VISVESWARAYA TECHNOLOGICAL UNIVERSITY
Jnana Sangama, Belagavi-590018

TECHNICAL SEMINAR REPORT ON

"COMPARATIVE STUDY ON VARIOUS FACIAL RECOGNITION ALGORITHMS"

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF ENGINEERING
In
COMPUTER SCIENCE AND ENGINEERING
Submitted By

Khushwinder Singh
(1BY18CS074)

Under The Guidance Of
Mrs. Durga Bhavani A
Assistant Professor
Department of CSE

BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Avalahalli , Yelahanka , Bengaluru – 560064.

2021-2022

VISVESWARAYA TECHNOLOGICAL UNIVERSITY
Jnana Sangama, Belagavi-590018

BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Avalahalli , Yelahanka , Bengaluru – 560064.

CERTIFICATE

This is to certify that the Seminar work entitled "Comparative Study On Various Facial Recognition Algorithms " has been carried out by Mr Khushwinder Singh, 1BY18CS074, a bonafide student of BMS Institute of Technology and Management in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi during the year 2021-2022. It is certified that all corrections/suggestions indicated for assessment have been incorporated in the report deposited in department library. The Seminar report has been approved as it satisfies the academic requirements in respect of Seminar work prescribed for the said degree.

Signature of the Guide	**Signature of the HOD**
Guide Name	**Dr. Bhuvaneshwari C M.**
Designation	**Professor& Head of department**
Department of CSE	**Department of CSE**
BMSIT & M	**BMSIT & M**

DECLARATION

I, Khushwinder Singh **[USN: 1BY18CS074]**, student of VIII Semester BE, in Computer Science and Engineering, BMS Institute of Technology and Management hereby declare that the Seminar entitled "Comparative Study On Various Facial Recognition Algorithms **"** has been carried out by me and submitted in partial fulfillment of the requirements for the *VIII Semester degree of* ***Bachelor of Engineering in Computer Science and Engineering*** *of Visvesvaraya Technological University, Belagavi* during academic year 2021-22.

Date: **Student Name:**

Place: **USN:**

ACKNOWLEDGEMENT

We are happy to present this Seminar report after completing it successfully. This seminar would not have been possible without the guidance, assistance and suggestions of many individuals. We would like to express our deep sense of gratitude and indebtedness to each and everyone who has helped us make this a success.

We heartily thank our **Principal, Dr. MOHAN BABU G N, BMS Institute of Technology & Management,** for his constant encouragement and inspiration in taking up this project.

We heartily thank our **Head of the Department, Dr. Bhuvaneshwari C.M, Department of Computer Science and Engineering, BMS Institute of Technology &Management,** for her constant encouragement and inspiration in taking up this project.

We gratefully thank our Seminar Guide, **Mrs. Durga Bhavani A, Assistant Professor, Department of Computer Science and Engineering,** for his/her guidance, support, and advice.

Special thanks to all the staff members of the Computer Science Department for their help and kind co-operation.

Lastly, we thank our parents and friends for the support and encouragement given to us in completing this precious work successfully.

Khushwinder Singh
1BY18CS074

ABSTRACT

In this Survey we try to compare the various facial recognition algorithms and then try to figure out which is the best facial recognition algorithm. The various algorithms will be compared based on their accuracy, their architecture, the dataset used to train them and also on their loss functions. We will compare various 2-D facial recognition algorithms such as VGG19, Facenet by Google, Deepface by Facebook, Alexnet and Local Binary Pattern Histogram (LBPH) and 3-D facial recognition algorithms based on LBP and amalgamation of various Techniques. We will also see the use of Big data in the training of facial recognition algorithms.

CONTENTS

List Of Figures

Chapter 1
Introduction

Biometrics are measurements of human characteristics that can be used to verify identity. These distinguishing characteristics are nearly impossible to perfectly spoof, copy, or duplicate, making them an ideal candidate for increasing the security of user authentication. Facial biometric data, in particular, have shown great promise for authentication purposes, owing to the ease with which user faces can be discerned and identified by systems.

More than ever before, facial recognition is in the spotlight. Previous historical events have caused a rapid increase in face recognition investments. Given the global COVID-19 epidemic, we may see increased investment in biometric technologies such as facial recognition. Because COVID-19 is so contagious, there is a strong emphasis on contactless interactions. Face recognition technology is still primarily used for security purposes. Facial recognition is widely acknowledged as one of the most accurate and simple methods of establishing individual identity in a variety of industries.

Face recognition has a number of advantages over other biometric modalities in certain situations. It is widely accepted, very familiar, and easily understood by the general public. As a result, it has a wide range of applications, including criminal identification, unlocking smartphones and laptops, home access and security, finding missing people, assisting blind people, recognising people on social media, disease diagnosis, real-time monitoring and management systems, and so on.

1.1 Context

There have been several algorithms and strategies to recognise faces in the context of face recognition over the years. It is constantly improving as a result of advances in Artificial Intelligence science (AI). Even in low-light situations, the accuracy of recognising complexion, age, gender, and ethnicity has improved. Significant advancements in power, cost, and hardware size enable a wide range of use cases in a variety of industries. With

So many alternatives, the issue isn't "should I adopt a face recognition system?" but "what is the best facial recognition system for me?" With so many possibilities, the question isn't "should I adopt a facial recognition system?" but "what is the best facial recognition system for me?"

1.2 Motivation

From business solutions to commercial solutions to household solutions, we look at how facial recognition is enhancing the way various industries work today—and, if relevant, the exciting prospects it holds for the future. Beyond simply unlocking phones and laptops, biometric technologies used in facial recognition applications can now detect faces more accurately than humans. While this makes the technology an obvious choice for security and identification, it can also be recycled and used imaginatively to serve a variety of businesses.

We now use facial recognition technology probably a few 100 times a day to unlock our phones. It has other applications such as Security companies using facial recognition to secure their premises, Immigration checkpoints using facial recognition to enforce smarter border control, Fleet management companies can use face recognition to secure their vehicles, Ride-sharing companies can use facial recognition to ensure the right passengers are picked up by the right drivers, IoT benefits from facial recognition by allowing enhanced security measures and automatic access control at home, Law Enforcement can use facial recognition technologies as one part of AI-driven surveillance systems, Retailers can use facial recognition to customise offline offerings and to theoretically map online purchasing habits with their online ones and many more. Seeing all this progress motivated me to do something in this domain and contribute to this growing industry of the 21st century.

Due to the numerous techniques and algorithms in the market, the developer usually gets confused about which facial recognition algorithm will be the best for their use case. Reading across various journals and papers does not always provide an accurate comparison between the algorithms. Hence we try to compare the various techniques present in 2-dimensional or 3 dimensional, basic or advanced, complex or simple in a single place.

Chapter 2
Literature Survey

Face detection has been the focus of study for the past two decades. Face detection is a hot topic in research, with a variety of classical and deep learning algorithms being used. Both detection and recognition are hot topics in the academic world these days. Many methods for face detection can yield good accuracy, implying that the field of face detection is nearly completed. As a result, there are additional obstacles to overcome in terms of recognition. Face recognition was studied using a holistic 171 learning method (eigenface, fisher face, SRC and CRC, etc.) and a local handmade approach (LBP, HD-LBP, etc.) from the early 1990s to the early 2000s.

Later in 2010, shallow and deep learning became popular, with deep learning being the most effective for facial recognition. The Deep Convolutional Neural Network FaceNet [9] is a deep convolutional neural network. On YouTube Faces DB [9], it reaches a new record accuracy of 99.63 per cent and gives 95.12 per cent. FaceNet is used to extract features in the proposed system, and it works by embedding the features into 128 dimensions. A support vector machine is used as a classifier after feature extraction. For facial recognition, support vector machines (SVMs) outperform Multi-Layer Perceptron (MLP) Classification.

Various other 2-dimensional recognition algorithms like Principle Component Analysis (PCA), Alexnet, VGG19, RESNET, Deepface by Facebook, and Local Binary Pattern (LBP) also evolved with time.

The rapid advancement of 3D sensors has shown a new face recognition path that may be able to overcome the fundamental limits of 2D technologies. The geometric information provided in 3D facial data could significantly increase recognition accuracy in difficult-to-recognize situations3. Many academics have shifted their attention to 3D face recognition, resulting in a new research trend.

2.1 2-Dimensional facial recognition algorithms

2.1.1 Alexnet

Convolutional Neural Network (CNN) is a relatively new established competent image recognition method that uses local receptive fields as neurons in the brain, weights sharing and linking information, and significantly reduces training constraints when compared to other neural networks.

Alexnet boosted CNN's popularity in computer vision by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). To achieve higher accuracy in CNN image classification, the development and application of deeper and more complex CNN has become a research trend.

The neural network has eight weighted layers, the first five of which are convolutional and the remaining three are fully connected. The output of the final fully-connected layer is fed into a 1000-way softmax, which generates a distribution over 1000 class labels. The network maximises the multinomial logistic regression objective, which is equivalent to maximising the average of the log probability of the correct label under the prediction distribution across training cases.

2.1.2 VGG

AlexNet was released in 2012 and improved on traditional Convolutional neural networks. We can think of VGG as a successor to AlexNet, but it was created by a different group called the Visual Geometry Group at Oxford, hence the name VGG. It carries and improves on some ideas from its predecessors and uses deep Convolutional neural layers to improve accuracy.

VGG stands for Visual Geometry Group. A VGG neural network (VGGNet) is one of the most used image recognition model types based on deep convolutional neural networks.

The VGG architecture became famous for achieving top results at the ImageNet challenge. The model is designed by the researchers at the University of Oxford.

While the VGG-Face has the same structure as the regular VGG model, it is tuned with facial images. The VGG face recognition model achieves a 97.78% accuracy on the popular Labelled Faces in the Wild (LFW) dataset.

VGG takes in a 224x224 pixel RGB image. For the ImageNet competition, the authors cropped out the centre 224x224 patch in each image to keep the input image size consistent.

2.1.3 Deepface by Facebook

DeepFace is the facial recognition system used by Facebook for tagging images. It was proposed by researchers at Facebook AI Research (FAIR) at the 2014 IEEE Computer Vision and Pattern Recognition Conference (CVPR).

DeepFace is a deep neural network used for face recognition. It follows the flow of detect, align, represent and classify to achieve the task. It consists of 3D Face Modelling, followed by a piecewise affine transformation. Later, a face representation is derived from a 9-layer Deep neural Network. The 2D images are warped into the 3D plane with the help of 67 anchor points.

It doesn't matter if the face is tilted, at an angle, or in bad lighting. DeepFace is unlike previous generations of facial recognition software which follow the convention steps: detect → align → represent → classify. Facebook's DeepFace employs 3D face modelling and derives a picture from a deep network of millions of parameters. DeepFace uses an algorithm that is capable of identifying a face with a 97.25% accuracy. The average human can do so with a 97.5% accuracy.

2.1.4 Facenet by Google

FaceNet is a face recognition system that was described by Florian Schroff, et al. at Google in their 2015 paper titled "FaceNet: A Unified Embedding for Face Recognition and Clustering." The FaceNet system can be used to extract high-quality features from faces, called face embeddings, that can then be used to train a face identification system. It is a system that, given a picture of a face, will extract high-quality features from the face and predict a 128 element vector representation of these features, called a face embedding. The model is a deep convolutional neural network trained via a triplet loss function that encourages vectors for the same identity to become more similar (smaller distance), whereas vectors for different identities are expected to become less similar (larger distance). The focus on training a model to create embeddings directly (rather than extracting them from an intermediate layer of a model) was an important innovation in this work. These face embeddings were then used as the basis for training classifier systems on standard face recognition benchmark datasets, achieving then-state-of-the-art results.

2.2 3-Dimensional facial recognition

2.2.1 Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is a simple yet very efficient texture operator that labels the pixels of an image by thresholding the neighbourhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, the LBP texture operator has become a popular approach in various applications. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Perhaps the most important property of the LBP operator in real-world applications is its robustness to monotonic grey-scale changes caused, for example, by illumination variations. Another important property is its computational simplicity, which makes it possible to analyse images in challenging real-time settings.

It should be noted that when using the histogram-based methods the regions do not need to be rectangular. Neither do they need to be of the same size or shape, and they do not necessarily have to cover the whole image. It is also possible to have partially overlapping regions. The two-dimensional face description method has been extended into a spatiotemporal domain. It depicts facial expression descriptions using LBP-TOP. Excellent facial expression recognition performance has been obtained with this approach. It can recognize the face of a person from both the front face and side face.

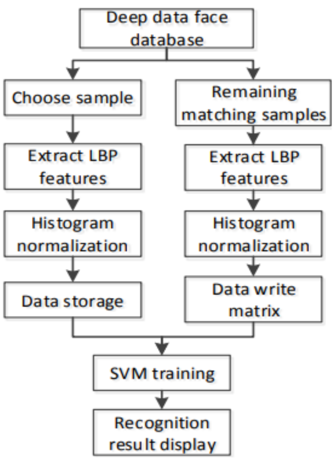

Fig 2.1 Schematic Diagram of 3D Face Recognition

Chapter 3
Methodology

3.1 General Methodology for facial Recognition

Algorithms work based on the following steps:

Step-1 Face detection: The camera detects and locates the image of a face.

Step-2: Face analysis: Next, an image of the face is captured and analysed

Step-3: Converting the image to data: The image is converted into analog information and embeddings are generated.

Step-4: Finding a match: Your faceprint is then compared against a database of other known faces.

3.2 Important algorithms used in Facial recognition

3.2.1 AlexNet

AlexNet is the name of a convolutional neural network (CNN) architecture, designed by Alex Krizhevsky in collaboration with Ilya Sutskever and Geoffrey Hinton, who was Krizhevsky's PhD advisor. AlexNet competed in the ImageNet Large Scale Visual Recognition Challenge on September 30, 2012. The network achieved a top-5 error of 15.3%, more than 10.8 percentage points lower than that of the runner up. The original paper's primary result was that the depth of the model was essential for its high performance, which was computationally expensive but made feasible due to the utilisation of graphics processing units (GPUs) during training.

The net contains eight layers with weights; the first five are convolutional and the remaining three are fully connected. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels. Our network maximises the multinomial logistic regression objective, which is equivalent to maximising the average across training cases of the log probability of the correct label under the prediction distribution. The kernels of the second, fourth, and fifth convolutional layers are connected only to those kernel maps in the previous layer which

resides on the same GPU (see Figure 2). The kernels of the third convolutional layer are connected to all kernel maps in the second layer. The neurons in the fully-connected layers are connected to all neurons in the previous layer. Response-normalisation layers follow the first and second convolutional layers. Max-pooling layers, of the kind described in Section 3.4, follow both response-normalisation layers as well as the fifth convolutional layer. The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer. The first convolutional layer filters the 224×224×3 input image
with 96 kernels of size 11×11×3 with a stride of 4 pixels (this is the distance between the receptive field centres of neighbouring neurons in a kernel map). The second convolutional layer takes as input the (response-normalised and pooled) output of the first convolutional layer and filters it with 256 kernels of size 5 × 5 × 48. The third, fourth, and fifth convolutional layers are connected without any intervening pooling or normalisation layers. The third convolutional layer has 384 kernels of size 3 × 3 × 256 connected to the (normalised, pooled) outputs of the second convolutional layer. The fourth convolutional layer has 384 kernels of size 3 × 3 × 192, and the fifth convolutional layer has 256 kernels of size 3 × 3 × 192. The fully-connected layers have 4096 neurons each.

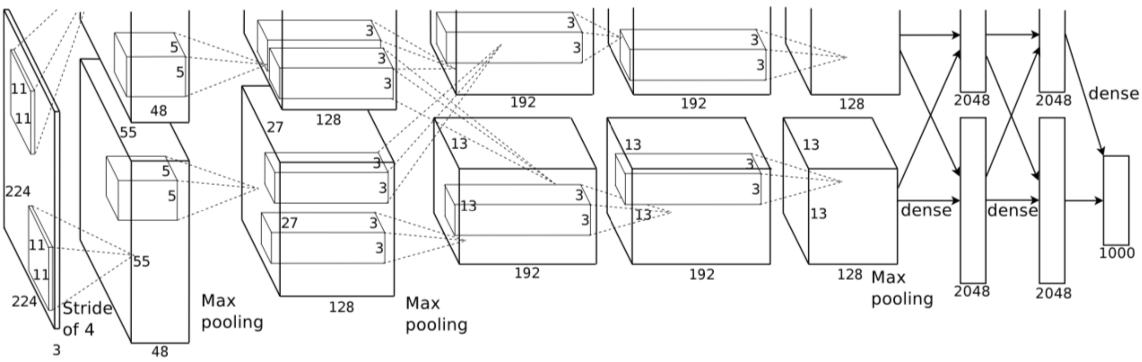

Fig 3.1 AlexNet Architecture

3.2.2 VGG-Face

VGG19 is a variant of the VGG model which in short consists of 19 layers (16 convolution layers, 3 fully connected layers, 5 MaxPool layers and 1 SoftMax layer). There are other variants of VGG like VGG11, VGG16 and others. VGG19 has 19.6 billion FLOPs.

The architecture of VGG Face is as follows:

A fixed-size (224 * 224) RGB image was given as input to this network which means that the matrix was of shape (224,224,3). The only preprocessing that was done is that they subtracted the mean RGB value from each pixel, computed over the whole training set. Used kernels of (3 * 3) size with a stride size of 1 pixel, this enabled them to cover the whole notion of the image. Spatial padding was used to preserve the spatial resolution of the image. Max pooling was performed over 2 * 2-pixel windows with side 2. This was followed by a Rectified linear unit(ReLu) to introduce non-linearity to make the model classify better and to improve computational time as the previous models used tanh or sigmoid functions that proved much better than those. Implemented three fully connected layers from which the first two were of size 4096 and after that, a layer with 1000 channels for 1000-way ILSVRC classification and the final layer is a softmax function.

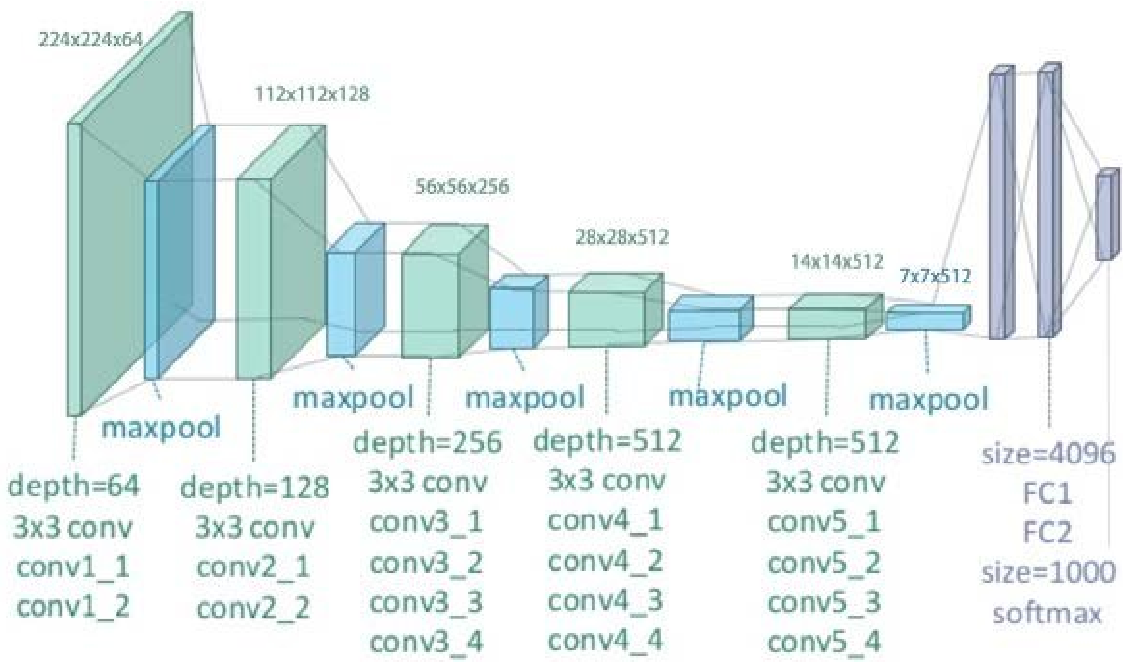

Fig 3.2 VGG-Face 19 Architecture

3.2.3 DeepFace

DeepFace is a deep neural network used for face recognition. It follows the flow of detecting, aligning, representing and classifying to achieve the task. It consists of a 3D Face Modelling, followed by a piecewise affine transformation. Later, a face presentation is derived from a 9-layer Deep neural Network. The 2D images are warped into 3D planes with the help of 67 anchor points.

It doesn't matter if the face is tilted, at an angle, or in bad lighting. DeepFace is unlike previous generations of facial recognition software which follow the convention steps: detect → align → represent → classify. Facebook's DeepFace employs 3D face modelling and derives a picture from a deep network of millions of parameters. DeepFace uses an algorithm that is capable of identifying a face with a 97.25% accuracy. The average human can do so with a 97.5% accuracy.

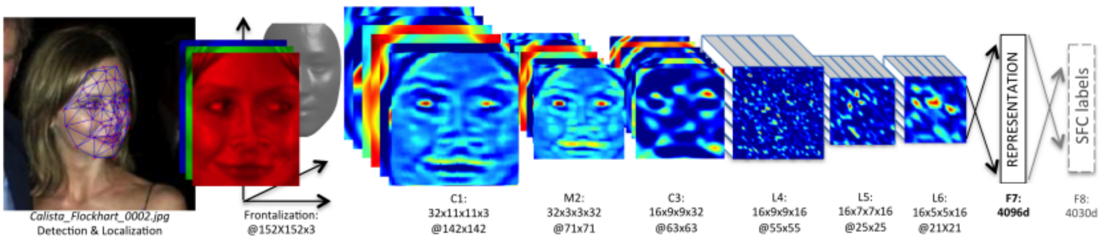

Fig 3.3 Architecture of DeepFace

DeepFace is trained for multi-class face recognition i.e. to classify the images of multiple people based on their identities.

It takes input into a 3D-aligned RGB image of 152*152. This image is then passed the Convolution layer with 32 filters and size 11*11*3 and a 3*3 max-pooling layer with the stride of 2. This is followed by another convolution layer of 16 filters and size 9*9*16. The purpose of these layers is to extract low-level features from the image edges and textures.

The next three layers are locally connected, a type of fully connected layer that has different types of filters in a different feature map. This helps in improving the model because different regions of the face have different discrimination abilities so it is better to have different types of feature maps to distinguish these facial regions.

The last two layers of the model are fully connected. These layers help in establishing a correlation between two distant parts of the face. Example: Position and shape of eyes and position and shape of the mouth. The output of the second last fully connected layer is used as a face representation and the output of the last layer is the softmax layer K classes for the classification of the face.

The total number of parameters in this network is approximately 120 million with most of them (~95%) coming from the final fully connected layers. The interesting property of this network is the feature map/vector generated during the training of this model is amazingly sparse. For Example, 75% of the values in the topmost layers are 0. This may be because this network uses a ReLU activation function in every convolution network which is essentially max(0, x). This network also uses Drop-out Regularization which also contributed to sparsity. However, Dropout is only applied to the first fully connected layer. In the final stages of this network, we also normalise the feature to be between 0 and 1. This also reduces the effect of illumination changes across. We also perform an L2-regularisation after this normalisation.

3.2.4 FaceNet

FaceNet is a Deep Neural Network used for face verification, recognition and clustering. It directly learns mappings from face images to a compact Euclidean plane. When an input image of 96*96 RGB is given; it simply outputs a 128-dimensional vector which is the embedding of the image.

This conversion into a simple Euclidean plane simplifies all the tasks as it is a simple distance calculation.

Triplet loss is used to learn the embeddings. The name is derived from its use of three images to learn the mapping namely anchor, positive and negative.

The loss function is such that it minimises the distance between the anchor and positive and simultaneously moves the anchor and negative further away.

The Loss L is

$$\sum_{i}^{N}\left[\|f(x_i^a)-f(x_i^p)\|_2^2-\|f(x_i^a)-f(x_i^n)\|_2^2+\alpha\right]_+$$

Where a is the anchor, p is positive and n is negative and α is the margin enforced between positive and negative samples.

The architecture is

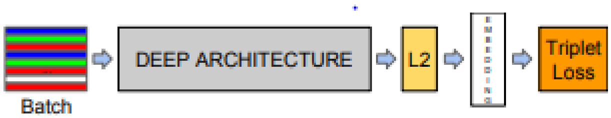

Fig 3.4 Architecture of FaceNet

It describes that the network consists of a batch input layer and a deep CNN followed by L2 normalisation, which results in face embedding. This is followed by the triplet loss during training. The figure represents the basic model of FaceNet. Squared L2 is used to directly correspond to embedding vectors. These embeddings are fine-tuned by minimising the triplet loss.

The most important process is the selection of a triplet. The triplets should be chosen in such a way that given an anchor image, select the “hardest” positive image (of the same person) (i.e. the one that’s furthest away in the dataset) and select the “hardest” negative image (of a different person) as in (i.e. the one that’s closest in the dataset). If this triplet doesn’t violate the condition, then none with that anchor will.

There are two types of Deep architectures used i.eThe Zeiler Fergus inspired architecture and the inception inspired architecture. Their practical differences lie in the difference in parameters and FLOPS. The best model may be different depending on the application.

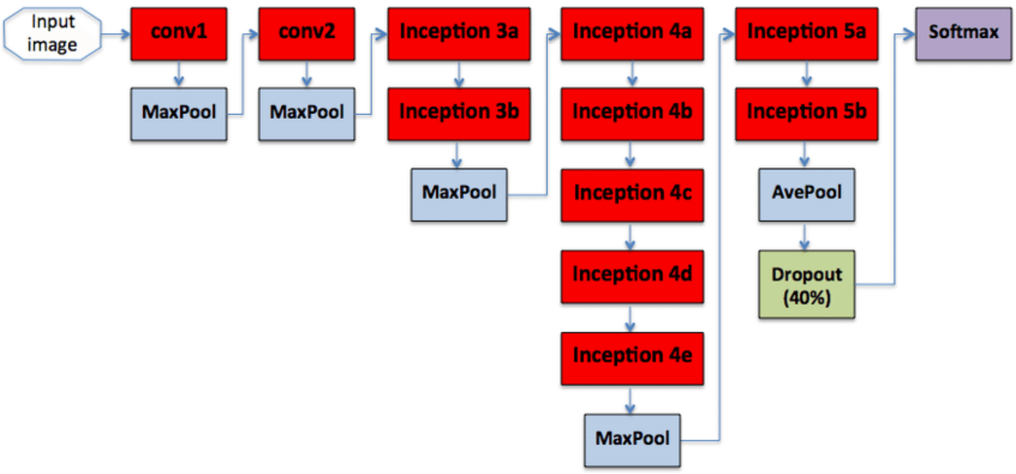

Fig 3.5 GoogleNet Architecture

Table 3.1 Details of the Facenet Architecture based on GoogleNet

type	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj (p)	params	FLOPS
conv1 (7×7×3, 2)	112×112×64	1							9K	119M
max pool + norm	56×56×64	0						m 3×3, 2		
inception (2)	56×56×192	2		64	192				115K	360M
norm + max pool	28×28×192	0						m 3×3, 2		
inception (3a)	28×28×256	2	64	96	128	16	32	m, 32p	164K	128M
inception (3b)	28×28×320	2	64	96	128	32	64	L_2, 64p	228K	179M
inception (3c)	14×14×640	2	0	128	256,2	32	64,2	m 3×3,2	398K	108M
inception (4a)	14×14×640	2	256	96	192	32	64	L_2, 128p	545K	107M
inception (4b)	14×14×640	2	224	112	224	32	64	L_2, 128p	595K	117M
inception (4c)	14×14×640	2	192	128	256	32	64	L_2, 128p	654K	128M
inception (4d)	14×14×640	2	160	144	288	32	64	L_2, 128p	722K	142M
inception (4e)	7×7×1024	2	0	160	256,2	64	128,2	m 3×3,2	717K	56M
inception (5a)	7×7×1024	2	384	192	384	48	128	L_2, 128p	1.6M	78M
inception (5b)	7×7×1024	2	384	192	384	48	128	m, 128p	1.6M	78M
avg pool	1×1×1024	0								
fully conn	1×1×128	1							131K	0.1M
L2 normalization	1×1×128	0								
total									7.5M	1.6B

3.2.5 Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is a relatively common algorithm used to classify texture features. The LBP operator represents the texture characteristics of the grayscale image. Its advantage is that it has strong robustness to the light, and it still has the same characteristics when the shooting angle and viewing angle change a lot. The most primitive local binary mode (LBP) is to use the pixel value of the most central point as the threshold value. The neighbouring pixels are binarized. Its expression is

$$LBP\ (x_c, y_c) = \Sigma_{p=0}^{p-1} 2^p\ s(i_p - i_c)$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The calculation process is to take each of the pixels in the image as the centre point and takes a 3 x 3 size area around the pixel as the neighbcentre. Compare the other eight pixel values in the adjacent area with the pixel values in the centre. If the pixel value is greater than or equal to the centre pixel value, set the pixel value to 1, otherwise, set it to 0. The 8-bit sequence of 0 and 1 will be generated to form an 8-bit unsigned integer binary value. After calculation, the binary number is the LBP eigenvalue in the centre of this window. The calculation process is shown in

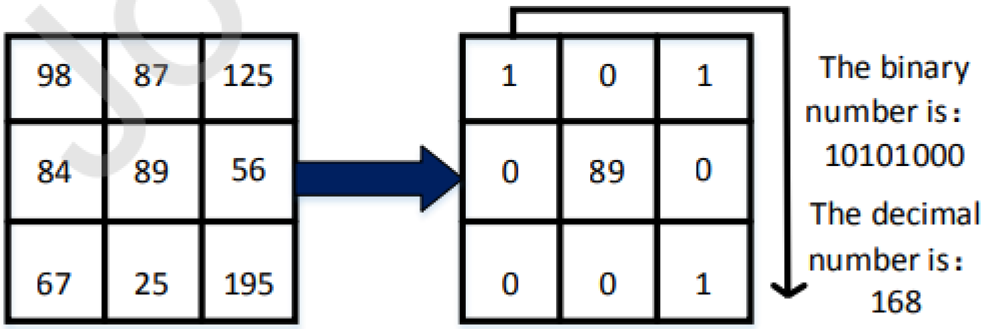

Fig 3.6 Schematic diagram of the LBP neighbourhood process

To make the LBP operator have grey invariance and rotation invariance. Expand the original limited area to any size. The original square neighbourhood is replaced by a circular neighbourhood. Since the radius of the circular neighbourhood is variable, there can be any number of pixels in it, and the circle has the characteristic of rotation invariance. Therefore, the improved local binary model operator has rotation invariance, which is called Circular LBP.

The value of each sampling point can be calculated by the following formula:

$$x_p = x_c + R\cos(\frac{2\pi p}{N})$$

$$y_p = y_c - R\sin(\frac{2\pi p}{N})$$

Where, (x_c,y_c) is the centre point of the area and (x_p,y_p) is a sampling point in the neighbourhood. The coordinates of any point in the neighbourhood can be calculated by the above formula.

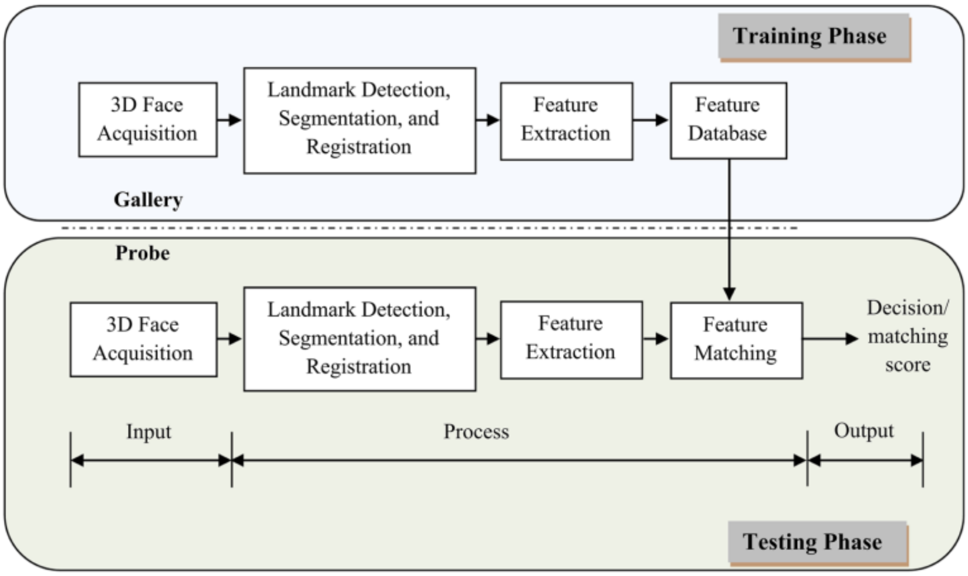

Fig 3.7 A general 3-d face recognition system

3.3 Training the model

After the embeddings are generated using the various algorithms mentioned above the models are then trained and classification of the embeddings is done using algorithms like SVM, K-Nearest Neighbour, Decision Tree, Bayesian Classification, etc. The trained algorithms are then tested and finally used in the application that it was intended for.

Chapter 4
Evaluation

The evaluation of the algorithms will be done based on the following parameters:

- **Architecture**: A CNN architecture is formed by a stack of distinct layers that transform the input volume into an output volume (e.g. holding the class scores) through a differentiable function. A small change in the architecture can make a huge difference to the final result of the output, hence it's an important parameter when evaluating the facial recognition algorithms.

- **The loss function used**: The loss function specifies how training penalises the deviation between the predicted output of the network, and the true data labels (during supervised learning). It plays an important role in the way the model is trained and corrected using the loss values. Various loss functions can be used, depending on the specific task. It is a salient feature of a training algorithm.

- **Dataset Used for training:** Various datasets can be used to train the model with the algorithms. Every dataset has different types of images of the faces in it. The images might be taken in different conditions and angles, and they may also capture different information like emotions, time of the image, angle of capturing, distance of the person, etc. which make it a prime parameter for evaluation.

- **Accuracy**: It is no doubt the most important parameter for evaluation. The accuracy is the only thing that is highlighted while the evaluation takes place. Almost all the developers adopt the algorithm with a high accuracy rate.

Chapter 5
Results

We will use the parameters defined above to arrive at the results of each algorithm.

5.1 2-D facial recognition algorithms

Table 5.1 Evaluation of algorithms

Sl. no	Name of Algo	Architecture	Dataset used	Loss function	Accuracy
1	AlexNet	8 layers (5 convolution layers, 3 fully connected layers)	ImageNet (15M Images)	cross-entropy	84.7%
2	VGG Face 19	19 layers (16 convolution layers, 3 fully connected layers, 5 MaxPool layers and 1 SoftMax layer)	VGGface (2.6M Images)	triplet-loss	98.95%
3	Deepface	8 layers (5 convolution layers, 3 fully connected layers)	Facebook(4.4 M Images)	soft-max	97.35%
4	Facenet	22 layers (3 convolution layers, 18 inception layers,1 linear layer)	Google(500M Images)	triplet-loss	99.63%

From the above table we can observe the following:

- Highest accuracy is obtained by Google's FaceNet followed by VGGface19
- The largest dataset used for training is Google(500 Million images) followed by Imagenet(15 Million)
- The most complex network is used by Facenet (GoogleNet-24 - 22 layers) followed by VGGface19 (19 layers).

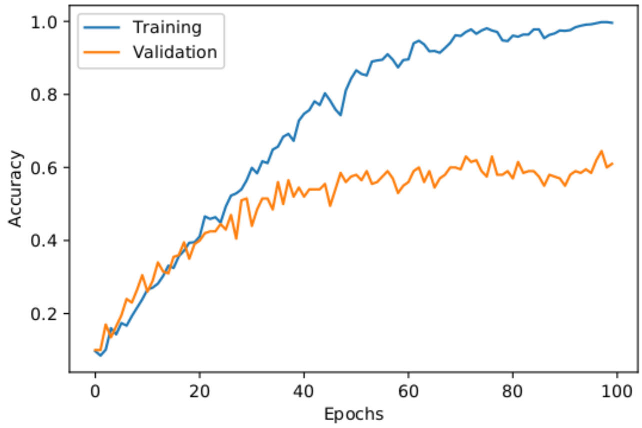

Fig 5.1 AlexNet Training and Validation Accuracy

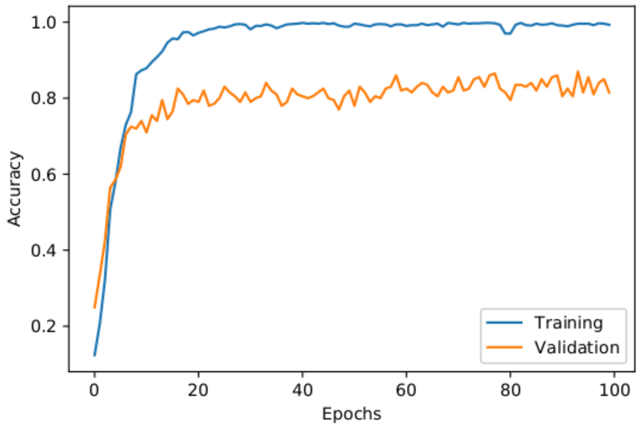

Fig 5.2 VGGFace 19 Training and Validation Accuracy

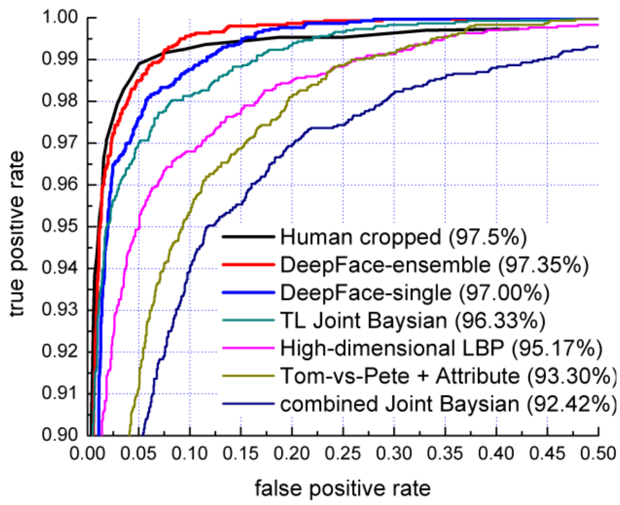

Fig 5.3 DeepFace True positive vs false positive rate

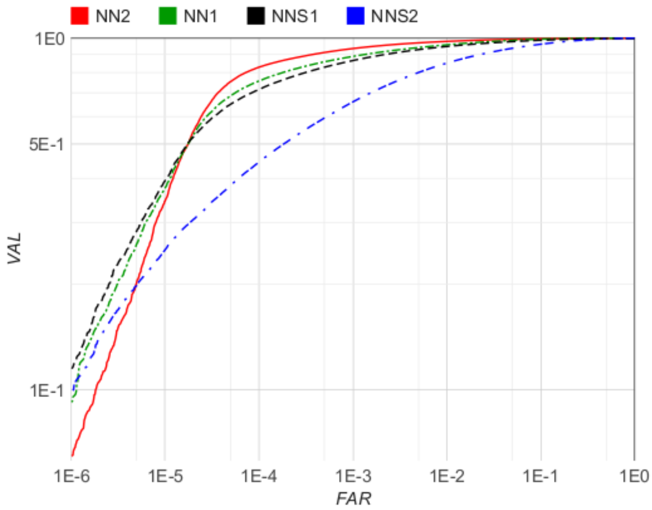

Fig 5.4 FaceNet True positive vs false positive rate

5.2 3-D facial recognition algorithms

Local Binary pattern (LBP) is used for 3-dimensional facial recognition models. This technique when clubbed with SVM gives the following result:

Dataset used: Texas 3DFRD(5 depth maps per person)

Accuracy: 96.83

Chapter 6
Conclusion

From the above results we can conclude that:

- Facenet has the highest accuracy of 99.63% in 2-D facial recognition algorithms.
- Facenet is also the most reliable as it is trained with a dataset of 500 Million images.
- VGG Face 19 is also pretty accurate but the dataset size is less.
- 3 Dimensional facial recognition captures more features than the 2-dimensional facial recognition algorithms but it is yet to reach the level of accuracy as 2-dimensional recognition algorithms.
- In the coming future with better computation and hardware, the demand for 3-dimensional recognition will increase as it is more reliable and trustworthy than 2-dimensional recognition.
- We already see big companies like Apple already shifting to 3-d technologies for facial recognition and it is predicted that more and more companies will shift to 3-d technology with improving accuracy of the algorithms.

References:

[1] Zhou, S., Xiao, S. 3D face recognition: a survey. Hum. Cent. Comput. Inf. Sci. 8, 35 (2018). https://doi.org/10.1186/s13673-018-0157-2

[2]Shi L, Wang X, Shen Y, Research on 3D Face Recognition Method Based on LBP and SVM, Optik (2020), doi: https://doi.org/10.1016/j.ijleo.2020.165157

[3]T. Ahmed, P. Das, M. F. Ali and M. -. F. Mahmud, "A Comparative Study on Convolutional Neural Network-Based Face Recognition," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225688.

[4]Chandra, Y.B., & Reddy, G.K. (2020). A Comparative Analysis Of Face Recognition Models On Masked Faces. International Journal of Scientific & Technology Research, 9, 175-178.

[5]Gwyn, Tony, Kaushik Roy, and Mustafa Atay. 2021. "Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study" Future Internet 13, no. 7: 164. https://doi.org/10.3390/fi13070164

[6] Suriya Priya R Asaithambi, Sitalakshmi Venkatraman, Ramanathan Venkatraman. Proposed big data architecture for facial recognition using machine learning[J]. AIMS Electronics and Electrical Engineering, 2021, 5(1): 68-92. doi: 10.3934/electreng.2021005

[7]A. Verma, P. Singh and J. S. Rani Alex, "Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition," 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), 2019, pp. 169-173, doi: 10.1109/IWSSIP.2019.8787215.

[8]T. Nyein and A. N. Oo, "University Classroom Attendance System Using FaceNet and Support Vector Machine," 2019 International Conference on Advanced Information Technologies (ICAIT), 2019, pp. 171-176, doi: 10.1109/AITC.2019.8921316.

[9]M. Wang, W. Deng, Deep Face Recognition: A Survey, Neurocomputing (2020), doi: https://doi.org/10.1016/j.neucom.2020.10.081

[10]F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[11]Paul, Liton & Suman, Abdulla. (2012). Face recognition using principal component analysis method. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 1. 135-139.

[12]Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. <i>Commun. ACM</i> 60, 6 (June 2017), 84–90. https://doi.org/10.1145/3065386

[13]Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[14]Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.