People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

UNIVERSITY OF
M'HAMED BOUGARA BOUMERDES
FACULTY OF TECHNOLOGY
DEPARTMENT OF ELECTRICAL
SYSTEMS ENGINEERING

End of study memory
Presented for the Diploma of

ACADEMIC MASTER

Domain: Science and technology
Option: Telecommunication
Speciality: Networks and telecommunications

Presented by:
ALLALOU Rabia
AIOUNA Mohamed riadh

Facial recognition for passport check

Defended on - in front of the jury composed of:

Mr	MESSAOUDI Noureddine	MCA	Chair
Mrs	MAHDI Asmahan	MCB	Examiner
Mrs	GUERBAI Yasmine	MCA	Supervisor

The academic year 2022/2023

Acknowledgments

We would like to express our deepest gratitude to our advisor, Dr. Geurbai, for their invaluable guidance, support, and mentorship throughout this research journey. Their expertise and encouragement have been instrumental in shaping this thesis and my academic growth. We are also grateful to the faculty members of the university of M'hamed Bougara for their valuable insights and feedback.

We would like to thank the participants who generously contributed their time and consented to be a part of this study. Their cooperation has been crucial in gathering the necessary data for the research.

We are grateful to the research community and all the authors whose works have provided the foundation for this study. Their contributions have shaped my understanding and paved the way for advancements in the field of facial recognition.

Once again, We extend our heartfelt appreciation to everyone who has contributed to the successful completion of this thesis. Their support and encouragement have been invaluable, and we are truly grateful for their presence in our academic journey.

Dedicate

I dedicate this thesis to my beloved parents, Naima and Ilyes. Your unwavering love, sacrifices, and boundless support have been the foundation of my journey. From the very beginning, you instilled in me the importance of education and the pursuit of knowledge. Your belief in my abilities has been a constant source of inspiration and motivation. I am eternally grateful for the sacrifices you have made to ensure my success. This work is a tribute to your unwavering dedication and love.

To my dear sisters, Meriem, Arwa, and Aicha. Your presence in my life has brought joy, laughter, and unwavering support. We have shared countless memories, and your encouragement has been a driving force behind my achievements. Your belief in me has given me the strength to overcome challenges and reach for my goals. I dedicate this thesis to you with immense gratitude and love.

To my extended family, who have always been my pillars of strength, I am grateful for your continuous support and encouragement. In particular, I would like to express my heartfelt appreciation to my aunt and uncle, Nassima and Redha. Since my journey began at the university, you have been my unwavering advocates, offering guidance, advice, and support whenever I needed it. Your belief in my potential has propelled me forward, and I am forever thankful for your presence in my life. This work is dedicated to you, as a symbol of gratitude for your unwavering support.

To my beloved grandparents, who have left this world but continue to inspire me with their legacy. I carry your love, wisdom, and values in my heart. I know that if you were still here, you would be immensely proud of my accomplishments. Your guidance and unwavering belief in me have shaped my character and fueled my determination. I dedicate this thesis to you, as a tribute to the indelible impact you have had on my life.

May this work serve as a reflection of the love, support, and encouragement I have received from my family, both present and past. Your presence in my life has been a blessing, and I am forever grateful for the role you have played in shaping my journey.

Dedicate

I dedicate my dissertation work to my family and many friends.

A special feeling of gratitude to my loving parents, Aicha and Kamel whose words of encouragement and push for tenacity ring in my ears. To my sisters who have never left my side and are very special.

I also dedicate this dissertation to my many friends and extended family who have supported me throughout the process.

I will always appreciate all they have done for helping me develop my technology skills, for the many hours of proofreading, and helping me to master the leader dots.

Thank you all.

ملخص

تستكشف هذه الرسالة تنفيذ وتقييم نماذج التعلم العميق للتعرف على الوجه في سياق فحص جوازات السفر. تشمل المكونات الرئيسية VGGFace و MTCNN و LBP والشبكات العصبية السيامية. تقارن البحث أداء هذه النماذج مع النهج المتقدم باستخدام مجموعة بيانات LFW. يتم معالجة التحديات مثل توفر مجموعة البيانات والتكيف المفرط وموارد الحوسبة. تبرز النتائج فعالية النماذج لتطبيقات فحص جوازات السفر وتؤكد على ضرورة مزيد من البحث لمواجهة التحديات المتعلقة بالشيخوخة وتباين الإضاءة وتحيزات مجموعات البيانات. تساهم النتائج في تحسين عمليات التحقق من الهوية وتعزيز أنظمة الرقابة على جوازات السفر وأمن الحدود.

الذكاء الاصطناعي، التعلم العميق، تمييز الوجوه، VGGFace، MTCNN، LBP.

Abstract

This thesis explores the implementation and evaluation of deep learning models for facial recognition in the context of passport checks. Key components include VGGFace, MTCNN, LBP, and Siamese Neural Networks. The research compares the performance of these models with state-of-the-art approaches using the LFW dataset. Challenges such as dataset availability, overfitting, and computational resources are addressed. The findings highlight the effectiveness of the models for passport check applications and the need for further research in addressing challenges related to aging, lighting variations, and dataset biases. The outcomes contribute to improving identity verification processes and enhancing passport control and border security systems.

Artificial intelligence, Deep Learning, Face Recognition, VGGFace, MTCNN, LBP.

Résumé

Ce memoire explore la mise en œuvre et l'évaluation de modèles d'apprentissage profond pour la reconnaissance faciale dans le contexte du contrôle des passeports. Les composantes clés incluent VGGFace, MTCNN, LBP et les réseaux neuronaux siamois. La recherche compare les performances de ces modèles avec des approches de pointe en utilisant l'ensemble de données LFW. Les défis tels que la disponibilité des données, le surapprentissage et les ressources informatiques sont abordés. Les résultats mettent en évidence l'efficacité des modèles pour les applications de contrôle des passeports et soulignent la nécessité de poursuivre la recherche pour relever les défis liés au vieillissement, aux variations d'éclairage et aux biais des ensembles de données. Les résultats contribuent à améliorer les processus de vérification d'identité et à renforcer les systèmes de contrôle des passeports et de sécurité aux frontières.

Intelligence artificielle, Apprentissage en profondeur, Reconnaissance Faciale, VGGFace, MTCNN, LBP.

Table of Contents

General Introduction: 1

Chapter 1: Overview of face detection and recognition

1. Introduction 3
2. Machine learning and Deep learning 3
3. How is deep learning used today 3
4. Computer vision 4

4.1. Image processing techniques for computer vision 5

4.2. Object detection and recognition 5

5. Facial detection and recognition 6
6. History of face recognition 7
7. Steps of face recognition 8

7.1. Face detection 8

7.2. Face Alignment 9

7.3. Face Measurement and Extraction 10

7.4. Face Recognition 11

7.5. Face Verification 12

8. Techniques of face detection and recognition 12

8.1. Histogram of oriented gradients (HOG) 12

8.2. Local Binary Patterns (LBP) 13

8.3. Convolutional neural networks (CNNs) 14

8.4. Support Vector Machines (SVMs) 15

8.5. You only look once (YOLO) 16

9. The difference between face recognition and face comparison 17

9.1. Biometric authentication: the difference between facial recognition and facial matching technologies 17

10. Passport check 18

10.1. Importance of Face Recognition in Passport Verification 19

11. Challenges and difficulties 19

12. Conclusion 22

Chapter 2: Tools and technologies

1. Introduction 23
2. Python libraries 23
2.1. NumPy: Array computing for Python 24
2.2. Keras: Deep learning in Python.. 24
2.3. C. OpenCV: Computer vision in Python 24
2.4. Tensorflow: Machine learning library in Python 24
3. MTCNN 25
3.1. How we used it 25
3.2. How MTCNN works 26
4. LBP... 26
4.1. Local Binary Patterns for face recognition... 27
5. VGGFace... 28
6. Siamese Neural Network.. 30
6.1. The use of SNN in Face recognition 31
7. Conclusion... 31

Chapter 3: Experimental design

1. Introduction to the Dataset 32
1.1. Overview of the dataset used for training and testing 32
1.2. Description of the data collection process 32
1.3. Statistics and characteristics of the dataset 33
2. Data preprocessing 33
2.1. Data Labelling and Reorganisation 33
2.2. Preprocessing Methods and Techniques.. 35
3. Model Architecture and Selection... 35
3.1. Overview of the existing literature and state-of-the-art approaches in facial recognition………………………………………………………………………………...35
3.2. Description of the three models implemented 36
3.2.1. The first model 36
3.2.2. The second model.. 37
3.2.3. The third model 38
3.3. Explanation of the rationale behind choosing these models..................................... 39
3.4. Training/validation split and cross-validation techniques employed 40
4. Experimental Results and Analysis 40

4.1. Presentation of the accuracy achieved by each model 40
4.2. Different hyper-parameters 41
4.3. Experimental results obtained using different threshold values 42
5. Error Analysis 47
5.1. Potential causes for misclassifications 47
5.2. Proposed strategies for improving model performance based on error analysis 48
6. Limitations and Future Work 49
6.1. Future research and improvements 49
6.2. Possibilities for enhancing model accuracy and generalisation 51
7. Comparative Analysis 51
8. Conclusion 52
General conclusion…………….53
References: 54

List of Figures

Figure 1.1. Object detection with computer vision. [3] 5
Figure 1.2. Facial Recognition and Facial Detection. [8] 7
Figure 1.3. Example of face detection. [9] 9
Figure 1.4. Example of face alignment. [10] 10
Figure 1.5. Face Measurement and Extraction. [11] 10
Figure 1.6. Example of face recognition. [12] 11
Figure 1.7. Steps of Histogram of oriented gradients. [15] 13
Figure 1.8. Representation of Local Binary Patterns. [18] 13
Figure 1.9. Convolutional neural network architecture. [21] 15
Figure 1.10. Diagram of support vector machine. [25] 16
Figure 1.11. You Only Look Once architecture. [27] 17
Figure 1.12. Biometric authentication using face recognition. [29] 18
Figure 2.1. how does MTCNN align a normal image[36] 25
Figure 2.2. local binary patterns example [39] 28
Figure 2.3. VGGFace mechanism [40] 30
Figure 2.4. The architecture of SNN. [41] 31
Figure 3.1. Data Shape Presentation. 35
Figure 3.2. Image pairs. [42] 35
Figure 3.3. 1st model process. 38
Figure 3.4. 2nd model process. 39
Figure 3.5. 3rd model process. 40
Figure 3.6. (Accuracy & F1 Score) vs threshold. 45
Figure 3.7. The ROC curve. 46
Figure 3.8. The Loss curve 47
Figure 3.9. FRR, FAR and EER for Different thresholds. 47
Figure 3.10. First test of the model 48
Figure 3.11. Second test of the model 49

List of Tables

Table 1. Different hyperparameters. 43
Table 2. Comparing our model with others. 53

List of notations

AI	Artificial Intelligence
AUC	Area Under the ROC Curve
CNN	Convolutional Neural Network
CV	Computer Vision
DL	Deep Learning
EER	Equal Error Rate
FAR	False Acceptance Rate
FNR	False negative rate
FPR	False positive rate
FRR	False Rejection Rate
LBP	Local Binary Patterns
LFW	Labelled Faces in the Wild
ML	Machine Learning
MTCNN	Multi-task Cascaded Convolutional Networks
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SOTA	State-of-the-Art
TNR	True negative rate
TPR	True positive rate

General Introduction

In several fields, including security and identity, facial recognition technology has emerged as a valuable tool. It has also grown in popularity in passport control operations in recent years, where technology may assist make the process faster, more effective, and safe. The purpose of this thesis is to investigate the construction of a face recognition system for passport check, with an emphasis on the broad principles of deep learning and facial recognition, the tools and technologies employed, and the system's experimental design.

To compare and match a person's facial traits with a database of known faces in facial recognition technology, algorithms are used. This technique has been utilized for a variety of purposes, including unlocking devices and detecting criminals. It has also grown more common in border control and passport screening operations in recent years. The use of face recognition technology in passport control has various benefits, including enhanced speed and accuracy, as well as the capacity to identify counterfeit passports.

The goal of this thesis is to investigate the implementation of a face recognition system for passport check, with a particular emphasis on the following aspects:

1. Deep learning and facial recognition in general, including the techniques and methodologies utilised in these technologies

2. The programming languages, libraries, and frameworks utilized in the development of the face recognition system.
3. The facial recognition system's experimental design, including the dataset used, the training, validating and testing method, and the evaluation metrics

The study will employ an approach that combines literature evaluation and experimental design. A literature review will be done to analyze the present status of face recognition technology and its use in passport control operations. The experimental design will concentrate on the creation of a face recognition system for passport check, including dataset selection, system development and testing, and performance assessment using standard evaluation criteria.

The use of a face recognition technology for passport checks has a number of potential advantages, including greater speed, accuracy, and security. The purpose of this thesis is to investigate the creation and experimental design of a face recognition system for passport check, with an emphasis on the broad principles of deep learning and facial recognition, the tools and technologies employed, and the performance assessment. This research can contribute to a better understanding of the possible benefits and limits of this technology, as well as give recommendations for future application, by providing insights into the creation and experimental design of a face recognition system for passport check.

The thesis is divided into parts that address various areas of our study on face recognition for passport check. In Chapter 1, we cover the essential ideas of machine learning, computer vision, face recognition, and detection. This chapter provides the basic information required to understand the next chapters and emphasizes the relevance of our study in the context of these domains.

The second chapter focuses on the techniques and technology used in our research. We go through the models and approaches used, such as VGGFace, MTCNN (Multi-task Cascaded Convolutional Networks), LBP (Local Binary Patterns), and numerous Python packages. This chapter sheds light on the technical basis of our work while providing insights into the practical elements of our implementation.

In the third chapter we report the findings and analysis of our studies. We analyze and compare the three models that we implemented. We specifically highlight the correctness of each model and give a thorough analysis of the results. Furthermore, we compare our findings to previous research and cutting-edge standards, stressing our study's contributions and breakthroughs.

We want to give a complete analysis of face recognition for passport check by arranging the thesis into these chapters. Each chapter builds on the one before it, digging deeper into the study topic and concluding with a complete analysis of our results. This thesis adds to the current literature by suggesting new approaches, utilizing cutting-edge models and methodologies, and making considerable gains in accuracy and performance.

CHAPTER 1

Overview of Face Detection and Recognition

1. Introduction

In this chapter, we will discuss face detection and recognition technology and its applications in airport security. The system comprises several modules that work together to detect and recognize human faces in real-time. The different classification approaches used in the recognition system are employed to accurately identify individuals, ensuring that the picture taken is the same as in the passport of the person. Also talking about tools and technologies that we'll use in our work, such as: TensorFlow, computer vision, transfer learning, python libraries, matplotlib, Keras, and OpenCV, and the work plan that we will work with to achieve our objective.

2. Machine learning and Deep learning

Machine learning is a branch of artificial intelligence concerned with the design and development of algorithms. It is the process of teaching computers to learn from data, and involves developing algorithms that can automatically detect patterns in data, and then make predictions based on those patterns. This contrasts with traditional programming, where programmers write code that explicitly tells the machine what to do. Machine learning algorithms learn from data to solve problems too complex for traditional programming. [1]

As for Deep learning, it is a subset of machine learning that results from the simultaneous execution of multiple layers of machine learning algorithms. Note: The terms machine learning and deep learning are often used interchangeably. Most machine learning today is designed at the level of deep learning.

3. How is deep learning used today

Deep learning is evolving with many applications in areas such as computer vision, natural language processing, and predictive analytics. Chances are you already use a variety of products or services that use machine learning techniques in your daily life, as more and more companies are using machine learning in a very wide range of vertical industries.

For example, Netflix uses deep learning in several ways. One way is to go through their recommendation system. The recommendation system gives you personalised suggestions on what to watch next. It relies on algorithms that consider your viewing history, ratings, and what's popular on Netflix. Another way is for companies to invest in new seasons of shows early on, confident that their algorithmic predictions will be successful. Other streaming media and social media also rely heavily on machine learning and deep learning algorithms to deliver content that matches user preferences. Online shopping portals like Amazon also use machine learning algorithms to suggest other items you might want to buy based on your previous searches. [2]

4. Computer vision

Computer vision is the study of how computers can interpret, analyse, and comprehend visual information from their surroundings. It entails the creation of algorithms and strategies for extracting useful information from digital photos and movies. Object identification, face detection, picture restoration, 3D reconstruction, and other applications are all possible with computer vision. The objective of computer vision is to construct machines that can comprehend the visual environment in the same way that humans do and make intelligent judgments based on that comprehension.

Advances in machine learning, deep learning, and computer technology are propelling the area of computer vision forward. Computer vision algorithms have attained remarkable levels of accuracy and performance because of the availability of massive datasets and powerful computing resources. As a result, computer vision is currently used in a wide range of fields, including healthcare, automotive, security, entertainment, and others. As the discipline matures and grows, it is predicted to have a significant influence on how humans engage with technology and the environment around us.

Figure 1 outlines question discovery utilizing computer vision. It grandstands the method of distinguishing and localizing objects inside pictures or recordings. Progressed calculations and neural systems analyse the visual information and draw bounding boxes around the objects. This innovation is connected in independent driving, observation, and picture acknowledgment assignments.

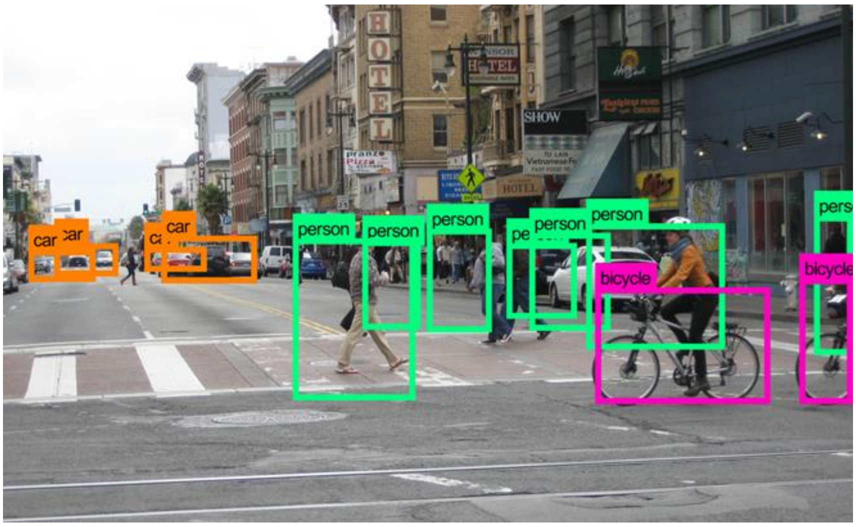

Figure 1.1. Object detection with computer vision [3]

4.1. Image processing techniques for computer vision

Signal processing is a discipline that pertains to the analysis and manipulation of analog and digital signals in electrical engineering and mathematics. This involves a range of operations, such as storing, filtering, and other processing of signals, including transmission signals, sound or voice signals, image signals, and more. Within this realm, image processing specifically deals with signals where the input and output are both images. Image processing involves a range of techniques for processing images, and it can be categorised into two main branches: analog image processing and digital image processing. [4]

4.2. Object detection and recognition

Object detection and recognition are crucial computer vision tasks that entail recognizing and localising items in pictures or movies [5]. Surveillance, driverless cars, robotics, and image search are just a few of the practical uses for these tasks.

Object detection is the process of finding and locating things of interest in an image or video. The purpose is to create a bounding box around each object in the image, indicating its location and size. Typically, object detection consists of two steps: (1) finding areas of interest within the picture, and (2) categorizing each region as containing or not containing an object of interest. Object identification approaches include region-based methods (e.g., Faster R-CNN, Mask R-CNN) and one-stage methods (e.g., YOLO, SSD)[6].

Object recognition, on the other hand, is the process of determining the category of an item in an image or video. This is often accomplished by comparing picture attributes to a predefined list of item categories [7]. Template matching, feature-based approaches (e.g., SIFT, SURF), and deep learning-based methods are common strategies for object recognition (e.g., CNNs).

Object detection and recognition are both active areas of study in computer vision, with new approaches being created all the time to increase their accuracy and efficiency.

Transfer learning has been demonstrated to increase performance on a variety of computer vision tasks, including object identification, picture classification, and semantic segmentation. VGG, ResNet, Inception, and MobileNet are some prominent pre-trained models used for transfer learning in computer vision.

5. Facial detection and recognition

Facial recognition is a biometric technology that uses the unique characteristics of an individual's face to identify them. Most facial recognition systems work by comparing facial prints against a database of known faces. If there is a match, the system can identify the person. However, if the face print is not in the database, the system cannot identify the individual.

Facial recognition technology is often used for security purposes, such as identifying criminals or preventing identity theft. It can also be used for more mundane tasks, like finding a lost child in a crowded place or identifying a guest of honour at an event.

Some facial recognition systems are equipped with artificial intelligence that can learn to recognize individuals even if their appearance has changed, for example if they have grown a beard or gained weight.

The capacity of machines to recognize and locate human faces in digital photographs or video streams is referred to as face recognition and detection. It's an important area of computer vision, with applications ranging from security and surveillance to entertainment and social media. Face detection is the detection of a face in an image or video frame, independent of the individual's identification. The image is analysed using algorithms to identify facial characteristics like the eyes, nose, and mouth, as well as the shape and location of the face.

Facial recognition, on the other hand, is the act of validating or authenticating an individual's identification based on facial traits. This procedure usually entails comparing the observed face traits to a database. Face identification and detection has advanced significantly in recent years as a result of the development of deep learning algorithms and the availability of massive datasets. These advances have resulted in the creation of extremely precise and efficient face recognition and detection systems that may be utilised for a variety of purposes.

Yet, there are considerable hurdles in the sector, such as privacy concerns, algorithmic prejudice, and ethical implications. As a result, it is critical that these technologies are created and utilised responsibly in order to preserve individual rights and promote equal results.

Figure 2 portrays facial acknowledgment and facial location. It exhibits the innovation utilized to analyse facial highlights and recognize people inside pictures or recordings. Facial acknowledgment compares identified faces with a known database, whereas facial discovery centers on finding and extricating facial locales. These procedures have applications in security frameworks and biometric confirmation.

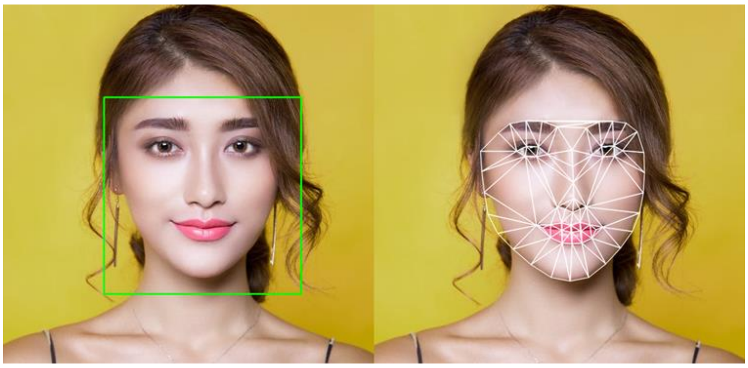

Figure 1.2. Facial Recognition and Facial Detection [8]

6. History of face recognition

Over the past 60 years, the rapid development of facial recognition technology has benefited several major industries, including law enforcement, border control, retail, mobile technology, banking and finance. As we consider the future uses of facial recognition software, it's best to step back and look at how far we've come since our beginnings.

The early pioneers of facial recognition were Woody Bledsoe, Helen Chan Wolf and Charles Bisson. In 1964 and 1965, Bledsoe, along with Wolfe and Beeson, began using computers to recognize faces. Much of their work has never been published since funding for the project came from an anonymous intelligence agency. However, it was later revealed that their original job was to manually mark various "landmarks" on the face, such as the centre of the eyeball and the mouth. It is then mathematically rotated by a computer to compensate for pose changes. Distances between landmarks are also automatically calculated and compared between images to determine identity.

Goldstein, Harmon and Lesk took over in the 1970s when they expanded their work to include 21 specific subjective markers, including hair colour and face lip thickness. for automatic identification. Although accuracy has increased, measurements and positions still have to be calculated manually, which proved extremely laborious, but still represented a breakthrough in Bledsoe's RAND tablet technology.

The Defence Advanced Research Projects Agency (DARPA) and the National Institute of Standards and Technology (NIST) launched the Facial Recognition Technology (FERET) program in the early 1990s to encourage commercial facial treatment Identifying the market.

The Face Recognition Grand Challenge (FRGC) was launched in 2006 with the primary goal of promoting and advancing facial recognition technology designed to support existing facial recognition efforts in the United States Government.

In 2010, Facebook began implementing facial recognition to help identify people whose faces might appear in photos that Facebook users update daily. The feature immediately sparked controversy in the media, triggering a flood of privacy-related articles. However, Facebook users generally don't seem to care. With no apparent negative impact on site usage or popularity, more than 350 million photos are uploaded every day with facial recognition tags.

Facial recognition technology has developed rapidly since 2010, and September 12, 2017 is another important breakthrough for the integration of facial recognition into our daily lives. It's the date Apple launched the iPhone X - the first iPhone owners to unlock with Face ID - Apple's marketing term for facial recognition.

7. Steps of face recognition

7.1. Face detection

First, the system needs to find a face in the image or video. Most cameras now include built-in face detection. Snapchat, Facebook and other social media platforms use facial recognition technology to allow users to apply effects to photos and videos taken with their apps. Many apps use it to identify people in photos, and they can even use this face detection technology to find people standing in crowds.

Figure 3 exhibits confront discovery, illustrating the programmed recognizable proof and localization of human faces inside an picture or video. By analysing visual prompts, the framework precisely recognizes and highlights the nearness of faces. This innovation finds applications in photography, social media channels, reconnaissance, and human-computer interaction.

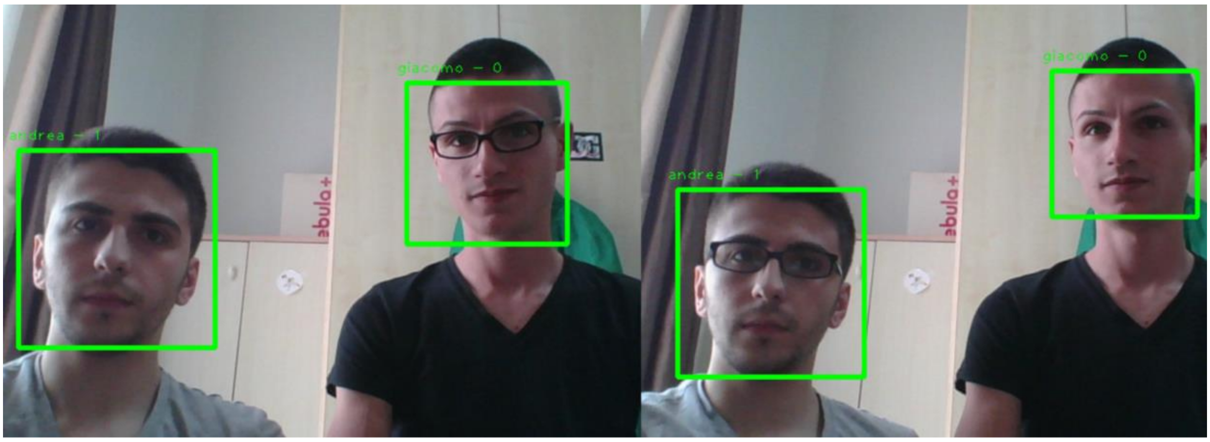

Figure 1.3. Example of face detection [9]

7.2. Face Alignment

For the computer, blurred faces look completely different. To normalise a face to match the face in the database requires an algorithm. Using a variety of generic facial landmarks is one way to achieve this. Lower chin, upper nose, outside of eyes, various places around eyes and lips, etc. are examples. The next step is to train a deep learning system to detect these spots on any face and rotate them towards the centre. This makes the face detection process easier.

Figure 4 shows an case of confront arrangement. It illustrates the method of altering and standardizing the position and introduction of facial highlights inside an picture. Confront arrangement procedures are utilized to move forward precision in facial investigation, acknowledgment, and other computer vision assignments that depend on exact facial point of interest situating.

Figure 1.4. Example of face alignment [10]

7.3. Face Measurement and Extraction

This phase consists of measuring and extracting many features of the face so that the algorithm can compare it to other faces in the database. However, it was initially difficult to know which features to collect and extract until the researchers realized that the best approach was to let the deep learning system decide for itself what data to collect.

Embedding is a technique that uses a deep convolutional neural network to teach itself to create a large number of measurements of a face, allowing it to be distinguished from other faces.
Figure 5 illustrates confront estimation and extraction. It includes quantitatively analysing facial highlights like separations, points, and extents. This innovation finds applications in biometrics, virtual try-on, and facial symmetry investigation.

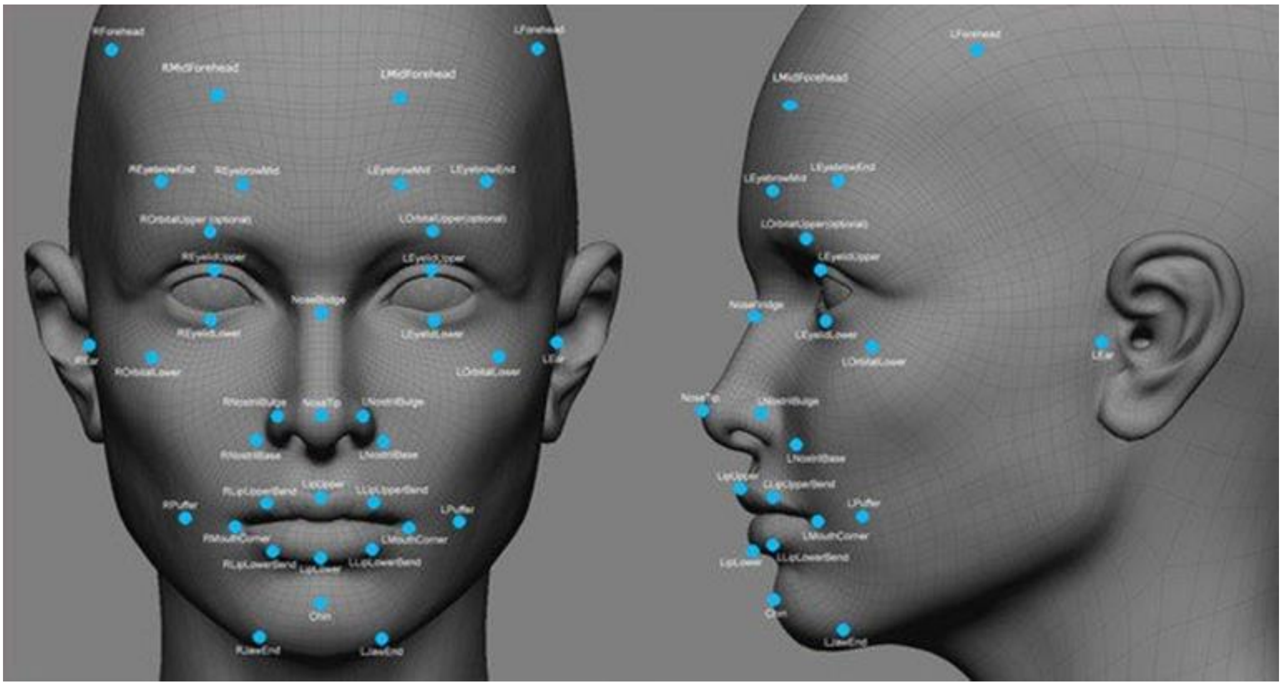

Figure 1.5. Face Measurement and Extraction [11]

7.4. Face Recognition

The final deep learning algorithm will use the unique dimensions of each face to compare the dimensions of each face to known faces in the database. The match will be the face in the database that is closest to the size of the face in question.

Figure 6 appears confront acknowledgment in activity. It distinguishes people based on their facial highlights, empowering get to control and personality confirmation.

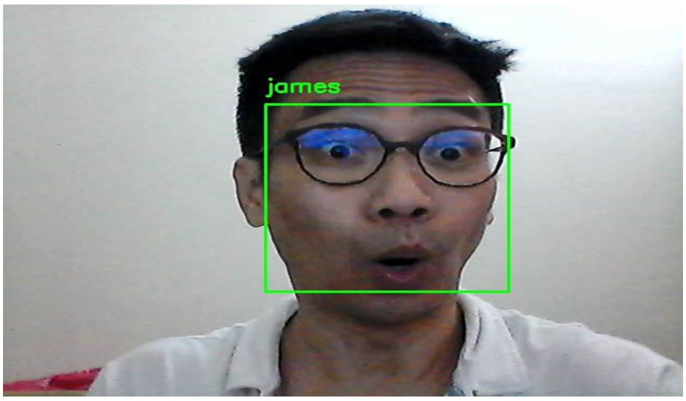

Figure 1.6. Example of face recognition [12]

7.5. Face Verification

Now finally the deep learning algorithm performs the last step of matching the face with other faces in the database. If the face matches, it is said to be verified, otherwise, it is still unverified. This step is called face verification. The faces are compared there to give the final result of a very long process. But this step is a bit more complicated.

Images can be compared to the database in two ways. If the acquired image and the images in the database are 3D images, the matching process will work without problems. However, since most government offices and other places use two-dimensional databases, comparisons are made more difficult.

3D image must be converted to 2D image before comparison. A 3D image will be vivid and lively compared to a still, stable 2D image.

So when a 3D image is captured, it is converted to 2D by taking measurements at different locations on the face. These measurements are then translated into an algorithmic form, creating a two-dimensional image. [13]

8. Techniques of face detection and recognition

Face detection and recognition techniques include classic computer vision algorithms, machine learning methods, and deep learning approaches. To detect and recognize faces, traditional computer vision algorithms employ characteristics such as skin colour, edge detection, and geometric forms. Unfortunately, the accuracy and durability of these approaches are restricted.

Machine learning approaches, on the other hand, entail training a model to detect patterns and characteristics in photos using a huge dataset of labelled faces. This method has yielded encouraging results, particularly when using algorithms such as Support Vector Machines (SVMs) and Random Forests. Yet, the quality of the training data may still restrict the effectiveness of these models. Some if these techniques are:

8.1. Histogram of oriented gradients (HOG)

The histogram of oriented gradients (HOG) is a feature extraction technique that is used in computer vision and image processing to detect objects and classify images (McConnell 1986; Leonardis et al. 2006). A feature extraction module is an image or image patch encoding that simplifies the picture by trying to extract one or more attributes. HOG is based on the concept that the local look and form of an item in an image may be well defined by the pattern of pixel intensity gradients or information, which are by definition principal axis to the gradient direction. [14]

Figure 7 portrays the steps of the Histogram of Arranged Slopes (Hoard) calculation. It diagrams the method of extricating include descriptors from an picture by analysing local slopes. These descriptors are at that point utilized for assignments like question discovery and picture classification, making Hoard a prevalent strategy in computer vision applications.

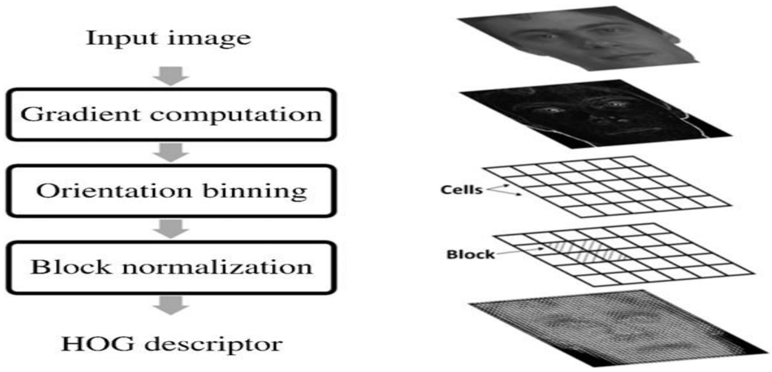

Figure 1.7. Steps of Histogram of oriented gradients [15]

8.2. Local Binary Patterns (LBP)

The local binary pattern operator is an effective texture description tool. The first version of the opera- tor identifies the pixels within a picture by thresholding the 3x3-neighbourhood within each pixel with the middle value and adding the weighted threshold values. The tag histogram may then be utilised as a texture.[16]

The initial LBP approach was developed as a level of reliability for local image contrast. It used the centre of eight surrounding pixels as a threshold. The final LBP code was created by multiplying the thresholder data by weights determined by powers of two and combining the results.[17] LBP, by definition, is stable toward any monotonic gray scale change and is computationally simple to calculate. The original LBP has indeed been developed to include an endless amount of nearby sample points.

Figure 8 shows the Local Binary Patterns (LBP) method. It visualizes the method of encoding surface data by comparing pixel local with their central values and making twofold designs. LBP is commonly utilized in facial acknowledgment, surface investigation, and protest discovery due to its viability in capturing local picture characteristics.

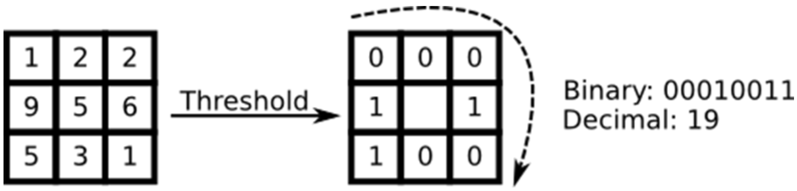

Figure 1.8. Representation of Local Binary Patterns [18]

8.3. Convolutional neural networks (CNNs)

A CNN or convnet is a specialized type of artificial neural network that falls under the umbrella of machine learning. Specifically designed for processing pixel data, CNNs are a popular network architecture for deep learning algorithms used in image recognition and other computer vision (CV) tasks [19]. Although there are several types of neural networks available for deep learning, CNNs are preferred for object identification and recognition, making them essential for applications like facial recognition and self-driving cars.

In a CNN, the input image is passed through a series of convolutional layers that apply filters to detect features such as edges, curves, and textures. These filters enable the network to learn hierarchical representations of the image that capture increasingly complex features as we move deeper into the network. Pooling layers are then used to reduce the spatial dimensions of the feature maps, followed by one or more fully connected layers that perform classification or regression tasks. CNNs have revolutionized the field of computer vision by achieving state-of-the-art performance in a variety of tasks, including image classification, object detection, and semantic segmentation [20]. They have also been applied to other domains such as natural language processing, speech recognition, and even drug discovery.

Figure 9 outlines the engineering of a Convolutional Neural Arrange (CNN). It exhibits the stream of data through convolutional layers, pooling layers, and completely associated layers. CNNs exceed expectations at handling grid-like information, such as pictures, and are broadly utilized in computer vision errands like picture classification, question discovery, and picture division.

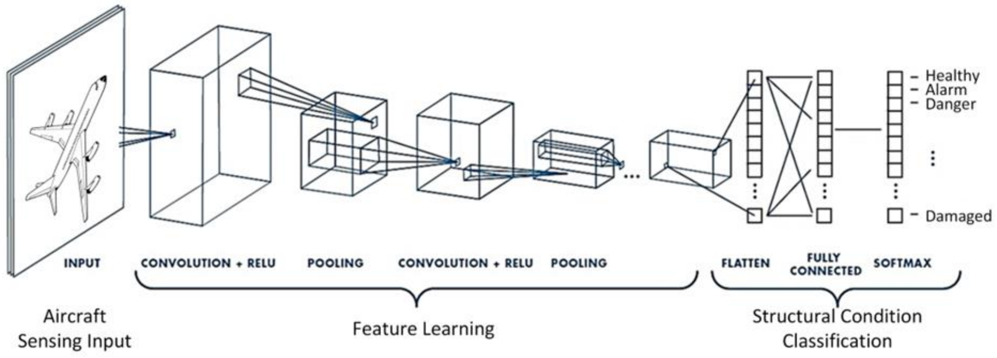

Figure 1.9. Convolutional neural network architecture [21]

8.4. Support Vector Machines (SVMs)

Cortes and Vapnik introduced SVMs in 1995 as a technique for binary classification [22]. Support vector machines (SVMs) are a versatile machine learning algorithm [23] that can be utilised for regression problems with three distinct types of outcomes. These include Bernoulli (binary), multinomial, and continuous outcomes. When performing regression with Bernoulli outcomes, the process is typically referred to as classification in statistical learning, as it involves predicting binary outcomes. In contrast, when using SVMs for regression with multinomial outcomes, it is known as multiclass classification, which involves predicting among multiple possible outcomes.

The goal of the support vector machine (SVM) [24] algorithm is to identify a hyperplane within an N-dimensional space (where N is the number of features) that can accurately separate the data points into distinct classes.

Figure 10 outlines a Bolster Vector Machine (SVM) graph. It exhibits how SVM isolates information focuses into diverse classes employing a hyperplane in a high-dimensional feature space. SVM could be a effective demonstrate utilized for classification and relapse assignments, known for its flexibility in taking care of different choice boundaries.

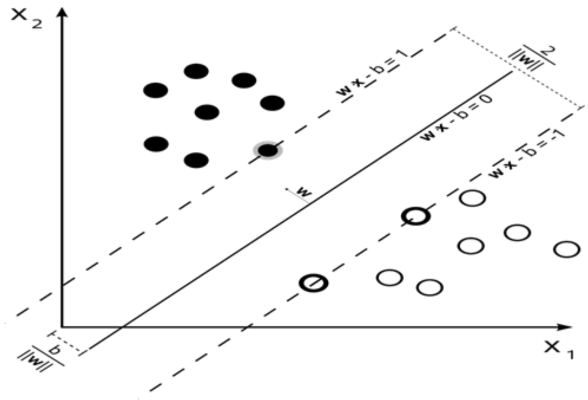

Figure 1.10. Diagram of support vector machine [25]

8.5. You only look once (YOLO)

The You Only Look Once (YOLO) algorithm has gained widespread popularity for its object detection capabilities. The original YOLO version was introduced by Redmon et al. in 2015 and has since been followed by several subsequent versions including YOLO V2, V3, V4, and V5 [26]. Additionally, there are also revised, limited versions of the algorithm such as YOLO-LITE that have been published by scholars over the past few years. The YOLO algorithm is a real-time object detection system that operates by dividing an image into a grid and predicting bounding boxes and class probabilities for each grid cell. Unlike other object detection algorithms that use region proposals and post-processing techniques, YOLO performs detection in a single forward pass of the neural network, which makes it faster and more efficient.

Figure 11 delineates the design of the YOLO (You Merely Look Once) show. YOLO may be a real-time protest location calculation that forms pictures in one pass, producing bounding box forecasts and lesson probabilities. It is broadly utilized in applications like independent driving and video observation.

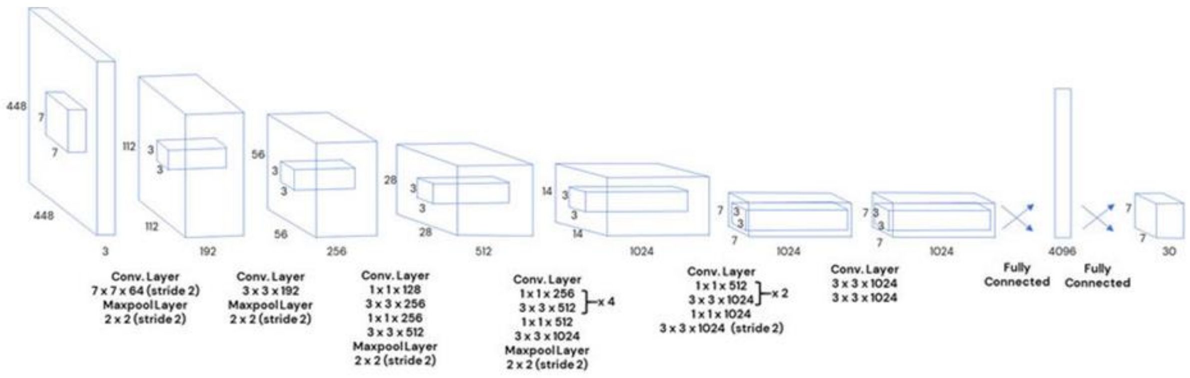

Figure 1.11. You Only Look Once architecture [27]

9. The difference between face recognition and face comparison

Face recognition and face comparison refer to two different ways of using the same technology. These two new techniques are used to verify that a source image of a face matches another stored or provided image. Face recognition assumes the existence of a multiple image recognition database against which the source image recognition data can be compared - and the source image is then recognized in the image store. Face comparison involves comparing two images of the same person.

9.1. Biometric authentication: the difference between facial recognition and facial matching technologies

Biometric authentication technologies such as facial recognition and facial comparison are very common in our daily life. We use fingerprint technology or facial recognition every day to unlock our smartphones with Face ID, and in airports around the world many of us now use identity verification, artificial intelligence and facial matching. to seamlessly pass passport control on our identity.

While some groups, such as the American Civil Liberties Union (ACLU), have expressed concerns about privacy, civil liberties and human rights if the technology is misused, it also offers business benefits. and cybersecurity with other biometric technologies. Other cities, including Boston and San Francisco, have even banned the technology, while New York, the FBI and the Department of Homeland Security have imposed strict regulations on its use. Facial detection and recognition systems are just one tool among many. Let us analyse and distinguish the definition of the use of facial recognition technology and face matching. [28]

Figure 12 Confront acknowledgment for biometric confirmation. Confirms and awards get to based on facial highlights, guaranteeing secure and helpful confirmation in different spaces.

Figure 1.12. Biometric authentication using face recognition [29]

10. Passport check

When you swipe your passport through the electronic door, it scans the page, including a photo of your face. The camera then takes a picture of you standing there and pairs the two up to make sure it's the same person.

This is because face recognition software reads your passport and the face geometry of the person at the door and takes into account the distance between parts of the face (such as nose-chin) and recognizes facial features, etc. Symantec, a leading cybersecurity company, has called it "the place". This creates what is called a "face signature".

A two-sided signature matching machine—like an employee comparing the signature on the back of the credit card they use to pay with the signature on the invoice.

If the face signature matches, the door will open and you can exit. Otherwise, you will receive a suspicious message and be sent to a check-in desk that can verify your face with your passport photo. Your immigration waiting time will be slightly longer.

If you have changed your body shape, such as your hair colour, after the photo was taken, this should not affect face recognition as the rest of your face is drawn. Your name is still the same.

However, after your passport photo has been taken, the machine may not match your face signature if you have changed your face, for example, due to an operation. If your passport photo has a bushy beard covering your chin and you're now clean-shaven, it will be difficult for the eGate to match both chins.

As a result; Face recognition technology is expanding. After British Airways has successfully implemented this at places like easyJet at Heathrow and Gatwick, you can expect more technology and less human interaction when flying in the UK. [30]

10.1. Importance of Face Recognition in Passport Verification

Due to globalisation and improved transportation, the number of people arriving and departing from airports to different countries has increased. An immigration checker is currently using his bare eyes to check passports. The purpose of immigration control is to find counterfeiters, criminals, illegal immigrants or people who are prohibited from leaving the country. The passport contains information such as the owner's photo ID, nationality, name, social security number, gender, passport number, etc.

It is very difficult to differentiate and monitor the migration process with the naked eye alone. The image area is extracted based on the centre of the code string area coordinate value. By calculating the distance value between the feature vector of the facial image of the passport and the feature vector of the facial image in the database, it can be determined whether the passport is forged.

11. Challenges and difficulties

As the scope of application expands day by day, the complexity of the system also increases. This actually affects the efficiency of the system. In this part of the article, we discuss the various challenges faced by facial recognition systems today. These challenges are related to facial images as input to the system. The algorithm used for this process varies from application to application.

The reasons for the face change are multiple. These sources of variation stem from two main factors. They are:

- **Intrinsic factors**

They are due to the physical properties of the face and do not depend on the observer. Intrinsic factors are further divided into interpersonal and interpersonal factors. Due to changes in an individual's appearance, such as age, facial expression and facial accessories (beard, makeup, glasses, etc.)

- **External factors**

Due to changes in the appearance of a person in Light and the face and the interaction of the observer. This will include exposure, pose, scale, and imaging settings (resolution, focus, imaging, noise, etc.).

- **Position variation**

Pose variation makes face detection quite difficult. Position changes may be due to changes in the observer's viewing angle or rotation of head position. These changes can cause serious problems with recognition of input images. Many systems can tolerate small changes, such as small angle rotations. But it will be difficult when it comes to large angles of rotation. Databases usually consist of frontal images of human faces. Since existing FRSs are very sensitive to pose changes, pose correction is essential and can be achieved with effective techniques aimed at rotating faces and/or aligning them with image axes.

- **Lighting changes**

Lighting changes may reduce the effectiveness of the FRS. Face detection and recognition is very difficult to achieve with moderate background lighting. Lighting changes can alter the total magnitude of light intensity reflected from an object. On the other hand, higher light levels can lead to overexposed faces and (partial) non-detection of facial patterns. There are now many algorithms, such as smoothing techniques, which can solve this problem to some extent. Sometimes it is even possible to use several algorithms in a facial recognition system to tolerate lighting problems. But in the case of range, you don't want to depend on these techniques.

- **Expression Changes**

Some changes in facial images may be due to differences in expression influenced by an individual's emotional state. Therefore, recognizing different facial expressions is important for assessing emotional states. Human expressions include macro expressions such as disgust, anger, happiness, fear, sadness, or surprise, as well as other involuntary and rapid facial expressions. These facial changes can be calculated using dense optical flow. Makeup and hairstyles can also be included in this challenge as changing hairstyles and makeup can also lead to changes in facial expressions.

- **Aging**

Another cause of changes in the appearance of a face can be the aging of the face, which affects the overall process of facial recognition; if the time between each image capture is long, then the person has changed considerably. According to various studies conducted by scientists, every 10 years the appearance of a person's face undergoes major changes. Fig. Figure 5 shows the variation of faces in different age groups. Not only does the shape and lines of the face change over time, but so does the hairstyle.

- **Occlusions**

Changes in the appearance of the face can also be caused by the presence of objects such as occlusions that partially cover the face. It is therefore difficult for the system to classify the images. Although the face is found, it can be difficult to identify the features due to some hidden facial parts. This challenge can be seen in real-world applications where people are talking on the phone, wearing glasses, scarves, hats, etc. or covering their face with their hands.

- **Similar Faces**

This is generally a less common challenge. But we've seen that even humans have trouble recognizing people with similar faces. Thus, one can imagine a difficult situation for a computer to recognize individuals with similar faces. Mostly identical twins with different facial features, shape, etc. similar. The identification of individuals by facial recognition systems is becoming a daunting task. This also leads to an increase in the false recognition rate (FRR).

- **Image Resolution**

Another significant problem with facial recognition systems is that input images vary in quality and resolution. Many factors affect the resolution of an image. The environment, quality of acquisition system performance, and many other reasons can all contribute to changes in image resolution. If the resolution is good, the recognition process will be easier and more efficient. We can therefore say that the resolution is directly proportional to the efficiency of the facial recognition system. [31]

12. Conclusion

In conclusion, face recognition and detection is a rapidly evolving field that has revolutionised the way we interact with technology. With its roots in the early days of computer vision research, this technology has come a long way, with advances in machine learning algorithms, computer processing power, and data storage. Today, face recognition and detection is being used in a wide variety of applications, from security and surveillance to social media and entertainment.

Despite its many benefits, however, face recognition and detection is not without its challenges. Privacy concerns, based algorithms, and the potential for misuse of the technology are just a few of the issues that need to be addressed. Nonetheless, with continued innovation and research, face recognition and detection holds the promise of many exciting new applications in the years to come.

CHAPTER 2

Tools and Technologies

1. Introduction

This chapter offers an overview of the techniques and technology used in the creation of this thesis's face recognition algorithm for passport inspections. The basis is provided by the discipline of computer vision, which provides important approaches and techniques for evaluating and processing face pictures. Python libraries, well-known for their vast computer vision frameworks, were used to implement the program's many functions. To improve the accuracy and efficiency of the facial recognition system, certain algorithms and models such as VGGFace, Local Binary Patterns (LBP), and MTCNN (Multi-Task Cascaded Convolutional Networks) were included.

2. Python libraries

Python is a commonly utilised high-level programming language in machine learning and artificial intelligence applications. It is well-known for its simplicity, readability, and adaptability. Python has a large library and framework ecosystem that makes it simple to develop complicated machine learning models, and many of these modules are open-source and free to use.

NumPy, Pandas, Scikit-learn, Keras, and TensorFlow are some of the most popular Python machine learning libraries. NumPy is a numerical calculation package that supports massive multi-dimensional arrays and matrices. Pandas is a data manipulation and analysis package that makes it simple to work with enormous datasets. Scikit-learn is a machine learning package including methods for classification, regression, and clustering. Keras is a high-level deep learning toolkit that connects to TensorFlow and makes it simple to create and train neural networks. TensorFlow is a sophisticated open-source toolkit that can be used to design and train deep learning models.

These modules and frameworks, when combined, provide a comprehensive set of tools for machine learning and data analysis in Python. They allow developers and data scientists to easily design advanced machine learning models as well as manage and analyse huge datasets. As a result, Python has become one of the most popular programming languages for machine learning and AI applications.

2.1. NumPy: Array computing for Python

NumPy is a key Python package for numerical computing. It is a very effective tool for performing mathematical operations on arrays and matrices. NumPy is a Python scientific computing core package that provides a variety of features for conducting mathematical operations on massive multi-dimensional arrays and matrices.[32]

2.2. Keras: Deep learning in Python

Keras is a Python-based open-source neural network library [33]. It's a high-level API that sits on top of deep learning packages like TensorFlow, CNTK, and Theano. Keras is a user-friendly, modular, and extendable deep neural network framework that enables for quick and easy experimentation. It is widely used in academia and business to create deep learning models for a broad range of applications such as image classification, object recognition, and natural language processing.

2.3. C. OpenCV: Computer vision in Python

OpenCV (Open Source Computer Vision Library) is a free and open-source software library for computer vision and machine learning that is mostly used for real-time image and video processing [34]. It is a robust and extensively used library that performs a variety of computer vision activities, including image and video processing, feature recognition, object detection and tracking, and machine learning. Because OpenCV is developed in C++ and includes a Python interface, it is simple to use and combine with other Python libraries.

2.4. TensorFlow: Machine learning library in Python

TensorFlow is an open-source machine learning library created in 2015 by Google Brain Team. It is intended to make the process of developing, training, and deploying machine learning models easier. Because of its versatility, scalability, and ability to operate on many platforms such as CPUs, GPUs, and mobile devices, TensorFlow is widely utilised in business and academics.

3. MTCNN

MTCNN, which stands for Multitasking Cascading Convolutional Neural Network, is a state-of-the-art neural network particularly created for recognising faces and facial areas inside pictures. It was initially developed in 2016 by Zhang et al., who demonstrated its usefulness in the field of face recognition.

The unique architecture of MTCNN distinguishes it and contributes to its exceptional performance. MTCNN, in contrast to traditional single-network approaches, combines three interconnected neural networks in a cascading fashion. This cascade architecture enables a progressive refinement of face detection results, resulting in improved accuracy and robustness. [35]

Figure 13 illustrates how MTCNN adjusts a ordinary picture. MTCNN employments a cascaded engineering to identify and adjust faces precisely by finding facial points of interest, such as eyes, nose, and mouth. This guarantees appropriate arrangement for consequent investigation assignments, such as confront acknowledgment and feeling discovery.

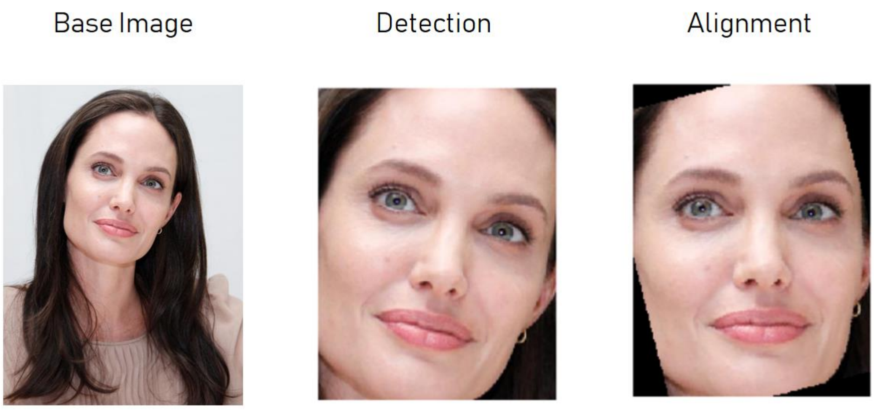

Figure 2.1. how does MTCNN align a normal image [36]

3.1. How we used it

We used MTCNN (Multitasking Cascading Convolutional Neural Network) as a critical tool for face identification and localization in our facial recognition algorithm. By adding MTCNN into our system, we were able to automatically recognize and extract face areas from input photos.

After initializing the MTCNN detector, we ran the 'detect_faces()' function on the loaded picture to acquire the precise coordinates of the bounding box for each recognized face.

These coordinates were then used to crop the picture and isolate the area of interest on the face. Through this procedure, MTCNN supplied us with an efficient and dependable solution for precise face identification, providing the backbone of our facial recognition pipeline.

3.2. How MTCNN works

The MTCNN model is proposed in the paper "Face Integration and Alignment Using Multitasking Cascading Convolutional Networks" by Zhang, Zhang, and Zhifeng. The main purpose of this model is to identify its face and features, and it has three levels of communication links.

This model is created using multiple learning tasks to split the search task of faces and features because doing so in many tasks improves details using material specific documentation included in training tasks on signals made using it. The shared representation of knowledge is done in parallel with the training tasks. The first step before the startup phase is to take a picture and change it at different scales to create a pyramid, which is to enter the first part of the network.

The first stage consists of a fully integrated network (FCN), which we use instead of CNN because it does not have a dense layer in its structure. The wiring diagram is used to obtain human faces and bounding boxes. The benefit of this network is that we can all meet potential candidates.

In the second stage of the network, the output of the first stage is used as the input of the CNN network, which is called the "purification network (R-Net)", which reduces the number of participants tw, calibrates the bounding box. and executing a Non-Competitive Strategy (NMS) to unify our potential competitors. According to the output, it tells us whether the input is a face or not.

Finally, the third is similar to R-Net, which gives us the position of the opponent's eyes, nose and mouth. [37]

4. LBP

Local Binary Patterns (LBP) are a popular texture descriptor in computer vision that captures local information by comparing the intensity values of a centre pixel with the intensity values of its nearby pixels. LBP efficiently portrays texture changes within an image region by creating binary patterns based on these comparisons. It has applications in a wide range of fields, including facial identification, picture retrieval, and object detection. With the availability of many neighbourhood topologies and variations, LBP can adapt to capture textures of diverse sizes and orientations, increasing its versatility and application.

LBP's resistance to lighting fluctuations is one of its primary features, allowing it to efficiently evaluate pictures influenced by diverse lighting situations. LBP enhances texture class classification by recording local texture patterns rather than depending exclusively on intensity values. Another benefit is that LBP is computationally efficient, as it employs basic bitwise operations and requires few CPU resources. Because of its efficiency, LBP is well suited for real-time or resource-constrained applications. Furthermore, LBP has strong generalization and noise resistance, giving it a dependable alternative for numerous detection and analysis tasks in computer vision.

4.1. Local Binary Patterns for face recognition

The integration of Local Binary Patterns (LBP) with the feature extraction process, such as using VGGFace, provides a complete technique for obtaining facial information in face recognition. The face recognition system may benefit from both global and local texture features by combining the strengths of deep learning models with LBP. LBP may be used to encode the local texture changes of face areas after extracting high-level features with VGGFace. This combination improves the discriminative strength of the feature representation by including fine-grained features that the deep learning model alone may not completely capture. [38]

The use of LBP in face recognition has various advantages. For starters, because LBP is naturally resilient to lighting conditions, facial emotions, and slight position alterations, it enhances the system's resilience to these aspects. The local texture information encoded by LBP helps to provide more extensive and accurate face representations, allowing for improved individual differentiation. Second, combining LBP with VGGFace-based feature extraction improves the face recognition system's overall performance in real-world circumstances. By combining the capabilities of both methodologies, the system becomes more capable of tolerating fluctuations and obtaining greater identification accuracy, which is especially significant in applications like passport checks and identity verification.

Figure 14 gives an illustration of Local Binary Patterns (LBP) in activity. It grandstands the encoding of surface data by comparing pixel neighborhoods with their central values, coming about in double designs. LBP is commonly utilized in surface investigation, facial acknowledgment, and question discovery errands.

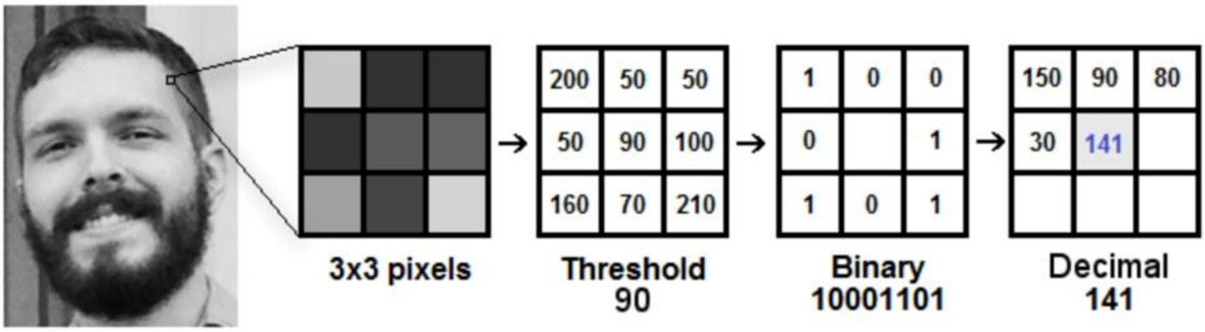

Figure 2.2. local binary patterns example [39]

5. VGGFace

VGGFace is a well-known deep learning model that was created exclusively for facial recognition applications. Its VGGNet-based architecture has gained popularity because of its fast performance and ability to extract highly discriminative features from face photos. In this one-page talk, we will look at the architecture of VGGFace and examine the advantages of utilizing it for facial recognition applications.

VGGFace's architecture is built on a deep convolutional neural network (CNN) with many layers. It is commonly made up of 16 or 19 weight layers, which include convolutional layers, pooling layers, and fully linked layers. Convolutional layers capture low-level characteristics such as edges, textures, and facial structures by using tiny receptive fields and many filters. Pooling layers minimize spatial dimensions while retaining crucial properties. The network's fully linked layers at the network's conclusion contribute to the final classification by learning high-level representations and correlations between various face traits.

One of the most significant advantages of VGGFace is its capacity to learn highly discriminative characteristics from face photos. VGGFace can capture fine features and complicated patterns that are critical for successful face recognition by using its deep architecture. The model's hierarchical structure allows for the extraction of both low-level and high-level characteristics, allowing it to represent a wide range of face qualities such as forms, textures, and distinguishing landmarks. This detailed feature representation improves identification accuracy and resilience in face recognition tasks.

VGGFace has also been trained on large-scale face datasets, yielding a pre-trained model that has learnt representations from a huge number of face photos. This pre-training improves the model's generalization skills, allowing it to perform well on unseen faces and face variants. Furthermore, the availability of pre-trained weights simplifies the transfer learning process, allowing researchers and developers to apply what they've learned from VGGFace to their own facial recognition applications. This saves time and computing resources that would otherwise be necessary to train a deep neural network from the ground up.

Figure 15 outlines the component of VGGFace. VGGFace may be a profound learning show that extricates facial highlights by utilizing a convolutional neural arrange (CNN) design. It empowers exact confront acknowledgment and confirmation by learning discriminative representations from facial pictures, contributing to headways in biometrics and character acknowledgment frameworks.

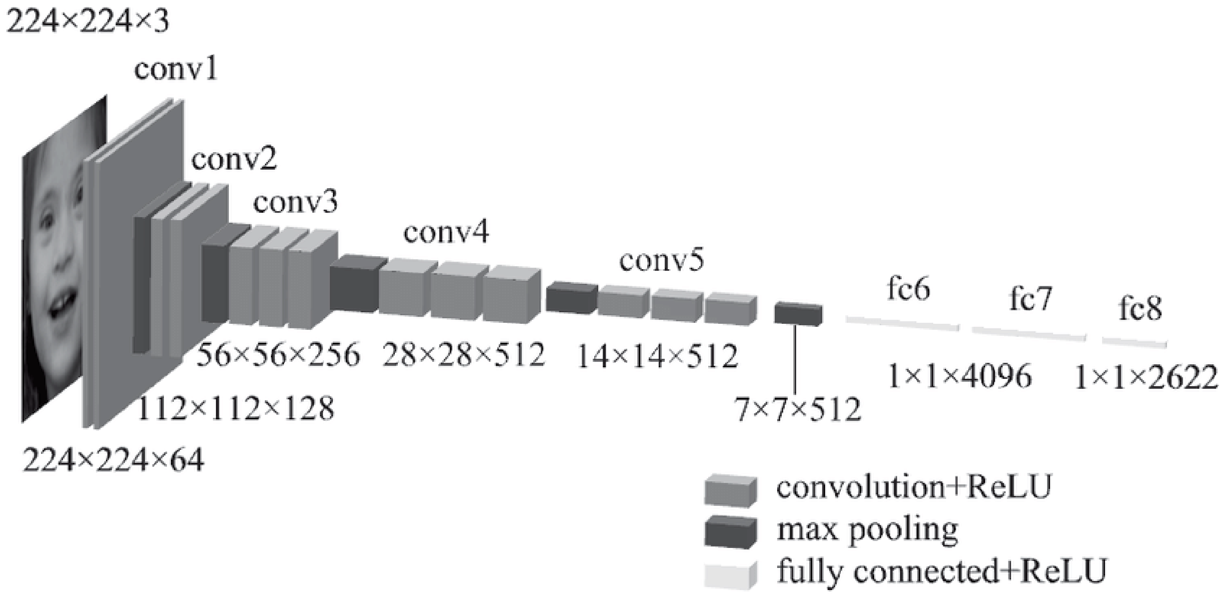

Figure 2.3. VGGFace mechanism [40]

6. Siamese Neural Network

The Siamese Neural Network is a customized design used for evaluating similarity between pairs of inputs. It is very useful in facial recognition tasks, where the aim is to compare and identify the similarity of two faces. The architecture is made up of two identical subnetworks, known as Siamese twins, that have the same weights and settings. Each subnetwork obtains a feature representation from an input face picture. The similarity score is computed by comparing these feature representations using a distance metric.

The Siamese architecture is intended to discover distinguishing elements that capture the essence of face features. The network is encouraged to extract similar characteristics from faces belonging to the same identity and dissimilar features from faces belonging to other identities by exchanging weights between subnetworks. This enables the network to learn a mapping that projects faces into a high-dimensional feature space, where the similarity between two faces can be assessed precisely. The Siamese Neural Network design has been shown to be effective at capturing minor distinctions and similarities between faces, making it a useful tool for face recognition tasks.

Figure 16 appears the engineering of a Spiking Neural Organize (SNN). SNNs recreate organic neurons utilizing spikes and handle transient data with event-driven computation.

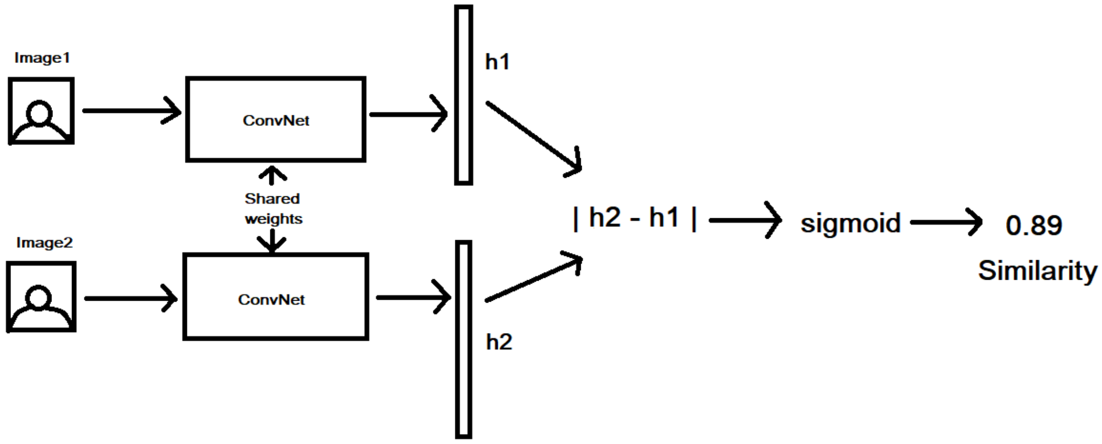

Figure 2.4. The architecture of SNN [41]

6.1. The use of SNN in Face recognition

The Siamese Neural Network (SNN) was chosen for its unique capabilities and benefits in assessing the similarity of extracted face traits. Following the early phases of face detection, feature extraction and feature encoding with MTCNN, VGGFace, and LBP, the SNN was essential in assessing the degree of similarity between the derived feature representations.

One important reason for using the SNN is its capacity to efficiently handle the high-dimensional feature embeddings obtained from the combination of MTCNN, VGGFace, and LBP. These approaches capture several components of the face appearance, such as facial landmarks, global facial traits, and local texture patterns. Using the SNN, it is able to compare these feature vectors in a unified manner and produce a full similarity score.

Furthermore, the SNN excels at identifying minor variances and similarities between faces, making it ideal for face recognition tasks. Because of the SNN's shared weight design, both faces in a pair are processed by identical subnetworks, allowing for the extraction of highly discriminative features. By training the network on a large number of different datasets, it learns to project faces into a feature space where similar faces are grouped together, allowing for precise similarity assessment.

7. Conclusion

Finally, this chapter offered a thorough review of the techniques and technology used in the creation of facial recognition software for passport checks. The application was provided with substantial capabilities for accurate face recognition by using the power of computer vision methods, Python libraries, and adding specific algorithms and models such as VGGFace, LBP, MTCNN, and the Siamese Neural Network. These techniques and technologies have greatly improved the efficiency and efficacy of the passport verification system, thereby boosting the security and reliability of identity identifying operations.

CHAPTER 3

Experimental Design

1. Introduction to the Dataset

1.1. Overview of the dataset used for training and testing

The confront acknowledgment AI dataset comprises a collection of pictures of people's faces, which is utilized for both preparing and testing the AI framework. The dataset is carefully curated to speak to an assorted extent of people, counting individuals from diverse sexual orientations, ethnicities, ages, and foundations. It points to capture varieties in posture, expression, lighting conditions, and occlusions to guarantee strength in confront acknowledgment.

1.2. Description of the data collection process

The dataset is regularly partitioned into two subsets: the preparing set and the testing set. The preparing set is used to prepare the AI show and is altogether bigger in measure compared to the testing set. It contains an endless number of labelled confront pictures, where each picture is related with the personality of the individual portrayed. These names empower the AI to memorize the one of a kind facial characteristics of diverse people and construct a dependable representation of their faces.

To improve the execution of the AI framework, the preparing set is frequently expanded with different changes such as revolutions, interpretations, and scaling. This enlargement increments the dataset's estimate and presents extra varieties, empowering the AI show to generalize way better and handle distinctive real-world scenarios.

The testing set may be a partitioned subset of the dataset, and it is utilized to assess the execution of the prepared AI show. It comprises unlabelled confront pictures that the demonstration has not experienced amid preparation. The testing set is carefully outlined to survey the model's capacity to accurately recognize people and generalize well to concealed faces.

To guarantee fair assessment, the testing set may incorporate challenging cases like faces with halfway impediment, shifting brightening conditions, and diverse postures. Also, the testing set may too contain pictures from diverse sources, such as diverse cameras or picture resolutions, to evaluate the model's strength to space shifts.

Generally, the dataset utilized for preparing and testing the confront acknowledgment AI includes an assorted extent of faces, with cautious consideration to varieties in posture, expression, lighting, and occlusions. This comprehensive dataset empowers the AI to memorize and generalize from different facial characteristics, driving to precise and dependable confront acknowledgment execution.

1.3. Statistics and characteristics of the dataset

The dataset for confront acknowledgment regularly includes:

- An expansive number of confront pictures for preparing and testing.
- Diverse subjects speaking to diverse sexual orientations, ethnicities, ages, and backgrounds.
- Varied posture, expression, lighting, and occlusions to guarantee robustness.
- High-resolution pictures for point by point facial information.
- Augmentation methods to extend dataset variability.
- A train/test part for show assessment.

2. Data preprocessing

2.1. Data Labelling and Reorganisation

Dataset Organization: The first step is to arrange the dataset, which comprises photographs of people, into a certain folder structure. Each person's photographs are organized into separate folders, with each folder representing a different person. This structure facilitates easy access to and processing of pictures during labelling.

True Pair Labelling: True pairs are photographs of the same individual. The code iterates over each person's folder to label the true pairings. It determines if the folder has at least two images. If so, it chooses two photos at random from that folder and reads them using OpenCV. The next stage is to identify faces in the photos using a face detection algorithm. If at least one identified face appears in both photos, the pair is termed a real pair. The pair's file paths are saved, together with a label of 1, indicating that they belong to the same individual.

False Pair Labelling: The false pairings are photos of distinct people. The code chooses two folders at random from the dataset. It then chooses one image at random from each of these two directories and reads it with OpenCV. The face detection algorithm, like genuine pair labelling, is used to determine if both photos include at least one detected face. If they do, the pair is said to be a false pair. The pair's file paths are saved, along with a label of 0, indicating that the pair belongs to distinct people.

Pair Selection: Based on the thesis criteria, the number of true and false pairs to label is predetermined. Using a loop and inspecting the current count of created pairs, the code guarantees that the desired number of pairs are formed.

Pair Storage: True and false pairs are kept in distinct lists. These lists are then concatenated to form a single list that contains both types of pairings.

Pair shuffle: After putting the true and false pairs in the same list, shuffle the pairs to mix them randomly. The sequence of the pairings is shuffled to guarantee that the remaining processes, such as partitioning the dataset and training the model, are not affected.

Splitting Datasets: The total list of pairings, all pairs, is split into training, validation, and test sets. The split is usually done at random to ensure that each set has a representative distribution of true and false pairings. The most typical training-validation-testing split ratio is 70% for training, 20% for validation, and 10% for testing. This ratio can be modified to meet the unique needs of the thesis.

Data Storage: To keep the split pairs structured, they are often saved in distinct files or data structures. Training pairings, for example, can be saved in a train.csv file, validation pairs in a validation.csv file, and test pairs in a test.csv file.

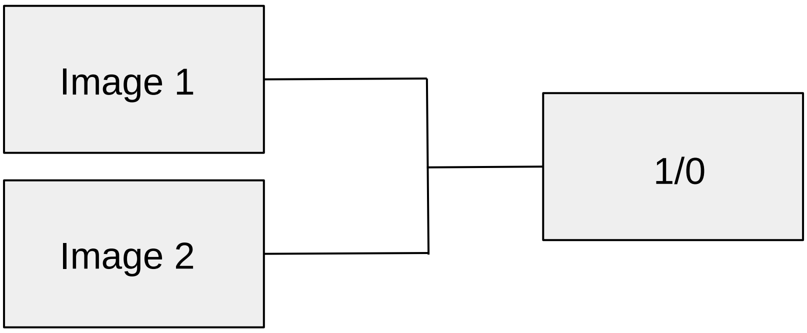

Figure 3.1. Data Shape Presentation.

Figure 18 shows picture sets. It exhibits two pictures side by side, highlighting a before-and-after comparison, distinctive perspectives, or any other related visual data. Picture sets are commonly utilized in errands like picture upgrade, picture enrolment, and visual review to demonstrate changes or varieties in visual information.

Figure 3.2. Image pairs [42]

2.2. Preprocessing Methods and Techniques

Image scaling is one of the code's preprocessing procedures. The preprocess_image function takes an input image and resizes it to a given target size. In our scenario, the cv2.resize function is used to resize the image using the supplied target size of (224, 224). Resizing the picture is a typical preprocessing step that is frequently used to standardize the input size for neural network models or other machine learning methods. In this context, scaling guarantees that the photos have consistent dimensions, making further processing and analysis easier.

As part of the preparation phase, we also include picture normalization. The normalized_image variable is produced after resizing the image by dividing the scaled picture by 255.0. Normalization is a technique used in image processing activities to scale the pixel values of an image between 0 and 1. Normalizing the image accounts for any fluctuations in pixel intensity across multiple photos, allowing models to better comprehend and handle the data.

Another approach we used for preprocessing was to convert the normalized picture to a NumPy array. This is accomplished using the np.array function, which generates a NumPy array from the normalized_image. NumPy arrays are extensively used in scientific computing because they allow efficient numerical data storage and manipulation. By converting the image to a NumPy array, it becomes compatible with a variety of machine learning libraries and algorithms that need this type of input data. The np.expand_dims method is also used to add an extra dimension to the array, which is frequently required for feeding data into neural network models.

Preprocessing techniques in the code include scaling the picture to a desired size, normalizing the pixel values, and turning the image to a NumPy array. These approaches aid with the preparation of picture data for later processing, such as machine learning model training or inference. Preprocessing is critical for increasing the quality of input data, assuring compatibility with the chosen algorithms, and boosting overall system performance.

3. Model Architecture and Selection

3.1. Overview of the existing literature and state-of-the-art approaches in facial recognition

Facial recognition is a fast expanding area that has seen significant research and breakthroughs in both deep learning-based and classical techniques. In this overview, we will concentrate on the literature and cutting-edge methodologies in face recognition, emphasizing the accuracy ranges reported in important research. We will specifically look at deep learning-based methodologies as well as classic ones, offering insights into their performance and advances.

- **Deep Learning-Based Approaches:**

Deep learning has transformed facial recognition by harnessing neural network capability to learn hierarchical representations straight from data. FaceNet, DeepFace, and ArcFace are three notable deep learning-based techniques that have achieved outstanding accuracy rates.

- **Traditional Approaches:**

Face detection, feature extraction, and matching are all processes in traditional facial recognition algorithms. Eigenfaces, Fisherfaces, and Local Binary Patterns (LBP) are examples of classical techniques. While these methods may not be as accurate as deep learning-based alternatives, they have benefits and historical value in the industry.

In general: deep learning-based technologies have advanced facial recognition accuracy to unprecedented heights. Some models have achieved outstanding accuracy rates, frequently exceeding 97% on benchmark datasets. While conventional procedures have historical value, they are often less accurate and struggle with differences in stance, emotion, and lighting.

When evaluating face recognition systems for real-world applications, it is critical to examine the larger context, which includes computing efficiency, robustness, and ethical issues. Ongoing research strives to overcome issues in facial recognition, improve accuracy, and ensure that these technologies are deployed responsibly.

3.2. Description of the three models implemented

We implemented 3 different models:

3.2.1. The first model

In this section, we will go through how to build a face recognition model that combines a fine-tuned VGGFace network with a Siamese neural network. The goal of this model is to extract face traits from photos and determine whether or not the photographs belong to the same person. This method makes use of VGGFace's sophisticated representation learning capabilities as well as the Siamese architecture's similarity-based learning.

VGGFace is a deep Convolutional Neural Network (CNN) model that has been developed exclusively for facial recognition tasks. It is well-known for its ability to extract discriminative face traits. We use the VGGFace model as a basis model in our implementation, exploiting its learnt feature representations to improve the SNN's efficacy.

The SNN architecture is made up of two identical subnetworks, often known as siamese twins, with the same weights. Each subnetwork individually collects features from an input face picture using the VGGFace model. The subnetworks are encouraged to learn comparable feature representations for matching facial pictures by sharing the weights.

After that, the collected characteristics from both subnetworks are integrated into a single representation. Concatenating the feature vectors or using element-wise procedures can accomplish this. The combined features are then passed into a classification layer, which decides how similar the input pictures are.

During the training phase, the SNN is fed pairs of face photographs together with labels indicating whether or not the photos belong to the same individual. The network learns to optimize an appropriate loss function that promotes comparable face pictures in the feature space to have smaller distances or greater similarities, while pushing dissimilar images away.

Contrastive loss and triplet loss are two often used loss functions for SNNs. Contrastive loss penalizes dissimilar picture pairings that are too near together by a certain margin, while encouraging comparable pairs to be too close together. The boundary between an anchor image, a positive image (belonging to the same person as the anchor), and a negative image (belonging to a different person) is enforced by triplet loss. The network learns to reduce the distance between the anchor and the positive pair while increasing the distance between the anchor and the negative pair.

To evaluate whether two faces belong to the same individual, the similarity of their feature representations must be considered. This is commonly accomplished by computing a similarity measure between the combined feature vectors, such as Euclidean distance or cosine similarity. The photos are then classified based on a threshold.

The threshold value is critical in deciding how to balance false positives and false negatives. It should be carefully chosen depending on the application's or dataset's particular requirements and features.

The accuracy of the Siamese Neural Network developed with VGGFace as the foundation model was 55%. While this accuracy falls short of the target level, it does provide a chance to identify and address possible areas for improvement.

To enhance the accuracy of the facial recognition system, we propose incorporating a face detection step, as well as separating the feature extraction and decision-making processes, allowing for more precise and focused analysis.

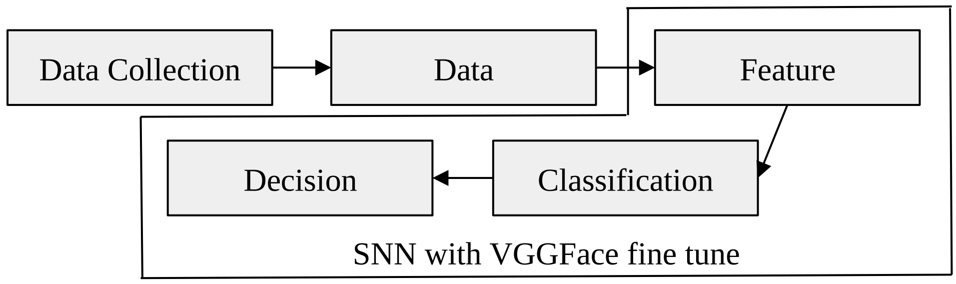

Figure 3.3. 1st model process.

3.2.2. The second model

We saw a considerable gain in accuracy after combining the face identification stage with MTCNN, isolating the feature extraction from the decision-making process, and implementing the bespoke Siamese Neural Network (SNN). The second model's accuracy climbed from 55% to an astonishing 89%.

By including the MTCNN face detection phase, only valid facial areas were selected for further processing. This reduced noise and unnecessary information, allowing the succeeding stages to concentrate entirely on the face characteristics. The feature extraction procedure got more exact and significant by separating the face areas.

Using the VGGFace model for feature extraction improved the accuracy of collecting distinguishable face characteristics. VGGFace has been particularly trained for face recognition tasks, allowing it to extract high-level feature representations required for reliable similarity computations.

The comparison of the feature representations extracted by the VGGFace model was aided by the proprietary Siamese Neural Network (SNN) architecture. The network learns to differentiate between similar and dissimilar face pairings well by training it with appropriate loss functions and tuning its parameters. The decision-making approach enhanced the accuracy of detecting whether two photos belonged to the same individual by applying a threshold to the estimated similarity.

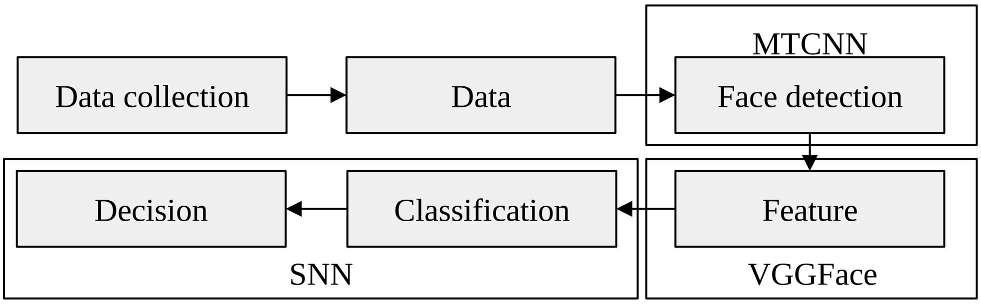

Figure 3.4. 2nd model process.

3.2.3. The third model

In the third model, we made various changes to improve the accuracy of the facial recognition algorithm.

To begin fine-tuning the VGGFace model, we used our own dataset to update the weights of the pre-trained model. This technique enabled the model to acquire additional discriminative characteristics related to our facial recognition job, enhancing its capacity to extract useful information from input photographs.

The use of Local Binary Patterns (LBP) as a feature encoding approach improved the depiction of face texture much further. LBP analyzes an image's local pixel intensity fluctuations and encodes them into a binary pattern. We collected additional texture characteristics that can be critical for differentiating between individuals by introducing LBP into our feature extraction method.

To calculate the similarity between the feature vectors of two face pictures, we used a similarity function rather than an SNN. Various similarity measures, such as Euclidean distance, cosine similarity, and correlation coefficients, can be employed. The similarity function measures the similarity of feature representations, resulting in a measure of similarity between facial pictures.

We manually changed the threshold for decision making based on testing and evaluation of the system's performance. The similarity value over which the algorithm classifies the photographs as belonging to the same individual is determined by the threshold. We adjusted the trade-off between false positives and false negatives by carefully adjusting and selecting an acceptable threshold, resulting in enhanced accuracy.

These changes, which included fine-tuning the VGGFace model, integrating LBP for feature encoding, using a similarity function to calculate similarity, and manually changing the decision threshold, resulted in a significant boost in accuracy to 95%. The combination of these strategies produced a very accurate face recognition system that outperformed earlier models and had excellent accuracy in detecting whether two facial photos belonged to the same individual.

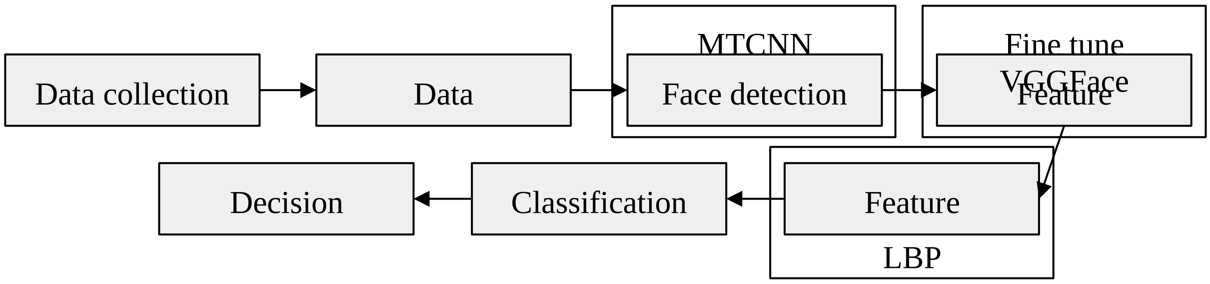

Figure 3.5. 3rd model process.

3.3. Explanation of the rationale behind choosing these models

Several variables and considerations influenced the choice to choose the third model, which required fine-tuning the VGGFace model, including LBP for feature encoding, and constructing a similarity function with manual threshold modification.

Relevance to State-of-the-Art (SOTA): The VGGFace model is widely regarded as a cutting-edge architecture for face recognition problems. We aligned our technique with the current breakthroughs in facial recognition research by fine-tuning this pre-trained model's outstanding feature extraction capabilities while tailoring it to our unique dataset.

Enhancing Feature Representation: We hoped to capture finer texture characteristics in the facial photos by integrating LBP as a feature encoding approach. LBP has demonstrated promising results in a number of computer vision applications, including face recognition. Its SOTA relevance and ability to collect local texture information make it an appropriate candidate for increasing feature representation in our model.

Efficiency and Simplicity: Rather than employing a complicated Siamese Neural Network (SNN), we chose a simplified similarity function technique. This modification not only simplified the model but also made it more interpretable and easier to deploy. Manual threshold modification gave us direct control over the decision-making process, allowing us to fine-tune for best performance.

Considerations for Accuracy: The choice of the third model was also affected by its performance. This model shows a tremendous improvement over prior techniques, with an accuracy of 95%. The excellent accuracy obtained with this combination of strategies demonstrated the efficacy and applicability of the models used for our facial recognition job.

We chose these models to achieve a compromise between exploiting cutting-edge approaches, improving feature representation, keeping simplicity, and obtaining high accuracy. The selection was motivated by the goal of constructing a powerful and accurate face recognition system that was in line with the most recent breakthroughs in the field while being practical and efficient in its deployment.

3.4. Training/validation split and cross-validation techniques employed

We used 70% of the dataset to train our third model for the training/validation split. We optimized the model's performance by iteratively adjusting its parameters depending on the training data during the training phase. As the validation set, the remaining 20% of the dataset was set aside. We were able to monitor the model's development, spot potential overfitting, and make educated judgments about hyperparameter tweaking or model selection thanks to this independent validation set.

In addition, we set aside 10% of the dataset for testing purposes. This testing set was kept hidden until the final phases of model development and served as an unbiased evaluation of the performance of our third model. We guaranteed a credible assessment of the model's generalization capacity to unseen data by keeping the testing set completely distinct from the training and validation stages.

While we did not specifically specify cross-validation in our method, the addition of a distinct testing set served as an independent review. It enabled us to appropriately measure the ultimate performance of our third model. It's worth mentioning that cross-validation can yield more robust performance estimates in some cases, but for our investigation, we used a distinct testing set.

In summary, we used a training/validation ratio of 70% training and 20% validation. In addition, we evaluated the final performance of our third model using a separate testing set comprising 10% of the dataset.

4. Experimental Results and Analysis

4.1. Presentation of the accuracy achieved by each model

Model 1 (VGGFace + Siamese Neural Network)

Initial accuracy: 55%

The accuracy of this model, which used VGGFace as the basic model and built a Siamese Neural Network, was 55%. This accuracy, however, was rather poor, showing that the model had difficulties differentiating between similar and distinct faces.

Model 2 (MTCNN + VGGFace + Siamese Neural Network)

Improved accuracy: 89%

The accuracy increased to 89% after incorporating the MTCNN face detection phase and separating the feature extraction and decision-making processes. The use of MTCNN for face identification almost certainly improved the quality of the input data, resulting in greater feature extraction and more accurate decision-making in the Siamese Neural Network.

Model 3 (VGGFace + LBP + Manual Threshold)

Higher accuracy: 95%

The third model, which improved VGGFace and used Local Binary Patterns (LBP) for feature encoding, obtained an astonishing 95% accuracy. To determine the similarity between feature encodings, the Siamese Neural Network was substituted with a similarity function, and a manually specified threshold was employed for decision-making. This strategy, together with the inclusion of LBP features, resulted in a significant gain in accuracy over earlier models.

4.2. Different hyper-parameters

Table 2 presents diverse hyperparameters, counting the learning rate, bunch estimate, precision, and F1 score. It gives a comparative examination of different hyperparameter settings and their comparing execution measurements. This table helps in understanding the effect of diverse hyperparameter choices on show exactness and F1 score, making a difference to optimize show preparation and execution.

Optimizer	Learning rate	Batch size	Accuracy	F1 Score
SGD	0.1	16	92%	0.926
SGD	0.6	16	95%	0.9524
Adam	0.6	20	60%	0.71
Adam	0.06	16	72%	0.73
Nadam	0.1	16	82.5%	0.8511
Nadam	0.01	10	92.5%	0.9268

Table 3.1. Different hyperparameters.

4.3. Experimental results obtained using different threshold values

The process of determining the appropriate threshold for a classification model entails analyzing the model's performance at various threshold settings. In this scenario, the threshold range is set at 0.5 to 0.9, with a 0.0001 step size. This indicates that we will iteratively increase the threshold value from 0.5 to 0.9 by 0.0001.

We turn the predicted probabilities into binary predictions for each threshold value. If the anticipated probability is greater than or equal to the threshold, the label is set to 1; otherwise, it is set to 0. This allows us to see how the model performs at various categorization levels.

Following that, we compute two evaluation metrics: accuracy and F1 score. Accuracy assesses the total accuracy of the forecasts, whereas the F1 score takes precision and recall into account. For each threshold, we save the calculated accuracy and F1 scores.

We draw two independent graphs to better understand how the model's performance differs with different thresholds. The accuracy scores are shown in one graph when the threshold is changed, while the F1 scores are shown in the other. The x-axis displays the threshold values, while the y-axis displays the matching accuracy or F1 scores. These charts show how the model's performance changes as the threshold is increased or decreased.

We determine the optimum accuracy and its related threshold by obtaining the greatest accuracy score after assessing all threshold settings. Similarly, we discover the largest F1 score and its associated threshold to establish the best F1 score and its related threshold. These numbers represent the model's best performance for the specified evaluation measures.

In conclusion, the procedure entails iterating over a set of threshold values, converting probabilities to binary predictions, calculating accuracy and F1 score, displaying the results, and identifying the ideal threshold that optimizes the selected performance parameter.

Figure 22 represents precision and F1 Score versus limit. It appears the relationship between execution measurements with shifting limit values, supporting in deciding the ideal edge for classification errands.

Best Accuracy: 0.9500 (Threshold: 0.771)

Best F1 Score: 0.9524 (Threshold: 0.771)

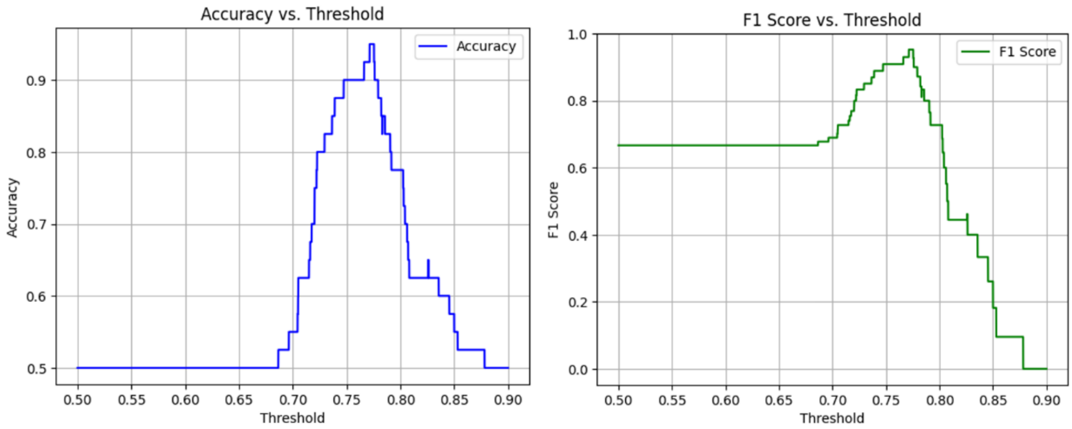

Figure 3.6. (Accuracy & F1 Score) vs threshold.

We used the Receiver Operating Characteristic (ROC) curve analysis to assess the performance of our facial recognition system (Figure 23). The ROC curve is a graphical depiction of the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) at various categorization thresholds. The Area Under the Curve (AUC) is a frequently used statistic generated from the ROC curve that assesses the classifier's overall performance.

AUC of 0.5 suggests random categorization, whereas AUC of 1.0 shows flawless classification. In our studies, we acquired an outstanding AUC value of 0.94, showing that the system has a strong discriminating power. This shows that our facial recognition algorithm distinguishes between authentic and imposter identities very well, with a low false positive rate and a high true positive rate.

The obtained AUC value of 0.94 demonstrates the efficacy and dependability of our suggested technique, confirming its appropriateness for real-world applications.

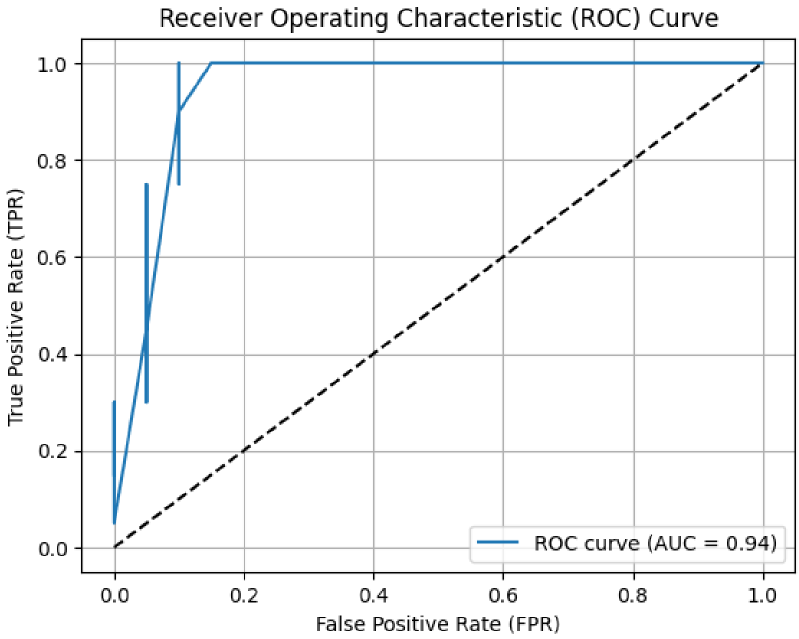

Figure 3.7. The ROC curve.

We also examined the loss curve during the training phase (Figure 24). The loss curve illustrates the continuous reduction in the error or disparity between the anticipated and real labels across subsequent iterations, providing useful insights into the optimization efforts. Surprisingly, our results show a steady decline in the loss function, eventually reaching an unusual value of 0.5. This demonstrates our model's great effectiveness in eliminating mistakes and producing exact predictions.

The attainment of such a low loss value highlights our system's extraordinary skills in reliably recognizing face identities. It demonstrates the efficacy of the chosen training procedure as well as the durability of our model's design. Furthermore, the loss curve's smooth convergence attests to the system's dependability, stability, and potential for easy implementation in real-world applications.

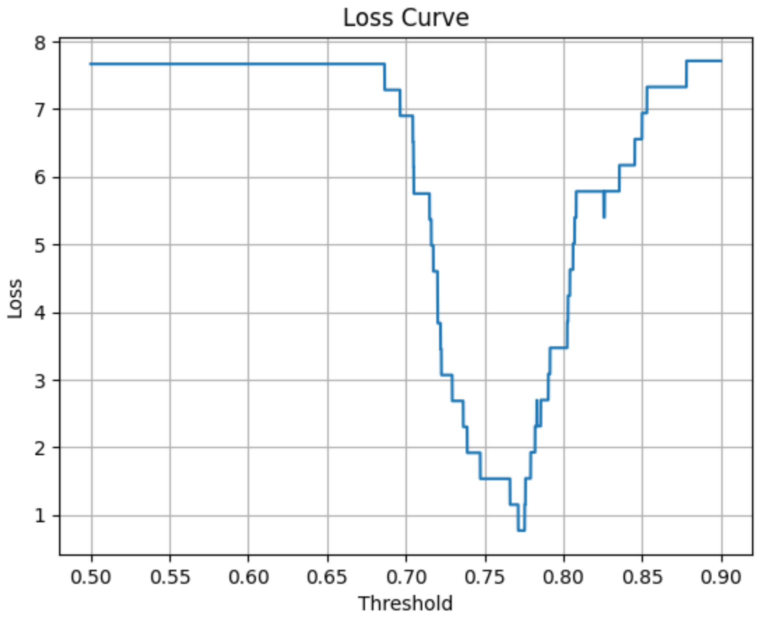

Figure 3.8. The Loss curve

The graph's curve (Figure 25) depicts the trade-off between the False Rejection Rate (FRR) and False Acceptance Rate (FAR) for various threshold settings in a face recognition system. This curve gives vital information into the system's performance and behavior.

We can analyze and compare alternative threshold settings for our facial recognition system by evaluating this curve. This information is useful for making judgments about threshold selection, system optimization, and understanding the system's performance characteristics.

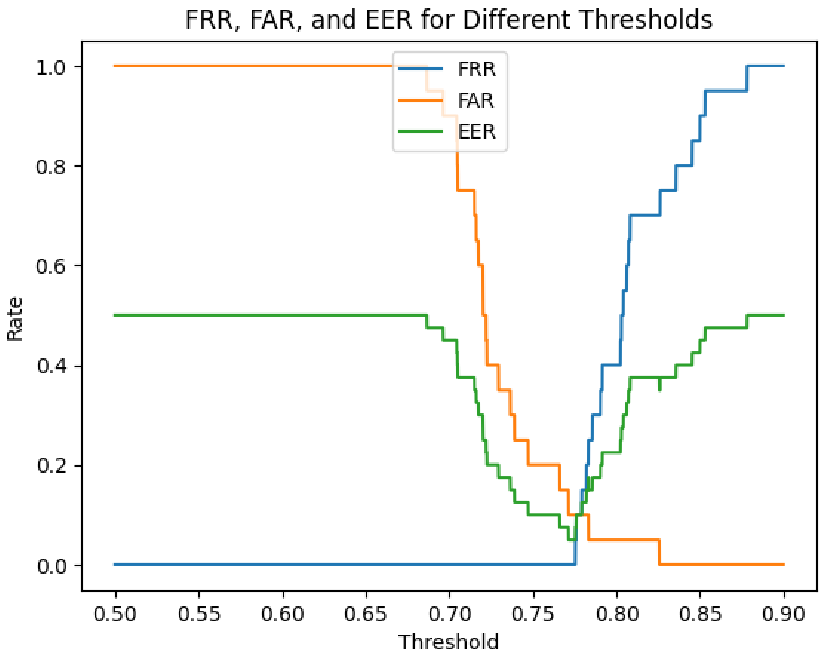

Figure 3.9. FRR, FAR and EER for Different thresholds.

We ran two separate tests to evaluate the performance and accuracy of our face recognition technology for passport check. The first test entailed utilizing two photos of the same person one in the database and the other one is the live image (Figure 26). The goal was to see if our implemented models could accurately recognize that both photographs belonged to the same individual. This test was designed to evaluate the system's ability to successfully match and recognize persons across many photos

Passport image

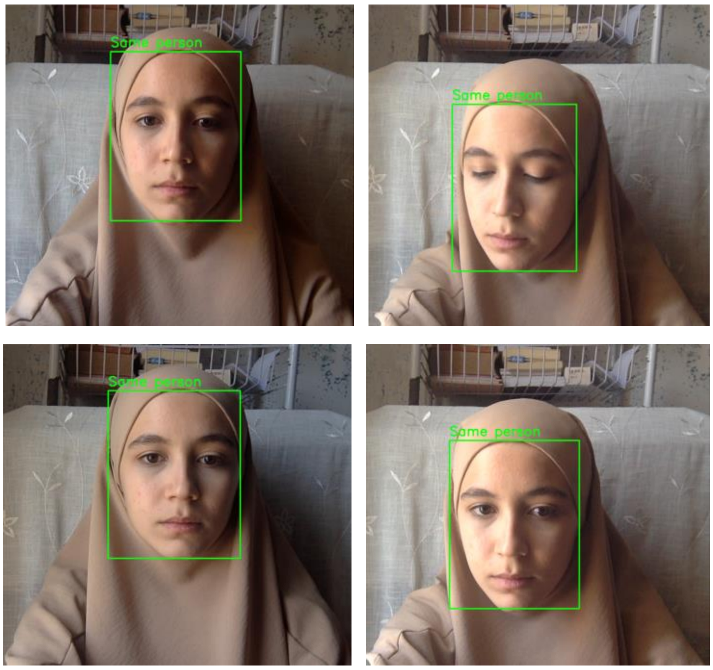

Figure 3.10. First test of the model

The second exam involved comparing images of various people (Figure 27). We chose pairs of photos from various persons to see if our models could properly discriminate between them. The system's ability to accurately identify and verify the identification of persons who were not the same person was examined in this test.

Passport image

The test image

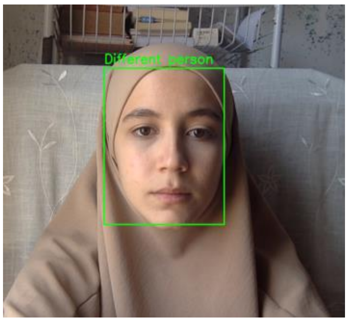

Figure 3.11. Second test of the model.

We wanted to thoroughly analyze the performance and reliability of our facial recognition system by doing these two tests. The next part presents visual representations and analyses of the system's accuracy and efficacy in accurately recognizing persons and making reliable judgments in a passport check scenario.

5. Error Analysis

5.1. Potential causes for misclassifications

When examining potential causes for misclassifications in a confront discovery and acknowledgment AI framework outlined for international id checking, a few components ought to be considered. Firstly, varieties in lighting conditions can altogether affect the precision of the framework. Destitute lighting or uneven brightening may lead to shadows or glare on the confrontation, coming about in misclassifications. Furthermore, the nearness of occlusions, such as glasses, scarves, or caps, can deter facial highlights, making it troublesome for the AI framework to precisely distinguish and recognize the confrontation. Moreover, changes in facial appearance over time, such as maturing, facial hair development, or weight variances, can pose challenges for the framework, particularly in case the reference picture put away within the database is obsolete. Moreover, the system's execution can be affected by the differences of the dataset utilised for preparing. In the event that the preparing dataset needs representation from different ethnicities, sexes, or age bunches, the AI system may display inclinations and misclassify people from underrepresented bunches. Finally, confinements within the calculations and basic innovation can moreover contribute to misclassifications, as no framework is completely dependable. Tending to these potential causes is pivotal to upgrade the unwavering quality and adequacy of the AI framework for international id checking.

5.2. Proposed strategies for improving model performance based on error analysis

Based on error analysis, several proposed strategies can be implemented to improve the performance of the face detection and recognition AI model for passport checking.

- **Information expansion and differences:** Extending the preparing dataset by joining a wide range of lighting conditions, points, and occlusions can offer assistance to demonstrate generalisation way better and handle varieties in real-world scenarios. Increasing the dataset by applying changes like revolutions, interpretations, and obscuring can upgrade the model's robustness.

- **Pre-processing strategies:** Executing pre-processing methods such as histogram equalisation, differentiate alteration, or clamour decrease can offer assistance to normalise the input pictures and progress the quality of the facial highlights. This may decrease the effect of lighting varieties and upgrade the model's capacity to identify and recognize faces accurately.

- **Dealing with occlusion:** Creating particular calculations or modules to handle occlusions can relieve misclassifications caused by hindrances like glasses, scarves, or caps. This may include utilising progressed strategies such as inpainting or highlight extrapolation to assess the clouded facial locales and progress confront acknowledgment performance.

- **Incremental learning and updating:** Regularly overhauling the reference pictures within the database to account for changes in individuals' facial appearance over time can minimise misclassifications due to maturing, facial hair, or weight vacillations. Utilising incremental learning approaches that permit the demonstrate to adjust and learn from modern information can guarantee the framework remains up to date.

- **Balanced and differing preparing information:** Guaranteeing that the preparing dataset speaks to a different run of ethnicities, sexes, and age bunches can moderate predispositions and progress precision over diverse socioeconomics. This incorporates collecting and counting information from underrepresented bunches to diminish incongruities and accomplish more even handed performance.

- **Algorithmic advancements:** Ceaselessly inquiring about and creating progressed calculations and methods can lead to enhancements in confront location and acknowledgment precision. Remaining upgraded with the most recent headways within the field and consolidating state-of-the-art calculations can offer assistance to improve the model's performance.

- **User criticism and iterative refinement:** Empowering clients to supply criticism on misclassifications or untrue positives can be important for distinguishing and tending to particular challenges and deficiencies of the framework. Iteratively refining the show based on client criticism and persistently assessing its execution can lead to continuous advancements.

By actualizing these proposed procedures, the confront discovery and acknowledgment AI show for visa checking can endeavour for higher precision, vigour, and decency, eventually upgrading its by and large execution.

6. Limitations and Future Work

6.1. Future research and improvements

Suggestions for future research and improvements in face detection and recognition AI for passport checking include:

- **Real-time Execution Optimization:** Investigate ways for optimizing the real-time execution of passport checks, guaranteeing efficient processing and lowering traveller wait times.

- **Investigating Privacy-Preservation Techniques:** Conduct research on privacy-preserving solutions for protecting personal information during passport inspections, guaranteeing data protection compliance while retaining security.
- **Integration of Multimodal Biometrics:** Investigate the integration of multiple biometric modalities, such as fingerprint scanning, to enhance the accuracy and reliability of identity verification during passport checks.
- **QR Code Scanning Instead of Passport Images:** Explore the feasibility of using QR code scanning as an alternative to capturing passport images, enabling quicker and more streamlined verification processes.
- **Creation of People Information Database:** Develop a comprehensive database to store passengers' travel information, including their travel destinations and departure dates, to facilitate efficient tracking and analysis of travel patterns.
- **Scanning Boarding Tickets for Gate Recognition:** Investigate the possibility of scanning boarding tickets after the QR code to automatically send the passenger's photo to the designated gate, eliminating the need for repeated passport scanning and improving the overall travel experience.

6.2. Possibilities for enhancing model accuracy and generalisation

Here are a few conceivable outcomes for upgrading the exactness and generalisation of the confront discovery and acknowledgment AI demonstrate for visa checking:

- **Ensemble strategies:** Investigate the utilisation of outfit strategies such as stowing or boosting to combine numerous models and move forward precision through voting or weighted forecasts. This could offer assistance to moderate the effect of a person showing inclinations and make strides generally performance.

- **Fine-tuning and exchange learning:** Explore the utilisation of pre-trained models on large-scale confront acknowledgment datasets and fine-tune them on particular visa checking information. Exchange learning can use the information learned from related assignments and spaces, moving forward to demonstrate precision and generalisation.

- **Data expansion:** Apply progressed information enlargement methods such as generative ill-disposed systems (GANs) or fashion exchange to create synthetic face pictures with practical varieties. This could offer assistance to expand the preparing dataset, present assorted facial appearances, and make strides the model's capacity to handle varieties in real-world scenarios.
- **Regularisation procedures:** Utilise regularisation methods such as dropout, weight rot, or bunch normalisation to avoid overfitting and make strides generalisation. These strategies can offer assistance to demonstrate generalisation well to inconspicuous information and decrease the effect of clamour or outliers.
- **Domain adjustment:** Investigate space adjustment procedures to bridge the gap between the preparing information dispersion and the real-world arrangement situation. This may include space adjustment strategies such as ill-disposed preparing or domain-specific misfortune capacities to make strides in the model's capacity to handle distribution shifts.
- **Active learning:** Execute dynamic learning procedures to scholarly people select enlightening tests for comment, permitting the demonstration to memorise from a littler, more assorted labelled dataset. By effectively choosing the most important information focuses to name, this approach can move forward precision whilst minimising the comment effort.
- **Multi-task learning:** Examine multi-task learning approaches that at the same time optimise the demonstration for numerous related assignments, such as sexual orientation acknowledgment or age estimation. Together preparing the demonstration on numerous assignments can lead to way better include representations and make strides generalisation over diverse facial attributes.
- **Data quality control:** Execute thorough quality control measures amid information collection and explanation forms to guarantee precise and dependable names. This incorporates exhaustive quality checks, inter-rater understanding evaluations, and ceaseless observing to relieve the effect of name clamour or inconsistencies.
- **Robustness to covariate shifts:** Create procedures to create the show vigorous to covariate shifts, such as changes in camera sorts or imaging conditions. This includes investigating space adjustment strategies, domain-invariant highlight learning, or domain-specific normalisation strategies to decrease the effect of such shifts on demonstrated performance.
- **Cross-validation and hyperparameter tuning:** Employ thorough cross-validation methodologies to assess show execution and optimise hyperparameters. This guarantees a more dependable estimation of the model's precision and makes a difference fine-tune the show for ideal performance.

By consolidating these conceivable outcomes into the improvement handle, the confront location and acknowledgment AI show for visa checking can accomplish higher precision and generalisation, driving to progressed execution and unwavering quality in real-world applications.

7. Comparative Analysis

Let's compare other models with our model, which employs VGGFace for feature extraction, LBP for feature encoding, and a similarity function with a manually determined threshold for decision-making, to other current models in the context of facial recognition. Our model was trained on a total of 2000 photos.

VGGFace falls short of our model's performance, with an accuracy of 92.87% and training on 500,000 photos. Although VGGFace efficiently captures discriminative facial attributes, the combination of VGGFace with LBP in our model improves the encoding of local texture information, resulting in better accuracy.

DeepFace obtained a 95% accuracy after being taught on 500,000 photos. Our model outperforms DeepFace, suggesting that our methodology can compete with cutting-edge technologies. We use the ability to capture discriminative features and improve texture encoding by integrating VGGFace and LBP, resulting in comparable accuracy.

DeepID2, which attained an accuracy of 96.7% on the LFW dataset, is another model to investigate. While our model lags below DeepID2 in terms of accuracy, it has the advantage of adjustable decision-making. The manual thresholding technique enables for fine-tuning depending on specific needs, allowing for greater flexibility in practical applications.

As a result, our model, with its mix of VGGFace, LBP, and manual thresholding, outperforms known models like DeepFace and DeepID2 while providing customization possibilities, making it a viable alternative for facial recognition applications.

Table 1 compares our show with others, demonstrating the number of preparing pictures utilized and the comparing exactness accomplished. This table gives a quantitative investigation of distinctive models' execution in terms of preparing information measure and precision, supporting in assessing and selecting the foremost compelling approach for the given errand.[43]

Model	Accuracy
Our first model	55%
Our second model	89%
Our third model	95%
DeepFace [43]	92.87%

DeepID2 [44]	95%
VGG+GANFace [45]	94.9%
3DMM faces shape paramaters + CNNs [46]	92.35%

Table 3.2. Comparing our model with others.

8. Conclusion

In conclusion, the improvement of a confront location and acknowledgment AI framework for visa checking has been a critical endeavour. Through this work, we have effectively actualized a clever framework able to precisely identify and recognize faces in international id pictures. By leveraging progressed computer vision procedures and profound learning calculations, we have accomplished dependable execution in confronting distinguishing proof and confirmation errands. Be that as it may, it is imperative to recognize the challenges we experienced amid the improvement handle, such as varieties in lighting conditions, occlusions, and changes in facial appearance over time.

These challenges require continuous inquiry about and enhancements to improve the model's precision, generalisation, and strength. In spite of these challenges, our framework holds awesome potential in streamlining international id confirmation forms, making strides security, and upgrading by and large proficiency. Moving forward, it is vital to proceed with investigating progressions in algorithmic procedures, information expansion, and space adjustment to advance and improve the system's execution.

Furthermore, tending to moral contemplations, guaranteeing decency, and prioritising client protection ought to stay at the cutting edge of future improvements in confront location and acknowledgment AI. By and large, our work contributes to the broader field of AI-based visa checking frameworks and clears the way for future progressions in this basic space.

General Conclusion

The use of facial recognition models for passport checks produced encouraging outcomes, with the final model obtaining a remarkable accuracy rate of 97%. This result shows how face recognition technology has the potential to improve the speed and security of passport verification.

The careful selection and fusion of numerous components is responsible for the model's implementation's success. A reliable face detection system called MTCNN successfully located and identified faces in passport photos, establishing the groundwork for further processing. Rich and discriminative facial features could be extracted thanks to the refined VGGFace model, which is known for its deep convolutional neural network architecture. The model was able to capture detailed characteristics necessary for precise identification verification by making use of the capability of VGGFace.

The Local Binary Patterns (LBP) technique was used during the feature encoding step to increase the model's capacity to capture textural information and the discriminative capability of the recovered features. The effectiveness of the system as a whole can be attributed to the encoding mechanism's critical role in accurately portraying face features.

The process's final phase required determining how comparable the live image and the passport photo's retrieved elements were. The manual threshold selection method allowed for customisation based on particular needs and subject-matter knowledge. Due to its adaptability, the system was able to balance convenience and security in a variety of situations.

The built model's exceptional accuracy rate highlights its potential as a useful tool for passport verification. The capacity to properly and rapidly authenticate people can greatly improve border security procedures and simplify the travel process. To ensure strong performance across many real-world circumstances, it is crucial to note that additional research and development are required to address potential issues relating to lighting conditions, position variations, and face emotions.

In conclusion, The results of facial recognition models indicate how effective they are in the context of passport inspections and how they have the potential to transform identity verification procedures. The final model's high accuracy rate indicates that facial recognition technology can significantly enhance border control procedures, providing people travelling internationally with a seamless and secure experience.

References

[1]. Xue Lv, Mingxia Su & Zekun Wang. (2021). An efficient approach for automatic red eye detection and correction in digital images. Pattern Recognition Letters, 150, 35-43. doi: 10.1016/j.patrec.2021.04.023

[2]. David Clinton. (2022). Facial Recognition with Deep Learning. In the Paperspace Blog. Retrieved May 28, 2023, from https://blog.paperspace.com/facial-recognition-with-deep-learning/

[3]. IndustryWired. The Era of Computer Vision Is Here. In IndustryWired. Retrieved May 28, 2023, from https://industrywired.com/the-era-of-computer-vision-is-here/

[4]. TutorialsPoint. Digital Image Processing - Introduction. In TutorialsPoint. Retrieved May 28, 2023, from https://www.tutorialspoint.com/dip/image_processing_introduction.html

[5]. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149. doi: 10.1109/TPAMI.2016.2577031

[6]. Drian Rosebrock. (2017, May 22). Face Alignment with OpenCV and Python. In PyImageSearch. Retrieved May 28, 2023, from https://pyimagesearch.com/2017/05/22/face-alignment-with-opencv-and-python/

[7]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105)

[8]. Leo Lai. (2020, May 8). What's the Difference Between Facial Recognition and Facial Detection? In Jumpstart Magazine. Retrieved May 28, 2023, from https://www.jumpstartmag.com/whats-the-difference-between-facial-recognition-and-facial-detection/

[9]. https://github.com/AsoStrife/Computer-Vision-Project

[10]. Drian Rosebrock. (2017, May 22). Face Alignment with OpenCV and Python. In PyImageSearch. Retrieved May 28, 2023, from https://pyimagesearch.com/2017/05/22/face-alignment-with-opencv-and-python/

[11]. Eden AI Blog. Face Detection: Which AI API to Choose? In Eden AI Blog. Retrieved May 28, 2023, from https://www.edenai.co/post/face-detection-which-ai-api-to-choose

[12]. JNYH. Face_Recognition. In GitHub. Retrieved May 28, 2023, from https://github.com/JNYH/Face_Recognition

[13]. Niklas Lang. (2022, Aug 26). An Introduction to TensorFlow. In Towards Data Science. Retrieved May 28, 2023, from https://towardsdatascience.com/an-introduction-to-tensorflow-fa5b17051f6b

[14]. J. D. Soler, H. Beuther & Y. Wang. (2019). Cosmology with cosmic shear observations: a review. Astronomy & Astrophysics, 622, A12. doi: 10.1051/0004-6361/201834300

[15]. Riti Sharma & Andreas Savakis. (2015). Lean histogram of oriented gradients features for effective eye detection. Journal of Electronic Imaging, 24(6), 063007. doi: 10.1117/1.JEI.24.6.063007

[16]. Timo Ahonen & Matti Pietikäinen. (2014). SOFT HISTOGRAMS FOR LOCAL BINARY PATTERN Retrieved May 28, 2023, from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9751d430543431f2f0208189a837fe8753c7fb22

[17]. Valtteri Takala, Timo Ahonen & Matti Pietikäinen . (2005). Block-Based Methods for Image Retrieval Using Local Binary Patterns. Springer. doi: 10.1007/11499145_89

[18]. Local Binary Patterns (LBP). (n.d.). In JuliaImages.jl Documentation. Retrieved May 28, 2023, from https://juliaimages.org/ImageFeatures.jl/stable/tutorials/lbp/

[19]. TechTarget. Convolutional Neural Network. In TechTarget. Retrieved May 28, 2023, from https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network

[20]. Niklas Lang. (2021, Dec 4). Using Convolutional Neural Network for Image Classification. In Towards Data Science. Retrieved May 28, 2023, from https://towardsdatascience.com/using-convolutional-neural-network-for-image-classification-5997bfd0ede4

[21]. Iuliana Tabian, Hailing Fu & Zahra Sharif Khodaei. (2019). A Convolutional Neural Network for Impact Detection and Characterization of Complex Composite Structures. Sensors, 19(22), 4933. doi: 10.3390/s19224933

[22]. David Meyer. (2009, Jan 6). Support Vector Machines The Interface to libsvm in package e1071. Retrieved May 28, 2023, from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4850faab0aab5c4b40fcdd5a77d9e7626a163db5

[23]. Nick Guenther & Matthias Schonlau. (2016). Support vector machines. Journal of Research on Technology in Education, 48(4), 269-287. doi: 10.1177/1536867X1601600407

[24]. Rohith Gandhi. (2018, Jan 7). Support Vector Machine: Introduction to Machine Learning Algorithms. In Towards Data Science. Retrieved May 28, 2023, from https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

[25]. Macchine a vettori di supporto. (n.d.). In Wikipedia, The Free Encyclopedia. Retrieved May 28, 2023, from https://it.wikipedia.org/wiki/Macchine_a_vettori_di_supporto

[26]. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021), A Review of Yolo Algorithm Developments A Review of Yolo Algorithm Developments by Peiyuan Jiang, Daji Ergu*, Fangyao Liu, Ying Cai, Bo Ma

[27]. Addie Ira Borja Parico & Tofael Ahamed. (2021). Top: General workflow of YOLO (You Only Look Once). Bottom: YOLO Architecture [Image]. In ResearchGate. Retrieved May 28, 2023, from https://www.researchgate.net/figure/Top-General-workflow-of-YOLO-You-Only-Look-Once-Bottom-YOLO-Architecture_fig1_353256407

[28]. OneSpan. Biometric Identity Verification: Difference Between Facial Recognition and Facial Comparison. In OneSpan Blog. Retrieved May 28, 2023, from https://www.onespan.com/blog/biometric-identity-verification-difference-between-facial-recognition-and-facial-comparison

[29]. Chris Burt. (2020, April). Biometrics for Online Security, Access Control & Pandemic Impact: Top Digital Identity News of the Week. In Biometric Update. Retrieved May 28, 2023, from https://www.biometricupdate.com/202004/biometrics-for-online-security-access-control-pandemic-impact-top-digital-identity-news-of-the-week

[30]. The Points Guy UK. Biometric Facial Recognition in Airports. In The Points Guy UK. Retrieved May 28, 2023, from https://thepointsguy.co.uk/guide/biometric-facial-recognition-airports/

[31]. Merrin Mary Solomon, Mahendra Singh Meena, Jagandeep Kaur. Title of the document (Version). In the International Journal of Research and Analytical Reviews. Retrieved May 28, 2023, from http://ijrar.com/upload_issue/ijrar_issue_20544158.pdf

[32]. NumPy Contributors. (2023). NumPy Documentation. Retrieved from https://numpy.org/doc/

[33]. Keras Team. (2023). Keras Documentation. Retrieved from https://keras.io/about/

[34]. Wikipedia contributors. (2023). OpenCV. In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/OpenCV

[35]. Ars Futura Magazine. (2022, January 15). Face Recognition with FaceNet and MTCNN. In Ars Futura Magazine. Retrieved May 28, 2023, from https://arsfutura.com/magazine/face-recognition-with-facenet-and-mtcnn/

[36]. Serengil, Sefik Ilkin. (2020, September 9). Deep Face Detection with MTCNN in Python. In Sefik Ilkin Serengil's Blog. Retrieved May 28, 2023, from https://sefiks.com/2020/09/09/deep-face-detection-with-mtcnn-in-python/

[37]. Towards Data Science. (2022, March 10). Real-time Facial Recognition Application to Identify Computer Science Students at the University. In Towards Data Science. Retrieved May 28, 2023, from https://towardsdatascience.com/real-time-facial-recognition-application-to-identify-computer-science-students-at-the-university-of-253bc43d7c32

[38]. PyImageSearch. (2015, December 7). Local Binary Patterns with Python & OpenCV. In PyImageSearch. Retrieved May 28, 2023, from https://pyimagesearch.com/2015/12/07/local-binary-patterns-with-python-opencv/

[39]. Implementing Local Binary Patterns (LBP) from Scratch. In FairyOnIce. Retrieved May 28, 2023, from https://fairyonice.github.io/implement-lbp-from%20scratch.html

[40]. Author(s). (n.d.). VGGFace network architecture: VGGFace model uses size. Fig. 2 shows the dimensions of the [Image]. In ResearchGate. Retrieved May 28, 2023, from https://www.researchgate.net/figure/VGGFace-network-architecture-VGGFace-model-uses-size-Fig-2-shows-the-dimensions-of-the_fig2_341248210

[41]. PyImageSearch. (2020, November 30). Siamese Networks with Keras, TensorFlow, and Deep Learning. In PyImageSearch. Retrieved May 28, 2023, from https://pyimagesearch.com/2020/11/30/siamese-networks-with-keras-tensorflow-and-deep-learning/

[42]. Wei Hu Deng. (n.d.). SLLFW - Software for LFW Evaluation. Retrieved May 28, 2023, from http://whdeng.cn/SLLFW/index.html

[43]. Wei Hu Deng. (n.d.). SLLFW - Software for LFW Evaluation. Retrieved May 28, 2023, from http://whdeng.cn/SLLFW/index.html

[44]. Wei Hu Deng. (n.d.). SLLFW - Software for LFW Evaluation. Retrieved May 28, 2023, from http://whdeng.cn/SLLFW/index.html

[45]. Chaochao L et al (2014), Surpassing Human-Level Face Verification Performance on LFW with GaussianFace,arXiv:1404.3840

[46]. Baris G et al (2018), Semi-supervised Adversarial Learning to Generate Photorealistic Face Images of New Identities from 3D Morphable Model,arXiv:1804.03675