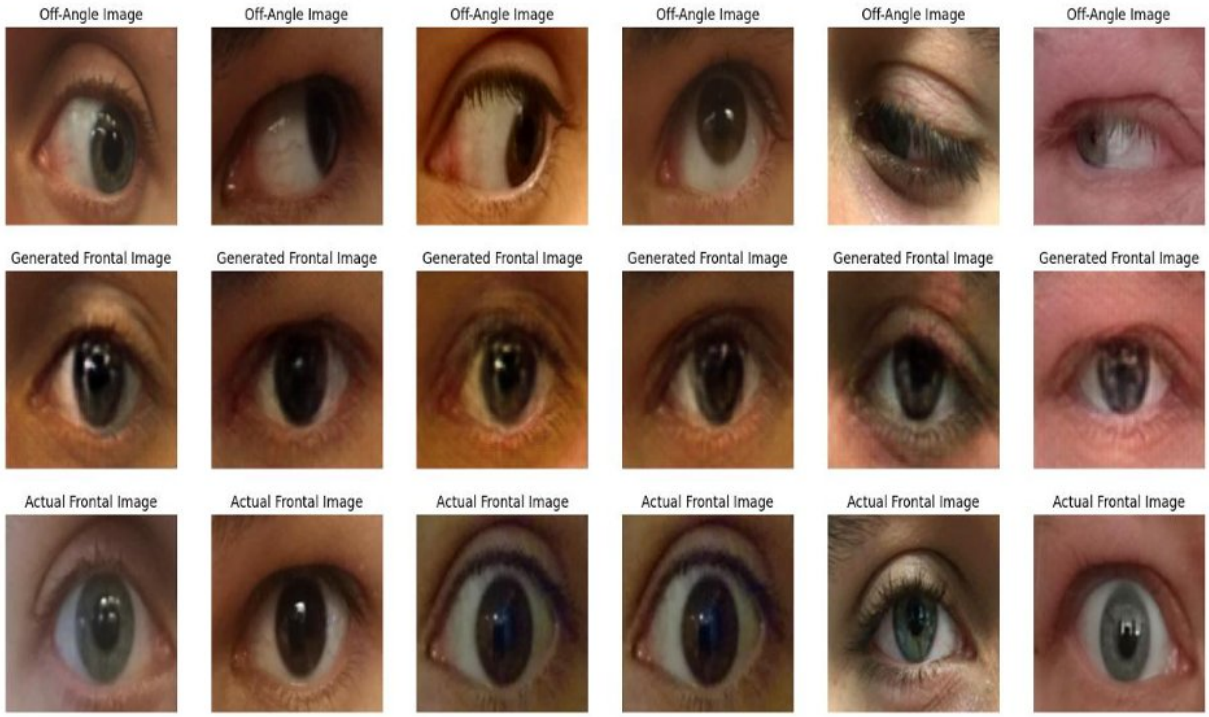
Yu, Y., Zhang, W., & Deng, Y. (2021). *Frechet Inception Distance (FID) for Evaluating GANs* (Issue 3).

Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). Self-attention generative adversarial networks. *International Conference on Machine Learning*, 7354–7363.

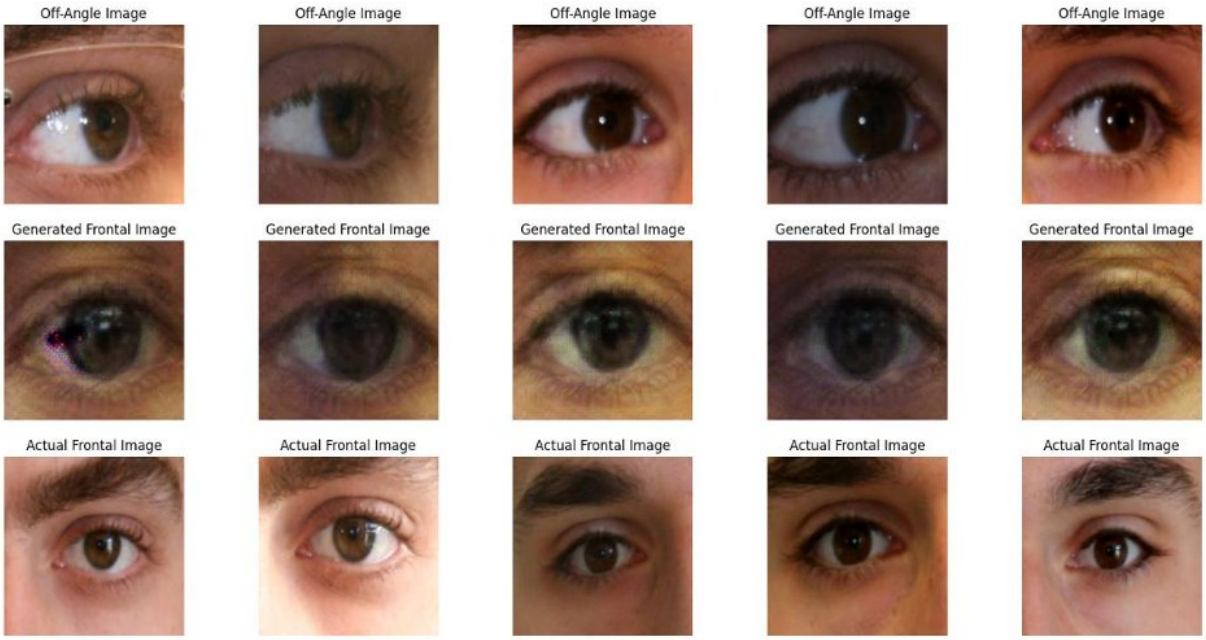
Zhang, Q., & Liu, Y. (2018). Improving brain computer interface performance by data augmentation with conditional deep convolutional generative adversarial networks. *ArXiv Preprint ArXiv:1806.07108*.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2223–2232.

Appendix A: MOBIUS Dataset Test Result



Appendix B: UBIRIS.v2 Dataset Test Result



Appendix C: Sample Code

Overall Generator Implementation:

```
class UNetGenerator(nn.Module):
    def __init__(self):
        super(UNetGenerator, self).__init__()

        # Encoder with Channel Attention
        self.enc1 = self.down_conv(3, 64)
        self.enc1_att = ChannelAttention(64)
        self.enc2 = self.down_conv(64, 128)
        self.enc2_att = ChannelAttention(128)
        self.enc3 = self.down_conv(128, 256)
        self.enc3_att = ChannelAttention(256)
        self.enc4 = self.down_conv(256, 512)
        self.enc4_att = ChannelAttention(512)

        # Bottleneck
        self.bottleneck = nn.Conv2d(512, 512, kernel_size=3, padding=1)

        # Decoder with Self-Attention
        self.dec4 = self.up_conv(512, 256)
        self.dec4_att = SelfAttention(256)
        self.dec3 = self.up_conv(256, 128)
        self.dec3_att = SelfAttention(128)
        self.dec2 = self.up_conv(128, 64)
        self.dec2_att = SelfAttention(64)
        self.dec1 = nn.Sequential(
            nn.ConvTranspose2d(64, 3, kernel_size=4, stride=2, padding=1),
            nn.Tanh())
    def down_conv(self, in_channels, out_channels):
        return nn.Sequential(
            nn.Conv2d(in_channels, out_channels, kernel_size=4, stride=2, padding=1),
            nn.ReLU(),
            nn.BatchNorm2d(out_channels)
        )

    def up_conv(self, in_channels, out_channels):
        return nn.Sequential(
            nn.ConvTranspose2d(in_channels, out_channels, kernel_size=4, stride=2, padding=1)
            nn.ReLU(),
            nn.BatchNorm2d(out_channels)
        )
    def forward(self, x):
    # Encode
        enc1 = self.enc1(x)
        enc2 = self.enc2(enc1)
        enc3 = self.enc3(enc2)
        enc4 = self.enc4(enc3)

    # Bottleneck
        bottleneck = self.bottleneck(enc4)

    # Decode
        dec4 = self.dec4(bottleneck + enc4)  # Skip connection
        dec3 = self.dec3(dec4 + enc3)  # Skip connection
        dec2 = self.dec2(dec3 + enc2)  # Skip connection
        output = self.dec1(dec2 + enc1)  # Skip connection
        return output
```

Appendix D: Model Training Complexity Analysis

Generator Training Complexity Analysis:

```
--- Model Training Complexity Analysis ---
Using device: cuda

--- Generator (UNetGenerator) ---
Trainable Parameters: 8,027,826
Non-Trainable Parameters: 0

--- Generator Architecture Table ---
```

Layer Name	Layer Type	Param #
enc1.0	Conv2d	3,136
enc1.1	ReLU	0
enc1.2	BatchNorm2d	128
enc1_att.fc1	Conv2d	260
enc1_att.fc2	Conv2d	320
enc1_att.relu	ReLU	0
enc2.0	Conv2d	131,200
enc2.1	ReLU	0
enc2.2	BatchNorm2d	256
enc2_att.fc1	Conv2d	1,032
enc2_att.fc2	Conv2d	1,152
enc2_att.relu	ReLU	0
enc3.0	Conv2d	524,544
enc3.1	ReLU	0
enc3.2	BatchNorm2d	512
enc3_att.fc1	Conv2d	4,112
enc3_att.fc2	Conv2d	4,352
enc3_att.relu	ReLU	0
enc4.0	Conv2d	2,097,664
enc4.1	ReLU	0
enc4.2	BatchNorm2d	1,024
enc4_att.fc1	Conv2d	16,416
enc4_att.fc2	Conv2d	16,896
enc4_att.relu	ReLU	0
bottleneck	Conv2d	2,359,808
dec4.0	ConvTranspose2d	2,097,408
dec4.1	ReLU	0
dec4.2	BatchNorm2d	512
dec4_att.query	Conv2d	8,224
dec4_att.key	Conv2d	8,224
dec4_att.value	Conv2d	65,792
dec3.0	ConvTranspose2d	524,416
dec3.1	ReLU	0
dec3.2	BatchNorm2d	256
dec3_att.query	Conv2d	2,064
dec3_att.key	Conv2d	2,064
dec3_att.value	Conv2d	16,512
dec2.0	ConvTranspose2d	131,136
dec2.1	ReLU	0
dec2.2	BatchNorm2d	128
dec2_att.query	Conv2d	520
dec2_att.key	Conv2d	520
dec2_att.value	Conv2d	4,160
dec1.0	ConvTranspose2d	3,075
dec1.1	Tanh	0
Total Trainable Parameters:		8,027,823

```
Estimated FLOPs: 4.59 GFLOPs
Average Inference Time (batch size 1, 100 runs): 4.054 ms
```

Discriminator Training Complexity Analysis:

```
--- Discriminator (MultiScaleDiscriminator) ---
Trainable Parameters: 8,294,229
Non-Trainable Parameters: 0

--- Discriminator Architecture Table ---
```

Layer Name	Layer Type	Param #
scale1.0	BatchNorm2d	6
scale1.1	Conv2d	3,136
scale1.2	LeakyReLU	0
scale1.3	Conv2d	131,200
scale1.4	LeakyReLU	0
scale1.5	Conv2d	524,544
scale1.6	LeakyReLU	0
scale1.7	Conv2d	2,097,664
scale1.8	LeakyReLU	0
scale1.9	Conv2d	8,193
scale2.0	BatchNorm2d	6
scale2.1	Conv2d	3,136
scale2.2	LeakyReLU	0
scale2.3	Conv2d	131,200
scale2.4	LeakyReLU	0
scale2.5	Conv2d	524,544
scale2.6	LeakyReLU	0
scale2.7	Conv2d	2,097,664
scale2.8	LeakyReLU	0
scale2.9	Conv2d	8,193
scale3.0	BatchNorm2d	6
scale3.1	Conv2d	3,136
scale3.2	LeakyReLU	0
scale3.3	Conv2d	131,200
scale3.4	LeakyReLU	0
scale3.5	Conv2d	524,544
scale3.6	LeakyReLU	0
scale3.7	Conv2d	2,097,664
scale3.8	LeakyReLU	0
scale3.9	Conv2d	8,193
Total Trainable Parameters:		8,294,229

```
Estimated FLOPs: 1.87 GFLOPs
Average Inference Time (batch size 1, 100 runs): 3.627 ms

--- Model Training Complexity Analysis Complete ---
```