Bima Journal of Science and Technology, Vol. 7 (4) Dec, 2023 ISSN: 2536-6041

DOI: 10.56892/bima.v7i4.546

An In-Depth Analysis of Face Recognition Models: A Comparative Study of Deep Learning Architectures on the Labeled Faces in the Wild Dataset

[1]*Suraj Salihu, [1]Suleiman Salihu Jauro, [1,4]Abdulrasheed Jimoh, [1]Bala Modi, [2,3]Nasiru Salihu, and[1,5]Usman Idris Ismail

[1]Department of Computer Science, Faculty of Science, Gombe State University, Gombe, Nigeria
[2]Center of Excellence in Theoretical and Computational Science (TaCS-CoE), Fixed Point Research Laboratory, Fixed Point Theory and Applications Research Group, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok 10140, Thailand
[3]Department of Mathematics, Faculty of Science, Modibbo Adama University, Yola, Nigeria
[4]Department of Computer Science, Federal College of Education (Technical), Gombe, Nigeria
[5]Department of Computer Science, Federal University of Kashere, Gombe, Nigeria

* Corresponding Author: surajsalihu@gsu.edu.ng

ABSTRACT

Artificial Intelligence is one of the important tools widely used in our society mainly Deep Learning (DL) which has quite a lot of applications due to its ability to learn robust representations from images and videos for recognition tasks. Convolutional Neural Networks (CNN) is a subset of DL heavily used by researchers following the breakthrough of AlexNet by winning the most difficult image classification standard ImageNet Large Scale Visual Recognition Challenge (ISLVRC) 2012 by decreasing the error by 10% from the winning algorithm of the 2011 on the benchmark. This paper compares the performance of ten face recognition models built using DL architectures on Labelled Faces in Wild (LFW) dataset: DeepFace, FaceNet and VGGFace were trained using triplet loss which have lower convergence but achieved higher accuracy. DeepID, DeepID2, DeepID2+ and DeepID3 were trained using sofmax loss, but learn features that are not discriminative enough because the linear transformation matrix's size grows the number of identities increases linearly, SphereFace reached good performance but the training process is unstable, ArcFace attained the state of art performance by introducing Additive Angular Margin Loss to mitigate the two main problems associated with the previous approaches.

Keywords: Face Recognition, CNN, Deep Learning, Face Recognition and Loss Function

INTRODUCTION

Human being has the ability to distinguish faces, but not lot of faces, large numbers of faces are hard for human being to recognize and memorize, however computers if well trained can potentially handle large number of faces and even distinguishes between faces of human being using distinct features in the individual faces. Computer Scientist in the middle of 1960s started working on the use of computers to identify and recognize human faces, since that period, governments and organisations started using facial recognition system for authentications and in recent years, researchers in computer vision have recently become quite interested in facial recognition networks. Deep learning procedures for facial identification are also becoming increasingly popular, particularly when AlexNet won the ImageNet Visual Recognition Challenge of ISLVRC by dropping the error with a margin of 10 % from the winning traditional algorithm of the 2011 (Krizhevsky, Sutskever, and Hinton, 2017), this great success brought a new perspective to image

DOI: 10.56892/bima.v7i4.546

classification as well face recognition. Though humans are fairly good in identifying and recognizing faces but are not extremely able to deal with a huge number of unknown faces (Turk and Pentland, 1991).

There are several biometric authentication methods available that support many forms of identification and verification; such as iris scan, retina scan, fingerprint scan, facial scan gait, voice print, keystrokes, hand geometry (Choudhary and Moriwal, 2020), (Kataria, Adhyaru, Sharma, and Zaveri, 2013) those regularly used biometrics have several disadvantages, Iris identification and verification are remarkably accurate, but costly during implementation on a large scale and is not well embraced by the public. Fingerprints are trustworthy and nonintrusive, but not appropriate for non-collaborative persons (Abate, Nappi, Riccio, and Sabatino, 2007). Face recognition is now a good option that balances societal acceptance and dependability while still maintaining safety and privacy (Jain, Nandakumar, and Ross, 2016). Face identification methods work in an unconstrained situation (Best-Rowden, Han, Otto, Klare, and Jain, 2014) and has the significant benefit of being able to work in a variety of settings with a large number of unknowing visitors. Face identification has become one of the most widely used biometric techniques because of these benefits (Kortli, Jridi, Al Falou, and Atri, 2020). Face detection and identification are cognitive abilities that form the foundation for our social interactions (Weigelt, Koldewyn, and Kanwisher, 2012). From birth, people engage in face-to-face interactions that contribute to their ability to recognize faces (Johnson, Rickel, and Lester, 2000). Numerous tools, technologies and strategies were developed to manage information security, however, one of the key problems is the requirement for adequate authentication. Instead of just examining whether a legitimate form of identification or key has been used, or whether the user entered authentic passwords or Personal Identification Numbers, face authentication methods establish the presence of an authorized individual, Jain (Ghali, Ali, and Yousif, 2020).

Apparently it has become crucial to develop biometric applications like facial recognition, biometric systems that differentiate between human being are based on biological characteristics and they are very attractive due to their flexibility and easy to use, faces are composed of diverse arrangements and characteristics, based on this, it has become one of the approach use for identification and verification, given its prospective in several applications and fields such as counter terrorism, surveillance, faceID, smart cards, border control, e-commerce, criminal justice systems, student ID, driver licenses, security systems, immigration, home security, image database research, and eventually, banking(Al-Kawaz, 2019). Face recognition technology as an identity (ID) is already available to people outside of phones, such as at airport check-ins, sport events, etc, etc in developed countries (Jridi, Napoléon, and Alfalou, 2018).

Problem Statement

Authentication systems such as the one used in identification card and signature has a lot of errors, particularly to an organization like institution of higher learning (Minaee, Abdolrashidi, Su, Bennamoun, and Zhang, 2023), security enforcements etc, where an unauthorized individual may get an identification card of authorized person and presented it as authentic to gain access to environment or signatures can be simply fabricated by someone else. To formulate a face identification problem, we need to have a face input image and a database of known faces; Face verification is a one-to-one (1:1) mapping between a query image and the known image in the gallery while face

DOI: 10.56892/bima.v7i4.546

identification is one-to-many mapping (1:N), Here, the unidentified face in the image is compared to every face in the database of known people, and a decision is made as a result of all the comparisons (Maltoni, Maio, Jain, and Prabhakar, 2009). This researcher study is proposing reviewing deep learning methods using CNN for Face identification systems since there is problem with hand craft feature engineering associated with the traditional approaches as Scale Invariant Feature Transform (SIFT) (Clemons, 2007), Local Binary Patterns (LBP) (Ahonen, Hadid, and Pietikäinen, 2006) etc.

GENERAL DESCRIPTION OF FACE RECOGNITION

Face identification is designed to recognize a face in real-time or videos (Shanthi and Sivalakshmi, 2023), whereas face detection aims to recognize and extract facial features in an image, different facial identification approaches have being developed and effectively implemented, many of which have one drawback or the other such as; occlusion, variations (view point, scale, intra class), deformation, background clutter, low recognition accuracy, high false acceptance and rejection rate, expression, pose, as well as illumination among others (Schroff, Kalenichenko, and Philbin, 2015), however the following are traditional steps in Face recognition systems; Using image acquisition as input and a face anti-spoofing unit, The security of the system is guaranteed by the inclusion of Adversarial Attack Detection. Facial landmarks and or Face are detected in the image, Pre-processing is then carryout on the image, such as alignment and noise reduction (Nguyen et al., 2022), following phase is feature extraction from the image using either a texture-based, model-based, or holistic approaches. The last step is identification or verification, as shown in Figure 1.

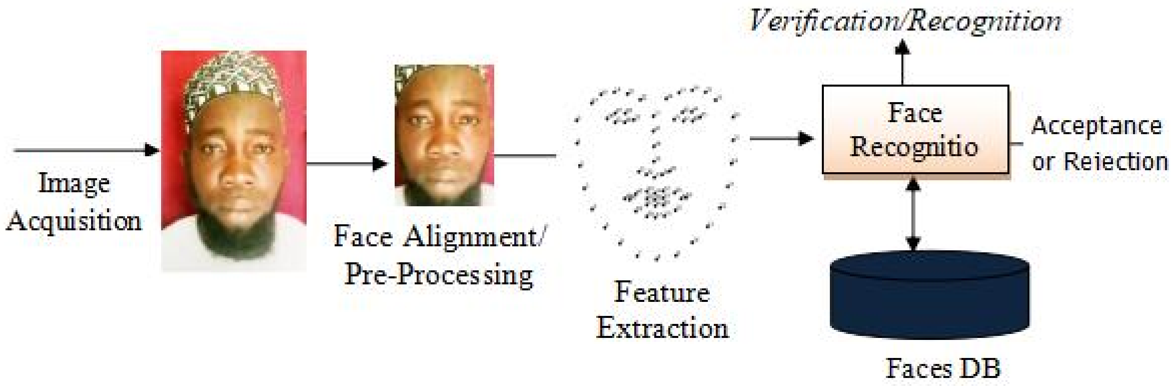

Figure 1: Face Recognition Process

The paper significantly contributes to the understanding of deep learning techniques in face recognition, providing valuable insights into training strategies and loss functions in face recognition system:

1. Conducts a thorough comparison of ten face recognition models built using Deep Learning on the Labelled Faces in the Wild dataset.
2. The paper compares models training procedures employ where triplet loss achieves higher accuracy despite lower convergence, and softmax loss, which may result in less discriminative features.
3. It provides insights into the performance of various face recognition models, offering a nuanced understanding of their strengths and limitations based on the chosen training strategies.

RELATED REVIEW

One of the most significant developments in computer vision is deep learning, it is transforming industries and businesses, already deep learning in many area has

DOI: 10.56892/bima.v7i4.546

surpass human level performance and capacity, e.g., time taken by car delivery, movie ratings prediction as well decision to approve loan applications (Agrawal, Gans, and Goldfarb, 2022). There is a competition among best technologies companies in the world and economies on the direction of deep learning in future. While on the area of medicine deep learning has proved the ability to enhance human lives through more precise diagnosis in the classification and detection of diseases such as cancer and discovery of new drugs, in 2018 a deep learning model Google AI surpass human in the grading prostate cancer by 70% to 61% compared to the US certified general pathologists' average accuracy (Nagpal et al., 2019),as well in the prediction of natural disasters, (Nevo et al., 2019).

Taigman et al (Taigman, Yang, Ranzato, and Wolf, 2014) introduced a technique called DeepFace with 200 million images from Facebook which uses Deep Neural networks (DNN) approach; the model had been successful due to the collection of a huge database containing multiply labelled samples. As wellSchroff et al(Schroff et al., 2015) introduced another model called FaceNet, which uses Labelled Faces in Wild (LFW)(Huang, Mattar, Berg, and Learned-Miller, 2008) benchmark for face identification and trained using GoogleNet inception network architecture with the triplet loss function, It had 13,233 photos of 5,749 persons, 1,680 of whom had two or more, the DNN had recorded accuracy rate of 99.63% very close to the human level on the LFW databases taken under unrestricted conditions, since then the accuracy of LFW dataset has reached 99.80% (M. Wang and Deng, 2021). On March 27, 2019, the Turing Award ‘Nobel Prize’ of computing awarded three researchers, Geoffrey E. Hinton, Yoshua Bengio, and Yann LeCun for their extraordinary contributions to the field of deep learning (Ulhaq, 2021).

Techniques of Face Recognition

Face recognition history dated back to 1950s (Zhao, Chellappa, Phillips, and Rosenfeld, 2003), however automatic face recognition researches begin in the 1970s. At the beginning, face recognition were based on distinct features between important areas of the face, (Kanade, 1977). Since then, researches andstudies on face recognition continued flourishing till early 1990s. Turk and Pentland (Turk and Pentland, 1991) developed first successful techniques known as Eigenfaces, Viola and Jones (Viola and Jones, 2001) developed Haar cascade detection algorithm, Dalal and Triggs (Dalal and Triggs, 2005) Histograms of Oriented (HOG). The image method of face recognition systems are divided into 4 major theoretical development phases as pointed out by (M. Wang and Deng, 2021).

Appearance Based or Holistic Technique

Appearance or subspace approaches take the entire face as the feature; they do not demand the extraction of feature points or areas of the face like the ears, nose, eyes, or mouth. The primary purpose of these techniques is to characterize a image as a pixel matrix, which is frequently translated into feature vectors (Blanz and Vetter, 2023). The feature vectors are then used as low-dimensional space, but these methods are sensitive to changes in lighting and facial emotions (Raju, Chinna Rao, Saikumar, and Lakshman Pratap, 2022), as well pose, these methods can be categories as either linear or non-linear for the mapping of face to a lower dimensional subspace, other techniques make use of linear subspaces and sparse representations (Wright, Ganesh, Zhou, Wagner, and Ma, 2008),

Local Feature Technique

After year 2000 face recognition techniques continued to gained popular, hand-crafted features are used to define the face, the techniques such as LBP (Ahonen et al., 2006), Histogram of Gradient Orientation (HOG) (Dalal and Triggs, 2005), SIFT

Bima Journal of Science and Technology, Vol. 7 (4) Dec, 2023 ISSN: 2536-6041

DOI: 10.56892/bima.v7i4.546

(Lowe, 2004), Binary Robust Independent Elementary Features (BRIEF) (Calonder, Lepetit, Strecha, and Fua, 2010), and (Bay, Ess, Tuytelaars, and Van Gool, 2008), uses a geometrical technique called the analytical technique, where the face image is expressed by a collection of characteristic vectors with small or low-dimensional regions (patches)and focuses on essential points of the face like ears, eyes, mouth and nose, It also emphasizes the effectiveness of the detectors of the key features of the facial image, which can handle occlusions and missing pieces, (Lv and Ping, 2013).

Learning Method Local Descriptor Techniques

It appeared shortly after the 2010s which learned the image discrimination using shallow methods (Lei, Pietikainen, and Li, 2014).

Deep Learning Techniques

Attention shifted to deep learning approaches following the achievement of AlexNet in the ImageNet challenge, it introduced a new outlook to face identification. In order to achieve performance comparable to that of humans on big datasets collected in unrestricted environments, an astonishing stability for face recognition has been achieved utilizing Convolutional Neural Networks (CNN) as in (Parkhi, Vedaldi, and Zisserman, 2015). CNN is the greatest instance of supervised learning, it allows outstanding architectural model inspired by the eye visual cortex for computer vision. One of the frequently investigated areas is the face recognition, and CNN has had considerable success in this area, becoming a power engine in the research area as shown in (Gao et al., 2020). Figure 2 shows some of the classification of Face Recognition Networks.

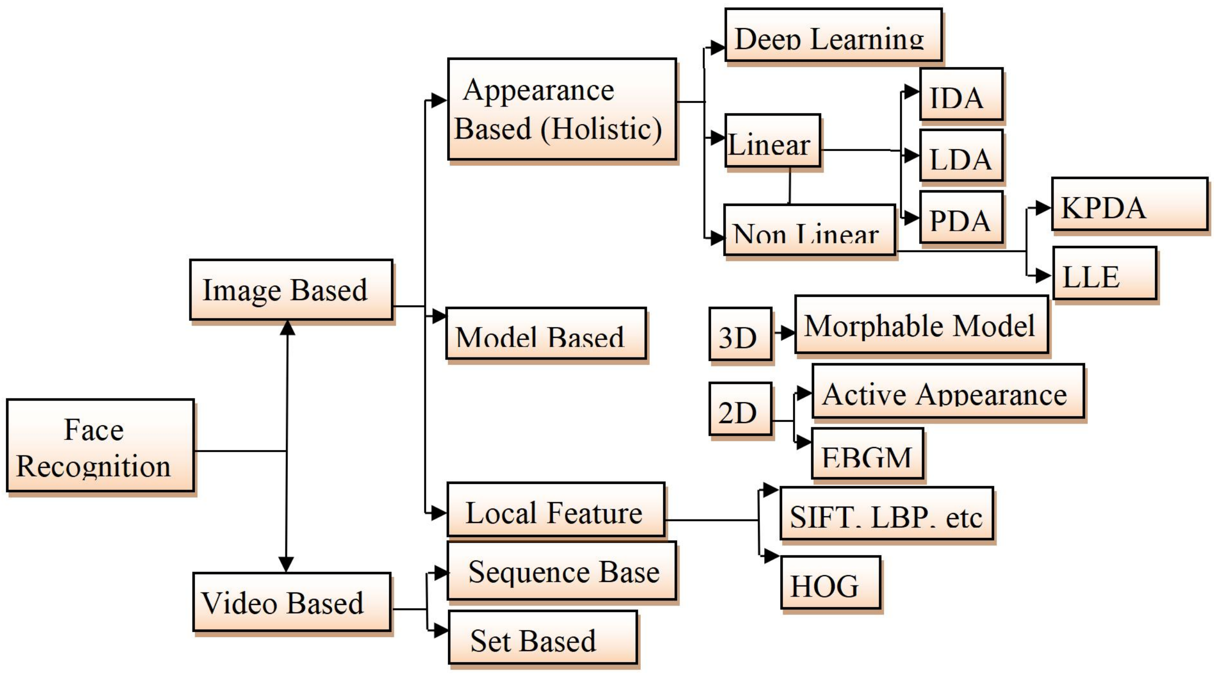

Figure 2: Classifications of Face recognition Systems from Literatures

CNN Training Method

Convolution is a process where we apply a matrix called kernel or filter or feature detector to an image to downsize it, or to maintain the same size, add several layers of padding. Convolution are also used to extract specific features from an image, such as a shape, lines, an edge and higher feature such as the facial structure from an image, (Singh, 2019). Contrary to ANN, Convolution operations are used by CNNs

DOI: 10.56892/bima.v7i4.546

in at least one of their layers, CNN architecture consists of several blocks made up of four main sections: a filter bank or kernels, a convolution layer and nonlinearity, a pooling layer and a Dense or FCL, each stage representing a set of arrays known as feature maps. Figure 3 depicts a general structure of CNN's architecture consisting of one or more FCL and several stacked convolutional stages, with a classification module as the final output.

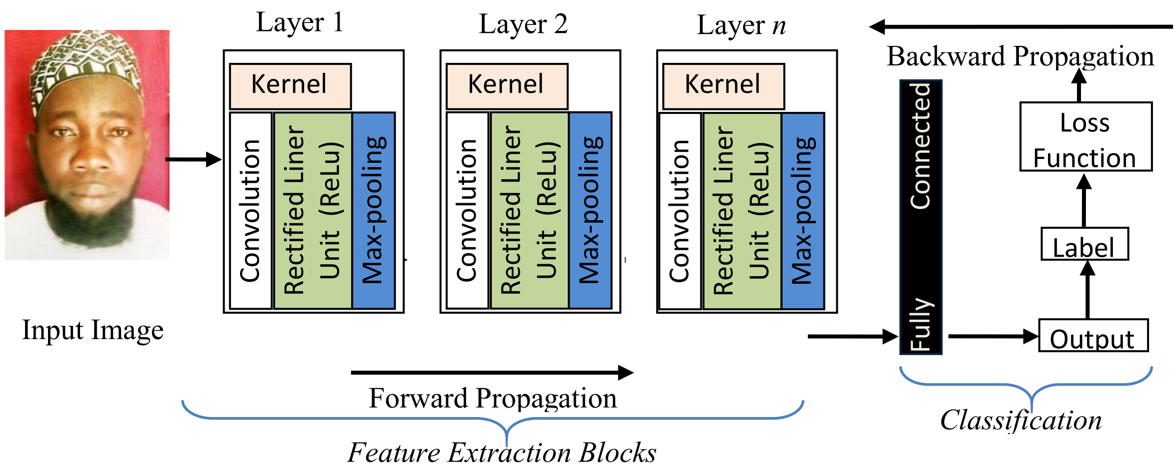

Figure 3: CNN Architecture depicted input layer, hidden layer and the output layers

The training procedure optimizes several layer parameter weight (w) and bias (b) as well hyper-parameters like learning rate (α), layer sizes, batch size etc of neural network (NN) to minimize differences among specified labels (y) using a training dataset, and the predicted results ($\hat{y}$), frequently use algorithm for training NN are gradient decent and the backpropagation, The following describes the backpropagation training procedure(Riedmiller, 1994):

1. Select image training data, often taken in a batch
2. Pass each batch via the network and get the result ($\hat{y}$).
3. Use a loss function L to calculate the error between the input labels (y) and the specified predictions ($\hat{y}$).
4. Back-propagate errors throughout the network.
5. Weights (W) should be updated to reduce error.
6. Repeat until you reach the limit of converged iterations.

Deep Learning Architectures for Face Recognition using CNN

Traditional classification architectures could have been better, but the major problem with these approaches is that we have fix number of classes; to address this problem different deep learning architectures for facial recognition surfaced as follows:

DeepFace Architecture

DeepFace is an Architecture of 9 CNN layers by three researchers Taigman et al (Taigman et al., 2014) from FaceBook AI Group. DeepFace is one of the first work that recorded greet achievement of 97.35% higher accuracy utilizing CNNs on the LFW dataset, lowering the error rate by 27% as well nearly achieving human-level performance, since then focuses has been sifted to face recognition research using deep learning based architectures. As shown in Figure 4, the first three layers are typical conv-pooling-conv layers settings, then 3 layers of locally connected, 2 layers of FCLs. To train the model, 4 million facial photos from a big face database were used with 4 thousand multi-class identities or subjects for face identification task experimental using Siamese structure that have same number of parameters, this specifically improved the gap between two (2) facial features in the top layer, however this requires twice the computation. While performing a face verification task, the training solely occurs on the two highest upper layers to avoid overfitting. The model also introduces 3D facial alignment, which is another contribution, face photos are often aligned with a 2D comparison transformation before being broadcast on CNNs, but, traditional 2D alignment unable to manage rotations well, the model was trained using softmax loss define as:

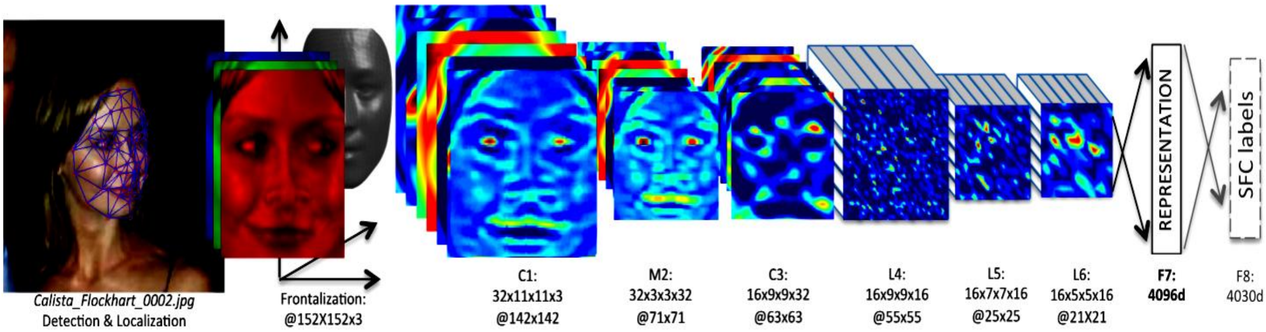

Figure 4: Outline of the DeepFace architecture, one conv-pooling-conv kernel front-end on input, 3 locally connected layers are next, then 2 FCLs. The feature maps for each layer are color-coded. There are higher than 120M parameters on the network greater than 95% originate in locally as well fully connected layers. (Figure from Taigman et al(Taigman et al., 2014))

DeepIdentity or DeepID Architectures

An Advanced Hidden Identity Feature called DeepID was suggested by Sun et al[1](Sun, Chen, Wang, and Tang, 2014) which is a CNN-based feature extraction technique whose learned their features from one large CNN and training a group of mini CNNs, the DeepID model learned, crops or patches of facial photos are fed into a single CNN as input where the features of each CNN has combined to create a powerful feature. The model was trained using both gray and RGB that were extracted near facial locations. DeepID is 2 (Grey and RGB) x 60 x 160 for crops) and (one network's feature length) pixels long, which equals 19,200. Each network has 2 FCLs, 3 max pooling layers, 4 convolutional layers. DeepID solely supervises CNN training by using identity information.

DeepID2 Sun et al[2](Sun, Wang, and Tang, 2015)is an expansion of DeepID, employs both verification and identification information to train a CNN with the goal of maximizing inter-class differences while minimizing intra-class variances. Sun et al[3](Sun, Wang, and Tang, 2014) Introduced DeepID2+ to enhance the performance of DeepID and DeepID2. Unlike DeepID and DeepID2 models, DeepID2+ provides supervisory feedback to the convolutional layers instead of just the higher ones.

On other hand DeepID3 Sun et al[4] (Sun, Liang, Wang, and Tang, 2015) inherited certain features from DeepID2+, including the addition of supervisory signals to early layers and neural weights that are not shared in the final layers of feature extraction. However, it is substantially deeper since several convolution/inception layers are stacked prior to each pooling layer. Unbroken convolution/inception assists in forming features while limiting the amount of parameters that have bigger receptive fields and more complicated nonlinearity.

The DeepID series; DeepID, DeepID2, DeepID2+, and DeepID3 extract strong characteristics from various local patches of the face.

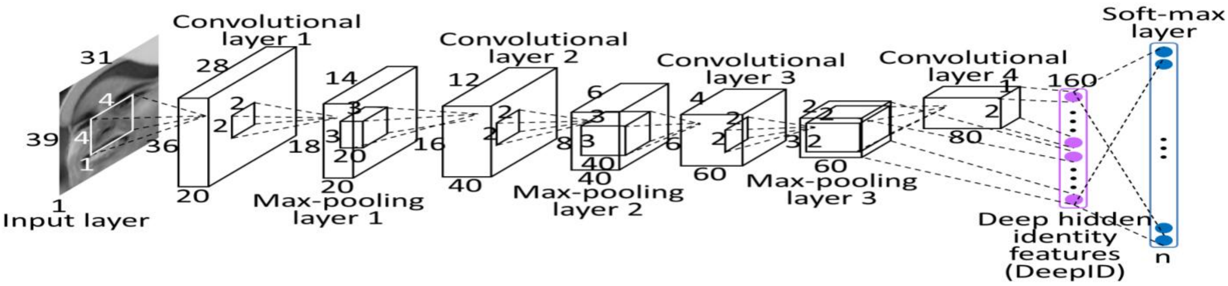

Figure 5: DeepID Architecture. The dimensions of every map for all input are; conv-max-pooling layers are represented by the length, width, and height of each cuboid. The 2D pooling dimensions of the conv-max-pooling layers are indicated by the little cuboids and squares inside respectively. The latest two fully connected layers' neuron numbers are indicated next to each layer(Figure from Sun et al[1] (Sun, Chen, et al., 2014)).

FaceNet

FaceNet is a model suggested by Google researchers Schroff et al (Schroff et al., 2015) that makes use of 128-dimensional representations produced by deep CNNs trained on 260M facial photos employing a triplet loss method as the last layer. In the triplet loss pipeline each training example consists of three images of an anchor, positive, and negative images. The negative image differs from the anchor image while the positive image is similar to it. The triplet loss function seeks to maximize the distance between the anchor and the negative image while minimizing the distance between the anchor and the positive image. Keeping the average loss across all triplets in the training set as low as possible is the goal of the triplet loss. The neural network is able to build a feature representation that captures the similarities and differences between images by minimizing the triplet loss. This feature representation may then be utilized for a variety of tasks, including image retrieval, clustering, and classification. The triplet loss function utilized in (Schroff et al., 2015) is defined as:

$$L\,(A,P,N) = \max\,(\|f(A)-f(P)\|^2 - \|f(A)-f(N)\|^2 + \propto, 0) \quad (1)$$

Where $f(A)$, $f(P)$and $f(N)$ are the embeddings of anchor (A), positive (P), and negative (N) samples, respectively produced by a neural network. $\|f(A)-f(P)\|^2$ represents the squared Euclidean distance between the embeddings of the anchor and positive samples. $\|f(A)-f(N)\|^2$ represents the squared Euclidean distance between the embeddings of the anchor and negative samples, whereα is a margin parameter that defines a minimum desired separation between the positive and negative pairs in the embedding space, max(· ,0) denotes the ReLU (Rectified Linear Unit) activation, ensuring that the loss is non-negative. If the expression inside the parentheses is negative, it's replaced by zero. The function is additionally appropriate for face confirmation, two different core frameworks were discussed based on (Zeiler and Fergus, 2014) architectures.

Visual Geometry Group Face(VGGFace)Architectures

Web-based human-in-the-loop automation was used to assemble the VGG Face from the University of Oxford by (Zeiler and Fergus, 2014), the model was built and

Bima Journal of Science and Technology, Vol. 7 (4) Dec, 2023 ISSN: 2536-6041

DOI: 10.56892/bima.v7i4.546

trained using VGG-16 framework (Simonyan and Zisserman, 2015) intended for image classification which includes of 16 layers, including 3 FCLs with 13 convolutional layers. The model requires a 224 × 224 input image and produces a 2622-dimensional feature vector that corresponds to the facial features in the input image. VGGFace contains Google search photos with a variety of age, pose and ethnicity, it contain 2.6M photos with more than 2.6K identities, at the time of its release, VGGFace demonstrated cutting-edge performance on a number of benchmarks for facial recognition. However, in 2018 another version the database called VGGFace2 by Cao, Q et al (Cao, Shen, Xie, Parkhi, and Zisserman, 2018)also from the University of Oxford was lunched containing 3.31 million photos representing 9131 identities was launched, in general 362 photos for each identity. In contrast VGGFace2 was built on the ResNet50 architecture(He, Zhang, Ren, and Sun, 2016), a more current and sophisticated architecture created especially for deep learning. There are two subclasses of the VGGface2: training set this includes 8631 classes and the test set, which consists of 500 classes.

In addition, there are two template annotations provided to enable evaluation of position as well age as follows;(1) Pose instance: 1.8K templates, 9K facial photos, and 5 faces for each template, each represents a standard pose (three-quarter view, frontal view, or profile view);(2) Age instance: 400 templates with 2,000 facial photos, as well 5 faces for each template. VGGFace2 has now overtaken it in terms of performance and accuracy, taking first place in a number of facial recognition benchmarks and challenges. VGGFace2 differs from VGGFace in that it includes improved data augmentation, a larger input image size, and improved training methods. These characteristics help it outperform VGGFace in terms of performance.

SphereFace

SphereFace is model proposed by (Liu et al., 2017),a deep hypersphere embedding approach using Angular variant of traditional softmax loss called Angular Margin as the loss function for the model CNNs architecture with a straightforward and innovative geometric interpretation, learn to identify distinguishing facial features. The faces lie on a manifold, and the learned features cover a hypersphere manifold in a discriminative manner. In order to avoid divergence at the beginning of training, an annealing optimization technique is used as well an Angular Margin Multiplicative (AMM) penaltyto simultaneously enforce increased inter-class disparity and intra-class coherence. AMM is a version of the popular softmax loss function that seeks to make the class centers farther apart in order to improve the model's ability to discriminate between classes.

The AMM penalty pushes the model to acquire additional discriminative features by adding an angular margin to the softmax loss. The penalty specifically increases the angle between each input's learnt feature representation and the weights of the corresponding class and can derive lower bounds for the loss function that can approximate the learning task. The technique first demonstrates the usefulness of angular margin in FR, increasing the trained model's capacity for discrimination. Having been trained using the CASIA dataset, which is open to the public, SphereFace obtains competitive outcomes on a number of benchmarks on LFW 99.42%.

ArcFace Architecture

Additive-Angular Margin (AAM) Loss used in Deep Face Recognition (ArcFace) method was suggested by (J. Deng et al., 2022). AAM is a different method that is applied while training NNs for classification and face recognition tasks and it is a variant of the softmax loss function, similar to the

DOI: 10.56892/bima.v7i4.546

MAM penalty, but it gives the learnt feature representation an additive margin. The model achieves cutting-edge performance benchmarks on several face-recognition tasks comprising substantial datasets for images and videos and it was presented to addressed the two primary kinds of problems with utilizing CNNs for training deep learning for facial identification first appeared in earlier studies in this field. Some learn an embedding directly from the FC layer, like the triplet loss in (Schroff et al., 2015), while others develop a classifier with many classes that can distinguish many identities contained in the practice set, like the softmax classifier used in DeepID series and VGGFace both utilizes triplet-loss-based techniques as well as softmax-loss-based techniques can successfully recognize faces using the extensive training data and complex DCNN structures.

Bima Journal of Science and Technology, Vol. 7 (4) Dec, 2023 ISSN: 2536-6041

DOI: 10.56892/bima.v7i4.546

Table I: Comparison of diverse Deep Learning Face-Recognition method performance on the Labelled Face in the Wild (LFW) Gallery

N/S	Model and Architecture	Authors	Input image	No. of Para	Description and Contributions	Training Dataset	Accuracy	Drawback
1	DeepFace/ CNN-9	Taigman et al (Taigman et al., 2014)	152x152x3	120M +	Applied piecewise affine transformation to generate features from a 3D face model and 3D face alignment to handle plane rotations and also Siamese network to enable training occurs only on the two topmost layers to prevent overfitting while using Softmax function as cost function .	FaceBook 129M+ Images 4K+ Subjects	97.35	It generates large embedding up to 1000 bytes which may require a lot of memory during training (Agarwal, Yan, Zhang, and Venkataraman, 2023). As well suffer the same drawback highlighted when using sotfmax function in DeepID series.
2	DeepID Series (DeepID, DeepID2, DeepID2+, DeepID3)CNN -9	(Sun, Chen, et al., 2014) Sun et al[2](Sun, Wang, et al., 2015) Sun et al[3](Sun, Wang, et al., 2014) Sun et al[4](Sun, Liang, et al., 2015)	39x 31x3	101M	DeepID: Each CNN uses the input of a face area, concatenates the features, and simultaneously classifies all identities, DeepID2: Is a group of 25 CNNs trained on diverse local patches; nevertheless, verification and identification signals are employed as supervision, and joint bayesian is used to generate a robust embedding space. DeepID2+ builds on DeepID2, combine identification and verification losses. DeepID3: Joint identification-verification supervision has been incorporated in the final layer as well as a few intermediate stages. These designs extracted robust characteristics from several local face patches.	CelebFaces+ (202 k, 10 k)	ID 97.45 ID299 .15 ID2+ 99.47 ID399 .53	It is hard to do face recognition task at scale, the sofmax loss learned features are not discriminative enough because the dimensions of the linear transformation matrix rise linearly with the quantity of identities (Benouareth, 2021); and secondly, for the closed-set classification challenge, the acquired features can be distinguishable, but they are not discriminatory enough for the open-set face identification task (R. Wang, Gao, Li, and Dong, 2023).

3	FaceNet/ GoogleNet	Schroff et al (Schroff et al., 2015)	96x 96x3	140M	Learn a mapping directly from photos to a compact Euclidean space and leaned extremely effective representation, the primary objective of the study is to project face into 128-D embedding where the margin will not be more than 0.2 by L2 Norm modeling with the Triplet Loss function to increase distances between samples of different classes while decreasing distances between samples of the same class.	Google 260 M images 8 M Subjects	99.63	Triplet loss causes slower convergence, and when the number of faces employing triplets increases (Osman, Dennis, and Elgazzar, 2021),(Lu and Lu, 2023), particularly in the case of huge datasets, resulting in an increased in number of iteration steps by significant amount.
4	VGGFace/ VGGNet	Parkhi et al (Parkhi et al., 2015)	64x 64x3		Combine a very deep CNNs with triplet embedding as well the VGGFace model performs better on a face with low resolution images matching in comparison with Microsoft's Residual and FaceNet Google's inception architectures.	VGGFace (2.6 M, 2.6 K) 2	98.95	Same as the problem associated with the use of triplet function highlighted in FaceNet above.
5	SphereFace/ ResNet-64	Liu et al (Liu et al., 2017)	112 x 96 x 3		SphereFace uses annealing as an optimization approach to prevent divergence at the start of the training procedure, as well as a MAM penalty to demand additional inter and intra classes compactness mismatch concurrently, resulting in a trained model with higher discriminative ability.	CASIA WebFace (494 k, 10 k) 2	99.42	The loss function needed a succession of approximations during computation, resulting in unstable network training (S. Wang, Teng, and Perdikaris, 2021).

Bima Journal of Science and Technology, Vol. 7 (4) Dec, 2023 ISSN: 2536-6041

DOI: 10.56892/bima.v7i4.546

6	ArcFace/ ResNet-100	Deng et al (J. Deng et al., 2022)	112 x 112 x 3	One of the most often used and successful loss functions is ArcFace, which can easily converge on any training dataset and does not require the use of other loss functions to get consistent results. ArcFace maximizes the decision boundary in angular space.	MS-Celeb-lM (5.8 M, 85 k) 2	99.83	The gradient curve and the non-monotonic logit (Wu, 2022), as well as an improper trend of loss value and utilize cosine function as the target logit (Alirezazadeh and Dornaika, 2023).

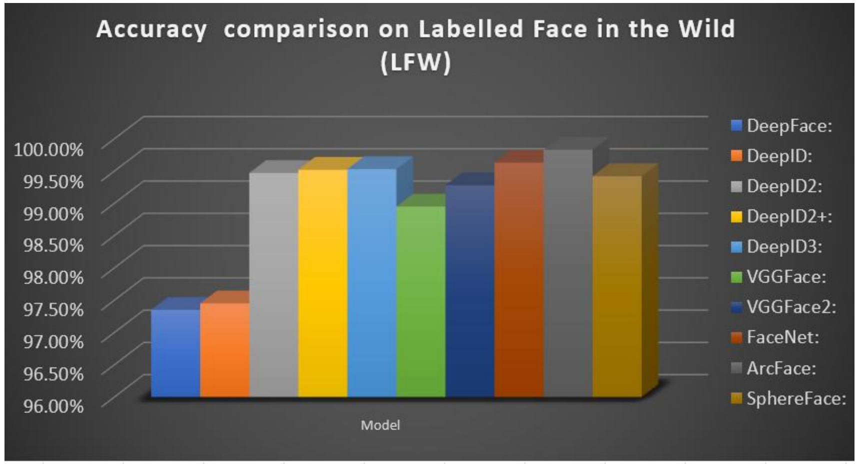

Figure 6: Shows the graphical representation of the Compared Deep Learning Face-Recognition methods

DOI: 10.56892/bima.v7i4.546

DISCUSSION

Facial recognition systems have witnessed substantial evolution, manifesting in various deep learning models employing diverse loss functions. Table 1 gave the summary of deep learning face recognition methods, face recognition systems are modelled as multi-class classification problem ((Krizhevsky et al., 2017), (Simonyan and Zisserman, 2015), (He et al., 2016), (LeCun, Bottou, Bengio, and Haffner, 1998) and (Szegedy, Ioffe, Vanhoucke, and Alemi, 2017)) on the surface, so the selection of the loss function is perhaps one of the most important element that will determine how well a model performs. In this comparative analysis, we delve into the intricacies of ten prominent models, scrutinizing their architectures, training datasets, accuracies, and inherent drawbacks. Each of the ten models under comparison uses a different loss function in order to optimize their model for recognizing faces. These functions aim to increase the distance across embeddings of distinct identities while reducing the gap between identical identity embeddings.

While evaluating these models, it is crucial to consider the trade-offs between accuracy, convergence speed, and memory requirements. DeepFace and the DeepID series, despite their lower convergence and memory issues, provide robust facial recognition. DeepFace (Taigman et al., 2014) of CNN-9 architecture recorded 97.35% accuracy by applying a piecewise affine transformation for feature generation from a 3D face model as shown in Figure 6, but it require a lot of memory during training due to the large embedding requirement, on the same note the optimization procedure can be slow which necessitates careful setting of hyper-parameters like the learning rate and margin. However, its drawback lies in memory-intensive training due to large embeddings.

DeepID2, DeepID2+ and DeepID3 attained accuracy of 99.15%, 99.47% and 99. 53% respectively, both of these methods were trained using sofmax, which learn features that are not discriminative enough because the number of identities allows the linear transformation matrix to increase linearly, on the other hand DeepID with accuracy of 97.45% require careful tuning of hyper-parameters and the optimization process is slow, while FaceNet and VGGFace achieved 99.63% and 98.95 both were trained using triplet loss and attained higher accuracy but record lower convergence in comparison with those trained using softmax.

GoogleNet introduced FaceNet, utilizing a compact Euclidean space mapping with triplet loss. Despite its impressive 99.63% accuracy, FaceNet faces challenges of slower convergence, especially with an increased number of faces. The triplet loss function used here can lead to overfitting and extended iteration steps. FaceNet and VGGFace, with triplet loss, exhibit superior accuracy but slower convergence. There are certain issues with loss of the triplet and that of the softmax. Regarding softmax loss: Instead of describing images' similarity and dissimilarity in an absolute meaning as in Contrastive loss (Hadsell, Chopra, and LeCun, 2006), the terms are only utilized in a relative sense. As the number of identities rises, the size of the linear transformation matrix grows linearly. (2) When it comes to the closed-set classification task, the learned features can be separated which lack sufficient discrimination with regard to the open-set facial recognition problem. On the case of triplet loss: (1) The optimization process is expensive an increase in the prevalence of face triplets always increased, particularly for large-scale datasets, which increases the number of iteration steps significantly which result to over fitting and slow convergence; (2) Mining semi-hard samples is also challenging task for efficient model training. Despite the fact that Sphereface pioneered the Angular Margin idea and their loss function may not required a number of approximate values to be calculated, this led to network training that is unreliable.

SphereFace (Liu et al., 2017) reached performance of 99.42%, addresses issues of the triplet loss but introduces new challenges in training stability, while ArcFace (J. Deng et al., 2022) attained the state of art performance by introducing Additive Angular Margin Loss to mitigate the two main problems associated with the preceding approaches and achieved 99.83% accuracy. ArcFace immediately maximizes the margin for the geodesic distance due to the perfect correlation in the normalized hypersphere between arc and the angle. ArcFace is not required to be paired with another function to get steady accuracy, and it quickly converges on any dataset.

These findings underscore the challenges and advancements in deep learning-based face recognition systems. Future research directions should focus on optimizing convergence speed, reducing memory requirements, and addressing the challenges posed by traditional loss functions.In conclusion, the comparative analysis reveals that the choice of the loss function profoundly influences the performance of facial recognition models. As we move forward, innovations like ArcFace's loss function pave the way for more robust and efficient face recognition systems, contributing to the broader landscape of artificial intelligence applications.

CONCLUSION

Deep learning has made significant advances in a variety of artificial intelligence problems (Goodfellow, Bengio, and Courville, 2016), this achievement may be attributed to parallel processing capabilities, considerably reduction in cost of hardwareand the most recent developments in machine learning research, like CNN, (L. Deng and Yu, 2014). Deep architecture allows deep learning to solve countless complex AI tasks, (Bengio, 2009). Numerous computer vision applications, including face recognition, natural language processing, audio processing, hand writing identification, visual object identification, speech recognition, object detection, and machine translation have been successfully applied by researchers using deep learning. CNNs' ability to decrease the number of ANN parameters and sparsity of connection has prompted developers and scientists to manage larger models to solve complicated tasks, which was not possible with traditional ANNs, on the same note fully connected network does not scale well on nonlinear data like image and audio, because of the number of trainable parameters. Face recognition systems is a popular topic of study due to its potential usefulness in a lot of real-world applications including access control, surveillance, homeland security and so on. CNNs achieved excellent results on Face-Recognition in an unconstrained environment, where an image is provided to the NN model as a raw pixel, unlike the traditional features, CNN learned robust features for verification, identification as well clustering tasks and are more resistant to complex intra-personal differences. The outcomes of this study shows that deep learning approaches for face identification outperformed the hand-craft face recognition algorithms like HOG, SIFT, LBP etc. Face identification models like FaceNet, DeepID variants and DeepFace are available for researcher, many datasets both public and private have been established and evaluated, however some of these dataset are skewed; IJB-A and LFW are utilized to train deep learning facial recognition systems by researchers to compare their results are heavily bias to male with 77 % and white with 83 % in LFW while IJB-A containing 20.4% of darker skin and 79.6% lighter skin images as well containing a lot of images of celebrities whom were taking by professional photographers. It is recommended to have benchmark dataset that will close up gap in all races (Asian and African).

REFERENCES

Abate, A. F., Nappi, M., Riccio, D., and Sabatino, G. (2007). 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, *28*(14), 1885–1906.

DOI: 10.56892/bima.v7i4.546

Agarwal, S., Yan, C., Zhang, Z., and Venkataraman, S. (2023). Bagpipe: Accelerating deep recommendation model training. In *Proceedings of the 29th Symposium on Operating Systems Principles* (pp. 348–363).

Agrawal, A., Gans, J., and Goldfarb, A. (2022). *Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence*. Harvard Business Press.

Ahonen, T., Hadid, A., and Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(12), 2037–2041. https://doi.org/10.1109/TPAMI.2006.244

Al-Kawaz, H. (2019). Facial identification for digital forensic. University of Plymouth.

Alirezazadeh, P., and Dornaika, F. (2023). Boosted Additive Angular Margin Loss for breast cancer diagnosis from histopathological images. *Computers in Biology and Medicine*, *166*, 107528.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. https://doi.org/10.1016/j.cviu.2007.09.014

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, *2*(1), 1–27. https://doi.org/10.1561/2200000006

Benouareth, A. (2021). An efficient face recognition approach combining likelihood-based sufficient dimension reduction and LDA. *Multimedia Tools and Applications*, *80*(1), 1457–1486.

Best-Rowden, L., Han, H., Otto, C., Klare, B. F., and Jain, A. K. (2014). Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security*, *9*(12), 2144–2157.

Blanz, V., and Vetter, T. (2023). A morphable model for the synthesis of 3D faces. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (pp. 157–164).

Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). BRIEF: Binary robust independent elementary features. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6314 LNCS, pp. 778–792). Springer. https://doi.org/10.1007/978-3-642-15561-1_56

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. In *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018* (pp. 67–74). IEEE. https://doi.org/10.1109/FG.2018.00020

Choudhary, S., and Moriwal, R. (2020). Face recognition based on PCA and BPNN. *J. Res. Sci. Eng*, *2*(10).

Clemons, J. (2007). Sift : Scale Invariant Feature Transform. *Transform*, *2*, 91–110. Retrieved from http://www.eecs.umich.edu/~silvio/teaching/EECS598/lectures/lecture10_1.pdf

Dalal, N., and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* (Vol. I, pp. 886–893). Ieee. https://doi.org/10.1109/CVPR.2005.177

Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., and Zafeiriou, S. (2022). ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 44, pp. 5962–5979). https://doi.org/10.1109/TPAMI.2021.3087709

Deng, L., and Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, *7*(3–4), 197–387.

Gao, G., Yu, Y., Yang, M., Chang, H., Huang, P., and Yue, D. (2020). Cross-resolution face recognition with pose variations via multilayer locality-constrained structural orthogonal procrustes regression. *Information Sciences*, *506*, 19–36. https://doi.org/10.1016/j.ins.2019.08.004

Ghali, F., Ali, N., and Yousif, A. (2020). Fingerprint Recognition. *IOP Conference Series: Materials Science and Engineering*, *928*(3), 51–96. https://doi.org/10.1088/1757-899X/928/3/032078

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Hadsell, R., Chopra, S., and LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 1735–1742). IEEE. https://doi.org/10.1109/CVPR.2006.100

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2016-Decem, pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

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

Jain, A. K., Nandakumar, K., and Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, *79*, 80–105.

Johnson, W. L., Rickel, J. W., and Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, *11*(1), 47–78.

Jridi, M., Napoléon, T., and Alfalou, A. (2018). One lens optical correlation: application to face recognition. *Applied Optics*, *57*(9), 2087. https://doi.org/10.1364/ao.57.002087

Kanade, T. (1977). *Computer recognition of human faces*. *Computer recognition of human faces* (Vol. 47). Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-5737-6

Kataria, A. N., Adhyaru, D. M., Sharma, A. K., and Zaveri, T. H. (2013). A survey of automated biometric authentication techniques. In *2013 Nirma university international conference on engineering (NUiCONE)* (pp. 1–6). IEEE.

Kortli, Y., Jridi, M., Al Falou, A., and Atri, M. (2020). Face recognition systems: A survey. *Sensors*, *20*(2), 342.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. In *Communications of the ACM* (Vol. 60, pp. 84–90). https://doi.org/10.1145/3065386

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2323. https://doi.org/10.1109/5.726791

Lei, Z., Pietikainen, M., and Li, S. Z. (2014). Learning discriminant face descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(2), 289–302. https://doi.org/10.1109/TPAMI.2013.112

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. (2017). SphereFace: Deep hypersphere embedding for face recognition. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Vol. 2017-Janua, pp. 6738–6746). https://doi.org/10.1109/CVPR.2017.713

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.00000296

64.99615.94

Lu, Z., and Lu, Y. (2023). A Balanced Triplet Loss for Person Re-Identification. *International Journal of Pattern Recognition and Artificial Intelligence*, *37*(01), 2256022.

Lv, X. F., and Ping, T. (2013). Realization of face recognition system based on Gabor wavelet and elastic bunch graph matching. In *2013 25th Chinese Control and Decision Conference, CCDC 2013* (pp. 3384–3386). IEEE. https://doi.org/10.1109/CCDC.2013.6561532

Maltoni, D., Maio, D., Jain, A. K., and Prabhakar, S. (2009). *Handbook of fingerprint recognition* (Vol. 2). Springer.

Minaee, S., Abdolrashidi, A., Su, H., Bennamoun, M., and Zhang, D. (2023). Biometrics recognition using deep learning: A survey. *Artificial Intelligence Review*, 1–49.

Nagpal, K., Foote, D., Liu, Y., Chen, P. H. C., Wulczyn, E., Tan, F., … Stumpe, M. C. (2019). Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *Npj Digital Medicine*, *2*(1), 48. https://doi.org/10.1038/s41746-019-0112-2

Nevo, S., Anisimov, V., Elidan, G., El-Yaniv, R., Giencke, P., Gigi, Y., … Matias, Y. (2019). ML for Flood Forecasting at Scale. *ArXiv Preprint ArXiv:1901.09583*. Retrieved from http://arxiv.org/abs/1901.09583

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., … Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, *223*, 103525.

Osman, Y., Dennis, R., and Elgazzar, K. (2021). Yield Estimation and Visualization Solution for Precision Agriculture. *Sensors*, *21*(19), 6657.

Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep Face Recognition, 41.1-41.12. https://doi.org/10.5244/c.29.41

Raju, K., Chinna Rao, B., Saikumar, K., and Lakshman Pratap, N. (2022). An optimal hybrid solution to local and global facial recognition through machine learning. *A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems*, 203–226.

Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. *Computer Standards and Interfaces*, *16*(3), 265–278.

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 07-12-June, pp. 815–823). https://doi.org/10.1109/CVPR.2015.7298682

Shanthi, K. G., and Sivalakshmi, P. (2023). Smart drone with real time face recognition. *Materials Today: Proceedings*, *80*, 3212–3215.

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Singh, H. (2019). *Practical Machine Learning and Image Processing For Facial Recognition, Object Detection, and Pattern Recognition Using Python-Himanshu Singh*. Springer. Retrieved from www.apress.com/978-1-4842-4148-6

Sun, Y., Chen, Y., Wang, X., and Tang, X. (2014). Deep learning face representation by joint identification-verification. *Advances in Neural Information Processing Systems*, *3*(January), 1988–1996.

Sun, Y., Liang, D., Wang, X., and Tang, X. (2015). DeepID3: Face Recognition with Very Deep Neural Networks. *ArXiv*

DOI: 10.56892/bima.v7i4.546

Preprint ArXiv:1502.00873. Retrieved from http://arxiv.org/abs/1502.00873

Sun, Y., Wang, X., and Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1891–1898). https://doi.org/10.1109/CVPR.2014.244

Sun, Y., Wang, X., and Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 07-12-June, pp. 2892–2900). https://doi.org/10.1109/CVPR.2015.7298907

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. In *31st AAAI Conference on Artificial Intelligence, AAAI 2017* (Vol. 31, pp. 4278–4284). https://doi.org/10.1609/aaai.v31i1.11231

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1701–1708). https://doi.org/10.1109/CVPR.2014.220

Turk, M., and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86. https://doi.org/10.1162/jocn.1991.3.1.71

Ulhaq, A. (2021). Deep learning, past present and future: An odyssey.

Viola, P., and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 1, pp. I–I). Ieee. https://doi.org/10.1109/cvpr.2001.990517

Wang, M., and Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, *429*, 215–244. https://doi.org/10.1016/j.neucom.2020.10.081

Wang, R., Gao, R., Li, Q., and Dong, J. (2023). Pig Face Recognition Based on Metric Learning by Combining a Residual Network and Attention Mechanism. *Agriculture*, *13*(1), 144.

Wang, S., Teng, Y., and Perdikaris, P. (2021). Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, *43*(5), A3055–A3081.

Weigelt, S., Koldewyn, K., and Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience and Biobehavioral Reviews*, *36*(3), 1060–1084.

Wright, J., Ganesh, A., Zhou, Z., Wagner, A., and Ma, Y. (2008). Demo: robust face recognition via sparse representation. *2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008*, *31*(2), 210–227. https://doi.org/10.1109/AFGR.2008.4813404

Wu, L. (2022). Learning a single neuron for non-monotonic activation functions. In *International Conference on Artificial Intelligence and Statistics* (pp. 4178–4197). PMLR.

Zeiler, M. D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8689 LNCS, pp. 818–833). Springer. https://doi.org/10.1007/978-3-319-10590-1_53

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