T.R.
SAKARYA UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

AN IMPROVED TRANSFER LEARNING BASED SIAMESE NETWORK FOR FACE RECOGNATION

Master's Thesis

Dalhm Ghalıb Halboos AL-SHAMMARI

Computer Engineering Department

JULY 2024

T.R.
SAKARYA UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

AN IMPROVED TRANSFER LEARNING BASED SIAMESE NETWORK FOR FACE RECOGNATION

Master's Thesis

Dalhm Ghalıb Halboos AL-SHAMMARI

Computer Engineering Department

Thesis Advisor: Prof . Dr. Devrim AKGÜN

JULY 2024

The thesis work titled "……." prepared by …… was accepted by the following jury on …./…. /……. by unanimously/majority of votes as a MSc /PhD THESIS in Sakarya University Graduate School of Natural and Applied Sciences, ……….. department, …………. programe

Thesis Jury

Head of Jury :	**Title Name SURNAME** Sakarya University	....
Jury Member :	**Title Name SURNAME** Sakarya University	....
Jury Member :	**Title Name SURNAME** Sakarya University	....
Jury Member : **(If neccessary)**	**Title Name SURNAME** …… University	....
Jury Member : **(If neccessary)**	**Title Name SURNAME** …… University	....

STATEMENT OF COMPLIANCE WITH THE ETHICAL PRINCIPLES AND RULES

I declare that the thesis work titled " AN IMPROVED TRANSFER LEARNING BASED SIAMESE NETWORK FOR FACE RECOGNATION ", which I have prepared in accordance with Sakarya University Graduate School of Natural and Applied Sciences regulations and Higher Education Institutions Scientific Research and Publication Ethics Directive, belongs to me, is an original work, I have acted in accordance with the regulations and directives mentioned above at all stages of my study, I did not get the innovations and results contained in the thesis from anywhere else, I duly cited the references for the works I used in my thesis, I did not submit this thesis to another scientific committee for academic purposes and to obtain a title, in accordance with the articles 9/2 and 22/2 of the Sakarya University Graduate Education and Training Regulation published in the Official Gazette dated 20.04.2016, a report was received in accordance with the criteria determined by the graduate school using the plagiarism software program to which Sakarya University is a subscriber, I accept all kinds of legal responsibility that may arise in case of a situation contrary to this statement.

(23/7/2024)

Dalhm Ghalıb Halboos AL-SHAMMARI

ACKNOWLEDGMENTS

All thanks to Allah, and peace and blessings be upon Mohammed (pbuh), his messenger. My sincere appreciation goes to my main supervisor, Prof. Dr. Devrim Akgün for seeing me through this work. During this work, I engaged with many researchers and academics that supported and encouraged me to accomplish this work; their views and tips are highly appreciated.

Finally, I am grateful to all my family members for their support and dua'a. In particular, I would like to thank my husband for her patience, encouragement, support and understanding.

Dalhm Ghalıb Halboos AL-SHAMMARI

TABLE OF CONTENTS

ABBREVIATIONS

ID	: Identification Card
FR	: Face Recognition
SNN	: Siamese Neural Network
VGG&16	: Visual Geometry Group-16
DARPA	: Defense Advanced Research Projects Agency
MIT	: Massachusetts Institute of Technologe
FRGC	: Face Recognition Grand Challenge
EBGM	: Elastic Bunch Graph Matching
PDM	: Point Distrbution Model
CLI	: Command Line Interface
DFFS	: Distance from the Free Space
LFW	: Labeled Faces in the Wild
AUC	: Area Under the Curve
FSL	: Few Shot-Learning
CNN	: Convolutional Neural Network
LSTM	: Long Short-Term Memory
LSCOM	: Large Scale Ontology for Multimedia
LBP	: Local Binary Patterns
GPU	: Graphics Processing Unit

LIST OF TABLES

Page

LIST OF FIGURES

Page

AN IMPROVED TRANSFER LEARNING BASED SIAMESE NETWORK FOR FACE RECOGNATION

SUMMARY

In the digital era, interest in algorithms and theories for face recognition systems FR have been growing rapidly. Criminal identification, video surveillance, unmanned and autonomous vehicles, and building access control, are just examples of real applications that are gaining attraction among industries. FR is currently a highly difficult and complex subject in deep learning DL, neural networks, pattern analysis, and computer vision domains. Different learning groups, including controller environment and uncontrolled environment, have debated this issue. FR is a novel artificial intelligence application that DL has discovered recently. Earlier, many efforts have been dedicated to building accurate and adaptive FR models. However, recognizing faces in unconstrained environments poses a significant challenge due to various factors such as head pose, age, illumination, and facial expression variations, and others. Therefore , the aim of this study is to develop an efficient FR approach based on a Siamese neural network SNNs and Transfer Learning methods TL. The proposed approach employs SNNs with Visual Geometry Group 16 VGG-16 as a background for efficient FR especially in the case of individuals having similar facial features, and for the purpose of precisely identifying individuals in varying environments. The proposed approach consists of several phases, first data gathering, second data pre-processing, third model building, then comparing the proposed VGG-16 with three more convolutional networks (EfficientNet, ResNetB0, and ConvNeXt algorithms) to keep sure that the proposed approach is robust. For this purpose, labeled faces in the wild LFW dataset were used for SNN with VGG-16. After training the networks, the SNNs with VGG-16 exhibited low loss and a high accuracy in FR. Performance of the architectures were measured using K-Fold cross validation for 5 partition. According to results, EfficientNet, RestNet50 and ConvNext produced 77.75% accuracy, 95% and 93.75 % accuracy respectively.On the other hand, SNN with VGG-16 exhibited a low loss and produced the best accuracy in FR with 96.25%.

YÜZ TANIMA İÇİN GELİŞTİRİLMİŞ AKTARIM ÖĞRENME TABANLI SIAMESE AĞI

ÖZET

Yüz tanıma, bir kişinin yüz özelliklerini analiz ederek kimlik doğrulaması veya tanımlama işlemi gerçekleştiren bir biyometrik teknoloji yöntemidir. Bu teknoloji, göz, burun, ağız ve çene gibi yüzün belirli bölgelerinin geometrik yapısını kullanarak bireylerin kimliklerini belirler. Yüz tanıma sistemleri, bir kamera aracılığıyla yüz görüntüsünü alır, bu görüntüden belirli noktaları veya bölgeleri tespit eder ve analiz eder. Çıkarılan yüz özellikleri, bir veri tabanında saklanan diğer yüz özellikleriyle karşılaştırılarak kişinin kimliği belirlenir veya doğrulanır. Yüz tanıma teknolojisi, güvenlik sistemleri, akıllı telefonlar, sosyal medya platformları ve kamu güvenliği gibi pek çok alanda yaygın olarak kullanılmaktadır. Günümüzde yüz tanıma (FR), bilgisayarlı görme, desen analizi, sinir ağları ve derin öğrenme gibi alanlarda karmaşık ve zorlu bir problem olarak kabul edilmektedir. Bilgisayarlı görme, bilgisayarların dijital görüntülerden veya videolardan anlamlı bilgiler çıkararak çevrelerini algılamalarını sağlayan bir disiplindir. Bu alanın önde gelen uygulamalarından biri de yüz tanımadır. Yüz tanıma, görüntü işleme tekniklerini kullanarak yüzlerin tespit edilmesi, yüz özelliklerinin çıkarılması ve bu bilgilerin analiz edilmesi süreçlerini kapsar. Desen analizi, verilerdeki belirli kalıpları veya yapıları tanıma ve sınıflandırma işlemidir ve yüz tanıma sistemlerinde kişinin yüzündeki benzersiz özellikler, yani gözler arasındaki mesafe, burnun şekli vb. üzerinden kimlik doğrulama yapılır. Bu konu, denetimli ve denetimsiz öğrenme gibi farklı yapay zeka topluluklarında sıkça tartışılmaktadır. Son yıllarda, derin öğrenmenin yükselişiyle birlikte yüz tanıma alanında önemli ilerlemeler kaydedilmiştir. Özellikle kısıtsız ortamlarda yüz tanıma, kafa pozisyonu, yaş, aydınlatma ve yüz ifadelerindeki farklılıklar gibi çeşitli zorluklara rağmen daha doğru ve uyarlanabilir modellerin geliştirilmesiyle mümkün hale gelmiştir.

Bu nedenle, bu çalışma, yüksek doğruluk ve düşük yanlış alarmlara sahip, verimli ve uyarlanabilir bir FR sistemini geliştirmeyi amaçlamaktadır. Önerilen sistemde, etkili yüz tanıma ve bireylerin, özellikle de benzer yüz özelliklerine sahip bireylerin doğru şekilde tanınması için arka uç olarak Siyam sinir ağı SNN VGG-16'yı kullanılmıştır. VGG-16, büyük ve çeşitli bir veri kümesi olan ImageNet üzerinde önceden eğitilmiş bir evrişimli sinir ağı modelidir ve genellikle transfer öğrenimi için kullanılır. Transfer öğrenmede daha önce VGG-16'da öğrenilen özellik çıkarıcı katmanlar kullanılmakta ve bu katmanların üzerine yeni sınıflandırma katmanları eklenmektedir. Tipik olarak, VGG-16'nın evrişimli katmanları dondurulur (ağırlıkları değişmez) ve yeni eklenen sınıflandırma katmanları yalnızca belirli bir veri kümesi üzerinde eğitilir. Bu yaklaşım, daha az veri ve daha az eğitim süresiyle yüksek doğruluk elde edilmesini sağlar ve modelin yeni görevlere iyi bir şekilde genelleştirilmesine olanak tanır. SIAM sinir ağları, iki veya daha fazla veri örneği

arasındaki benzerliği veya korelasyonu bulmak için kullanılan özel bir sinir ağı türüdür.

Aynı ağırlıklara sahip iki paralel sinir ağından oluşur ve bu iki ağ, girdi çiftleri üzerinde çalışır ve her girdi için bir özellik vektörü elde edilir. Üretilen özellik vektörleri genellikle iki girdi arasındaki benzerliği ölçmek için Öklid mesafesi gibi bir benzerlik ölçüsü kullanılarak karşılaştırılır. İki girdi arasındaki benzerlik karar süreci için ikili çapraz entropi kullanılır. Bu yapı, girdilerin aynı dönüşüme uğramasını ve karşılaştırılmasını sağlar. Siam sinir ağları, özellikle yüz tanıma, imza doğrulama ve one-shot learning gibi benzerlik ölçümünün kritik olduğu alanlarda kullanılır. Örneğin, yüz tanıma uygulamalarında, iki yüz resminin aynı kişiye ait olup olmadığını belirlemek için kullanılır. Eğitim sürecinde, ağ, çeşitli yüz çiftleri (aynı kişiye ait veya farklı kişilere ait) üzerinde eğitilir ve her yüz resmi için bir özellik vektörü çıkarılır. İki yüz resmi karşılaştırılır ve aralarındaki mesafe hesaplanarak karar verilir.

Bu ağların en büyük avantajı, aynı ağırlıkları paylaşan paralel ağlar sayesinde verimli öğrenme sağlamasıdır. Ayrıca, sınırlı veri ile etkili sonuçlar elde edilebilir, bu da özellikle tek seferde öğrenme (one-shot learning) ve birkaç adımda öğrenme (few-shot learning) senaryolarında önemlidir. Yeni ve görülmemiş örnekleri tanıma yeteneği yüksek olan Siam sinir ağları, benzerlik ölçümünün kritik olduğu birçok uygulama için güçlü ve esnek bir araçtır.

Sunulan çalışmada yüz tanıma algoritmalarının gerçek dünya koşullarındaki performansını değerlendirmek için kritik bir kaynak olup, yüz tanıma alanındaki ilerlemeyi ölçmek için standart bir platform sağlayan LFW veri seti kullanılmıştır. LFW veri seti, yüz tanıma algoritmalarının performansını değerlendirmek için iki ana görev sağlar: yüz doğrulama, belirli iki yüz görüntüsünün aynı kişiye ait olup olmadığının belirlenmesi ve yüz tanıma bir yüz görüntüsünün kimliğinin belirlenmesi. Veri kümesi, etiketli yüz çiftlerini ve her yüzün ait olduğu kişinin kimliğini içeren meta verilerle birlikte gelir. Ayrıca veri seti “View 1” ve “View 2” olarak adlandırılan iki farklı değerlendirme protokolü içermektedir. Modeli eğitmek ve seçmek için "Görünüm 1" kullanılırken, son performansı değerlendirmek için "Görünüm 2" kullanılır. LFW veri seti yüz tanıma teknolojisinde önemli bir standart haline geldi. Araştırmacılar ve geliştiriciler bu veri setini algoritmalarının doğruluğunu ve sağlamlığını değerlendirmek için kullanıyor.

VGG-16 tabanlı bir SNN modeli, LFW veri seti üzerinde eğitildiğinde, genellikle yüz tanıma görevlerinde düşük kayıp ve yüksek doğruluk oranları sergilemiştir. LFW veri seti, yüz tanıma algoritmalarının gerçek dünya koşullarındaki performansını değerlendirmek için kritik bir kaynak olarak kabul edilir. Farklı bireylere ait, çeşitli koşullarda çekilmiş 13,000'den fazla yüz görüntüsü içeren bu veri seti, yüz tanıma teknolojilerinin zorlu ve değişken senaryolarda nasıl işlediğini test etmek için standart bir platform sağlar. Bu sayede, yüz tanıma alanındaki ilerlemeyi nesnel bir şekilde ölçmek mümkün hale gelir.

LFW veri kümesini SNN'ye giriş olarak kullanmak için ön işleme tabi tutulması gerekir. Ön işleme adımları görüntülerin yeniden boyutlandırılmasını içerir ve bu amaçla görüntüler 62 x 47 piksel boyutlarına yeniden boyutlandırılmıştır. Görüntülerin yeniden boyutlandırılması hesaplama karmaşıklığını azaltmak ve ağın yüzlerin temel özelliklerine odaklanmasını sağlar. LFW veri seti sınırlı sayıda örnek içerdiğinden, önerilen çalışma, veri seti görüntülerinin sayısını artırmak ve aşırı uyumu ortadan kaldırmak için LFW veri setini 5 sete bölerek Katmanlı K-Katlı çapraz doğrulama tekniği kullanmıştır. K-katlı çapraz doğrulama ile veri setinin farklı bölümlerini eğitim ve doğrulama için kullanarak, modelin farklı veri parçaları üzerinde nasıl performans gösterdiğini görmeyi sağlar. Böylece, modelin performansı hakkında daha güvenilir bir değerlendirme yapılabilir. Daha sonra önerilen yaklaşımın sağlam EfficientNet, ResNetB0 ve ConvNeXt algoritmaları olduğundan emin olmak için önerilen VGG-16'yı üç evrişimli ağla daha karşılaştırılmıştır. Sonuçlara göre EfficientNet, RestNet50 ve ConvNext sırasıyla %77.75, %95 ve %93.75 doğruluk değerleri üretmiştir. VGG-16 kullanılan SNN, düşük kayıp sergilemiş ve %96.25 ile yüz tanımada en iyi doğruluğu üretmiştir..

1. INTRODUCTION

1.1. Overview

Systems that use behavioural traits Like the signature, gait, walking style, speaking patterns, and facial dynamics—some of which are often referred to as As biometrics that are soft —as well as systems that utilise static physiological elements like the fingerprint recognition, iris, and palm prints —are occasionally used as biometrics in addition to the growing need for biometric systems that can be used for person identification and verification in emerging technologies (Adjabi et al., 2020). Face has been one of the key biometric qualities, with applications in a wide range of fields, including law enforcement, security, education, finance, etc. The human face contains valuable information about a person's identity, age (Waelen, 2023), gender (Menon et al., 2021), race (Bhatta et al., 2023), facial expressions that convey emotions and mental states (Mohsin et al., 2018). Table 1.1 expresses the areas of FR techniques.

Table 1.1. The areas of face recognition techniques.

Item	Application area	In-depth explanation
1-	Security, health	• user authentication • human robot interaction • information security • recognizing faces from surveillance videos • login to electronic devices • health related applications • smart cities, cars • robotic assistants
2-	Law enforcement	• border monitoring • tracking suspects passports • illegal event analysis • national id cards • immigration
3-	Entertainment, marketing, education,	• video analytics • photo management • online learning • student follow-up, user engagement • advertising campaigns

Human face patterns and behaviour analysis is a research area that combines engineering, psychology, and neuroscience (Taskiran et al., 2020). It is an interdisciplinary field that studies various aspects of the human face and how it relates to behaviour. Unlike other biometric traits, FR do not always require the person's cooperation and can be conducted in a discreet manner. This makes it highly suitable for surveillance applications. İn addition, FR can utilise both physical (static) features as well as dynamic features of the face, which makes it well-suited for behavioral biometrics (Hsieh et al., 2022).

Recognizing faces in unconstrained environments poses a significant challenge due to various factors such as head pose, age, illumination, and facial expression variations (Hsieh et al., 2022). Changes in appearance can also occur as a result of make-up, facial hair, or accessories such as glasses or scarves. one additional challenge in FR is the resemblance between individuals, such as relatives or twins (Koley et al., 2022). As a result, although humans can almost readily and effectively perceive faces, computers find it harder to do. To recognise faces in connection to contextual information, the human visual system supports sophisticated brain pathways for processing both static and dynamic characteristics of the faces (Duan & Zhang, 2020).

1.2. Problem Background

Various FR techniques have achieved significant success in controlled environments, and security functions (adjabi et al. 2020; gerig et al. 2018). Due to the nonstable nature of facial images and the fact that it influences how the same person appears in multiple face images in real-life situations, these techniques commonly fail. Therefore, it is necessary to modify these techniques or design new suitable techniques, systems or models. The issues associated with illumination, pose variation, ageing, low resolution, plastic surgery, expression, as well as the fact that face acquisition processes may go through a broad range of modifications, have made the majority of FR algorithms yet to achieve the ideal level of recognition accuracy (lim et al. 2023; oloyede, hancke, and myburgh 2020). Therefore, it is important for any proposed solution to focus on enhancing the efficacy of the recognition models while reducing the complexity of resource utilisation during the recognition process. Figure 1.1 illustrates the scenario that gives rise to the problem addressed in this research. Figure 1.1 depicts the challenges associated with FR and the limitations of current recognition solutions.

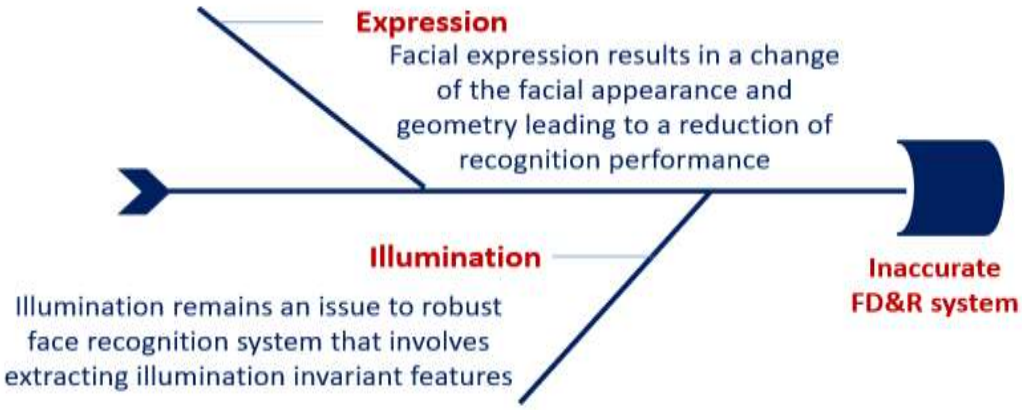

Figure 1.1. The challenges associated with FR and the limitations of current recognition solutions

The scenario depicted in Figure 1.1 emphasizes the key requirements that should be taken into account when designing FR solutions. These requirements include the need for image preprocessing techniques such as normalization, transformation, and augmentation .

Illumination refers to the variations in lighting that can cause a face image to appear differently. The extraction of illumination invariant features continues to be a challenge for robust FR systems. the use of illumination has significant effects on an image, resulting in changes to its location, shadow shape, and contrast gradients (Koley et al., 2022). Humans have the ability to recognise faces even when there are changes in lighting. However, FR systems struggle with this task and testing or training these systems is also challenging when lighting conditions vary. Therefore, it is necessary to employ image processing methods in order to Normalize face photos taken in various lighting scenarios.

The general face structure of people may vary significantly as they age, and their distinctive facial features may also alter. As a result, FR techniques commonly fail in these types of situations (Moustafa et al., 2020). In literature, it has been observed that there is a significant decrease in peormance in FR-based ageing when there is a large age difference between the target image and the query image (Li et al. 2016). Therefore, the query image must be represented using sparse representation classification to ensure the acquisition of the residuals for every class, with each pixel being determined in each class's residual as either occluded or not.

Humans are capable of displaying a variety of expressions on their faces at all times, unless they are frozen in a static position. These expressions are utilised for communicating diverse feelings and mental states to others. Facial emotion alters the geometry and appearance of the face, which decreases the accuracy of facial recognition (z. chen, feng, and zhang 2022). Therefore, it is necessary to apply customised pre-processing techniques that can only extract features specific for certain expressions-from the face image.

It has been observed that FR techniques are unable to identify individuals' faces accurately after they undergo plastic surgery (Adjabi et al., 2020). In this scenario, the facial image of a person is completely altered, resulting in a completely different individual. The alteration of the human face through cosmetic surgery changes the skin texture of the same person in-between photographs of the same person; this increases the difficult of face identification (Sabharwal & Gupta, 2022). Therefore, it is necessary to apply the local binary pattern, compared to the full facial image, in important areas of the image. The approach is based on a concept that only that local binary patterns (lbp) containing essential information, such as corners and edges, will be valuable for facial recognition following plastic surgery.

1.3. Problem Statement

Recognising facial images is a challenging task due to various factors such as illumination, ageing, facial expressions, and plastic surgery. This study trys to solve the challenges of face expression and illumination vactos. The ability to recognise face images using computational methods relies extensively on different image parameters, including quality, lighting, and non-stationary behaviour. Several techniques have been proposed for extracting illumination-invariant features. However, these techniques haven't yet achieved the optimum solutions. Additionally, these solutions have high computational complexity.

The main questions about how to increase the accuracy of the FR systems are as follows:

1. What are the challenges of the existing solutions regarding expressions and illumination-based face recognition systems?
2. How could the recognition of expressions and illumination-based image faces be improved?
3. How to evaluate the results that the proposed system has achieved?

1.4. Research Aim

The purpose of this study is to maintain computational complexity while designing and developing adaptive, high-accuracy expressions and illumination-based face recognition systems.

1.5. Research Objectives

The goals of this study are to investigate the main and related questions follows:

1. To investigate the state-of-the-arts problems of the FR systems and the techniques used for solving the illumination and expression issues.
2. To propose SNN with VGG-16 as backbone techniques that can improve the accuracy of the proposed expressions and illumination-based face recognition system.
3. To evaluate the accuracy of objective 2 and compare the achieved results with efficient loss, ResNet50 and ConvNext networks and with related state-of-the-art achievements.

1.6. Research Significance

This research is significantly important from both theoretical and practical perspectives for the domain of the FR system. The following are the motivations that inspired this investigation:

1. This system doesn’t require physical contact like the one necessary for fingerprint recognition scanners. The no-physical contact feature provides more sanitary conditions and protects the health of the users.
2. The research findings are expected to drive the advancement of the proposed FR system across various fields and applications. These include security and law

enforcement, education, finance, marketing, entertainment, and human-computer interaction.

1.7. Research Scope

The imitations of the proposed study are listed as following:

1. LFW dataset has been used with SNN with VGG-16 as a backbone algorithm to recognize the face from the dataset.
2. Improve the recognition of the proposed system using several pre-processing procedures like image normalization, image resize and image transformation process.
3. The evaluation process is crucial for the proposed system as it relies on standard statistical indices such as loss function, and accuracy.

1.8. Research Contribution

The contributions of the proposed study can be explained as follows:

1. Developing a new recognition system for expression and illumination-based face images through training SNN with VGG-16 as a backbone algorithm with the LFW dataset.
2. Decreasing the features of face image dimensions reduces computational complexity and ensures that the SNNs algorithms focus on the essential features of the faces rather than colour information.
3. Comparing the proposed VGG-16 with EfficientNet loss, ResNetB50 and ConvNext algorithms to keep sure that the proposed FR model is robust.

1.9. Research Outline

The proposed research contains six chapters organized as in Figure 1.2.

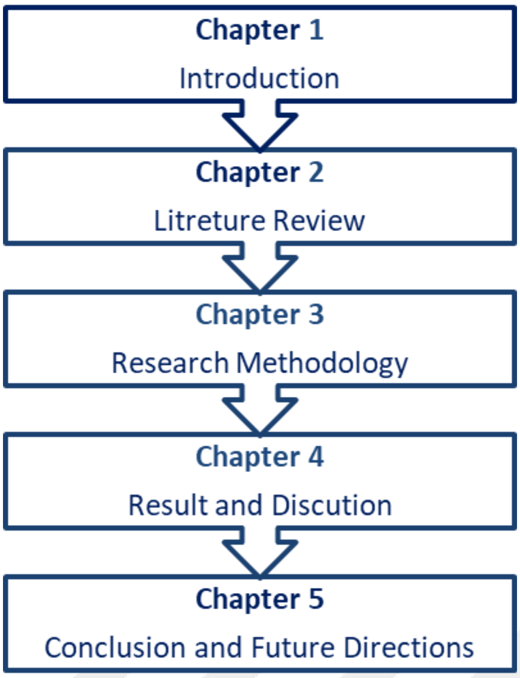

Figure 1.2. Thesis organizations

1. Chapter one: the introduction of this thesis offers a comprehensive description of the field of work and provides a summary of the details that relate to the developed idea. The research questions and problems were addressed by focusing on the state-of-the-arts.
2. Chapter two: this section offers an in-depth examination of previous studies that have investigated the objectives and scope of this thesis. It also explains the significance of FR systems and the different types they come in. this section also includes a list of the historical developments of the FR system. Additionally, the different types of classification for FR images were also discussed. The techniques used to address the current issues are identified and thoroughly examined. The discussion also included the limitations of the FR systems. The summary for this chapter has finally been provided.
3. Chapter three: this section of the thesis discusses the research methodology and conceptual framework of the proposed scheme. It includes an overview of the research concept, a problem description, the function of the proposed scheme, and the instrumentation and peormance measurement methods used. The purpose of this chapter is to propose a research solution for the problems identified in the recognition of face image features.
4. Chapter four: This chapter provides a detailed analysis of the simulation environment, including the results and analysis for each scenario used to

demonstrate the effectiveness of the proposed scheme. The outcomes of the simulation environment for the different scenarios were based on SNN with VGG-16 as a backbone, Moreovewr, EfficientNet , ResNetB0 and ConvNet have beem compared with the VGG-16 to keep sure that the proposed model is robust. and various tools were utilised. The possibility of improvement is achieved by comparing results with the current solutions to the exact same problem and demonstrating how the proposed solution is effective in reducing the loss level.

5. Chapter five: this section provides a summary of the research work, including the research achievements, contributions, and a brief discussion on suggestions for future works.

2. LITERATURE REVIEW

2.1. Introduction

A general overview of this thesis was introduced in chapter 1. This chapter reviews existing works related to the topic of this study. It starts with an overview of FR systems, history and application. The classification approaches of FR systems are then mentioned in this study. The main techniques that have been utilised for feature detection and classification have also been reviewed in this chapter. This chapter also details the datasets used in this thesis. Moreover, the evaluation metrics that have been used to evaluate the proposed systems are also listed in this chapter as models. This chapter has also highlighted the state-of-the-art in FR systems. The limitations of existing approaches are then mentioned. Finally, his chapter ends with a summary. Figure 2.1 depicts the flow of this chapter.

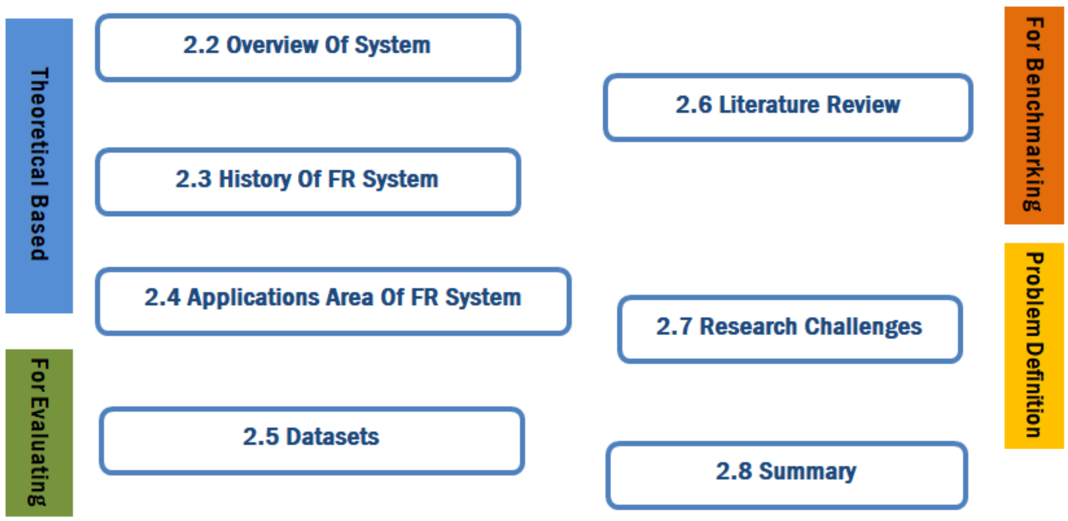

Figure 2.1. The flow of this chapter, where the numbers indicate the related section and subsection.

2.2. Overview of FR Systems

One might approach FR systems as an identification problem or as a verification problem. Face identification is the process of matching a given face with other faces in a database. It is often referred to as the 1:n matching problem. A determination is made by means of a comparison between the unidentified face and every face present in the database of recognised identities. The task is classified as closed-set if the individual is known to exist in the database. Otherwise, the term "open-set" is used. The 1:1 matching problem is the term used to describe face verification often. The identification of the query face is ascertained through comparison with the facial data of the identity that is asserted within the database. It is either confirmed or rejected based on this comparison (Bhandari et al., 2023).

This section presents a comprehensive overview of FR systems found in the literature. The main focus will be on the history, image databases utilised, as well as the applications employed for FR systems.

2.3. History of FR system

This section provides a summary of the key historical stages that have played a significant role in the development of FR technology.

Year of 1964: Bledsoe et al. conducted a study on computer programming for FR. The proposed method involves a semi-automatic approach, wherein operators are requested to input twenty computer measures. These measures may include dimensions such as the size of the mouth or the eyes.

Year of 1977: the model was improved via adding 21 additional markers (e.g., hair color, lip width).

Year of 1988: artificial intelligence was introduced as a means to improve upon existing theoretical tools that had demonstrated numerous weaknesses. Linear algebra, a branch of mathematics, has been employed to analyse images in a novel manner, enabling the discovery of techniques to simplify and manipulate them without relying on human annotations.

Year of 1991 : eigenfaces, the first effective application of FR technology was presented by Alex Pentland and Matthew Turk of the Massachusetts Institute of Technology (MIT). This technology utilises the statistical principal component analysis (pca) method (Adjabi et al., 2020).

Year of 1998: To promote progress in this field, The FR technology was developed by the Defense Advanced Research Projects Agency (DARPA). programme. This programme offered a substantial and demanding database consisting of 2400 images for 850 individuals, with the aim of encouraging collaboration between industry and academia.

Year of 2005: the FR grand challenge (FRGC) competition was created with the goal of promoting and advancing FR technology that can be used to enhance existing FR initiatives.

In 2011, there was a significant acceleration in technology thanks to the emergence of deep learning. deep learning is a machine learning technique that relies on artificial neural networks (Phillips et al., 2005). The computer selects the points that will be compared. It learns more effectively when it has a larger supply of images.

Year 2014: Facebook developed an internal algorithm called deep face, which enables the platform to accurately identify and recognise faces ("Deepface: Closing the Gap to Human-Level Peormance in Face Verification," 2014). The social network claims that a performance is achieved with its approach. level close to 97% of the human eye.

The advancement of facial recognition technology today has led to increased investments in various applications, including commercial, industrial, legal, and governmental sectors. Figure 2.2 explains the primary stages in the history of FR.

Figure 2.2. the primary stages in the history of FR (Adjabi et al., 2020).

2.4. Applications area of the FR system

The application of FR technology has significantly increased in various industries due to its ability to improve security, convenience, and personalization.

1. Security and surveillance

FR technology has found significant applications in the domains of surveillance and security (regin, rajest, and shynu 2023). FR technology is used by law enforcement organisations throughout the world (Qinjun et al., n.d.) To identify people in surveillance films, assisting in the investigation of crimes and enhancing public safety. The technology is also commonly used in access control systems, which allow for safe admittance into buildings or rooms by confirming people's identities.

2. Smartphone authentication

FR has grown into a standard feature in device authentication with the introduction of smartphones equipped with cutting-edge cameras and processing power (nachmani et al. 2022). Apple's face id, for example, utilises an advanced 3d FR system to securely unlock iphones, offering an easy alternative to passwords or pins.

3. Social media

Social media platforms such as Facebook utilise FR software to automatically tag individuals in photos. These platforms may improve user experience and connection among users by identifying and memorising the face data of users and suggesting tags for persons in recently submitted photographs (Awan et al., 2022).

4. Healthcare

FR technology is currently being studied for various applications in the healthcare industry. FR is utilised by certain systems to ensure the identity and security of patients. research is currently underway to identify diseases or medical conditions that can potentially impact FRS (Bisogni et al., 2022).

5. Retail

FR technology is utilised in the retail industry to improve both customer experience and security measures. Certain stores utilise FR technology to identify vip customers upon their arrival. This enables the staff to offer personalised service tailored to their specific needs and preferences. FR technology is also utilised for the purpose of detecting and deterring shoplifting, thereby making a valuable contribution to loss prevention efforts (Zhong et al., 2021).

These applications only scratch the suace of the potential uses of FR technology. More industries, including banking, tourism, entertainment, and others, are expected to use technology as it develops. While FR technology offers impressive benefits and applications, it also poses significant challenges that must be addressed as the technology advances. These challenges include concerns about accuracy, as well as issues related to privacy and ethics. It is crucial to carefully manage these challenges to ensure responsible use of the technology.

2.5. Dataset

In this thesis, the LFW dataset was utilised to implement the proposed methodology and achieve the results.

2.5.1. Labeled faces in the wild LFW dataset.

The LFW dataset (jalal and tariq 2017) comprises 13,323 web photos featuring 5,749 celebrities. These photos are further categorised into 6,000 face pairs across 10 different splits. Peormance is evaluated using three different protocols based on the mean recognition accuracy as follows: a) the restricted protocol – this approach permits using both same and not-same labels during the phase of training; b) the unrestricted protocol - allows access to additional training pairs during the training phase; c) an unsupervised setting – requires no training on LFW images.

2.6. Literature Review

Several studies have investigated systems based on SNNs and deep learning approaches. This section provides a list of state-of-the-art studies that have been reviewed in relation to the domain.

This paper focuses on the study of (Heidari and Fouladi-Ghaleh 2020) employed a siamese network architecture with two similar CNNs and transfer learning from the VGG-16 model to do facial recognition. They extracted features using a VGG network model that had previously been trained on the ImageNet dataset.To determine the degree of similarity between images, the Euclidean distance is used. The researchers employed network training with a contrastive loss function to lower similarity between pairs of photographs of the same person while boosting similarity between pairs of images of different people.The proposed model outperforms similar algorithms trained on short datasets in terms of accuracy (95.62%).

In a study presented by (Nandy & Mondal, 2019), the authors suggest fine-tuning pretrained parallel SqueezeNet networks and employing a similarity metric to integrate results and forecast kinship relationships. The cosine similarity metric surpasses L1 and L2 metrics in terms of accuracy due to its efficient and straightforward learning of the goal function. The L1 and L2 norms restrict data associations to linear and quadratic patterns, which may be inaccurate. However, cosine similarity describes relationships in converted space. SqueezeNet's squeeze and excitation blocks represent channel-wise relationships, allowing for precise feature response calibration. The SqueezeNet architecture used in the Siamese Network branches aids in understanding distinguishing characteristics from raw facial data.

In a study presented by (Zeng et al., 2022) , the authors offer a unique face forgery detection framework based on the modified Siamese network, which collects and aligns the features of the image's face area and backdrop and then uses their similarity to assess the image's validity. This framework not only provides excellent robustness and generalization performance, but it also fully utilizes the image background's feature information. We examine our method on a variety of datasets, demonstrating its practical effectiveness, particularly its ability to produce excellent results on low-quality datasets.

(Pang et al., 2021),the used SNNs as the research framework, with the addition of SE-Block and TAM. SE-CNN could use spatial feature information and channel correlation to change the retrieved feature weights according to contributions, similarly to a channel attention method. TAM may update the objective state by changing the weights of samples from the current and historical frames. The experimental results indicate that the suggested technique is robust in the application of target tracking, meeting real-time tracking requirements while successfully addressing the occlusion problem. However, there is still a problem with deviation because the target's pace is too rapid in some video sequences.

In a study presented by (Song & Ji, 2022) the authors provide a novel architecture that combines frequency feature sensing, LBP, and Siamese neural networks. By altering the parameters, the network can converge quicker and require less iterations to get the same recognition rate.

In a study presented by (Hayale et al., 2019) provides a novel strategy for FER is performed utilizing deep Siamese neural networks and supervised learning.The loss function. This approach creates a robust FER system by dynamically modulating verification signals over identification/classification signals. The recognition signal optimizes the distance across features between classes, whereas the verification signal minimizes the gap between features within the same class. Researchers assessed the approach using the AffectNet dataset and found encouraging results compared to previous deep learning models.

In a study presented by (Z. Zhang & Peng, 2019) presents the construction of deep and broad network topologies for Siamese trackers. This is inspired by the reality that simply replacing backbones with existing strong networks does not result in improvements. The study found that significant reasons include receptive field size, network padding, and stride. Experiments on five benchmarks show the efficacy of the suggested designs, resulting in competitive performance across five datasets.

2.7. Research Challenges

FR is considered to be a complex biometric system within the field of pattern recognition. This complexity arises from the challenges posed by variations in the appearance of facial images (jin and tan 2017). The peormance of FR techniques in controlled environments has been significant, however, they have been ineffective due to the unstable structure of face images. This is especially true in real-life scenarios, where images of the same person can appear different (y. zhang et al. 2016). To date, most FR algorithms have not achieved optimal levels of recognition accuracy due to various issues. This is due to the wide range of variations associated with the image acquisition process (lim et al. 2023). The FR system faces several challenges, including low resolution, expression fluctuation, lighting, plastic surgery, occlusion, and ageing (Mohsin et al., 2018). This section provides a more detailed description of the various issues associated with FR systems in unconstrained environments, also known as FR in the wild. A comprehensive presentation is provided, highlighting recent approaches to each issue.

1. Illumination-based FR challenge

Illumination refers to the variations in lighting that can cause a face image to appear differently. These changes in lighting can significantly impact the way a face is perceived. The change in lighting, as depicted in Figure 2.3, appears to be more significant than the variations among individuals. This can result in FR systems failing to correctly classify faces when comparing images. The problem of illumination persists in robust FR systems, which require the extraction of illumination invariant features. the use of illumination has a significant impact on an image, resulting in changes to the location, shape of shadows, and the reversal of contrast gradients (zhuang et al. 2013). Humans have the ability to recognise face images under different lighting conditions, which is considered normal. However, FR systems still struggle with this task. Previous research studies have shown that different image representations—such as edge maps, filtered images with gabor filters, and grey-scale images—have shown that altering illumination conditions has a larger impact on the image compared to changing the identity of the face (ma, song, and qian 2014)..

Figure 2.3. Changes in illumination adapted using the YALE face dataset (mohsin et al. 2018)

In addition, it is worth noting that testing or training can be affected by different lighting conditions. Scholars have paid more attention to the failure of FR systems in an illumination scenario over the last decade. As a result, various algorithms have been developed to address this issue. There are various approaches to handling the issue of illumination in proposed algorithms. Firstly, facial images with shades of lighting variations are first normalized using image processing techniques (Sinha & Barde, 2022). Gamma intensity correlation, logarithmic transformation, and histogram equalisation have been proven to effectively enhance images in various lighting conditions.

Additionally, researchers have employed the 3d face model approach to address the illumination problem (k. wang et al. 2017). Images of the front of faces taken in various lighting settings show an illumination cone that forms inside the subspace. Estimating these values in a low dimensional subspace using a generative model on the training set of data is comparatively easy. But employing a 3D face model necessitates a greater quantity of training examples. It might also be essential to identify the light's source, which is a method that isn't ideal for real-world circumstances. Using the third method, features are extracted from the areas that are illuminated and then sent to the recognition step. Using these methods, numerous algorithms have been presented up to this point. Nonetheless, the goal of continuing research needs to be done to reach its full potential. (mohsin et al. 2018).

2. Pose variation-based FR challenge

The presence of pose variation in the input images, as shown in Figure 2.4, can cause the poor peormance of a FR system. Face images are mostly captured from a frontal view in real settings, such as passport control. However, the peormance of FR system may be inadequate in uncontrolled situations when faces may appear from various perspectives as a result of rotation (zhao et al. 2003). Pose variation refers to the ability to recognise face images that are presented in different poses. While this may seem simple, it is a tedious task for computers in applications such as surveillance (Drira et al., 2013). Due to the rising frequency of terrorist attacks worldwide, the majority of airports have implemented surveillance camera systems. During the scanning procedure, the faces of terrorists are photographed and archived for further comparison with the faces of actual travellers. The terrorist's photos appear to have been taken at an earlier time and in different stances, which facilitates recognition of

the images using basic FR algorithms. In real-life scenarios where the dataset does not include face images with different poses, the system's peormance will be suboptimal. This is because the system will not be able to accurately match different poses of the same individual, resulting in false non-matches (k. wang et al. 2017). When a FR system lacks pose tolerance, it ceases to be passive and non-intrusive. Therefore, it is crucial to have FR systems that can peorm optimally even when faced with different poses that may differ from the face image contained in the dataset. Different FR techniques most face the problem of obtaining distinct features that are not affected by changes in pose in situations involving pose variation..

Figure 2.4. Changes in pose adapted using the orl dataset

3. Expression-based FR challenge

Human beings have the inherent ability to display a wide range of facial expressions, unless the face image remains in a static mode (goyal et al. 2018). These expressions in Figure 2.5 represent various emotions and mental states experienced by an individual. In addition, the issue of FR is not solely regarded as a matter of identity or verification. It is also utilised in medical applications, where specific facial expressions can be associated with certain ailments (tsai and chang 2018). research has revealed that when individuals communicate, only 7% of the information they convey is transmitted through language, while speech accounts for 38%, and facial expressions play a significant role at 55% (savran and sankur 2017). based on this, it can be inferred that a significant amount of valuable information can be gathered by observing an individual's expression in order to detect their consciousness and mental activities. facial expressions can cause changes in the appearance and geometry of the face, which can ultimately reduce the ability to recognise someone. the need to address this issue has inspired researchers to develop various approaches and schemes like the investigations presented in (ge et al. 2022; zeng et al. 2022).

Figure 2.5. different face expression using scface dataset (guo and wang 2012).

4. Low resolution-based FR challenge

The problem of low resolution in FR systems arises when the test face image has been significantly degraded. This leads to the loss of important information on the facial image among different individuals. This issue particularly impacts FR systems, particularly in applications like surveillance (mamatkulovich 2023). Therefore, FR systems still face challenges when it comes to comparing images of lower resolution with those of higher resolution. Chu et al. (2017) conducted a study wherein the authors investigated the application of low-resolution FR using a single sample per individual. The study also suggested a cluster-based regularised simultaneous discriminant analysis technique that incorporates intercluster and intracluster matrices to regularise the matrices that depict the differences between and within classes. Unsupervised clustering is used to estimate cluster-based scatter matrices. Using the scface dataset, the researchers ran tests and were able to reach an 82.36% recognition rate.

(Zhang et al. 2015) presented coupled marginal discriminant mappings, a useful approach that focuses on distance metric learning. They employ a method that creates a single space from the data points of the initial low- and high-resolution features, enabling categorization. Using the ar face dataset, the researchers found that 84.75% of the faces could be recognised. Table 2.1 provides a concise overview of recent approaches that have been taken to address the problem of low resolution.

Table 2.1 provides a concise overview of recent approaches that have been taken to address the problem of low resolution

Reference	Method	Dataset	Remarks
(chu et al. 2017)	Cluster-based discriminant analysis method	scface	Their method regularizes matrices between and within classes by combining intercluster and intracluster matrices. Cluster-based scatter matrices are estimated via unsupervised clustering.
(p. zhang et al. 2015)	Coupled marginal-based discriminant mappings method	ar	The data points from both the original low and high-resolution features are integrated into a single space, and classification is conducted.

2.8. Summary

FR is currently a highly difficult and complex subject in DL, NNs, pattern analysis, and computer vision; it has been discussed in several research communities, including both controlled and uncontrolled environments. This chapter covered the comprehensive studies of the FR models. The utilized data sets with the research challenges have been listed in this chapter, as well the classifications –based FR models have been explained well.

3. RESEARCH METHODOLOGY

3.1. Introduction

In chapter 2, related existing FR were explored, and the research gaps that led to the design and development of an adaptive and efficient FR system were identified. This chapter explains the research methodology adapted to design and develops the proposed model. The chapter starts off by highlighting the issues that this research addresses and offering the appropriate solutions. Then, an overview of the research framework is introduced. The dataset that was used in the thesis are illustrated. The design considerations for each stage of the methodology are then described. Finally, the peormance evaluation metrics and result analysis measures are outlined.

3.2. An Overview of Research Framework

The aim of this research is to improve the efficiency of existing FR approach. Like any research, implementing a study requires a well-designed research plan. This plan serves as a work plan and ensures that the evidence gathered is sufficient to determine the response to the research objective. Generally, the proposed model involved several phases, including: investigating the literature of the existing studies; designing and developing the SNNs with VGG -16 as a backbone; discussing the experimental results; evaluating the achievements and finally writing the thesis as shown in Figure 3.1.

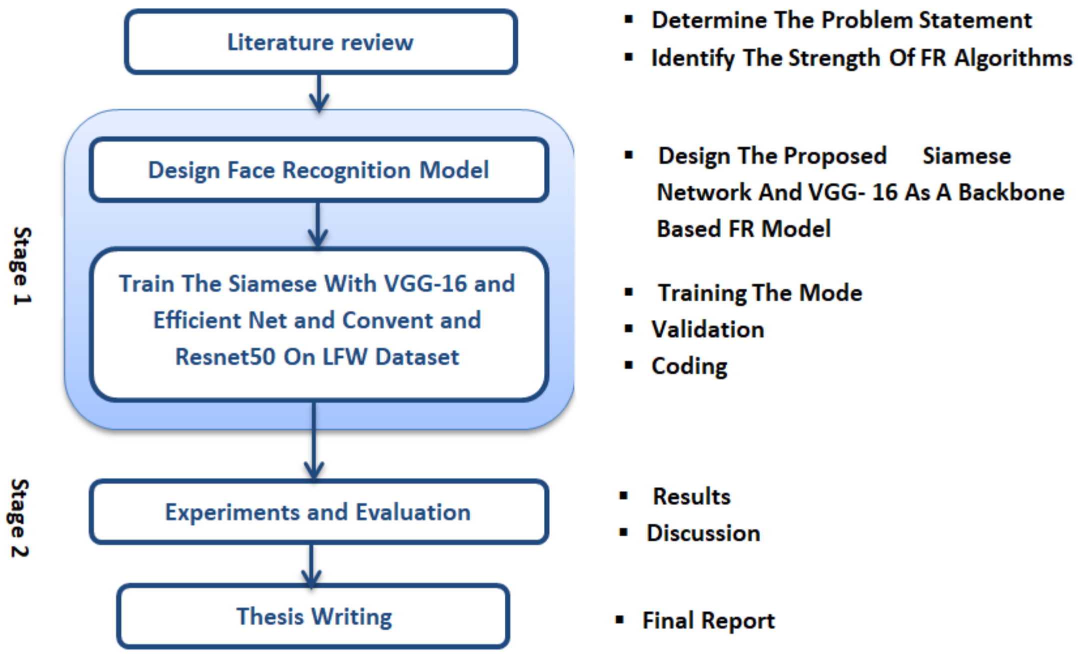

Figure 3.1. The general framework of the proposed FR model.

The first main stage focuses on designing the SNN with VGG-16 with three more DL algorithms Efficient , ResNetB0 and ConvNext -based FR using the LFW dataset. The second main stage of this study involves evaluating stages 1. The following sections provide a detailed explanation of each section.

The proposed model begins to function after datasets are obtained that is appropriate for the SNN with VGG-16 as a backbone. Data pre-processing is the process of after data gathering, photos are resized and transformed into the most appropriate format for networks. SNN with VGG-16 as a backbone are focused on peorming FR. After networks were recently learned, the peormance of the system and its capacity to identify and recognise faces are evaluated using new input data (testing data). In real-life scenarios, the input data is obtained from digital cameras or video captures installed within the organization that intends to utilise this system for FR. Figure 3.2 illustrated the proposed FR model.

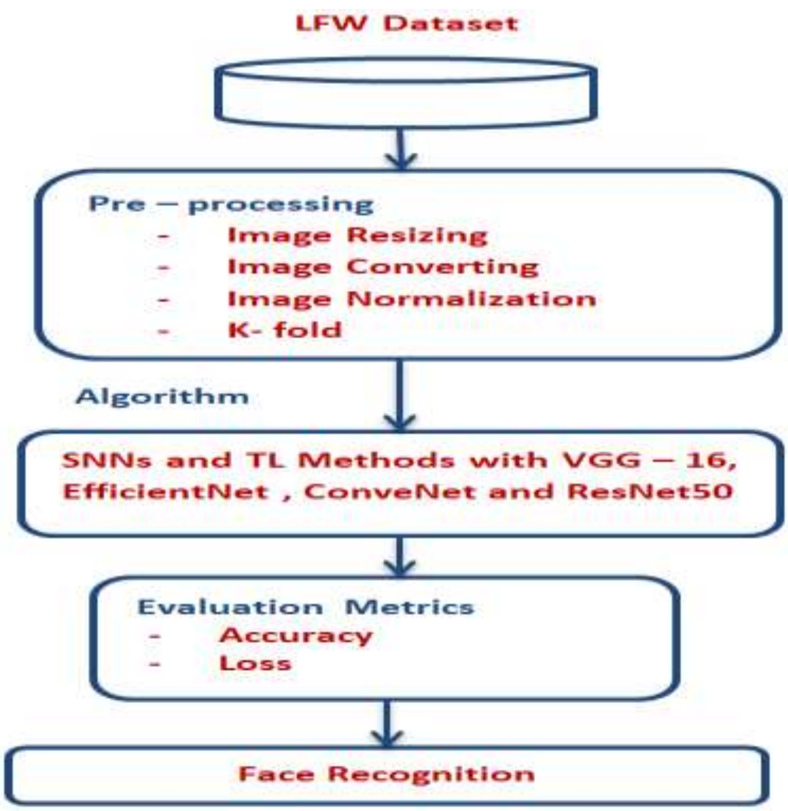

Figure 3.2. The proposed FR model.

3.2.1. Materials and Methods

This section consists of three actions: designing the SNN with pre-trained networks VGG-16, EfficientNetB0, ConvNeXtBase and ResNet50 as a backbone based FR model and providing an explanation of the LFW dataset. Finally, the SNN with VGG-16, EfficientNetB0, ResNet50 and ConvNextTiny model was trained using the same dataset..

A. Transfer Learning

A Transfer Learning (Torrey et al. 2010). A particular techniques used to transfer previously learned knowledge gained from a related domain to a target field. Since our data set is compact with around 13,000 samples, to avoid training issues, we perform data transfer. There are specific steps to transfer the particular knowledge. First, we load the model we want to transfer in this study use VGG-16, EfficientNetB0 , ConvNeXtBase and ResNet50 , and then determine how many layers the previous training requires to keep. Next, we use the new model to be the backbone of the SNN model. The training set used in this second stage is the LFW in the next stage, the pre-trained network model. It is used for LFW testing. Finally, accuracy is calculated based on the task Transfer learning. Architectural flow plan for transfer of learning.

While selected to train the final two layers of the model without freezing it. This approach was selected to optimize the model's adaptability and responsiveness to the dataset, leading to consistent performance and learning in this study, the pre-trained models VGG-16, EfficientNetB0, ConvNeXtBase, and Wide ResNet-50 were used. Pre-trained models require input photos adjusted in mini-batches of 3-channel RGB images of form (3 × H × W), with H and W predicted to be 62 x 47. The final classification layers were replaced with fully linked ones. ReLU activation functions and a binary cross-entropy loss were also utilized. The basic layers from training were frozen, while the changed layer was fine-tuned using the LFW dataset. The model was trained for 120 epochs with a batch size of 32. Adam (Kingma et al. 2014). was utilized as the model optimizer with a training rate of 1e-4. To improve the model's efficiency in terms of computation time and performance, model improvement approaches were applied.

B. Siamese neural network algorithm (SNNs)

A SNNs (chicco 2021) refers to a type of neural network architecture that consists of two or more subnetworks that are identical to each other. In this context, 'identical' refers to having the same configuration, parameters, and weights. The parameter updating is mirrored across both sub-networks. Artificial neural networks are utilised to determine the similarity between inputs by comparing their feature vectors. As a result, these networks find application in numerous domains (serrano and bellogín 2023).

Neural networks are nearly peect at every task in the deep learning era of today, but they need more data to do so. However, in certain cases such as FR obtaining more data may not always be feasible. To address these challenges, a novel neural network architecture known as SNNss has been developed (naser et al. 2023). It utilises a limited number of images to improve its predictions. Moreover, SNN have gained popularity in recent years due to their remarkable ability to learn from minimal amounts of data.

In this study, the Siamese network consists of 2 identical neural networks, each with VGG-16 as the backbone, that receive the input images. The last layers of these networks are subsequently connected to a binary cross-entropy loss function (ruby and yendapalli 2020), responsible for measuring the similarity between the two inputs. To ensure simultaneous learning of identical features for both input images, the twin networks share the same configuration and parameters (heidari and fouladi-ghaleh 2020).

Two scenarios are presented in this study; first scenario is having 2 images of the same person “positive pair”. Second scenario is feeding 2 images of different individuals “negative pair”. Through training, the network becomes capable of differentiating between positive and negative pairs, which is also used to compare a new, unseen image with a stored image for authentication purposes. As an output, the network gives out a metric that quantifies how much the two input images are different. This metric is then compared with a threshold to decide whether the two images are of the same person or not. The threshold is determined based on the distribution of distances observed during the training phase. To summarize, the SNN with VGG-16 as the backbone presents a reliable and effective approach for FR. Its ability to handle diverse factors like pose, lighting, and expression renders it well-suited for practical applications, including the specific scenario described in this study. Figure 3.3 explain the architecture of SNN with VGG-16 model as a backbone.

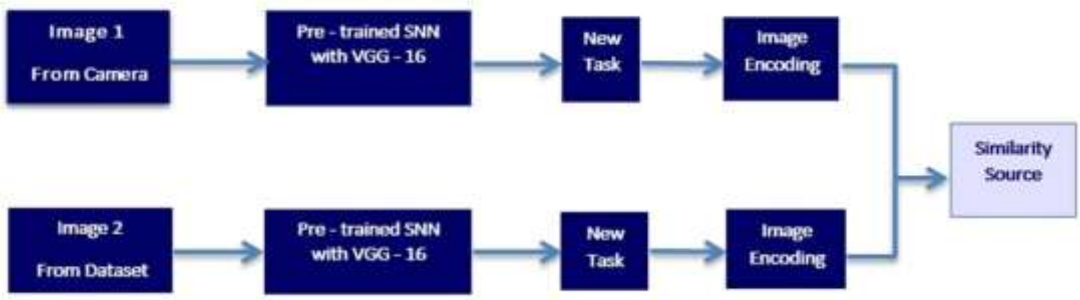

Figure 3.3. The architecture of SNN with VGG-16 model as a backbone..

C. VGG-16 network

VGG-16 (tammina 2019), named after the visual geometry group (VGG) at oxford, is a convolutional neural network (CNN) architecture that has proven to be one of the most effective vision model architectures to date. It consists of 16 layers with weights, making it a relatively extensive network with a total of 138 million parameters. The architecture of VGG-16 is simple and incorporates the most important features of convolutional neural networks. It uses small 3x3 convolution filters and a stride of 1, which are in the same padding, and a max pool layer of 2×2 filters for stride 2. This arrangement of convolution and max pool layers is consistent throughout the whole architecture. The network also includes three fully connected layers. The simplicity of the VGG-16 architecture is its main attraction, and its ability to handle diverse factors like pose, lighting, and expression makes it well-suited for practical applications (simonyan and zisserman 2014). Figure 3.4 illustrates the general architecture of the VGG-16 network.

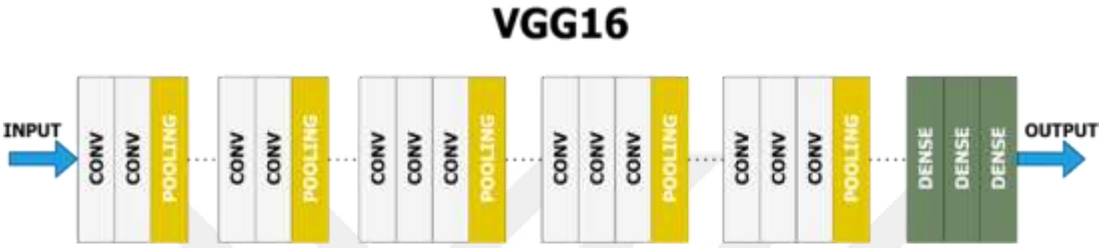

Figure 3.4. The general architecture of the VGG-16 network.

To perform the transfer learning process, we freeze the layers to preserve the weights that were previously trained on ImageNet dataset, then separate the VGG-16 head, which consists of three fully connected layers and the final CNN, and replace them with SNN layers to perform the recognition process by adding relu as activation function .Figure 3.5.

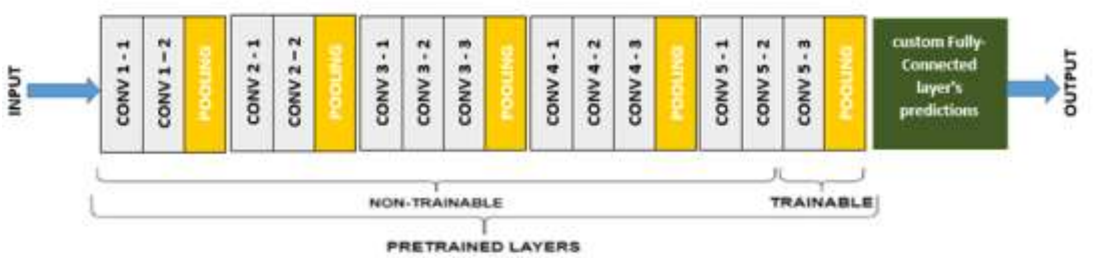

Figure 3.5 Adding New layers on the top of the VGG-16 model

D. EfficientNetB0

EfficientNetB0 (Tan et al. 2019) is a convolutional neural network architecture, It solves the challenging task of reconciling accurate modeling with speed of computation, with a target of reaching state-of-the-art performance with significantly fewer the parameters and utilized resources than previous versions. The design makes certain that all network parameters modify in a consistent manner, improving performance while decreasing computation weight. The primary feature of EfficientNet is its composite scaling mechanism, which constantly changes the network's depth, width, and accuracy according to a set of predefined scaling factors. EfficientNet includes a hierarchy of layers formed comprised of repeating blocks that gradually improve the network's depth, width, and resolution. EfficientNet-B0 that used in this study has 7 layers in each block. EfficientNet-B0 has fewer layers compared to EfficientNet-B7 has 19 layers in each block. Once the architecture lengths, a total number of layers increases significantly. After construct the SNN, EfficientNetB0's layers should be frozen except for the last layers. Train the SNNs using the LFW dataset while updating the unfrozen layers' weights. Figure 3.6 shows the general architecture of the EfficientNetB0. Figure 3.7 shows the transfer learning.

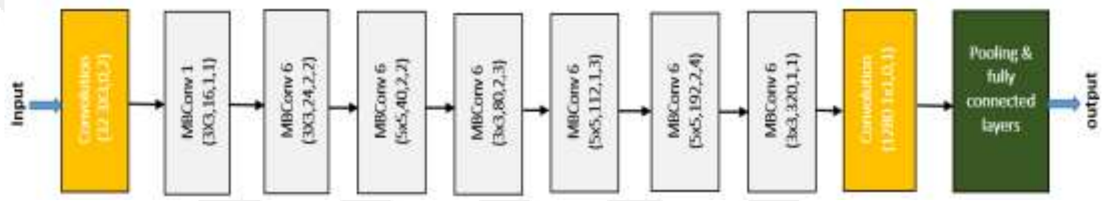

Figure 3.6 The general architecture of the EfficientNetB0 model

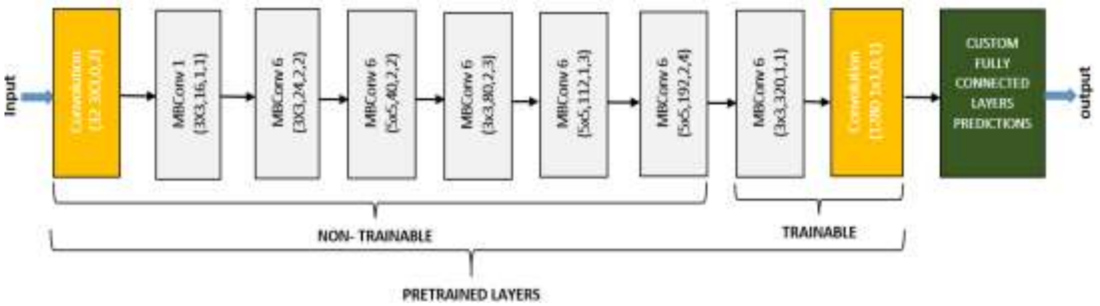

Figure 3.7 Adding New layers on the top of the EfficientNetB0 model

E. ResNet50

ResNet50 is an architecture for convolutional neural networks that was first presented by (He et al. 2016). It belonged to the ResNet (Residual Network) family, which has been developed to use residual learning to solve the degradation issue that extremely deep networks face. Furthermore, ResNet50 was developed using a fundamental component architecture and has 50 layers. Figure 3.8 and 3.9 show the trainable model.

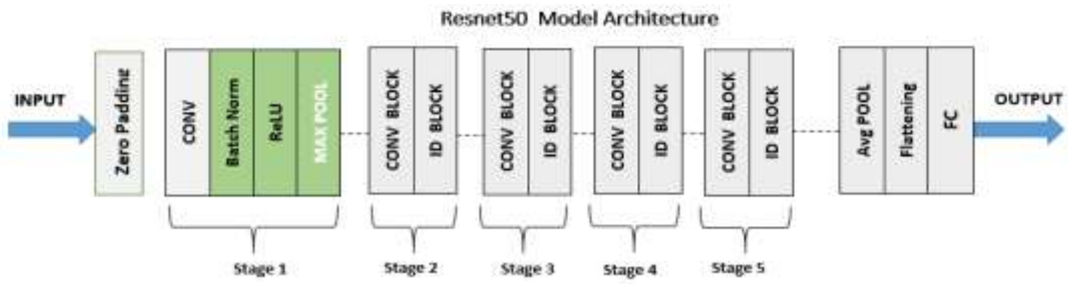

Figure 3.8 The general architecture of Baseline ResNet50 model.

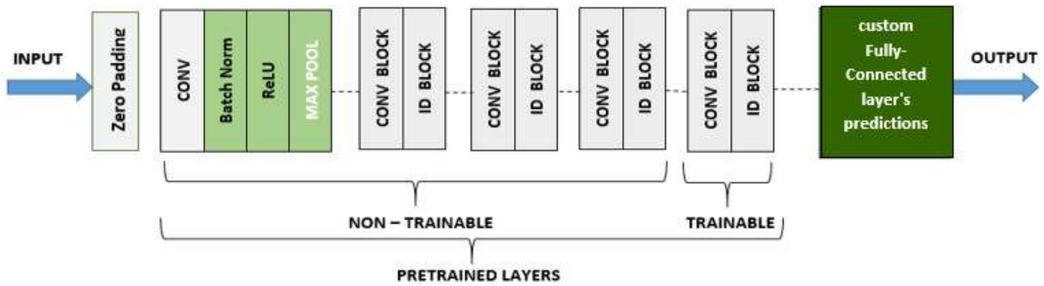

Figure 3.9 Adding New layers on the top of the. Baseline ResNet50 model

F. ConvNeXtBase

ConvNets (He et el. 2016) another developed categorized image method. The ConvNeXtBase architecture is a convolutional neural network (CNN) design that improves traditional CNNs by using structured set convolutions. It is proposed as part of the "Aggregated Residual Transformations for Deep Neural Networks" (ResNeXt) (Liu et al. 2022), framework, which seeks to increase model performance and efficiency. The number of layers in a ConvNeXt network varies based on the version and architecture configuration, with a preference for residual block composition and pooling convolutions over a fixed overall number of layers. Consider using a moderately deep architecture, such as ResNeXt-50 or ResNeXt-101 (Liu et al. 2022), for FRtasks that use faces tagged in the Wild LFW dataset, which comprises a very limited number of face photographs per subject (typically approximately 10). These architectures find a mix between model complexity and computational efficiency, making them ideal for workloads that require medium-sized datasets, such as LFW. After loading the pre-trained ConvNeXt model, drop its last fully connected layer. This layer is typically used for classification tasks and is not required for feature extraction in the SNN weights. Figure 3.10 (He at el. 2016)..

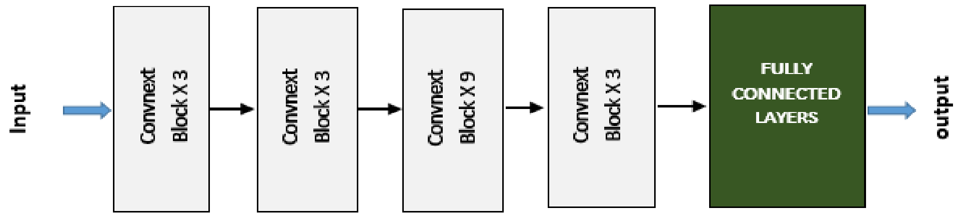

Figure 3.10. The general architecture of ConvNeXtBase model

G. Labelled faces in the wild LFW dataset

The labelled faces in the wild LFW dataset (jalal and tariq 2017) is a collection of face photographs specifically gathered for the purpose of studying unconstrained FR. The dataset comprises more than 13,000 facial images that have been gathered from various sources on the internet. The name of each person in the image was labelled on their respective face in this data set. The LFW dataset consists of four distinct sets of images. These sets include the original images as well as three types of aligned images. These aligned images are specifically designed to evaluate algorithms in various conditions. the dataset utilises funnelled images (g. b. huang et al. 2008), LFW-a, and deep funnelled images (sohn, lee, and yan 2015) for alignment purposes. Most face verification algorithms tend to yield better results when using deep funnelled and LFW-a images compared to funnelled images and original images. Figure 3.11 illustrates the LFW image dataset.

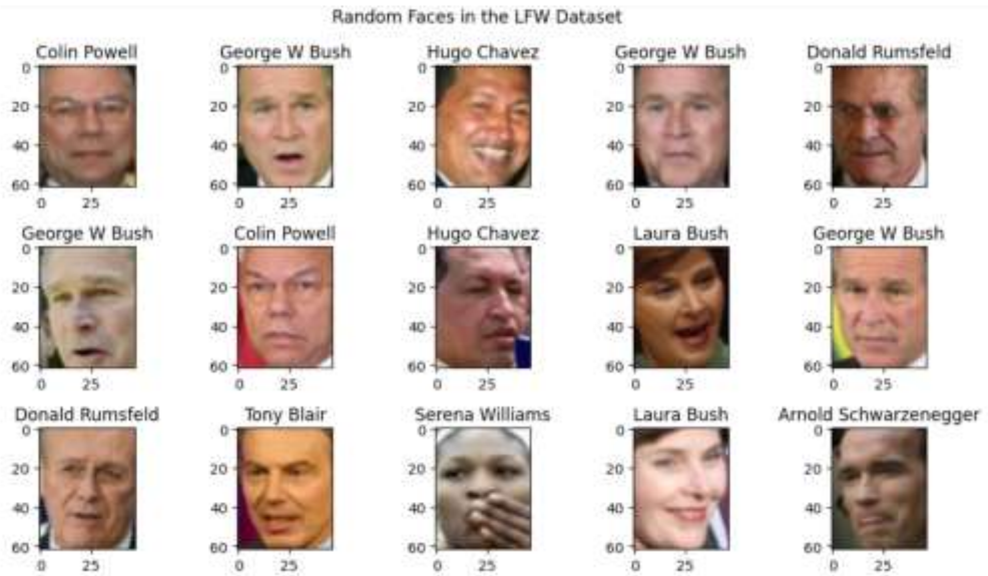

Figure 3.11. An example of the LFW image dataset.

H. Training SNN with transfer learning models

A commonly used neural network for determining the comparison across two identical input is called the SNN. This study explores the possibility of two individuals with comparable facial characteristics. To ensure the highest accuracy in facial recognition and identification, it is essential to incorporate the SNN with VGG-16 as the backbone.

The SNN with VGG-16 as the backbone have been trained using the LFW dataset, which contains images of 1680 individuals with two or more distinct image available. The main objective is to recognize the individuals. The LFW data set is specifically created for studying the challenge of FR in real-life scenarios, where conditions are unpredictable and include various factors such as pose, lighting, and expression variations. These conditions are commonly encountered in everyday situations.

In order to use the LFW dataset as input for the SNN, it is necessary to preprocess it. The pre-processing steps include resizing the images. The images have been resized to dimensions of 62 pixels by 47 pixels (see equation 3.4). Images are scaled to reduce computational complexity and ensure that the network concentrates on the main elements of the faces. The specific size was chosen because the dimensions of the head are always wider than its width. Since this study using VGG-16 which is a color-based model, the step to convert images to grayscale has been omitted.

$$\text{resized image} = \text{resize}(\text{image},(62,47)) \tag{3.4}$$

Since the LFW dataset contains limited number of samples (13,000), the proposed study utilised the Stratified K-Fold cross-validator (Stone, M. 1974) through splitting the LFW dataset into 5 sets to increase the number of dataset images and to a void the overfitting, allowing us each fold has a chance to appear in the training set (k-1) times, making sure each observation in the dataset appears in the dataset and enabling the model to learn the fundamental information distribution better. The number of observations (n= 5). Figure 3.12 depicts the process of splitting the utilised dataset into 5 loops.

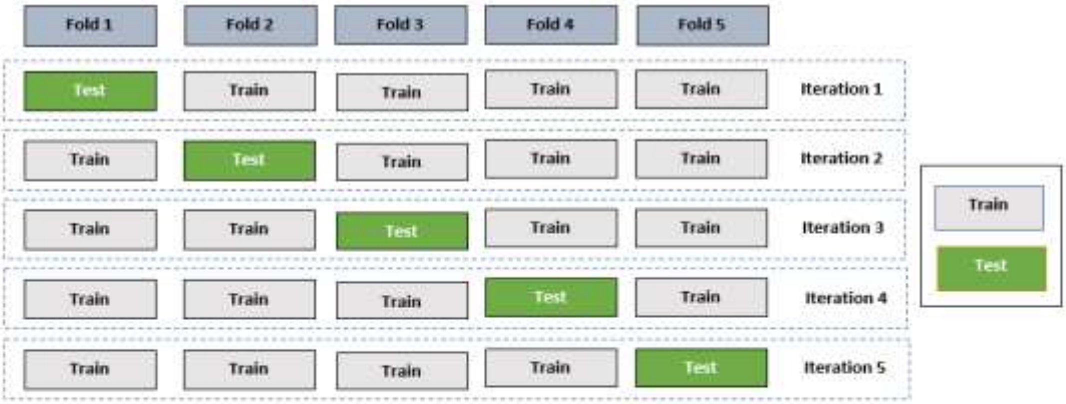

Figure 3.12 The process of splitting the dataset into 5 loops.

3.2.2. Evaluation metrics

To evaluate the accuracy of the proposed FR system, this study used accuracy, loss function metrics frequently employed by the body of existing research—to assess the accuracy of the FR systems (z. chen, feng, and zhang 2022; xiaopeng li, xiang, and li 2023). These measures are well-known and commonly used in the existing research (grother, ngan, and hanaoka 2019; howard et al. 2022). Moreover, computational complexity, and memory utilisation were used to evaluate the efficiency of the proposed scheme based on computation theory as it is commonly practised by the related works (zulfiqar et al. 2019). The following equations (3.5) and (3.6) and were used to determine the various parameters: accuracy and loss .

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.5}$$

$$L(y,p) = -(ylog(p) + (1-y)\log(1-p)) \tag{3.6}$$

where TP, TN, FP, FN denote the true positive, true negative, false positive and false negative respectively. While, Log is logartem.

3.3. Tools Used for Thesis

The proposal has been carried out using a personal computer that had the following specifications:

i. Intel® core™ i7 4210h cpu, operating at 2.90 ghz, and vga 2g nvidia gtx.

ii. Microsoft windows® 7 home premium edition (64-bit).

iii. 4 gb pc3-10600 (ddr3-1333) ram.
iv. 1 tb (7200 rpm) sata-4g hard disk.

Meanwhile, the following software will be used for the preparation of data samples, the implementation of algorithms, and the analysis of the result

i. Python opencv library with tensor flow library (windows 64 bit) for:

   a) Implementing the proposed SNNs in the initialization and implementation the recognition phases.
   b) Plotting graphs.

ii. Google Colab Pro + :

   a) GPUs = K80, P100, T4
   b) CPU = 2 x vCPU
   c) RAM = 32GB

3.4. Summary

In today's world, there is a growing need for an efficient and user-friendly employee logging system that saves time. This is particularly important as technology continues to advance and traditional logging systems become less convenient. Fingerprint logging systems are no longer considered a reliable option due to various issues, including forgery, deformations, and the potential transmission of diseases. Relying on facial recognition-based systems can provide a safer and more precise method for logging employees.

This study proposed a system that integrates a Siamese network for FR technology. SNN had been trained using the LFW data set. It demonstrated minimal cross-entropy loss and good accuracy, indicating its capacity to detect and distinguish humans in varied positions, lighting situations, and facial expressions.

4. THE PROPOSED RECOGNITION SCHEME

4.1. Introduction

The first chapter of this work presented the research problem, aim, and objectives, while the second chapter delved into the literature related to existing FR approaches. In Chapter 3, the methodology adopted to conduct this research was described. This chapter presents the design and implementation details of the proposed FR scheme in this work, in addition to the explanation of the experimental evaluation. In addition to the introductory section (4.1), the next section provides an overview of the proposed methodologies (4.2), an explanation of the proposed system (4.3), results analysis and comparison with the related works (4.4), and a summary of the chapter (4.4).

4.2. Overview of the Utilized Methods

The adopted deep learning algorithms for achieving the proposed FR system in this work are discussed in this section. The proposed study provide a comprehensive overview of the utilized VGG-16, and Siame Network in sections 4.2.1..

4.2.1. The Siamese neural network algorithm SNN

The SNN was first used for signature verification by Bromley et al. (1993) when time-delay neural networks were deployed as sub-networks for similarity comparison; since then, many SNNs have been proposed (Kumar, Carneiro, & Reid, 2016). Most of the accomplishments of SNN have come in the last few years; however, it has been noted that these networks needs relatively high processing power since they are based on DL architectures (Qi et al., 2020; Zeghidour et al., 2016). Regrettably, studies on deep learning techniques such as Siamese networks have slowed recently due to the decline in computing capacity. However, advancements in hardware equipment's computational capacity have recently allowed for more studies on these networks. Siamese network research has garnered increasing attention in recent years (McLaughlin et al., 2016).

The Siamese network is occasionally viewed as a restructured framework built upon a few fundamental networks. But in practical applications, a Siamese network is required (J. Chen et al., 2020; You et al., 2020). The following is a list of Siamese network advantages;

(i) Siamese networks excel in solving matching issues like object tracking, face matching, re-identification, change detection, and product recommendation. The Siamese network compares input instances to acquire knowledge. A Siamese network can easily learn input similarity and use that information to guide it through matching challenges.

(ii) Siamese networks can identify distinguishing characteristics for tasks below by exploring the feature domain. Siamese networks are frequently employed in regression and classification problems to develop efficient feature representations for decision-making.

(iii) Learning with labeled and unlabeled data is made possible by Siamese networks. It acquires knowledge through comparison rather than explicit labeling. In practical use, labels are not always universal. The Siamese network is crucial to overcoming this kind of obstacle.

(iv) Siamese networks can be applied to few-shot learning (FSL) to improve data.

The Siamese network produces more training data via comparing training pairs than by using the original instances directly, which is crucial for enhancing FSL's accuracy. The general Siamese network framework is shown in Figure 4.1. As per (Synnaeve & Dupoux, 2014), there are two branches and mostly three components in the basic Siamese network; these components include features visualization, sampling comparison, and decision-making processes. The typical Siamese framework in Figure 4.2 has two branches $f1_{(.)}$ and $f2_{(.)}$ sharing the same network structure and weight information W, with $f1_{(.)}$ and $f2_{(.)}$ being the two input instances x_1 and x_2, respectively. The feature extraction results are given as $f_1(x_1)$ and $f_2(x_2)$.

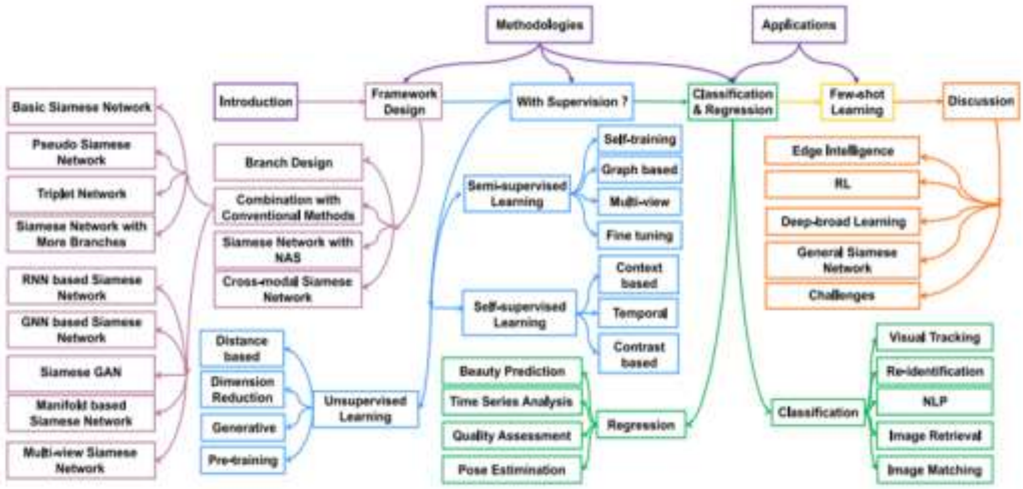

Figure 4.1. The general framework of the Siamese network (Li, Chen, & Zhang 2022)

Comparison head $C_h(.)$ was used to compare $f_1(x_1)$ and $f_2(x_2)$ and the result is given as a comparison score C_{score}

$$C_{score} = C_h(f_1(x_1), (f_2(x_2)). \quad (4.1)$$

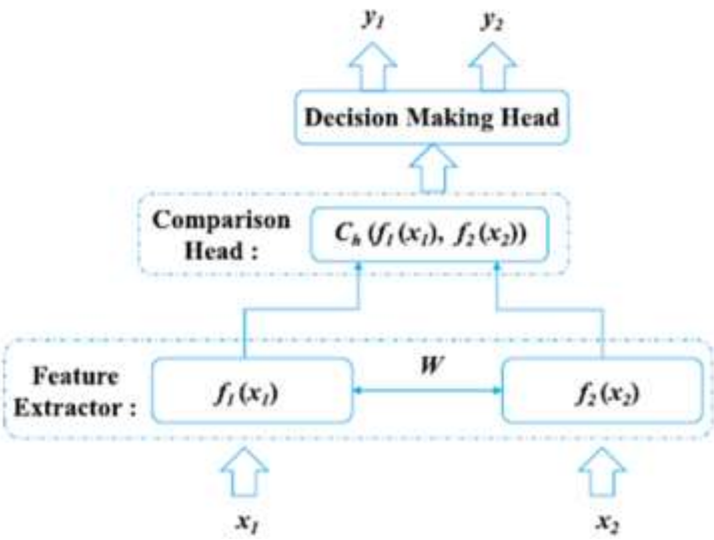

Figure 4.2. General network structure of the Siamese network.

Two major types of comparison- similarity and ranking, are the two main comparison mechanisms used in this Siamese network in this work. Similarity comparison is commonly used in classification, where it compares the priorities of samples. In regression, on the other hand, it primarily compares the degree of similarity between inputs by calculating distance D, where a smaller value of D indicates a higher degree of similarity. The ranking comparison results are presented as a ranking score R_{score}, as determined by the following equations:

$$C_{score} = \{D, R_{dcore}\} \tag{4.2}$$

C_{score} is then further used to calculate comparison loss L_{com} for model optimization

$$L_{com} = C_{loss}(C_{score}, M) \tag{4.3}$$

C_{loss} can be differently designed like equations (4.2), (4.3). $M > 0\, is$ is a hyper-parameter for guiding the optimization process. In similarity comparison, for instance, the optimization will proceed if $D < M$ for negative pairs (x_1, x_2) until $D \geq M$. On the other hand, optimization will continue in ranking comparison if $|R_{score}| < M$, until $|R_{score}| \geq M$. In decision-making, the SNNs predicts labels $y_1$and y_2 for $x_2$and x_2 according to C_{score} with varying mechanisms such as specially designed sub-networks and threshold-based methods.

In this work, binary entropy was applied in the SNN for FR to decrease the loss function during training. This encourages the network to provide probabilities, for example, the likelihood that two input images show the same person that correspond to the actual labels supplied during training determined by the following equations:

$$L(y, p) = -(ylog(p) + (1 - y)log(1 - p)) \tag{4.4}$$

If y = 1,the loss simplifies to $-log(p)$, encouraging p to be close to 1 (high confidence in positive class).

If y = 1,the loss simplifies to $-log(1 - p)$, encouraging p to be close to 0 (high confidence in negative class).

Practicaly, the proposed loss function is ideal for binary choice problems and is frequently used in machine learning due to its success in training models to make accurate probabilistic predictions.

4.3. Results Discussion and Analysis

This section is dedicated to presenting the results that were obtained during the training process of the proposed networks: SNNs with VGG-16 network as the backbone. Following the training the networks, they were tested with new images to evaluate how accurately they could recognize individuals. The performance of the models is evaluated based on how effectively they can distinguish various persons or recognize an identical individual in varied facial expressions, positions, lighting, etc.

4.3.1. Evaluation of the SNN with VGG-16 as backbone

The evaluation of the VGG-16-based SNN also yielded promising results, demonstrating its suitability for FR in the proposed systems. As shown in the Figure 4.3 below, as the number of training epochs rose, the binary cross-entropy loss reduced to lower levels. This decrease in loss implies that the network successfully learnt the distinctive traits of each face, resulting in a minimum divergence between the anticipated and real outputs. This effective learning makes the network a robust tool for FR in the proposed systems. Figure 4.3 depicts the binary cross entropy loss for the SNN.

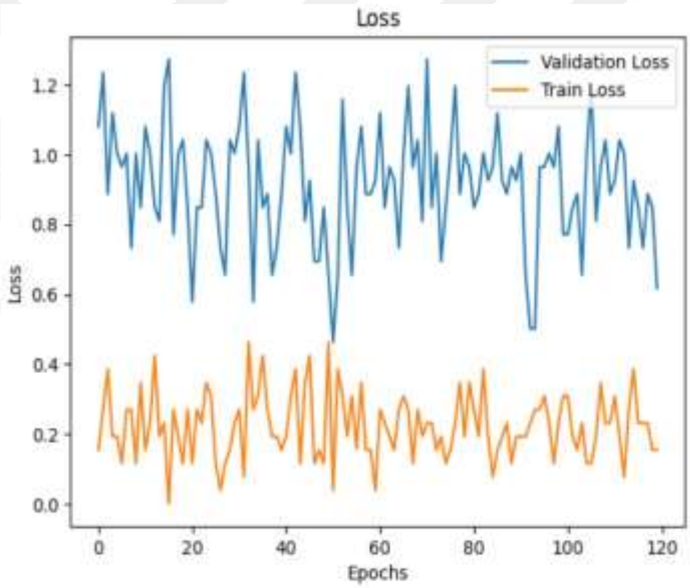

Figure 4.3. The Binary Cross Entropy Loss for SNN.

Furthermore, an accuracy curve was produced, revealing a high accuracy score of 96% approximately on the validation set. This number represents the ability of the model to accurately categorize the majority of the input imageThe high accuracy result makes the network appropriate for the job of the proposed system. Figure 4.4. Illustrates the accuracy of SNN with VGG-16 as the backbone

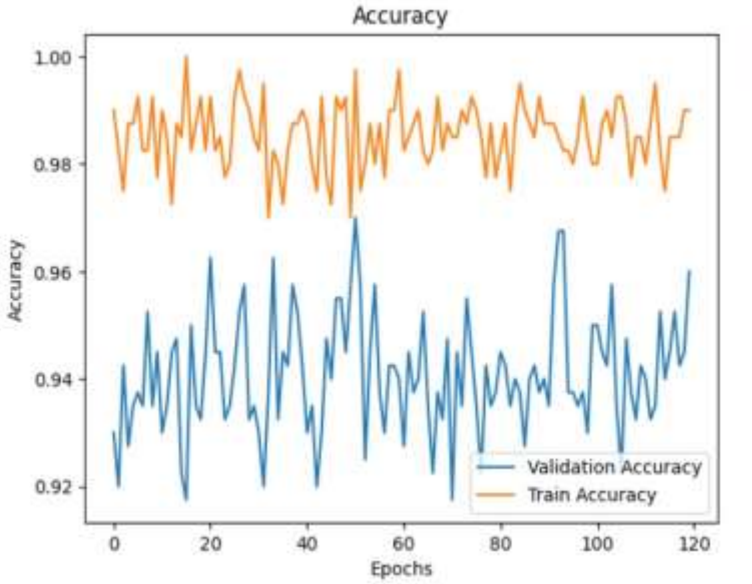

Figure 4.4. The accuracy of the SNN with VGG-16 as the backbone.

When training the SNN with VGG-16 as a backbone, it was tested on new test images to find out how well the network performed. When given with three images of the same individuals but with varied facial expressions, the network nevertheless properly identified the three images as the same person. This displays the network's ability to correctly recognize and categorize similar faces in the dataset. Figure 4.5 depicts the test model for a random sample.

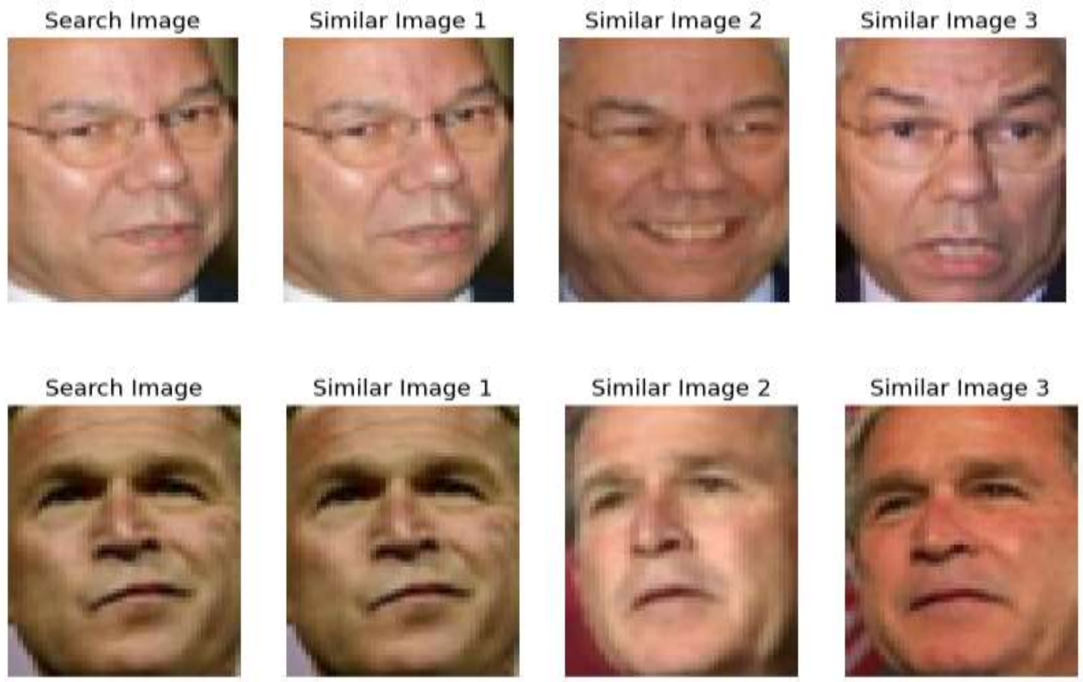

Figure 4.5. The test model on Random Sample.

As seen in Figure 4.6 below illustrates the results which had been achieved during training the proposed system using the suggested SNN with Efficient New algorithm. The achieved results explain that the proposed system has gain 77.75% as a validation accuracy. Figure 4.7 depicts the results gained from training the proposed SNN with ConvNext, it had ganied 93.75% asa validation accuracy and As seen in Figure 4.8 show SNN with ResNet-50 has 95 %.

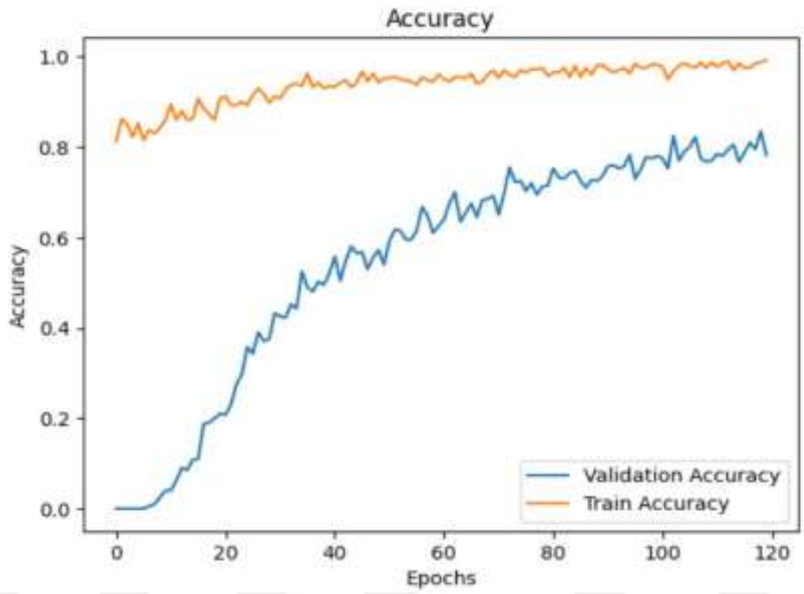

Figure 4.6 Siamese network with Efficient

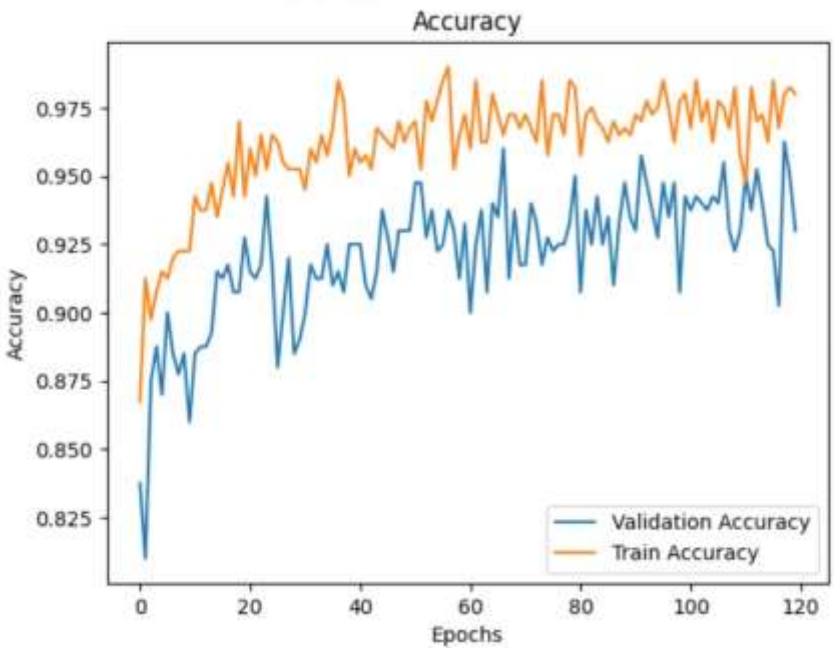

Figure 4.7 SNN with ConvNextTiny algorithm

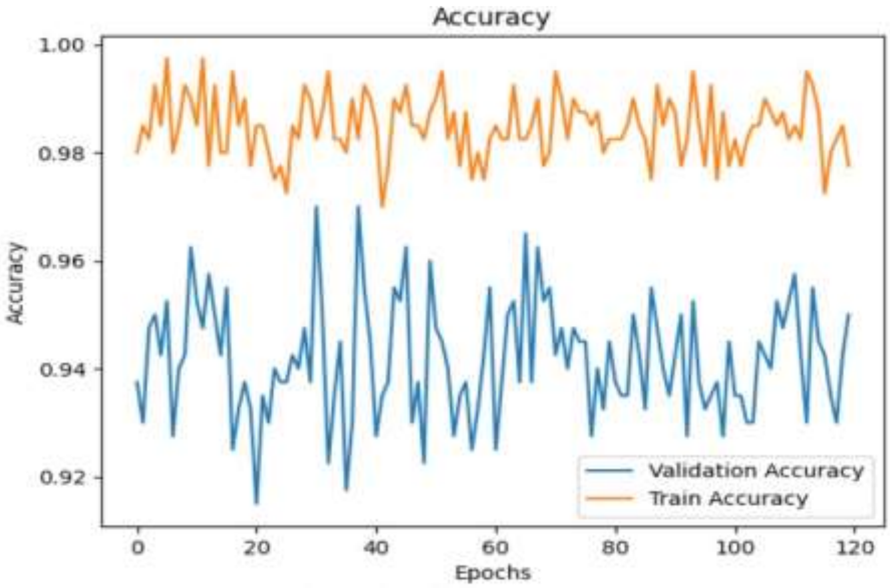

Figure 4.8 SNN with ResNet50 algorithm.

When the results obtained by the SNN with VGG-16 as the backbone are combined, it demonstrates the effectiveness of the integrated model in recognizing and identifying faces. The approach, which combines the Siamese network with VGG-16, is an effective method for logging employees within an organization with high accuracy and precision. This approach could potentially be a valuable tool for enhancing security and efficiency in workplace settings.

4.3.2. Comparison of FR peormance

In this research, a Siamese network with VGG-16 as a backbone has been used for the FR model. These approaches achieved an accuracy rate of 96.25% during the evaluation set, which was in compared with different studies is a higher degree accurate, or close to similarly excellent to them. For instance, in Table 4.1, the accuracy score of the proposed Siamese network with VGG-16 as a backbone is (96.25) This is much greater accuracy than certain approaches, including DLB (88.50), CFN+APEM (87.50), L-CSSE-KSRC (92.02), and SiameseFace1 (94.80). Other approaches, such as weighted PCA-EFMNet (95.00) and Siamese-VGG (95.62) offer respectable results, but their accuracy is still lower than the suggested Siamese network using VGG-16 as a backbone. The single technique that surpasses the suggested approaches in accuracy is CosFace, achieving 99.73 accuracy. However, despite achieving a lower accuracy compared to CosFace, the proposed SNN with VGG-16 is suited to run on simple hardware as this research has been trianed on the whole LFW dataset. While the CosFace study had been training utilzing only 5MB from the LFW dataset.

Table 4.1. Siamese Network with VGG-16 as a backbone peormance in comparison to other Methods.

Reference	Method	Face Recognition ACC (%)
Chong et al [31]	DLB	88.50
Xiong et al [32]	CFN+APEM	87.50
A. Majumdar, R. Singh [33]	L-CSSE+KSRC	92.02
J. Zhang, X. Jin, [34]	SiameseFace1	94.80
B. Ameur, M. Belahcene [35]	Weighted PCA-EFMNet	95.00
M. Haidari and K.Fouladi-Ghaleh [16]	Siamese-VGG	95.62
H. Wang *et al.* [36]	CosFace	99.73
SNN with VGG-16 as a backbone		**96.25**

To keep this study more robust and efficient, three more comparision studies have been utilized with VGG-16, these are Efficient loss, ResNet50 and ConvNext networks. When evaluating the SNN’s performance, it's essential to consider both accuracy and loss simultaneously. In this study, VGG-16 serves as the backbone architecture and to keep the proposed system more robust, VGG-16 have been benchmarked with ConvNext , ResNet50 and EfficientNet. The performance of these networks are evaluated based on metrics such as accuracy and loss, providing insights into their effectiveness for FR tasks , the performance of the networks across the five folds into which the database is divided is compared in terms of accuracy, loss and time as shown in Table 4.2 .

Table 4.2. Comparsion studies of the propsed SNN with VGG-16 with with ConvNext , ResNet50 and EfficientNet in terms of accuracy, loss and time.

Methods	Accuracy (%)	Loss (%)	Time (sec)
SNN with VGG-16 as a backbone	**96.25**	**57.84**	**570.63**
SNN with ResNet50 as a backbone	95	96.41	1129.87
SNN with ConvNextTiny as a backbone	93.75	74.21	3328
SNN with EfficientNetB0 as a backbone	77.75	103.9	542.76

By comparing the performance of different backbone architectures, this study aims to identify the most suitable model for enhancing the efficiency of FR approach. As seen in Figure 4.9 and 4.10 illustrate comparision studies between VGG-16 , ResNet50 and Efficient loss and ConvNext networks. Based on the graph in Figure 4.19, it is not possible to definitively say which model is the best. However, the train VGG-16 model appears to have the lowest loss. Based on the accuracy graph in the image, the Validation VGG-16 Accuracy appears to be the highest throughout the epochs, which suggests it might be the best model for this task.

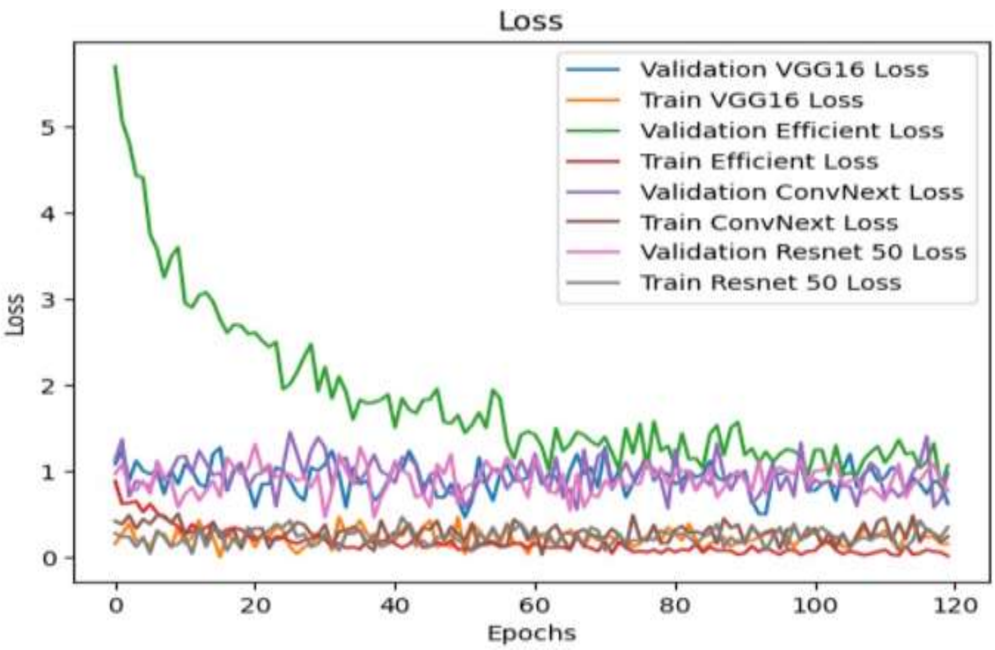

Figure 4.9 Comparative study of validation process between VGG-16 lass, Efficient loss and ConvNext loss and ResNet50

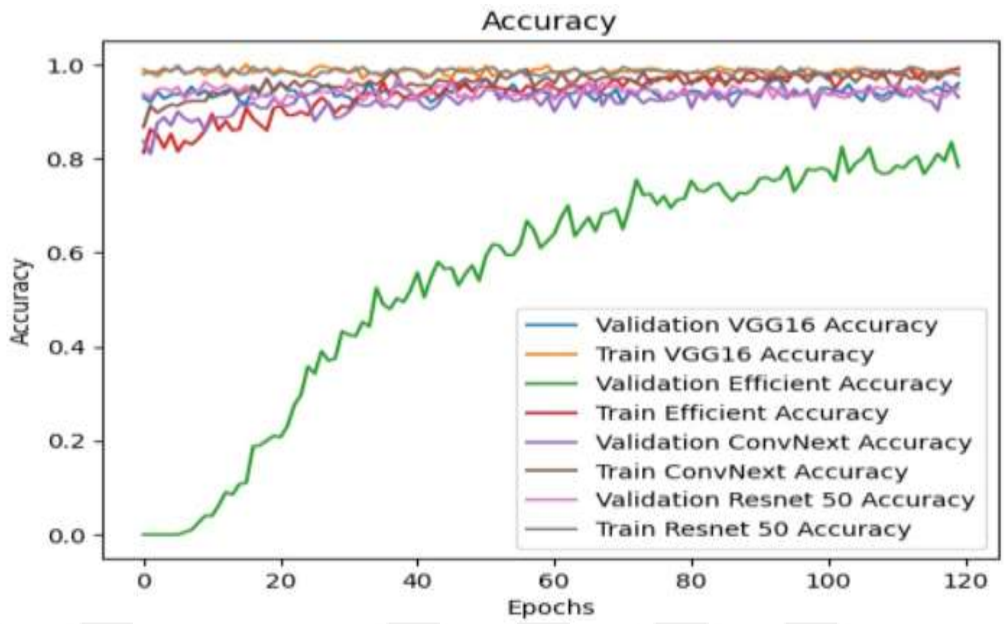

Figure 4.10 Comparative study of validation process between VGG-16 accuracy, Efficient accuracy and ConvNext accuracy

4.4. Summary

Many studies have been reported on FR but the issue of FR in complex scenarios remains unresolved. A FR-based system was proposed in this work by integrating a Siamese network with VGG-16 as the backbone for FR. The SNN with transfer learning models as the backbone. Various transfer learning architectures such as VGG-16, EfficientNet, RestNet50, and ConvNext have used for the experiments. 5-fold cross validation has been used to evaluate the performance of the architectures on LFW dataset. Comparative results show that EfficientNet, RestNet50 and ConvNext backbones achieved 77.75% accuracy, 95% and 93.75 % accuracy respectively. VGG-16 produced the best accuracy in FRwith 96.25%. This demonstrates its capacity to recognize and distinguish persons in various stances, lighting situations, and facial expressions.Thus, the integrated system combining the Siamese network with VGG-16 offers a promising solution for the problem of FR in complex environments.

5. CONCLUSION AND FUTURE WORK

5.1. Research Conclusion

The involvement of neural networks in deep learning-based FR models can unlock the capabilities that were earlier not feasible in the realm of science fiction; hence, it can be one of the ways of achieving secure, convenient, and efficient FR in various domains since it will encourage the development of more personalized and safe experiences in our daily lives. considering the advancements in this field and the need for ethical standards that ensure humanity's dignity and respect, this study proposes a FR-based system based on the combination of the Siamese network with VGG-16 (FR). Networks was trained on datasets (the LFW dataset) and the tests showed that the networks were suitable for the proposed tasks as they achieved good peormances in terms of the considered metrics; they were also able to distinguish people in different poses, facial expressions, and lighting conditions. Hence, the proposed system that integrated the Siamese network with VGG-16 could be described as a potential solution for FR systems.

5.2. Future Works

Advances are continuously made in the field of deep learning-based FR systems, and as such, many research directions and advancements are anticipated as follows:

1. Improved accuracy and robustness: scholars should focus on the development of more sophisticated architectures, novel training techniques, and larger, more diverse datasets that will improve the accuracy and robustness of future models in specific conditions, such as pose variations, low lighting, occlusions, and variations in facial expressions.
2. Multimodal fusion: multiple modalities such as infrared imaging, depth data, visible light imagery, and audio cues can be integrated for better FR systems in different environments.

3. Real-time peormance: the optimization of models for real-time peormance should be prioritized as it can enable more efficient FR on various platforms, including embedded systems and mobile devices.
4. Multi-modal and cross-modal FR: since face information can be captured most easily and affordably using the visual (rgb) modality, we concentrated on face identification using images and videos for this study. Nevertheless, we would want to draw attention to the busy and fascinating research fields of multi-modal FR, face anti-spoofing, and FR using additional modalities (3d, near-infrared, thermal infrared, drawings) (h. zhou et al. 2014). Researchers are becoming interested in heterogeneous FR, which attempts to match face photos taken with several modalities (peng et al. 2016). Forensic security, for instance, face drawings to picture matching is a major issue. research into the related, intriguing, and difficult problems of synthesizing a face shot from a sketch and vice versa (h. zhou et al. 2014).

Overall, the future of FR using deep learning approaches holds great promise for advancing technology's capabilities while addressing important societal concerns such as privacy, fairness, and transparency. Continued research and innovation in this field will undoubtedly lead to further breakthroughs and applications in diverse domains.

REFERENCES

Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: A review. *Electronics*, *9*(8), 1188.

Alyoubi, W. L., Shalash, W. M., & Abulkhair, M. F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, *20*, 100377.

Ameur, B., Belahcene, M., Masmoudi, S., & Hamida, A. Ben. (2018). Weighted PCA-EFMNet: A deep learning network for Face Verification in the Wild. *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 1–6.

Awan, M. J., Khan, M. A., Ansari, Z. K., Yasin, A., & Shehzad, H. M. F. (2022). Fake profile recognition using big data analytics in social media platforms. *International Journal of Computer Applications in Technology*, *68*(3), 215–222.

Bhandari, S. G., Rodrigues, S., Thejas, P. C., & Nausheeda, B. S. (2023). ANALYSIS OF FACE RECOGNITION USING LBPH ALGORITHM: A REVIEW. *Redshine Archive*, *2*.

Bhatta, A., Albiero, V., Bowyer, K. W., & King, M. C. (2023). The gender gap in face recognition accuracy is a hairy problem. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 303–312.

Bisogni, C., Castiglione, A., Hossain, S., Narducci, F., & Umer, S. (2022). Impact of deep learning approaches on facial expression recognition in healthcare industries. *IEEE Transactions on Industrial Informatics*, *18*(8), 5619–5627.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a" siamese" time delay neural network. *Advances in Neural Information Processing Systems*, *6*.

Chen, J., Yuan, Z., Peng, J., Chen, L., Huang, H., Zhu, J., Liu, Y., & Li, H. (2020). DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 1194–1206.

Chen, Z., Feng, X., & Zhang, S. (2022). Emotion detection and face recognition of drivers in autonomous vehicles in IoT platform. *Image and Vision Computing*, *128*, 104569.

Chicco, D. (2021). Siamese neural networks: An overview. *Artificial Neural Networks*, 73–94.

Chong, S.-C., Teoh, A. B. J., & Ong, T.-S. (2017). Unconstrained face verification with a dual-layer block-based metric learning. *Multimedia Tools and Applications*, *76*, 1703–1719.

Chu, Y., Ahmad, T., Bebis, G., & Zhao, L. (2017). Low-resolution face recognition with single sample per person. *Signal Processing*, *141*, 144–157.

Deepface: Closing the gap to human-level performance in face verification. (2014). *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1701–1708.

Drira, H., Amor, B. Ben, Srivastava, A., Daoudi, M., & Slama, R. (2013). 3D face recognition under expressions, occlusions, and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(9), 2270–2283.

Duan, Q., & Zhang, L. (2020). Look more into occlusion: Realistic face frontalization and recognition with boostgan. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 214–228.

Ge, H., Zhu, Z., Dai, Y., Wang, B., & Wu, X. (2022). Facial expression recognition based on deep learning. *Computer Methods and Programs in Biomedicine*, *215*, 106621.

Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schönborn, S., & Vetter, T. (2018). Morphable face models-an open framework. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 75–82.

Goyal, S. J., Upadhyay, A. K., Jadon, R. S., & Goyal, R. (2018). Real-life facial expression recognition systems: a review. *Smart Computing and Informatics*, *77*(1), 311–331.

Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test (fvrt): Part 3, demographic effects*. National Institute of Standards and Technology Gaithersburg, MD.

Guo, G., & Wang, X. (2012). A study on human age estimation under facial expression changes. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2547–2553.

Hadad, A. A.-A., Khalid, H. N., Naser, Z. S., & Taha, M. S. (2022). A Robust Color Image Watermarking Scheme Based on Discrete Wavelet Transform Domain and Discrete Slantlet Transform Technique. *Ingenierie Des Systemes d'Information*, *27*(2), 313.

Hayale, W., Negi, P., & Mahoor, M. (2019). Facial expression recognition using deep siamese neural networks with a supervised loss function. *Proceedings - 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019*, 1–7. https://doi.org/10.1109/FG.2019.8756571

Heidari, M., & Fouladi-Ghaleh, K. (2020). Using Siamese Networks with Transfer Learning for Face Recognition on Small-Samples Datasets. *Iranian Conference on Machine Vision and Image Processing, MVIP*, *2020-Febru*(February). https://doi.org/10.1109/MVIP49855.2020.9116915

Howard, J. J., Laird, E. J., Rubin, R. E., Sirotin, Y. B., Tipton, J. L., & Vemury, A. R. (2022). Evaluating proposed fairness models for face recognition algorithms. *International Conference on Pattern Recognition*, 431–447.

Hsieh, C.-F., Liu, P.-H., Lai, J.-X., & Sun, S.-H. (2022). Implementation of a Smart Checkout System Based on Face Recognition Using YOLO v3-tiny. *2022 IEEE 11th Global Conference on Consumer Electronics (GCCE)*, 903–904.

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*.

Jalal, A., & Tariq, U. (2017). The LFW-gender dataset. *Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13*, 531–540.

Jin, X., & Tan, X. (2017). Face alignment in-the-wild: A survey. *Computer Vision and Image Understanding*, *162*, 1–22.

Koley, S., Roy, H., Dhar, S., & Bhattacharjee, D. (2022). Illumination invariant face recognition using fused cross lattice pattern of phase congruency (FCLPPC). *Information Sciences*, *584*, 633–648.

Kumar BG, V., Carneiro, G., & Reid, I. (2016). Learning local image descriptors with deep siamese and triplet convolutional networks by minimising global loss functions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5385–5394.

Li, X., Xiang, Y., & Li, S. (2023). Combining convolutional and vision transformer structures for sheep face recognition. *Computers and Electronics in Agriculture*, *205*, 107651.

Li, Y., Chen, C. L. P., & Zhang, T. (2022). A survey on siamese network: Methodologies, applications, and opportunities. *IEEE Transactions on Artificial Intelligence*, *3*(6), 994–1014.

Li, Z., Gong, D., Li, X., & Tao, D. (2016). Aging face recognition: a hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing*, *25*(5), 2146–2154.

Lim, Y.-S., Lee, S.-H., Cheong, S.-J., & Park, Y.-H. (2023). A long-distance 3D face recognition architecture utilizing MEMS-based region-scanning LiDAR. *MOEMS and Miniaturized Systems XXII*, *12434*, 87–91.

Ma, X., Song, H., & Qian, X. (2014). Robust framework of single-frame face superresolution across head pose, facial expression, and illumination variations. *IEEE Transactions on Human-Machine Systems*, *45*(2), 238–250.

Majumdar, A., Singh, R., & Vatsa, M. (2016). Face verification via class sparsity based supervised encoding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1273–1280.

Mamatkulovich, B. B. (2023). A DESIGN OF SMALL SCALE DEEP CNN MODEL FOR FACIAL EXPRESSION RECOGNITION USING THE LOW-RESOLUTION IMAGE DATASETS. *MODELS AND METHODS FOR INCREASING THE EFFICIENCY OF INNOVATIVE RESEARCH*, *2*(19), 284–288.

McLaughlin, N., Del Rincon, J. M., & Miller, P. (2016). Recurrent convolutional network for video-based person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1325–1334.

Menon, S., Geroge, A., Aswathy, N., & James, J. (2021). Custom Face Recognition Using YOLO. V3. *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, 454–458.

Mohsin, A. H., Zaidan, A. A., Zaidan, B. B., bin Ariffin, S. A., Albahri, O. S., Albahri, A. S., Alsalem, M. A., Mohammed, K. I., & Hashim, M. (2018). Real-Time Medical Systems Based on Human Biometric Steganography: a Systematic Review. *Journal of Medical Systems*, *42*(12). https://doi.org/10.1007/s10916-018-1103-6

Moustafa, aAmal A., Elnakib, A., & Areed, N. F. F. (2020). Age-invariant face recognition based on deep features analysis. *Signal, Image and Video Processing*, *14*, 1027–1034.

Mustafa, S. M. N., Zehra, S. S., Baber, A., & Siddiqui, M. A. (2023). Fingerprint generation and authentication though Adaptive convolution generative adversarial network (ADCGAN). *2023 7th International Multi-Topic ICT Conference (IMTIC)*, 1–5.

Nachmani, O., Saun, T., Huynh, M., Forrest, C. R., & McRae, M. (2022). “Facekit”—Toward an Automated Facial Analysis App Using a Machine Learning–Derived Facial Recognition Algorithm. *Plastic Surgery*, 22925503211073844.

Nandy, A., & Mondal, S. S. (2019). Kinship verification using deep siamese convolutional neural network. *Proceedings - 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019*, *May*. https://doi.org/10.1109/FG.2019.8756528

Naser, Z. S., Khalid, H. N., Ahmed, A. S., Taha, M. S., & Hashim, M. M. (2023). Artificial Neural Network-Based Fingerprint Classification and Recognition. *Revue d’Intelligence Artificielle*, *37*(1).

Oloyede, M. O., Hancke, G. P., & Myburgh, H. C. (2020). A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications*, *79*, 27891–27922.

Pang, H., Xuan, Q., Xie, M., Liu, C., & Li, Z. (2021). Research on target tracking algorithm based on siamese neural network. *Mobile Information Systems*, *2021*. https://doi.org/10.1155/2021/6645629

Peng, C., Gao, X., Wang, N., & Li, J. (2016). Graphical representation for heterogeneous face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(2), 301–312.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W. (2005). Overview of the face recognition grand challenge. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, *1*, 947–954.

Qi, F., Yang, X., & Xu, C. (2020). Emotion knowledge driven video highlight detection. *IEEE Transactions on Multimedia*, *23*, 3999–4013.

Qinjun, L., Tianwei, C., Yan, Z., & Yuying, W. (n.d.). Facial Recognition Technology: A Comprehensive Overview. *Academic Journal of Computing & Information Science*, *6*(7), 15–26.

Regin, R., Rajest, S. S., & Shynu, T. (2023). Principal Component Analysis for ATM Facial Recognition Security. *Central Asian Journal of Medical and Natural Science*, *4*(3), 292–311.

Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng*, *9*(10).

Sabharwal, T., & Gupta, R. (2022). Deep facial recognition after medical alterations. *Multimedia Tools and Applications*, *81*(18), 25675–25706.

Savran, A., & Sankur, B. (2017). Non-rigid registration based model-free 3D facial expression recognition. *Computer Vision and Image Understanding*, *162*, 146–165.

Serrano, N., & Bellogín, A. (2023). Siamese neural networks in recommendation. *Neural Computing and Applications*, 1–13.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Sinha, A., & Barde, S. (2022). Illumination invariant face recognition using MSVM. *AIP Conference Proceedings*, *2455*(1).

Sohn, K., Lee, H., & Yan, X. (2015). Learning structured output representation using deep conditional generative models. *Advances in Neural Information Processing Systems*, *28*.

Song, C., & Ji, S. (2022). Face Recognition Method Based on Siamese Networks Under Non-Restricted Conditions. *IEEE Access*, *10*(3), 40432–40444. https://doi.org/10.1109/ACCESS.2022.3167143

Synnaeve, G., & Dupoux, E. (2014). Weakly supervised multi-embeddings learning of acoustic models. *ArXiv Preprint ArXiv:1412.6645*.

Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), 143–150.

Taskiran, M., Kahraman, N., & Erdem, C. E. (2020). Face recognition: Past, present and future (a review). *Digital Signal Processing*, *106*, 102809.

Tsai, H.-H., & Chang, Y.-C. (2018). Facial expression recognition using a combination of multiple facial features and support vector machine. *Soft Computing*, *22*, 4389–4405.

Waelen, R. A. (2023). The struggle for recognition in the age of facial recognition technology. *AI and Ethics*, *3*(1), 215–222.

Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., & Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5265–5274.

Wang, K., Chen, Z., Wu, Q. M. J., & Liu, C. (2017). Illumination and pose variable face recognition via adaptively weighted ULBP_MHOG and WSRC. *Signal Processing: Image Communication*, *58*, 175–186.

Xiong, C., Liu, L., Zhao, X., Yan, S., & Kim, T.-K. (2015). Convolutional fusion network for face verification in the wild. *IEEE Transactions on Circuits and Systems for Video Technology*, *26*(3), 517–528.

You, W., Zhang, H., & Zhao, X. (2020). A Siamese CNN for image steganalysis. *IEEE Transactions on Information Forensics and Security*, *16*, 291–306.

Zeghidour, N., Synnaeve, G., Versteegh, M., & Dupoux, E. (2016). A deep scattering spectrum—deep siamese network pipeline for unsupervised acoustic modeling. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4965–4969.

Zeng, D., Lin, Z., Yan, X., Liu, Y., Wang, F., & Tang, B. (2022). Face2exp: Combating data biases for facial expression recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20291–20300.

Zhang, J., Jin, X., Liu, Y., Sangaiah, A. K., & Wang, J. (2018). Small Sample Face Recognition Algorithm Based on Novel Siamese Network. *Journal of Information Processing Systems*, *14*(6).

Zhang, P., Ben, X., Jiang, W., Yan, R., & Zhang, Y. (2015). Coupled marginal discriminant mappings for low-resolution face recognition. *Optik*, *126*(23), 4352–4357.

Zhang, Y., Lu, Y., Wu, H., Wen, C., & Ge, C. (2016). Face occlusion detection using cascaded convolutional neural network. *Biometric Recognition: 11th Chinese Conference, CCBR 2016, Chengdu, China, October 14-16, 2016, Proceedings 11*, 720–727.

Zhang, Z., & Peng, H. (2019). Deeper and wider siamese networks for real-time visual tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 4586–4595. https://doi.org/10.1109/CVPR.2019.00472

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, *35*(4), 399–458.

Zhong, Y., Oh, S., & Moon, H. C. (2021). Service transformation under industry 4.0: Investigating acceptance of facial recognition payment through an extended technology acceptance model. *Technology in Society*, *64*, 101515.

Zhou, H., Mian, A., Wei, L., Creighton, D., Hossny, M., & Nahavandi, S. (2014). Recent advances on singlemodal and multimodal face recognition: a survey. *IEEE Transactions on Human-Machine Systems*, *44*(6), 701–716.

Zhuang, L., Yang, A. Y., Zhou, Z., Shankar Sastry, S., & Ma, Y. (2013). Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3546–3553.

Zulfiqar, M., Syed, F., Khan, M. J., & Khurshid, K. (2019). Deep face recognition for biometric authentication. *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 1–6.

CURRICULUM VITAE

Name Surname : DALHM GHALIB HALBOOS AL-SHAMMARI

EDUCATION:

- **Undergraduate** : 2020, Imam Ja,afar Al-Sadiq , Faculty of Engineering, Computer Engineering Department
- **Graduate** : 2024, Sakarya University, Computer and Information Engineering Department, Computer Engineering Program

PROFESSIONAL EXPERIENCE AND AWARDS:

Employee at the Iraqi Ministry of Oil / Oil Training Institute (2016)