T.C.

KIRSEHIR AHI EVRAN UNIVERSITY

INSTITUTE OF SCIENCES

DEPARTMENT OF ADVANCED TECHNOLOGIES

SIMPLE SECURITY APP USING REAL-TIME FACIAL RECOGNITION

ADNAN ABDULLAH YOUNUS
AL HAMMADI

MASTER'S THESIS

KIRŞEHİR / 2022

T.C.

KIRSEHIR AHI EVRAN UNIVERSITY

INSTITUTE OF SCIENCES

DEPARTMENT OF ADVANCED TECHNOLOGIES

SIMPLE SECURITY APP USING REAL-TIME FACIAL RECOGNITION

ADNAN ABDULLAH YOUNUS
AL HAMMADI

MASTER'S THESIS

Supervised by
Asst. Prof. Dr. MEMDUH KÖSE

KIRŞEHİR / 2022

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

ADNAN ABDULLAH YOUNUS AL HAMMADI

20.04.2016 tarihli Resmi Gazete'de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, Kırşehir Ahi Evran Üniversitesi'nin abonesi olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

ÖNSÖZ

Yüksek Lisansa / Doktoraya başlamamda ve yüksek lisans / doktora ders sürecinde kendisini tanıdığım günden bu yana gösterdiği sakin ve sabırlı hali ile her zaman bana örnek olmasının yanı sıra bir bilim adamının nasıl çalışması gerektiğini kendisinden öğrendiğim değerli danışmanım **Dr. Öğr. Üyesi Memduh KÖSE**' ye büyük bir içtenlikle teşekkür ederim. Tezimin her aşamasında gerek sorularımla gerekse alt ayda bir yapılan tez izleme komitesi sunumlarında tezin şekillenmesinde ve nihai hale gelmesinde katkıları olan değerli jüri üyelerim **Dr. Öğr. Üyesi Serkan KESER** ve **Dr. Öğr. Üyesi Filiz SARI** 'ya teşekkürlerimi içtenlikle sunarım.

Son olarak eğitim hayatım boyunca bana her türlü maddi ve manevi desteği sunan çok kıymetli aileme sonsuz şükranlarımı sunarım.

Aralık, 2022 ADNAN ABDULLAH YOUNUS AL HAMMADI

CONTENTS

Page No

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviations	Explanation
CNN	: Convolutional Neural Network
PSO	: Particle Swarm Optimization
ECG	: Electrocardiogram
ANN	: Artificial Neural Network
SIANN	: Space Invariant Artificial Neural Networks
PCA	: Principle Component Analysis
EP	:Evolutionary pursuit
NFL	:National Football League
VFR	:Video Face Recognition
TBE	:Trunk-Branch Ensemble
SVM	: Support vector machine
FRGC	: Face Recognition Grand Challenge
WCNNs	: Wasserstein convolutional neural networks
VIS	: Visible
NIR	: Near Infrared
MDS	: Multi Dimensional Scaling
CCA	: Canonical Correlation Analysis
2DNPE	: Two-Dimensional Neighbourhood Preserving Embedding
LVP	: Local Vector Pattern
BR	: Bilateral Recognition
CBFD	: Compact Binary Face Descriptor
PDVs	: Pixel Difference Vectors
LDA	: linear discriminant analysis
VC	: Computer Vision
RNN	: Recurrent Neural Network
LSTM	: Long Short-Term Memory

ÖZET

YÜKSEK LİSANS TEZİ

GERÇEK ZAMANLI YÜZ TANIMA KULLANAN BASİT GÜVENLİK UYGULAMASI

ADNAN ABDULLAH YOUNUS

Kırşehir Ahi Evran Üniversitesi

Fen Bilimleri Enstitüsü

İleri Teknolojiler Anabilim Dalı

Danışman: Dr. Öğr. Üyesi MEMDUH KÖSE

Yüz tanıma, güvenlik uygulamaları ve ticari uygulamalar da dahil olmak üzere birçok alanı ve disiplini kapsayan, görüntü analizi araştırma alanının amacına en uygun olan önemli bir araştırma problemidir. Yüz tanıma uygulamalarının farklı disiplinlerin konusu olması bu alanda yapılmış ve yapılacak çalışmalara olan ilgiyi artırmaktadır. Bu çalışmada, gerçek zamanlı yüz tanıma için derin evrişimli sinir ağı kullanan yeni bir yapı sunuyoruz. Önerilen yapı, gerçek zamanlı yüz tanıma gerçekleştirmek için düşük sayıda katman kullanacaktır. Önerilen uygulama amaç, transfer öğrenme tekniklerini kullanarak yeterli gerçek zamanlı performans elde etmektir. Düşük seviyeli sistemlerde düşük maliyetli uyarlanabilir güvenlik sistemlerinde bu uygulama gerçekleştirilebilir.

Aralık 2022, 41 Sayfa

Anahtar Kelimeler: Derin Öğrenme, Konvolüsyonel Sinir Ağları, Gerçek Zamanlı Yüz Tanıma.

ABSTRACT

M.Sc. THESIS

SIMPLE SECURITY APP USING REAL-TIME FACIAL RECOGNITION

ADNAN ABDULLAH YOUNUS AL HAMMADI

Kirsehir Ahi Evran University

Graduate School of Sciences and Engineering

Advanced Technologies Department

Supervisor: Asst. Prof. Dr. MEMDUH KÖSE

Facial recognition is an important research problem that best fits the purpose of the image analysis research field, covering many fields and disciplines, including security applications and commercial applications. The fact that face recognition applications are subject to different disciplines increases the interest in the studies that have been done and will be done in this field. In this work, we present a new structure that uses deep convolutional neural network for real-time face recognition. The proposed structure will use a low number of layers to perform real-time face recognition. The proposed application is to achieve sufficient real-time performance using transfer learning techniques. This application can be realized in low-cost adaptive security systems in low-level systems.

December 2022, 41 Pages

Keywords: Deep Learning, Convolutional Neural Networks (CNN), Real Time Face Recognition.

1. INTRODUCTION

It is obvious that the everyday output of data has greatly expanded in light of the recent technological developments that have taken place. Scientists have been given the opportunity to develop innovative facial recognition technology as a result of this growth, which coincides with an increase in both the processing power and speed of computers. Since the 1970s, face recognition technology has been utilized in the field of image processing, making it one of the most investigated fields of biometrics. The purpose of developing software for face recognition is to give computers vision capabilities that are comparable to human sight. This will allow computers to accurately detect and identify human faces that are contained within digital images. The establishment of these infrastructures is part of an endeavor to make the globe a more secure and hospitable place for people to live in.

The geometric method serves as the basis for the vast majority of standard practices. These methods are effective at identifying a person based on the facial characteristics that are most distinctive about them. In classical approaches, the Eigenfaces [1, 2] algorithms are some of the most well-known and frequently used ones. When there are few restrictions to work within, conventional approaches perform admirably; however, when there is more room for maneuver, they struggle. This is because specialists and researchers in these fields decide which facial characteristics to use in the analysis. As a direct consequence of this, the outcomes of the elections might be accurate or might only be partially accurate. In order to address problems of this nature, more trustworthy methods have been developed [3].

Traditional methods of facial recognition have been displaced in recent years by modern technologies, which have recently gained favor as a result of the restrictions outlined above. Deep learning approaches serve as the foundation for modern procedures. Unlike more traditional methods, which need manual selection, the features or embeddings used in a deep learning model are chosen automatically throughout the training process. The ability of deep learning methods to be trained on incredibly large data sets in order to understand how to derive the most relevant attributes is its fundamental advantage. Deep learning techniques, in which students learn on enormous data sets by experimenting with a variety of various training methodologies, are responsible for the great levels of accuracy that these techniques may achieve. As a result, we were inspired to create our model with the help of deep learning's Convolutional Neural Networks (CNN) [4].

The accuracy and effectiveness of deep learning models are directly proportional to the size of the data sets that are used to train them. At this point in time, the most advanced deep learning-based face recognition systems have been trained using millions of photographs. The generation of such massive data sets from scratch is an extremely difficult task. Researchers in this situation have access to a wide variety of tools for improving the quality of their data as well as helpful models [5].

The most secure form of system is one that is based on the recognition of biometric features. One of the reasons why biometric systems are so useful in this scenario is that the security parameter employed by these systems is unique to each user and cannot be stolen or copied. Today's most common biometric recognition systems include fingerprinting, vein scanning, and facial recognition. However, there are numerous biometric recognition technologies in use today. Similarly, to how vein recognition involves the use of specialized equipment, fingerprint recognition necessitates the use of costly and specialist technology in order to obtain images. In order for this device to receive photos, the user must establish physical contact with it. When people are forced to spend extended periods of time in close quarters with one another, germs and diseases have a larger possibility of spreading. This raises the likelihood of people being ill. Users are not voluntarily using these systems because of these and other comparable issues. Cameras are required to accomplish the recognition activities required by biometric facial recognition systems. Because cameras have grown so common in modern life, most personal computers now include built-in cameras as standard features. There is no longer any need to purchase additional gear in order to use facial recognition technology because all modern computers, tablets, and smartphones already include a camera. Furthermore, customers should not be concerned about their health because no physical touch is required [6].

1.1. The purpose of the Face Recognition Systems

One of the most typical obstacles that occurs when deploying face recognition algorithms is that the elements that need to be matched or recognized in real-world applications frequently come from a range of locations. The process of searching entails numerous processes, one of which is the extraction of information from an unknown environment. To conduct a successful search for an individual's identification across many environments, the learning data set must include information about cross-context associations than individuals with similar genetic traits, the identity of the confirmed technique used is far from the genetically search and gallery IDs, the gallery being searched is big, and the identity of the verified technique used is not in the learning set Furthermore, the gallery being searched has a significant number of photographs. When capturing photographs in the real world, the conditions, light, and chromaticity levels vary greatly. Furthermore, the camera's image modifications for the face, hair, and hat that cover the facial features of identities can be rather significant.

The recently discovered image processing technology for identifying people based on their faces has a wide range of possible uses. Each person has a distinct face structure. Engineers used this data to create face recognition technology. It streamlines the process of identifying people by emphasizing the characteristics that are unique to each individual. Our camera network uses cutting-edge surveillance technology, and our biometric access control system ensures that a secure atmosphere is maintained in any setting. In comparison to other biometric systems, this one has proven to offer the best level of dependability because it is based on an established method. A facial recognition system stands out not just for the countless benefits it delivers, but also for the one-of-a-kind traits it provides.

Businesses will be able to control who enters and exits the facility at specific times thanks to the development of facial recognition technology. After the device has been deployed, an image of each individual who is allowed to pass is captured in three dimensions and then transferred to the computer system.

Furthermore, any personally identifiable information about the subjects of the images will be saved by the system. The device's memory makes it straightforward to obtain all workers' profiles, including any accompanying photos. During the handoff, the device recognizes and compares the facial characteristics of any personnel who come into touch with it to the information recorded in its memory. When it achieves the required data, the door will be unlocked. Nonetheless, if there is a discrepancy in the information, the individual's access is restricted. This strategy prevents intruders from entering the premises.

Face recognition is presently the fastest biometric authentication method available. It is a sophisticated technology that verifies a person based on their appearance. As a result, it has become a device that is gaining popularity as a means of attracting attention to itself. Problems like misplacing a card or forgetting a password can be avoided by using this strategy. Using facial recognition technology for staff management has grown increasingly widespread in recent years. In the future, these systems will be appropriate for a wide range of varied applications. It is commonly used in public places such as classrooms, cafeterias, and dining halls, as well as libraries, museums, municipal halls, and the military. Using a technology capable of facial recognition has a lot of advantages. These devices rely on people' unique facial algorithms to function effectively. Each individual's face is represented by a unique algorithmic property. These gadgets operate by applying an algorithm to different regions of a person's face, most notably the forehead, nose, and chin. Because this algorithm acts slightly differently on each face, it is difficult for two persons to substitute for one another in a face reading scenario. This is by far the most significant advantage of adopting facial recognition technology.

While face recognition technology is most typically used for employee tracking, it is also increasingly being used in other contexts such as the cafeteria, the membership roster, the student body, the bathroom, and so on. Another example is. The ringing function of the face recognition system is simply one of many valuable functions. The bell-ringing feature is particularly beneficial in manufacturing environments where workers must move as a group. You may also ring the bell during scheduled breaks. The built-in battery life is just one of the benefits of adopting face recognition technology. Because they feature backup batteries, they can be used even when the power goes out. Furthermore, the storage capabilities of these devices range from thousands to tens of thousands of bytes. That is, it makes no difference how many individuals are present in the area you wish to use.

1.2. Overview

Because of the importance of having access to high-quality information, sophisticated database management systems are in high demand. The development of today's information systems relies heavily on automatic image recognition.

Humans and other animals have a fundamental feature in that we can recognize one another. A image is simply a description of what is being displayed. In certain ways, it outperforms the human visual recognition system, which is a complicated computer in its own right. The recognition process is divided into two main categories, each of which corresponds to a separate set of visual features. The first of these is the ability to recognize solid objects. This falls under the umbrella term "sensory recognition," and it includes both visual and aural image recognition. Picture recognition, identification, spatial classification, and transitory images are examples of this. Text, fingerprints, real-world objects, and photographs are examples of spatial images. There is a link between snapshots, voice waves, ECG waves, time series, and target signatures.

The ability to detect abstract objects is the second skill to employ. An existing argument or solution to the problem, on the other hand, is easily identified. This practice of recognizing abstract objects is known as conceptual recognition. Human recognition of concrete images may be characterized as a psychophysiological challenge, requiring interactions between the individual and the external information. Recognizing anything necessitates the use of prior knowledge of the input data as well as the context in which it was presented to us, both of which are based on previous experiences. Priority information is used to calculate relative probabilities that can be linked to a group of known statistical individuals. As a result, the problem with picture recognition is a lack of variance or individuality in the entering data from various populations. We may view it as a discriminating issue through the lens of attribute discovery.

1.3. Outline of Thesis

The thesis is divided into five parts as follows:

The first part contains a general description of the aims, objectives and contribution of the thesis. In the second part, information about the theoretical information about the subject is given. In the third chapter, after the methodology of the thesis is written, the results of the chapter are included. In the fifth chapter, the conclusion of the thesis is written.

2. BASIC CONCEPTS

2.1. Face Recognition

Within the effects given to us by fourth-generation information technology [7], which forecasts that users will communicate with computers rather than humans, there is evidence of an increase in studies on the human aspect in our daily lives. Personal recognition and identification are at the heart of the developing smart systems' interactive content. Human error, password theft, and password expiration are just a few of the issues that plague these systems and have spurred researchers to focus more heavily on biometric solutions.

The term "biometrics" originates from the Greek terms "bios" (life) and "metrics" (measurement). Biometrics, in this context, refers to methods of authentication based on the analysis of observable human biological or behavioral traits. Fingerprinting, hand and finger geometry, facial recognition, iris and retinal scanning, and vascular pattern identification are all examples of physical biometric procedures. Speaker and voice recognition, as well as signature verification, are under the category of behavioral biometrics. With the rise of security issues in modern technology and the proliferation of biometric recognition systems, facial recognition systems have become increasingly important in the 21st century.

Researchers have given facial recognition a high priority under pattern recognition since it is considered to be the most efficient and cost-effective method.

In recent years, face recognition technologies have seen widespread use. Its use has increased, especially in high-security settings (airports, police offices, banks, sports fields, business entry and exit monitoring of corporate companies). Researchers have produced a plethora of methods, and many different solutions have been proposed to the issue of ambient variation, which is a major challenge for facial recognition technologies. The evolution and current applications of facial recognition technology are covered in the next section.

2.2. History of Face Recognition Systems

Scientists' theoretical and practical interest in face recognition systems has been growing in tandem with the development of computer vision. In spite of the widespread adoption of other biometric authentication methods (fingerprint scanners, iris scanners, etc.), researchers have focused on perfecting facial recognition. Since its inception in the early 1960s, the field of computer facial recognition has advanced significantly. It has been suggested that we can use our faces' individual traits to solve certain problems. As the first of its sort, Matthew Turk and Alex Pentland's "Eigen faces" approach was created in 1987. The most widely used methods of facial recognition are as follows:

2.2.1. Neural Network

Neural networks are a set of algorithms that attempt to replicate the way the human brain works in order to discover hidden patterns and connections in data. Neuronal networks can refer to biological or synthetically created networks of neurons. Since neural networks are flexible, they can produce the greatest possible outcome even if the input criterion is modified. The use of neural networks, a notion with origins in AI, is becoming increasingly common in the creation of new trading systems [7].

2.2.2. Convolutional Neural Network (CNN)

Convolutional neural networks (CNNs) are a type of artificial neural network (ANN) that is widely utilized in the field of deep learning for image processing. CNN's are also known as Shift Invariant or Space Invariant Artificial Neural Networks because they are based on the shared-weight architecture of convolution kernels or filters that slide along input features and construct translation-equivariant outputs known as feature maps (SIANN). CNN's substitute normal matrix multiplication with a mathematical process known as a convolution in at least one of their layers. They are optimized for processing pixel data for the purposes of image recognition and processing [7].

2.2.3. Deep Sparse Auto-Encoders

A sparse auto-encoder can be thought of as an artificial neural network that uses the same unsupervised machine learning concepts as many other similar systems. The use of auto-encoders, a type of deep network, is based on the principles of unsupervised machine learning. Furthermore, it can be utilized for dimensionality reduction and model reconstruction via backpropagation. When compared to a Neural Network, it is easier to understand, put into practice, and justify in one's own mind [8].

2.2.4. Support Vector Machine

It is a supervised machine learning algorithm that can do classification or regression. Structure, data requirements, training duration, algorithm iterations, parameter optimization, and sensitivity to initial randomization of weights differ from Neural Networks [8].

2.2.5. Principle Component Analysis (PCA)

It can function as a statistical tool within a neural network. We can significantly reduce the dimensionality of a data set by paying close attention to its underlying dependencies [9].

2.2.6. Evolutionary Pursuit (EP)

Evolutionary Pursuit (EP) is worth investigating as a cutting-edge and adaptable technique to image encoding and categorization. The difficulty for EP, which tries to increase the learning machine's generalization skills, is finding a compromise between minimizing the empirical risk faced during training and narrowing the confidence interval for reducing the promised risk for future testing on unseen images [9].

In the Mexican elections of 2000, a mechanism was put in place to prevent duplicate voting by using face recognition technology. In addition, the final game of the NFL (National Football League) in the United States in 2001 resulted in the arrest of 19 offenders. Driver's licenses and identification cards issued by the United States' government use a facial recognition technique to ensure that only one person is listed on each card. The fundamental configuration of a face recognition system is depicted in Figure 2.1.

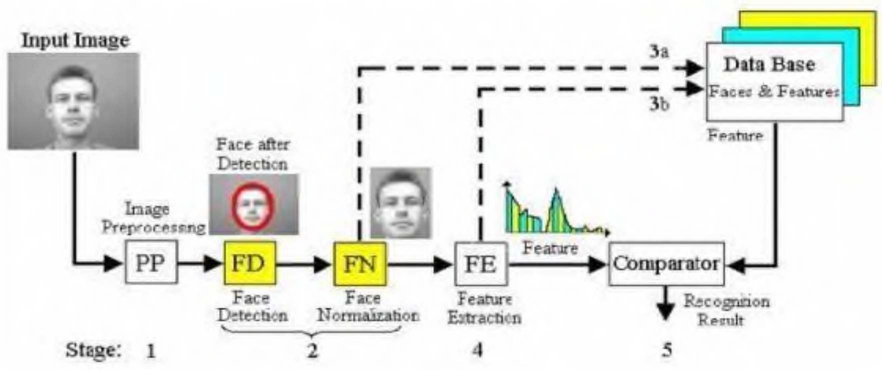

Figure 2.1. Structure of face recognition system [10]

2.3. Literature Review

To solve the challenges of video face identification, the authors of [7] provide a comprehensive framework based on CNN (VFR). To begin, we address a gap in available real-world video training data by artificially blurring training data composed of clear still pictures in order to build blur-resistant face representations. During training, CNN is taught to ignore blur automatically by being exposed to both unaltered and artificially blurred material. Second, we introduce a Trunk-Branch Ensemble CNN (TBE-CNN) model that extracts supplementary information from holistic face pictures and patches trims around facial components to improve CNN feature resilience to position changes and occlusion. TBE-CNN is an end-to-end model that maximizes feature extraction by combining the power of the trunk and branch networks at the lower and intermediate levels. Finally, we propose a modified triplet loss function to improve TBE-learned CNN representations' discriminative capabilities. Systematic experiments have confirmed the efficiency of the indicated strategies. TBE-CNN outperforms state-of-the-art algorithms on the PaSC, COX Face, and YouTube Faces video face databases in particular. The supplied approaches also resulted in first place in the BTAS 2016 Video Person Recognition Evaluation.

The author models high-resolution heterogeneous face generation in [9] as a synergistic combination of two sub-components: texture inpainting and position correction. The inpainting module uses synthesis and inpainting to transform NIR picture textures to VIS ones. By translating any orientation in the NIR images to a frontal orientation in the VIS images, the corrected images pair NIR and VIS textures. A warping procedure is used to combine the two components into a complete deep network. Both a fine-grained discriminator and a waveletbased discriminator seek to improve image quality. To ensure synthesis results, a unique 3Dbased posture correction loss, two adversarial losses, and a pixel loss are used. We demonstrate that by including the correction component, the process of heterogeneous face synthesis can be optimized from a one-to-many unpaired image translation to a one-to-one paired image translation, lowering spectral and postural disparity and maximizing recognition accuracy. Extensive experimental results show that, in addition to providing high-resolution VIS face images, our network also aids in the improvement of heterogeneous face recognition accuracy.

The proposed method in [10] may cause image distortion during processing and produces poor results in natural light. We present an innovative way to resolve the issue of illumination in facial recognition in this paper. To begin, we calculate a score based on the percentage difference between the query input and the people in the gallery classes for the top and second most similar. Based on the score, the recognition is subsequently applied to the most appropriate photographs (raw or edited). Experiments on the ORL, CMU-PIE, and Extended Yale B face databases show that our adaptive method outperforms the sum and maximum of similarities, two common fusion operators.

The author discusses the implications of recent improvements in automated face recognition for the field of forensic face recognition [11]. Advances in forensic face identification brought about by studies of facial ageing, facial marks, forensic sketch recognition, face recognition in video, near-infrared face recognition, and soft biometrics will be presented. Finally, we suggest future research areas for the use of facial recognition technology in forensics and examine the field's existing limits.

The author proposes a new method for learning invariant features between NIR and VIS facial photos using Wasserstein convolutional neural networks (WCNNs) in [12]. The lower layers of the WCNN are taught using publicly available VIS face images, while the upper layer is divided into three sections: an NIR layer, a VIS layer, and a shared NIR/VIS layer. The first two layers are dedicated to learning features that are unique to a given modality, whereas the NIR-VIS shared layer is designed to learn a modality-independent feature subspace. The Wasserstein distance is implemented in the NIR-VIS common layer to quantify the dissimilarity between disparate feature distributions. We use W-CNN learning to create deep feature representations of diverse faces that are consistent across all of them. This technique involves minimizing the Wasserstein distance between the NIR and VIS distributions. This paper introduces a correlation prior on fully-connected WCNN layers to reduce the size of the parameter space and avoid over-fitting on small-scale heterogeneous face data. This prior works in practice due to a low-rank limitation in a fully connected network. During the testing phase, the joint formulation results in efficient computing for heterogeneous data, as well as alternating minimization for deep feature representation during the training phase. The WCNN method outperforms state-of-the-art methods in extensive trials using three difficult NIR-VIS face recognition databases.

The authors of [13] analyze the efficacy of computational algorithms as models of human face processing by comparing how computers and humans interpret facial features. We compared model- and human-generated estimations of the similarity between pairs of faces to assess the congruence between several automatic facial recognition systems and human perceivers. We employed multidimensional scaling to illustrate the subjects' reaction patterns (MDS). Finally, the model's responses were projected into this domain. The results revealed a shared bimodal structure among the subjects and the majority of the models. The bimodal subject structure revealed differences in decision-making methods evaluating similarity. The bimodal structure of the models was linked to a variety of elements, including the representations themselves and the distance measurements actually used.

In [14], the author creates a framework for face tracking that makes use of an offline detector, an online tracker, and an online recognizer. We focus on the recognizer as the central part of our framework because it is responsible for selecting the target face from the set of detected and tracked faces with the highest degree of confidence. Accurate recognition is especially challenging in surveillance circumstances due to the lack of available prior information about the identities and the frequent changes in facial positions, necessitating an online formulation. We use the Canonical Correlation Analysis (CCA) technique to project the extracted characteristics of faces to a latent space, and then we use incremental online support vector machine (SVM) training to improve the accuracy of identity identification across a wide range of poses (LASVM). Compared to other state-of-the-art face trackers, our person-specific face tracking performs far better in experiments.

In [15], the authors provide a novel approach to extracting the characteristics for face recognition; they term it two-dimensional neighbourhood preserving embedding (2DNPE). Extensive tests are conducted utilizing ORL and the Yale face database to test and assess the novel approach. According to the results of the experiments, the 2DNPE approach is more efficient and produces superior results when it comes to facial recognition.

In [16], the author reveals an innovative two-step process for identifying people's faces. In the first phase, we leverage a recently suggested feature called Local Vector Pattern (LVP) and combine it with a weighting method to calculate the distance between an input face image and each enrolled face image, ultimately narrowing down the pool of candidates from M to a single winner. To get to the final face recognition result from among the M candidates, the second stage employs a new feature-point Bilateral Recognition (BR) technique. The forward and backward recognition components of bilateral recognition each use the feature points discovered in their respective reference images to independently search for the equivalent feature points in the matched images. Matching and detection of feature points are used to create geometrical models that are then compared. The final recognition result is based on a bilateral recognition score, which is calculated by adding the forward and backward recognition scores. Based on experimental results on the popular Feret face databases, it is clear that the suggested algorithm produces a high-quality recognition result and outperforms two other well-known face recognition methods.

The author of [17] describes a system for automatic 3D face identification that can distinguish expression deformations from interpersonal differences and recognize faces with no expressions at all. The training set is used to learn expression distortions and interpersonal diversity distortions. The deformations are then linearly combined to create a new face with the desired expression. By adjusting the linear combination coefficients, the synthesized face can be matched to a target face when it is received. Following the matching procedure, coefficients indicating interpersonal differences are chosen as features for recognition. Our experiments on the FRGC v2.0 database yielded good results.

The suggested face recognition procedure in [18] uses a combination of the Haar Cascades and Eigenface algorithms, which allows for simultaneous detection of up to 55 different faces. This refined method of facial recognition successfully identified several faces

The authors of [19] propose a compact binary face descriptor (CBFD) feature learning approach for facial identification and representation. To extract pixel difference vectors (PDVs) in local patches, we first compute the difference between each pixel and its neighbouring pixels in each face image. We next unsupervised learn a feature mapping to project these PDVs into low dimensional binary vectors, with the following aims in mind: 1) maximize the variance of all binary codes in the training set; 2) minimize the loss between the original real-valued codes and the learned binary codes; and 3) uniformly distribute the binary codes across each learnt bin, resulting in compact binary codes by removing redundancy in PDVs. Finally, we merge all of these binary digits into a single histogram feature to represent each facial image. We also present a coupled CBFD (C-CBFD) method for narrowing the modality gap of heterogeneous faces at the feature level to make our approach effective for heterogeneous face identification. Extensive experimental results on five popular face datasets show that our strategies exceed state-of-the-art face descriptors.

There is a proposal for a multi-modal face recognition algorithm in [20]. This person is then identified by a new feature Matrix that was reconstructed by a portion of the feature vector of each two-dimensional facial image following feature extraction using 2DPCA and dimensional reduction. Based on CAS-PEAL Face Database experiments, this strategy achieves a greater recognition rate than methods that rely on a single positive face under the same conditions.

As a first step, the author's framework in [21] transforms the pose-invariant face identification problem into a partial frontal face recognition problem. Following that, we develop a dependable patch-based face representation system for use with the generated partial frontal faces. A transformation dictionary is learned for each patch using the suggested multi-task learning technique. We can use the transformation dictionary to map the properties of different poses onto a more discriminative metric space. Finally, rather than being done system-wide, face matching is done patch by patch. Extensive and systematic trials on the FERET, CMUPIE, and Multi-PIE databases show that the proposed method consistently outperforms single taskbased baselines and state-of-the-art pose problem solutions. We are able to achieve state of the art performance on the tough LFW data set by expanding on a previously presented technique, making it a good candidate for use in unconstrained face verification applications.

The author presented a novel method for determining weights for the weighted sum rule in [22]. A comparable novel way to determine consequences for the weighted sum rule is proposed in this paper. It is demonstrated that recognition accuracy can be improved, and the advantage of decision fusion over feature fusion is advocated. We achieve much better results in the challenging AR database than standard and feature fusion methods using a validation accuracy weighting scheme and the nearest-neighbour discriminant analysis dimension reduction method. This is due to decision fusion's ability to merge different features into a single result.

Because of conflicts in their goals, the author of [23] employs methodologies in which feature selection and subspace learning are not ideally coupled. In this paper, we provide a new and effective feature selection method for linear discriminant analysis (LDA). We use the Fisher criterion to identify the most discriminative and relevant characteristics to improve face recognition performance. This ensures that the feature selection and classifier learning goals are in sync (both follow the Fisher criterion). The efficiency of the suggested strategy has been demonstrated experimentally on the FRGC v2.0 face database.

3. MATERIAL AND METHOD

3.1. Areas Image Processing

Medical imaging and imaging enhancement techniques such as computed tomography, ultraviolet light, gamma rays, x-rays, and the microscope; cardiography; and scintigraphy are used. Automation study involves car proximity sensors, autonomous automobiles that respond to traffic lights, and line-following robots to eliminate background noise from historical images. Traffic speed detectors, license plate readers, and live TV broadcast compression techniques are just a few examples. For use in space research utilizing satellite and microwave radar imaging. Facial recognition software for smartphones, tablets, and PCs is now being actively developed. In the corporate world and the military industry, picture processing techniques are used in the iris, fingerprint, and face recognition devices and security camera applications for target tracking, object detection, face recognition, image enhancement, and fingerprint recognition, respectively. Many other fields can be used as examples, including physics, art, animal husbandry, and agriculture [24].

3.2. Face Recognition in Image Processing

A human brain finds it easy to recognize a person's face. Experiments have shown that babies as early as three days old can differentiate their own faces from those of others. It can't be too difficult for a computer, can it? There is clearly still a lot we don't know about human recognition. To what extent do external (head shape, hairline) or internal (eyes, nose, mouth) characteristics influence accurate facial recognition? How does the human brain process visual information, and how do we inspect images? Individual neurons in the human brain respond to specific, local aspects of an environment, such as lines, edges, angles, and motion, according to David Hubel and Torsten Wiesel. Because we don't see the world as a collection of unrelated bits, the visual brain must figure out how to synthesize data from multiple sources into meaningful patterns. The ability to extract significant information from an image, translate it into a usable format and classify it is one of the most important parts of automatic face recognition. Geometric features may allow for more accurate facial identification. This is the most natural technique of recognition. In this post, we will look at a feature vector, which was one of the first types of automatic facial recognition technology (distance between points, angle between them , ...). To conduct the recognition, the Euclidean distance between the feature vectors of the probe and the reference image was used. The method is inherently robust to fluctuations in illumination, but it comes at a hefty expense. This method has a disadvantage in that exact registration of marker points is difficult even with powerful algorithms. Geometric facial recognition has made some progress in recent years. Despite the use of a 22-dimensional feature vector, experiments on enormous data sets demonstrated that geometric features do not provide enough information for face recognition [25].

The given Principal Component Analysis method approaches face recognition from a more holistic perspective, considering a face image as a point within a higher dimensional image space and offering a reduced representation of that space in which classification is simplified. By identifying the most variable axis, Principal Component Analysis is utilized to determine the subspace with the fewest dimensions. This reform focuses on organizational restructuring. While ideal, it excludes all categories. Consider a situation in which external causes contribute to variance. Because the most different axes do not have to have any identifying information on them, classification is difficult if they have the greatest variation. As a result, we employed Linear Discriminant Analysis to create class-based projections. The fundamental goal is to maximize dissimilarity between classes while limiting dissimilarity inside a class. Recently, various strategies for extracting local features have evolved. Because only a small portion of a picture must be represented in order to reduce the high dimensionality of the input data, the extracted features should be less subject to difficulties such as partial occlusion, poor lighting, and data scarcity. For local feature extraction, three techniques are used: Gabor Wavelets, the Discrete Cosine Transform, and Local Binary Patterns. Because spatial information has the potential to be beneficial, the ideal strategy for preserving it when using a local feature extraction is now under investigation [26].

3.3. Deep Learning

Deep learning, a subfield of machine learning, is built on artificial neural networks. We choose to focus on Computer Vision to make our deep learning discussion relevant to the thesis topic. As we've seen, it's vital to strip faces of their faces or features in order to solve the problem of face recognition. The same is true for visual object recognition. The human face has many identifying features, but it is also possible to learn about other aspects of a person's or thing's look by analyzing images or films of them. Manual removal is a time-consuming and demanding operation. As a result, different methods have been developed to assist in dealing with this issue [27].

These high-level features make it easy for people to identify and locate everyday objects such as autos, homes, buses, dogs, and so on. The car, for example, has four wheels, is built of metal, and has a windscreen. This level of image recognition is difficult for a machine to attain. To summarize, the current options cannot adequately address the status quo concerns. Deep learning can successfully extract these more sophisticated data from photos and videos despite the numerous obstacles they bring. Deep learning, in its most basic form, uses neural networks with hidden layers to learn. There could be tens, hundreds, thousands, or even millions of layers. When an image is sent into a deep learning network, the network combines basic information such as edges, shapes, and corners to produce complex images such as faces, vehicles, tables, and so on. Capable of learning new skills, such as object identification. The image is transmitted from layer to layer in a deep neural network, and at each stage, the network learns a single simple attribute about the image before utilizing that feature to forecast the object in the image. The name "deep learning" comes from the greater feature extraction of deep layers [28]. The deep learning process is depicted graphically in Figure 3.1.

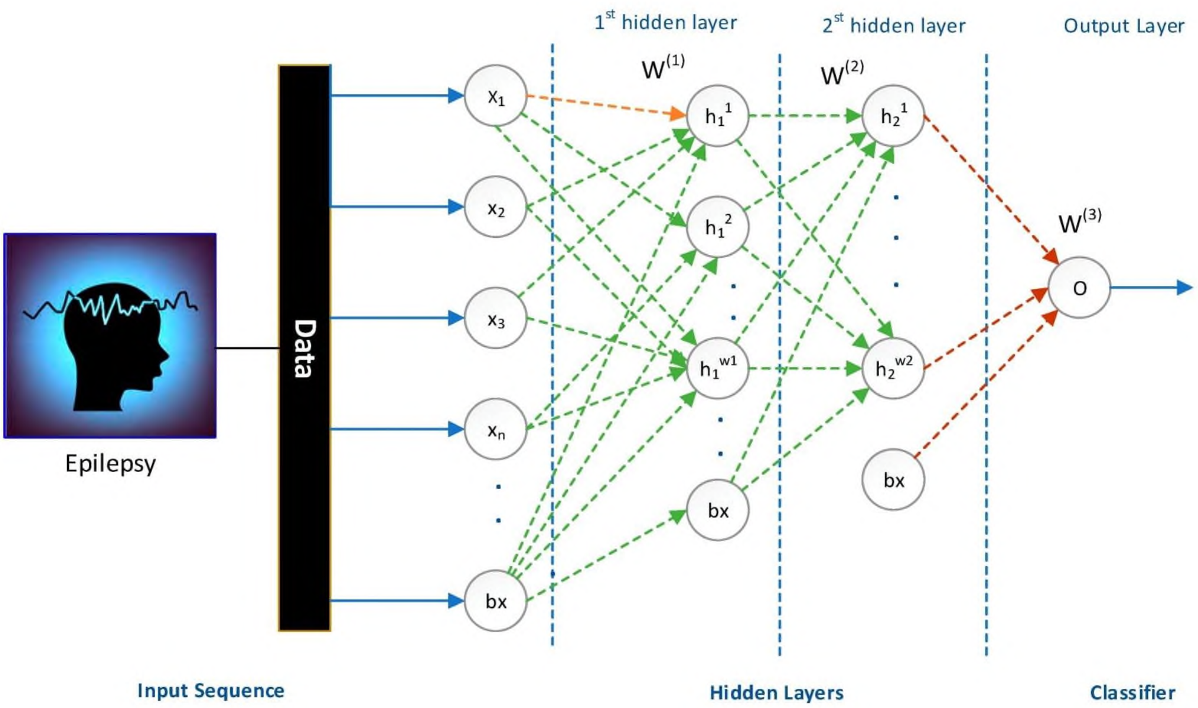

Figure 3.1. Deep learning for image recognition [28].

The DNN are a popular variant of the ANN, with the DNN structure containing more hidden layers than a standard neural network. The hidden layer's complexity is optimized for the particular problem. The DNNs typically employ multi-layer perception networks with at least three hidden layers that have been trained via supervised learning. As the number of hidden layers in such systems increases, so do the weight and neuron count. The vast amount of data and processing power required to train existing networks is directly related to the network's complicated structure and nodes. As we will learn in Chapter 4, insufficient computer resources prevent us from training a DNN from the beginning. Both ANNs and the Backpropagation approach have a long history. However, in the new millennium, Adequate training data was sparse prior to the emergence of the internet and social media. There were also not enough powerful computers available for ANN training. The good news come in 2006. Hinton's (2006) research demonstrates how to train an ANN by treating each backpropagation layer as a Constrained Boltzmann machine. Despite previous studies on the issue of ANN training, this study attracted renewed interest in the field and its prospective applications. The advent of GPUs, which can process both data and graphics, has resulted in an increase in the number of entertaining lessons. As a result, novel techniques have emerged [29].

3.3.1. Convolutional Neural Network

The CNNs, unlike DNNs, are built on the same multilayer perceptron architecture. CNN is commonly used in the field of Computer Vision (CV) for problem-solving and detecting objects in pictures. CNN is controlled by the visual cortex in both animals and humans. It's a doppelganger. Recently, successful applications of pre-existing networks have been discovered in disciplines such as computer vision and image processing. CNN was dubbed "convolutional" because it employs a linear process in mathematics known as "convolution.. Figure 3.2 displays the structure of a typical CNN, which is made up of many layers that all work together to establish what a picture should be. The initial stage in the process is to use an image as input. Neurons in the first hidden layer may detect fundamental shapes and patterns such as facial borders and the profile of an ear or nose. The image is subsequently passed on to a deeper layer with more granular learning capabilities. As the number of layers rises, neurons begin to learn complex items such as faces and cars. CNN, on the other hand, operates on a somewhat different notion than standard ANN. A single neuron in a layer will focus on a specific component of the image, while other neurons in the layer will search the entire image for that and similar features. It employs a distinct set of neurons for this function [30]. We can see in Figure 3.2 how the experimental CNN infrastructure model was conceived. The layers in this architecture are as follows: input, convolution1, pooling1, convolution2, pooling2, completely connected, and Softmax regression classification. Figure 3.2 shows that when comparing CNNs to more conventional neural networks, the addition of convolutional and pooling layers for feature extraction is the most notable distinction. These two levels are presented in this work.

3.3.1.1. Network Structure

The CNN infrastructure model designed in this experiment is shown in Figure 1. It includes the input layer, convolution layer 1, pooling layer 1, convolution layer 2, pooling layer 2, fully connected layer and Softmax regression classification layer. we find that the main difference between CNN and other traditional neural networks is that the processing of convolutional and pooling layers is added to extract image features. .

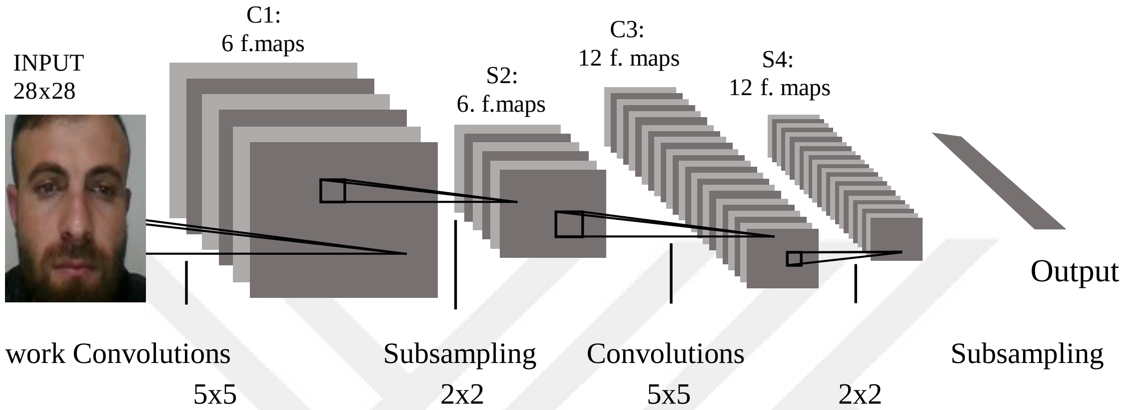

Figure 3.2. CNN Structure [35]

3.3.1.2. Convolutional Layer

Layers of components extracted from the input image are represented by the convolutional layer, which is also known as the feature extraction layer. Each neuron's input is tied into the preceding layer's local receptive field, allowing for the extraction of locally- layer signifies that the original image is convolutionally operated upon to produce a set of derivative images. Figure 3.2 depicts the convolution procedure using formula (35). The target image is placed second from the left after the convolution kernel value is arbitrarily chosen. Choose the picture matrices from left to right, top to bottom, and then multiply the result by the convolution kernel. Each convolution acts like a sieve to remove unsuitable areas of a picture throughout the feature extraction process. The greater the activation value, the greater the degree to which the screening conditions are satisfied. Convolution, like the filter effect, helps us extract numerous information from the input. Because convolution does not change the image's size, the treated image remains the same dimensions. Convolution process mathematical expression [35] .

The mathematical expression of the convolution process [35]

[redacted] (3.1)

Where n_in is the number of input matrices or the dimension of the last dimension of the tensor, b is the offset of the neurons on the convolutional layer feature map, X_k represents the kth input matrix, and W_k represents the convolution kernel k sub-convolution kernel matrices, s(i,j) is the value of the corresponding position element of the output matrix corresponding to the convolution kernel W.

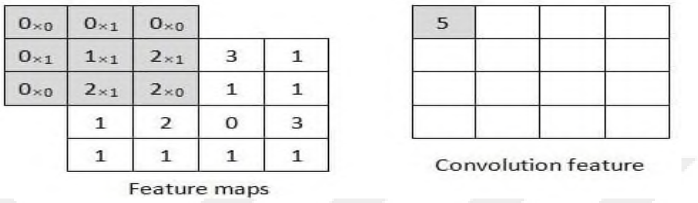

Figure 3.3 Pooling layer convolution operation process

Convolutional layers are typically followed by a pooling layer [31]. The preceding layer's image stacks and parameters will be compressed in this one. The issue of network incompatibility is so mitigated. On this layer, you can make a 23 or 33 frame that will float over the image at a predetermined step size. Each time through, the minimum and maximum values are preserved while the rest are eliminated. The processed image will be roughly 75% of the original's size. Miniature representations of the network shorten the learning curve and lessen the likelihood of memorization.

3.3.1.3. Fully Connected Layer

Every neuron in this layer communicates with every neuron in the layer underneath it. It is the final section of the ESA design and, in all cases, serves as the output layer [32]. The equation in describes a full-layer connection (3.2).

$$E(W) = \left[\sum_{i=1}^{m}\sum_{r=1}^{N} 1\{y^{(i)} = r\} \, log \, \frac{\exp\left(w_r^T x^{(i)}\right)}{\sum_{j=1}^{N} \exp(w_r^T x^{(i)})}\right] \qquad (3.2)$$

3.3.2. Long-Short Term Memory Model

A Neural Network uses a series of algorithms to uncover patterns similar to those found in the human brain. They are skilled at detecting numerical patterns in vectors that must be converted into quantifiable data. They make sense of a machine's raw input in the form of sensory data by using perception, labeling, and grouping. An artificial neural network is made up of several closely connected computer components known as neurons. They work together to discover a solution. A Recurrent Neural Network (RNN) is a sort of feed forward neural network that is a generalization with built-in memory. The previous process's output is used as an input for the current one. RNNs use this memory state to process input sequences. This restricts their application to certain domains, such as unsegmented, linked handwriting recognition or emotion identification. To generate the output, all of the hidden layers are trained with the same set of parameters. In comparison to other neural networks, this simplifies the parameters [33].

RNNs cannot process long sequences when using the tanh activation function. A recurrent neural network is also difficult to train. To overcome this issue, researchers created Long ShortTerm Memory Networks, a type of RNN. The Long Short-Term Memory (LSTM) network is a sort of repeating neural network that allows for the recovery of previously stored data. They excel at time series classification, processing, and prediction despite different time lags. Backpropagation is used to train the model. LSTMs are beneficial for limiting the amount of error that can be back propagated through time and layers. They enable recurrent networks to learn across multiple time steps by keeping the error relatively steady. This opens the door to establishing distant links between causes and their effects. The gated cell in an LSTM contains data that a recurrent network would not ordinarily handle. A cell, like a computer's memory, may be written to, stored in, and read from. Cell gates govern what is stored and when reading, writing, and deleting are permitted. These gates, on the other hand, are analog and operate by sigmoid multiplication. LSTM models profit from the fact that they do not require any preconditions, such as making a data set stationery.

While neural networks may simulate nonlinear processes, they require a large amount of input data to do so. The essential structure of both forms of recurrent neural networks is a succession of repeating modules in a neural network. Regular RNNs, such as a single tanh layer, will have a different structure than LSTMs, which will have many hidden layers. Within this repeating module, this differentiation will exist. There are now four layers in a neural network, rather than just one highly communicating layer. Data is either permitted to pass through the gates, which are modeled after the neural network's sigmoidal layer, or it is discarded [34].

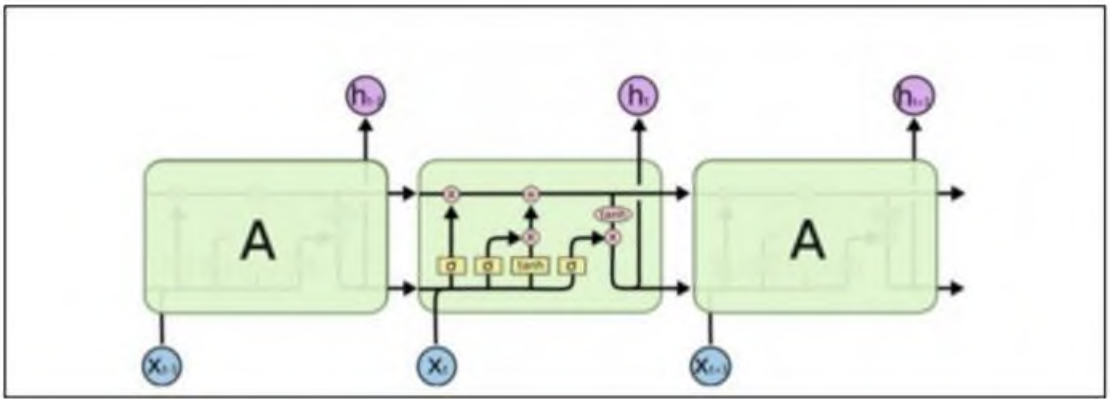

Figure 3.4. Structure of long-short term memory model4

As the sigmoid layer thickens, the range of data that may be stored in each cell increases from zero to one. A prediction with a value of zero usually means "let nothing go through," whereas a prediction with a value of one means "allow everything to go through." Forget about Gate. Every LSTM comprises three types of gates, all of which are concerned with monitoring the current state of every cell in the network. The resulting range is between zero and one, with zero indicating "completely ignore this scenario" and one indicating "definitely hold this."

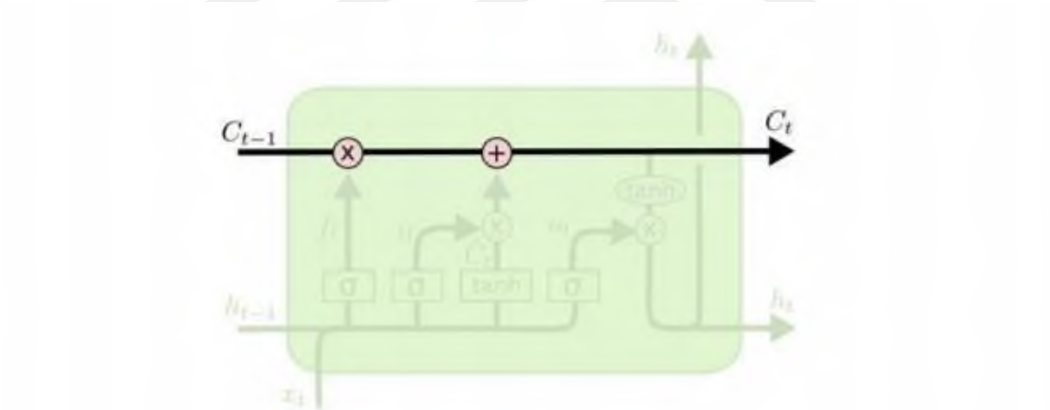

Figure 3.5. Long short memory cell structure.

Memory Gate is in charge of determining what fresh information should be put into the cell's storage. A sigmoid and a tanh layer are used in this method. The first sigmoid layer, often called the "door layer entry," is responsible for determining which variables are adjusted. A tanh layer, which comes before this phase, generates a vector that can be applied to the state to supplement it with additional possible values. The output gate of a particular cell is in charge of controlling everything that passes through it. The resulting number considers not only the filtered data but also any freshly provided information as well as the cell's current state. Figure 5 shows a diagram of the formation of a cell.

3.3.3. Reinforcement Learning

Reinforcement learning is a sort of deep learning that defines how a software agent should respond in its natural environment. This is the definition of reinforcement learning. This technique entails subjecting the agent to a series of actions, observing it, rewarding it, and attempting to predict the correct path it will take until an acceptable manner or forecast, also known as a policy, is established. For some time now, the integration of machine learning approaches into brain tumor detection and recognition has been a main emphasis, with the goal of enhancing performance and accuracy. When compared to machine learning technologies, the amount of time and effort required to manually detect brain tumors is substantially more. Machine Learning techniques appear to be a good fit for this business since they can improve accuracy and performance in the process of diagnosing brain tumors while using fewer resources and less time [35].

3.3.4. Auto-Encoder

An auto-encoder is a type of artificial neural network that is trained using unsupervised methods. The auto-encoder is designed to learn a sample vector that represents the data appropriately—frequently used in the context of dimensionality reduction. Its fundamental design incorporates an encoder and a decoder. The information is absorbed by the encoder module's thymic vector, which is subsequently utilized by the decoder module to generate new data. Auto-coding ANNs are classified into variable auto-encoders and deep productive models. Face recognition and word meaning are only two of the numerous practical applications aided by the usage of auto-encoders.

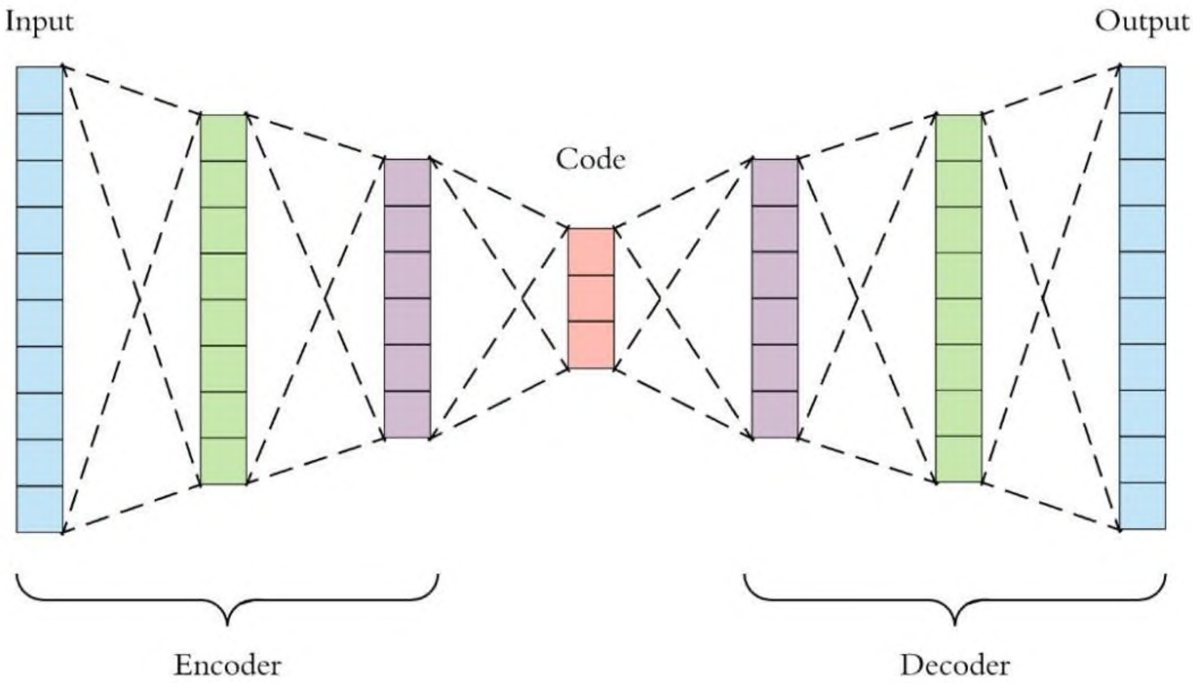

Figure 3.6. Long short memory cell structure [36]

$$\dot{x} \quad = (x) \qquad (3.3)$$

[redacted] (3.4)

n [redacted] (3.5)

Here n(1)i refers to the ith neuron at the first layer for the architecture, M is an activation function, w_i, and b_i refer to weight matrix and the bias parameter, respectively.
The final mathematical model is illustrated in Equation [36]

$n_{w,b}$ [redacted] (3.6)

3.4. Convolutional Neural Network Based Deep Auto-Encoders and PSO

This study described a novel approach to facial identification using deep auto-encoders within a convolutional neural network with PSO. This study intends to accomplish this by merging two well-known deep learning approaches. We used google.net, a previously trained CNN, for feature extraction. Feature extraction with a deep sparse auto-encoder connected to a Google.net. The first auto-encoder was able to decrease the feature set and enhance accuracy by rejecting less significant parts and extracting more relevant ones. Furthermore, the first auto-output encoder was linked to the second. The second auto-output encoder's was then connected to the SoftMax. SoftMax was utilized in this study to sort the collected features into classes that represented human groups. Figure 7 depicts the flowchart for the study.

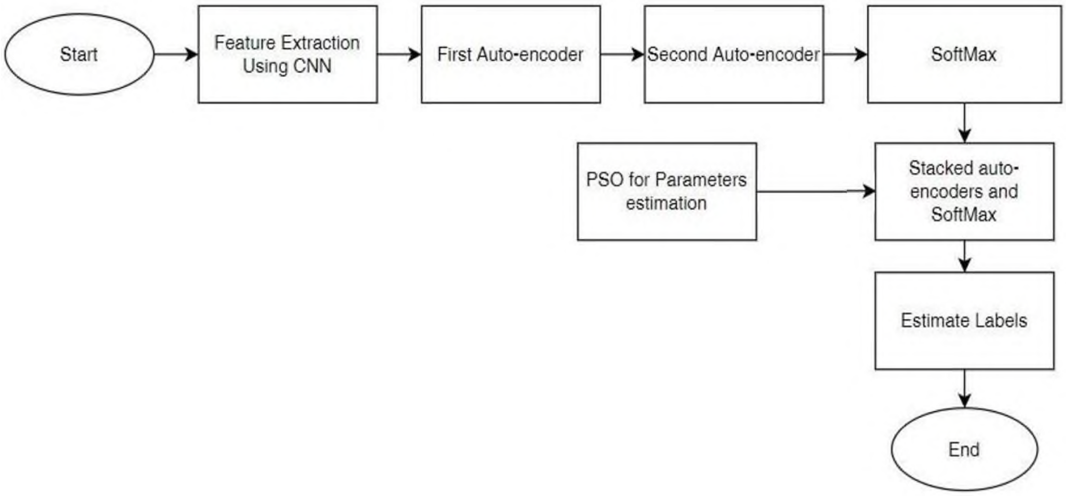

Figure 3.7. Convolutional neural network based deep auto-encoders and pso

3.5. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) method is a stochastic population-based optimization algorithm, firstly introduced by Kennedy and Eberhart in 1995 [36]. In PSO algorithm, each member of the population is called a ‘‘particle’’, and each particle ‘‘flies’’ around in the multidimensional search space with a velocity, which is constantly updated by the particle’s own experience and the experience of the particle’s neighbors or the experience of the whole swarm. It has already been applied in many areas, such as function optimization, artificial neural network training, pattern classification and fuzzy system control. The advantages of PSO are that PSO is rapidly converging towards an optimum, simple to compute, easy to implement and free from the complex computation as in genetic algorithm (e.g., coding/decoding, crossover and mutation). However, PSO does exhibits some disadvantages: it sometimes is easy to be trapped in local optima, and the convergence rate decreased considerably in the later period of evolution; when reaching a near optimal solution, the algorithm stops optimizing, and thus the accuracy the algorithm can achieve is limited.

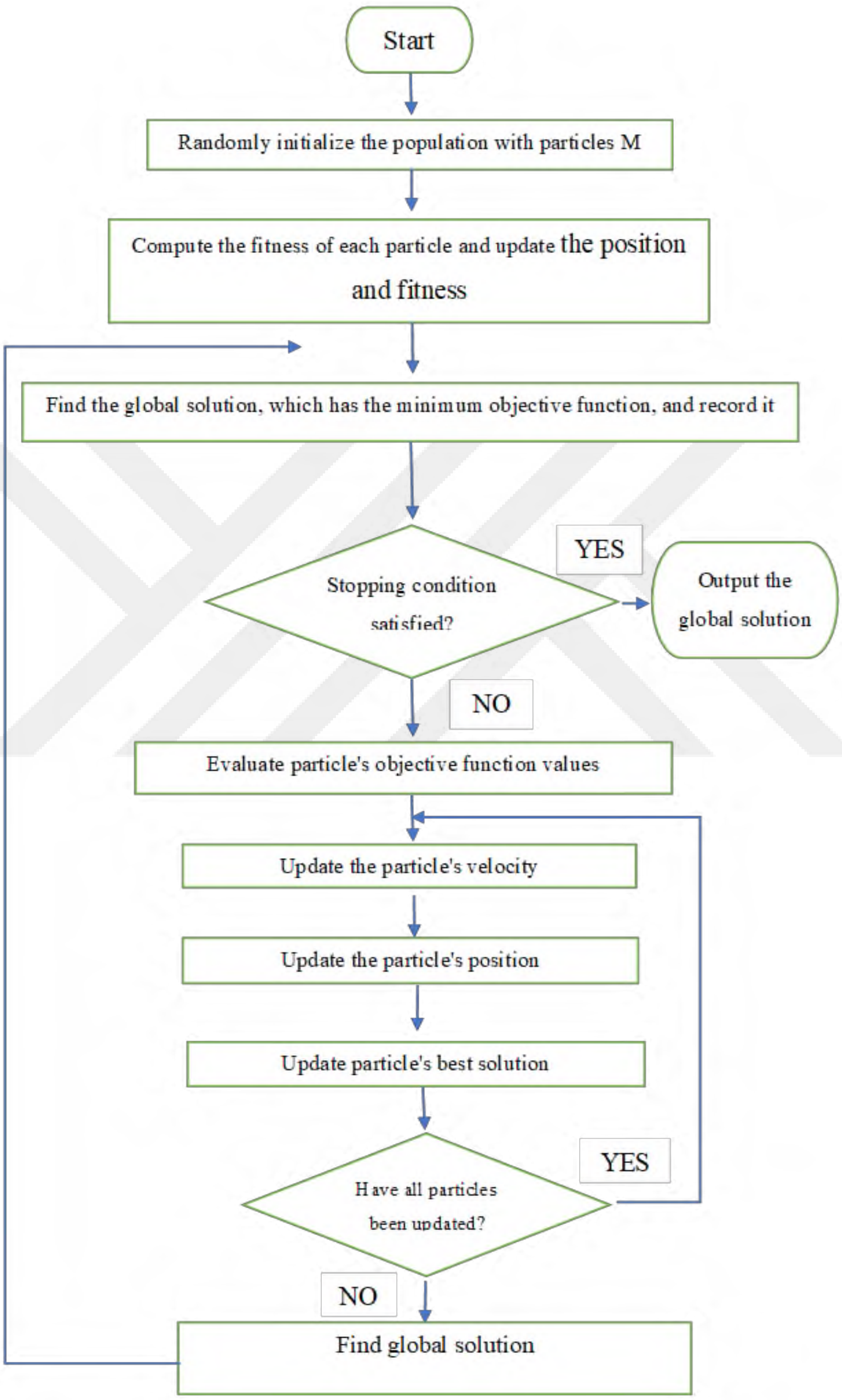

Figure 3.8. Shows the flowchart of PSO techniques [37].

4. EXPERIMENTS AND DISCUSSION

Experiments are run under various situations, the results are compared to specified metrics, and any similarities or discrepancies across the experiments are highlighted. Experiments were carried out using MATLAB 2018, which was used to implement them. The dataset properties specified in Chapter 3 are fed into the suggested MATLAB2018-based system.

The contents of the information folder were accessed and read. The following step requires the user to select a training-to-test ratio within the code, as shown in the MATLAB window that can be found below. The Flowchart of this step is shown in Figure 4.1 below.

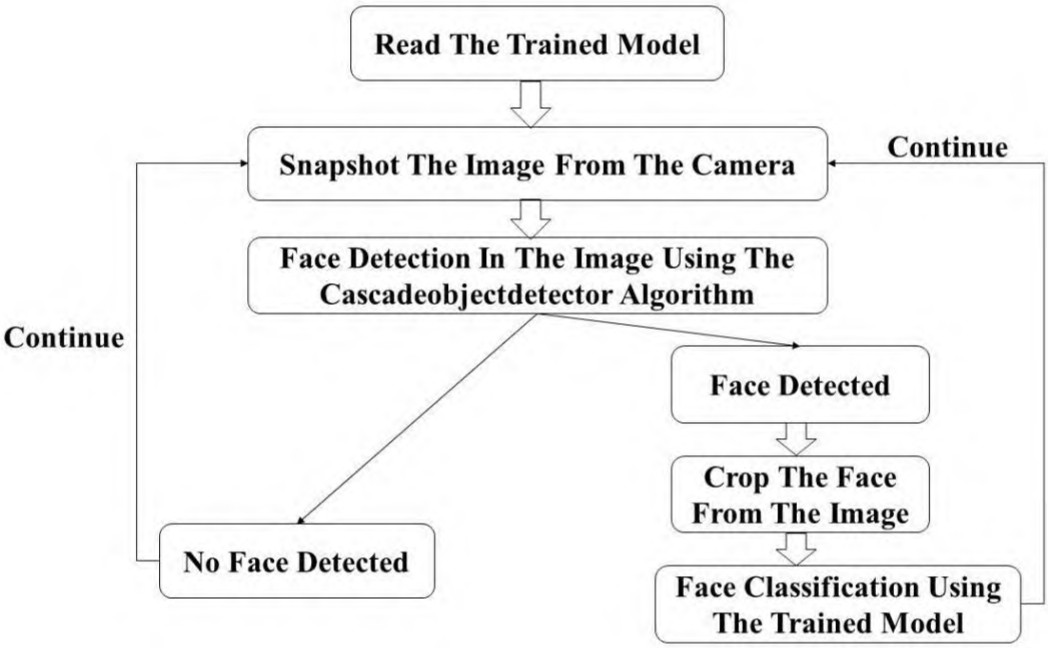

Figure 4.1. Face detection flowchart

The CNN section of AlexNet was followed by three entirely connected layers. These layers' objective was to combine the high-level retrieved characteristics from the last levels of CNN into a single vector array. The number of classes was expressed by the number of completely linked neurons. As shown in the code below, the classifiers get features from both the training and testing sets at layer 7. The CNN section of AlexNet was followed by three entirely connected layers are illustrated in the Figure 4.2.

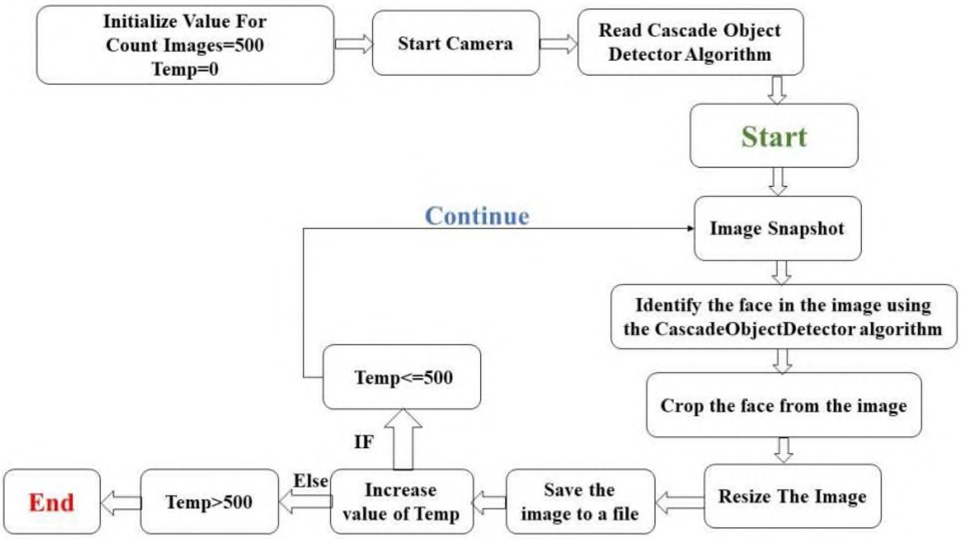

Figure 4.2. The CNN section of AlexNet

Below is the MATLAB code of the PSO that was used to estimate the auto-encoders classifier, the performance of which was affected by these parameters. PSO code that was used to estimate the auto-encoders classifier is concluded in [36]. The accuracy will calculated by using the script accuracy = mean (Ypred==YTest)*100. This section lead to calculate the accuracy of the model from degree 100%. When all was said and done, the script described below was put to use, and the estimated results may be seen in Figure 4.3.

Figure 4.3. Estimated Images

In this study, we demonstrated and tested a CNN-based approach to facial recognition. In a Googlenet like fashion, the suggested network is based on pre-trained models. Auto-encoder and SoftMax were inserted into the system's classification layer to better categorize faces in the Faces dataset.

4.1. Evaluation Parameters

In this section, we compute four statistical parameters related to the face recognition problem in order to evaluate the effectiveness of the solution that we have proposed. The accuracy of a measurement can be understood by how closely it corresponds to the actual value being measured. The efficiency of the method that was suggested will be evaluated based on these features. After the parameters have been estimated, a comparison may be made between them, and the results of well-known earlier research found in the literature.

4.2. Dataset and Capturing System

The data set was obtained from images taken from 20 people. There are 500 images of each person. Face profiles were taken from 5 directions, up, down, right, left and in front of directions. There are about 100 pictures from each direction. A total of 10000 image data sets were created in this way. Pictures were taken of the person sitting in front of the laptop camera at a distance of about 1m. The data set was collected on different days under room lighting conditions. Data capture software was created within the scope of the thesis. In Figure 4.5, the shape of the data collection mechanism is given. A random sample of the figures that make up this data set is given in figure 4.6. Each figure is in bmp standard with 227*227 pixels and 24-bit depth resolution. The creation time of the data set per person takes around 40 seconds on a mid-segment computer. The dataset included a mix of young, middle-aged men and women. Relatives with beard, mustache, long hair and different facial structures were preferred, which would increase the classification difficulty of the algorithms.

Figure 4.4. Capturing system

Figure 4.5. Dataset sample

4.3. Implementation

We use Random Subsampling to validate our methods on the dataset because it is more convenient. The datasets utilized in the presentation of the CNN-based deep sparse auto-encoder are listed in Table 4.1. The precisions are determined by solving equation 1 for x.

$$\text{Acc}= \frac{TP+TN}{Total} \tag{4.1}$$

We employed the CNN+PSO+deep auto-encoder method, which is among the most powerful approaches to classification, to make our classifications. There were first constructed face databases that separated out facial features from other areas. These datasets were used in a CNN+PSO+deep auto-encoder training procedure. After a period of training, the SVM technique was used to correctly classify the image to be identified. Our face database photos were used in conjunction with CNN and deep auto-encoder techniques to complete the recognition procedure. To get to the recognition success rate, first an auto-encoder training database was built up using both face and non-face photos. From these newly generated databases, constant descriptors independent of lighting, orientation, and perspective were extracted using the CNN technique. We then classified the collected IDs as either "face" or "not face" and trained an auto-encoder based on the results. By applying the auto-encoder technique to each individual image in the face database, the recognition success rate may be calculated. The results are shown in the table below. The process of adding new data and training the network takes up to 2 minutes. Testing the developed software takes 30-35 seconds. For training, 30% of the 20 classes for each aspect is sufficient for success. The remaining 70% dataset was used for testing. Each test was repeated 100 times to prevent the bias of the test. Classification performance was obtained as 99% for the proposed method.

The dataset should not be over-expanded in order to run on low-cost systems. However, using less data sets negatively affects classification performance.

Training on single CPU.
Initializing input data normalization.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:01	12.50%	2.9373	0.0010
5	50	00:00:55	100.00%	0.0007	0.0010
10	100	00:01:50	100.00%	0.0003	0.0010
14	150	00:02:44	100.00%	0.0002	0.0010
19	200	00:03:39	100.00%	3.6469e-05	0.0010
20	220	00:04:01	100.00%	0.0002	0.0010

Training finished: Max epochs completed.

Figure 4.6. The results visualized

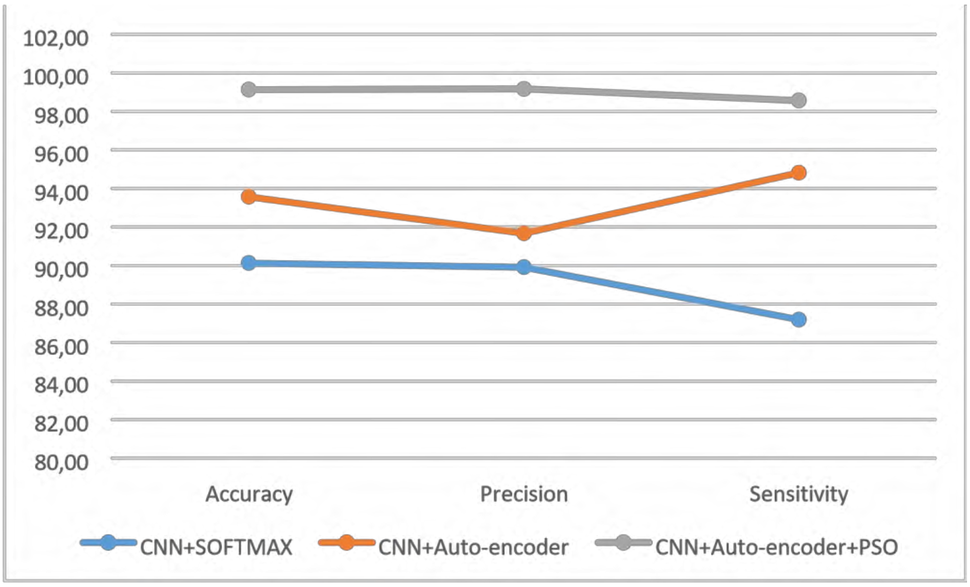

Figure 4.7. Comparison of Results.

As can be seen in Table 4.7, the current method results compare favourably with those of various other facial recognition studies published in recent years.

Table 4.1. Comparisons with previous studies

Methods	Data size	**Accuracy Results (%)**
LBP&MLP [7]	3000	90.010
ORB&KPCA&TBF [8]	1038	84.10
ART&NNC-SVM [9]	8600	94.86
PCA&ED [12]	10000	94.4
SURF&SVD [16]	2404	95.23
SURF&PCA [17]	5500	95.2
CNN&Deep Autoencoder&PSO	10000	99

Developed Matlab codes are given in Appendix 1-4.

5. CONCLUSION

This work focuses on two of the most essential ways for overcoming the problem of facial recognition that are currently available. The core components of the proposed facial recognition system include CNN, auto-encoder, and the PSO algorithm. Because my research shown that these algorithms produced more accurate results, they quickly gained popularity. Furthermore, these methods have been proposed since a reliable face recognition system should not be prone to the following issues:

- Alterations to the nature of the light source
- Based on the tilt of the head and the proportions of the face
- From minute characteristics like the spectacles and beards on his face
- From the background of the face

We concluded during this experiment that the facial background as well as the head had a significant impact on the effectiveness of the suggested face recognition system. Your level of recognition will be proportionate to the distance ratio you choose. For example, research has shown that keeping a distance ratio of roughly 0.7 yields the optimum results. Other factors, in addition to the proximity-to-distance ratio, might surely contribute to an increased identification success rate. It was also discovered that the quality of the image databases used to train auto-encoder algorithms outperformed the number of samples used in the training procedure. With this study, systems with fast reaction times in commercial use have been approached very quickly. This study shows that face recognition can be performed with high success in cloud systems, low-level systems, systems with low resolution cameras.

This study can be carried out on low-cost and low-level systems such as low-cost Raspberry Pi, ESP32-Cam, Asus Thinker PCs.

REFERENCES

[1] Goldstein, A. J., Harmon, L. D., and Lesk, A. B., Identification of human faces", Proc. IEEE 59, pp. 748-760, (1971).

[2] Haig, N. K., "How faces differ - a new comparative technique", Perception 14, pp. 601615, (1995).

[3] Rhodes, G., "Looking at faces: First-order and second order features as determinants of facial appearance", Perception 17, pp. 43-63, (1988).

[4] Kirby, M., and Sirovich, L., "Application of the Karhunen-Loeve procedure for the characterization of human faces", IEEE PAMI, Vol. 12, pp. 103-108, (1990).

[5] Sirovich, L., and Kirby, M., "Low-dimensional procedure for the characterization of human faces", J. Opt. Soc. Am. A, 4, 3,pp. 519-524, (1987).

[6] Terzopoulos, D., and Waters, K., "Analysis of facial images using physical and anatomical models", Proc. 3rd Int. Conf. on Computer Vision, pp. 727-732, (1995).

[7] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1002-1014, 1 April 2018, doi: 10.1109/TPAMI.2017.2700390.

[8] W. Zhao, R. Chellappa and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, 1998, pp. 336-341, doi: 10.1109/AFGR.1998.670971.

[9] R. He, J. Cao, L. Song, Z. Sun and T. Tan, "Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1025-1037, 1 May 2020, doi: 10.1109/TPAMI.2019.2961900.

[10] Corresponding author. E-mail addresses: zhujh@xjtu.edu.cn (J. Zhu), 343033704@qq.com (Z. Jiang). 1 Jihua Zhu and Zutao Jiang contributed equally to this work. https://doi.org/10.1016/j.ins.2019.03.024 0020-0255/© 2019 Elsevier Inc. All rights reserved.

[11] K. Jain, B. Klare and U. Park, "Face recognition: Some challenges in forensics," 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, 2011, pp. 726-733, doi: 10.1109/FG.2011.5771338.

[12] R. He, X. Wu, Z. Sun and T. Tan, "Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 7, pp. 1761-1773, 1 July 2019, doi: 10.1109/TPAMI.2018.2842770.

[13] A. J. O'Toole, Yi Cheng, P. J. Phillips, B. Ross and H. A. Wild, "Face recognition algorithms as models of human face processing," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, Grenoble, France, 2000, pp. 552-557, doi: 10.1109/AFGR.2000.840689.

[14] Zhaowei Cai, Longyin Wen, Dong Cao, Zhen Lei, Dong Yi and S. Z. Li, "Personspecific face tracking with online recognition," 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, 2013, pp. 1-6, doi: 10.1109/FG.2013.6553730.

[15] H. Du, S. Wang, J. Zhao and N. Xu, "Two-dimensional neighborhood preserving embedding for face recognition," 2010 2nd IEEE International Conference on Information

[16] Management and Engineering, Chengdu, 2010, pp. 500-504, doi: 10.1109/ICIME.2010.5477710.

[17] Y. Huang and S. Chen, "A geometrical-model-based face recognition," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 3106-3110, doi: 10.1109/ICIP.2015.7351375.

[18] X. Li and F. Da, "3D Face Recognition by Deforming the Normal Face," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 3975-3978, doi: 10.1109/ICPR.2010.967.

[19] T. Mantoro, M. A. Ayu and Suhendi, "Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods," 2018 6th International Conference on Multimedia Computing and Systems (ICMCS), Rabat, 2018, pp. 1-5, doi: 10.1109/ICMCS.2018.8525935.

[20] J. Lu, V. E. Liong, X. Zhou and J. Zhou, "Learning Compact Binary Face Descriptor for Face Recognition," in IEEE Transactions on Pattern Analysis and Machine

Intelligence, vol. 37, no. 10, pp. 2041-2056, 1 Oct. 2015, doi: 10.1109/TPAMI.2015.2408359.

[21] C. Chang and C. Huang, "Application of active appearance model for dual-camera face recognition," 2012 International Conference on Information Security and Intelligent Control, Yunlin, 2012, pp. 333-336, doi: 10.1109/ISIC.2012.6449774.

[22] C. Ding, C. Xu and D. Tao, "Multi-Task Pose-Invariant Face Recognition," in IEEE Transactions on Image Processing, vol. 24, no. 3, pp. 980-993, March 2015, doi: 10.1109/TIP.2015.2390959.

[23] B. Topçu and H. Erdogan, "Decision Fusion for Patch-Based Face Recognition," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 1348-1351, doi: 10.1109/ICPR.2010.333.

[24] Z. Lei, S. Liao and S. Z. Li, "Efficient feature selection for linear discriminant analysis and its application to face recognition," Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, 2012, pp. 1136-1139.

[25] Yan Cui, Jielin Jiang, Zuojin Hu, Xiaoyan Jiang, Wuxia Yan, Min-ling Zhang, Neighborhood kinship preserving hashing for supervised learning, Signal Processing: Image Communication, Volume 76, 2019, Pages 31-40, ISSN 0923-5965, https://doi.org/10.1016/j.image.2019.04.003..

[26] F. Dornaika, Y. El Traboulsi, Joint sparse graph and flexible embedding for graphbased semi-supervised learning, Neural Networks, Volume 114, 2019, Pages 91-95, ISSN 0893-6080, https://doi.org/10.1016/j.neunet.2019.03.002.

[27] He Di, Xu Ke, Zhou Peng, Zhou Dongdong, Surface defect classification of steels with a new semi-supervised learning method, Optics and Lasers in Engineering, Volume 117, 2019, Pages 40-48, ISSN 0143-8166, https://doi.org/10.1016/j.optlaseng.2019.01.011.

[28] Karim, A. M., Güzel, M. S., Tolun, M. R., Kaya, H., & Çelebi, F. V. (2019). A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing. Biocybernetics and Biomedical Engineering, 39(1), 148-159. doi:10.1016/j.bbe.2018.11.004.

[29] Victor de Abreu Campos, Daniel Carlos Guimarães Pedronette, A framework for speaker retrieval and identification through unsupervised learning, Computer Speech

& Language,Volume 58, 2019, Pages 153-174, ISSN 0885-2308, https://doi.org/10.1016/j.csl.2019.04.004.

[30] Eric S. Jackson, Naomi H. Rodgers, Derek B. Rodgers, An exploratory factor analysis of action responses to stuttering anticipation, Journal of Fluency Disorders, Volume 60, 2019, Pages 1-10, ISSN 0094-730X, https://doi.org/10.1016/j.jfludis.2019.03.001.

[31] Jihua Zhu, Zutao Jiang, Georgios D. Evangelidis, Changqing Zhang, Shanmin Pang, Zhongyu Li, Efficient registration of multi-view point sets by K-means clustering, Information Sciences, Volume 488, 2019, Pages 205-218, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2019.03.024.

[32] Xianping Du, Feng Zhu, A novel principal components analysis (PCA) method for energy absorbing structural design enhanced by data mining, Advances in Engineering Software, Volume 127, 2019, Pages 17-27, ISSN 0965-9978, https://doi.org/10.1016/j.advengsoft.2018.10.005.

[33] B. Richhariya, M. Tanveer, EEG signal classification using universum support vector machine, Expert Systems with Applications, Volume 106, 2018, Pages 169-182, ISSN 09574174, https://doi.org/10.1016/j.eswa.2018.03.053.

[34] Qiuling Hou, Jinxin Zhang, Liming Liu, Yiju Wang, Ling Jing, Discriminative information-based nonparallel support vector machine, Signal Processing, Volume 162, 2019, Pages 169-179, ISSN 0165-1684, https://doi.org/10.1016/j.sigpro.2019.03.026.

[35] 2019 High Performance Computing and Computational Intelligence Conference
Journal of Physics: Conference Series 1395 (2019) 012006
IOP Publishing doi:10.1088/1742-6596/1395/1/012006

[36] Received: 27 September 2020 / Revised: 3 November 2020 / Accepted: 5 November 2020 / Published: 9 November 2020Sensors 2020, 20(21), 6378; https://doi.org/10.3390/s20216378

[37] Homero de Leon-Delgado, Rolando J. Praga-Alejo, David S. Gonzalez-Gonzalez, Mario Cantú-Sifuentes, Multivariate statistical inference in a radial basis function neural network, Expert Systems with Applications, Volume 93, 2018, Pages 313-321, ISSN 09574174, https://doi.org/10.1016/j.eswa.2017.10.024.

[38] YANG Jucheng, LIU Na, FANG Shanshan,et al. Review of face recognition methods based on deep learning[J]. Journal of Tianjin University of Science and Technology, 2016, 31(06): 1 10.

APPENDİX

AP 1. Matlab face detection code

The contents of the information folder have been accessed and read. The next step requires the user to define the training-to-test ratio within the code, as shown in the MATLAB window which can be found below.

```
clc
clear all
close all
warning off;
cao=webcam;
faceDetector=vision.CascadeObjectDetector;
c=750;
temp=0;
while true
    e=cao.snapshot;
    bboxes =step(faceDetector,e);
    if(sum(sum(bboxes))~=0)
    if(temp>=c)
        break;
    else
    es=imcrop(e,bboxes(1,:));
    es=imresize(es,[227 227]);
    filename=strcat(num2str(temp),'.bmp');
    imwrite(es,filename);
    temp=temp+1;
    imshow(es);
    drawnow;
    end
    else
        imshow(e);
        drawnow;
    end
end
```

AP 2. Code combining high-level features

AlexNet's CNN section was followed by three fully connected layers. The goal of these layers was to combine the high-level properties retrieved from the last levels of the CNN into a single vector array. The number of categories was expressed as the number of fully connected neurons. As shown in the code below, classifiers obtain features from both the training and test sets at layer 7. The CNN section of AlexNet is illustrated with three fully connected layers.

```
clc
clear all
close all
warning off


allImages=imageDatastore('dataset','IncludeSubfolders',true,
'LabelSource','foldernames');
opts=trainingOptions('sgdm','InitialLearnRate',0.001,'MaxEpochs',20,'MiniBa
tchSize',64);
layers = [
    imageInputLayer([227 227 3])

    convolution2dLayer(3,8,'Padding','same')
    batchNormalizationLayer
    reluLayer

    maxPooling2dLayer(2,'Stride',2)

    convolution2dLayer(3,16,'Padding','same')
    batchNormalizationLayer
    reluLayer

    maxPooling2dLayer(2,'Stride',2)

    convolution2dLayer(3,32,'Padding','same')
    batchNormalizationLayer
    reluLayer

    fullyConnectedLayer(20)
    softmaxLayer
    classificationLayer];
myNet1=trainNetwork(allImages,layers,opts);
  save myNet1;
```

AP 3. MATLAB Code Complete for PSO, Auto Classifier

Below is the MATLAB code for that was used to estimate the autoencoder classifier, whose performance was affected by these parameters. The code that was used to estimate the autoencoder classifier is completed

```
clc;close;clear
c=webcam;
load myNet1;
faceDetector=vision.CascadeObjectDetector;
while true
    e=c.snapshot;
    bboxes =step(faceDetector,e);
    if(sum(sum(bboxes))~=0)
     es=imcrop(e,bboxes(1,:));
    es=imresize(es,[227 227]);
    label=classify(myNet1,es);
    image(e);
    title(char(label));
    drawnow;
    else
        image(e);
        title('No Face Detected');
    end
end
```

AP 4. Object detector cascade experiments

a) ImageDataStore prepared which is important to understand and deal with images by cnn technique

```
%The images prepared and will be read from the location that saved on it.
dataFolder = 'E:\Adnan_thesisDataSet\';

%read the categories which mean read the folders names which represented each class
such as ahmad, ali,  directoryContent = dir(dataFolder); %dir lists files and
folders in the current
folder.  dirFlags = [directoryContent.isdir]; subFolders

= directoryContent (dirFlags); categories = cell (1,

length

(subFolders)-2); for k = 3 : length (subFolders) categories (k-2)

= [subFolders (k).name];  subPath=fullfile (dataFolder,

subFolders (k).name, '\'); end

%ImageDataStore prepared which is important to understand and deal with images by
cnn technique.

Imds = imageDatastore (fullfile (dataFolder, categories), 'LabelSource',
'foldernames'); tbl=countEachLabel

(imds)

% split the data on train and test sets. train used to train the model and the test
used to test.

[indsTrain, imdsTest] = splitEachLabel (imds, 0.7, 'randomized');
```

b) The CNN section of AlexNet was followed by three entirely connected layers. These layers' objective was to combine the high-level retrieved characteristics from the last levels of CNN into a single vector array. The number of classes was expressed by the number of completely linked neurons. As shown in the code below, the classifiers get features from both the training and testing sets at layer 7.

```
 featuresTrain = activations (net, imdsTrain, layer, 'OutputAs',

'rows');  % feature used in test that extracted from layer 7.

featuresTest = activations (net, imdsTest, layer, 'OutputAs', 'rows');
% prepare the labels which represented the categories of train and test of the
images.
```

```
numTrainImages1 = numel (YTrain)

numTestImages2 = numel (YTest) numClasses

= numel (categories)
```

CURRICULUM VITAE

Kişisel Bilgiler	
Adı Soyadı	ADNAN ABDULLAH YOUNOS AL.HAMMADI
Doğum Yeri	
Doğum Tarihi	
Uyruğu	☑ T.C. ❐ Diğer:

Eğitim Bilgileri	
Lisans	
Üniversite	UNIVERSITY OF MOSUL
Fakülte	Education
Bölümü	Computer Science
Mezuniyet Yılı	2012

Yüksek Lisans	
Üniversite	Kirşehir Ahi Evran Üniversitesi
Enstitü Adı	Graduate School of Natural and Applied Sciences
Anabilim Dalı	Computer Engineering
Programı	Part of the requirements for obtaining a master's degree in computer engineering
Mezuniyet Tarihi	2022