Advances in Technology Innovation, vol. 7, no. 4, 2022, pp. 279-294

Learning Representations for Face Recognition: A Review from Holistic to Deep Learning

Fabian Barreto[1*], Jignesh Sarvaiya[2], Suprava Patnaik[3]

[1]Department of Electronics and Telecommunication, Xavier Institute of Engineering, Mumbai, India

[2]Department of Electronics, Sardar Vallabhbhai National Institute of Technology, Surat, India

[3]School of Electronics, Kalinga Institute of Industrial Technology, Bhubaneswar, India

Received 18 August 2021; received in revised form 07 January 2022; accepted 08 January 2022

DOI: https://doi.org/10.46604/aiti.2022.8308

Abstract

For decades, researchers have investigated how to recognize facial images. This study reviews the development of different face recognition (FR) methods, namely, holistic learning, handcrafted local feature learning, shallow learning, and deep learning (DL). With the development of methods, the accuracy of recognizing faces in the labeled faces in the wild (LFW) database has been increased. The accuracy of holistic learning is 60%, that of handcrafted local feature learning increases to 70%, and that of shallow learning is 86%. Finally, DL achieves human-level performance (97% accuracy). This enhanced accuracy is caused by large datasets and graphics processing units (GPUs) with massively parallel processing capabilities. Furthermore, FR challenges and current research studies are discussed to understand future research directions. The results of this study show that presently the database of labeled faces in the wild has reached 99.85% accuracy.

Keywords: learning representations, deep learning, autoencoders, variational autoencoders

1. Introduction

In the modern world, automatic face recognition (AFR) is embedded into smart e-commerce applets for better personalization and marketing of commodities, such as hair styling and digital makeup. Consumer-based photography has become a new trend in selecting a range of products that suit consumers' needs, with social media platforms providing facial recognition services to attract diverse users. Conventional facial recognition (FR) requirements are limited to basic security and access control applications, and are implemented in more advanced ways. Examples include accessing historical data and using cloud-based database identification and closed-circuit television (CCTV) video-supported tracking, leading to better enforcement of the law. Facial identification has become essential for forensics, surveillance, border control, lie detection, and access ID verification.

FR, in its various dimensions, is currently a research area in computer vision, and is the process of detecting and locating faces from a background, normalizing face images, and performing face verification (FV) or face identification (FI). There are two separate tasks for face matching while conducting FR, namely: FV and FI. In FV, one determines whether a given test image is from the same person being verified, while the FI aims to recognize the facial images of persons already enrolled in the database [1]. To verify genuineness, the output of FR is either "yes" or "no," which may be a result of the class number corresponding to the input image. In the FV, the input image is assumed to be a sample from a known possible class of inputs.

* Corresponding author. E-mail address: frfabiansj@xavier.ac.in

Tel.: +919833916407

Regarding face detection (FD), in 2001, Viola and Jones [2] used Haar-like features to detect human faces. A 24 × 24 pixel can have over 160,000 Haar-like features. The framework used the concept of integral images to perform intensive computation and the adaptive boost (AdaBoost) algorithm to select the best features from different subsets.

Wang et al. [3] categorized FD and recognition development into four broad representation learning types: holistic, handcrafted, shallow, and deep learning. Traditionally, FR techniques have been divided into two major categories: geometric and photometric techniques. Here, geometric techniques find distinct features and spatial positioning to form a template that is used to compare and eliminate variances in face images. Photometric approaches are distilled out and use hidden statistical properties that account for the entire input of facial images. Popular photometric approaches include principal component analysis (PCA) using the eigenface algorithm and linear discrimination analysis (LDA) using the Fisherface algorithm. The holistic approach uses low-dimensional representations in the form of a manifold or a linear subspace. However, this approach is limited by variations such as face appearances that introduce different statistical distributions, which are difficult to manage.

The early twentieth century saw a transition to handcrafted local feature-based methods. Inherent face changes are managed through local descriptors, such as local binary patterns (LBPs), Gabor filters, and histograms of oriented gradients (HOGs). These local features help remove redundant and meaningless information from raw representation; thus, they provide greater robustness than existing methods and are greatly invariant to transformation. The limitations of these approaches are that they suffer from a lack of compactness, distinctiveness over a large sample space, and acceptance for real-time applications, as well as being slow and susceptible to poor generalization.

Shallow representation learning, with a one- or two-layer representation, improved the distinctness of the codebook. Noticeable shallow approaches included the learning-based (LE) approach, discriminant face descriptor (DFD), feature vector, and PCANet. However, these approaches were not robust to the complex non-linear nature of the face.

Deep learning (DL) is a revolutionary approach that has changed the facial recognition landscape. In 2012, AlexNet achieved state-of-the-art (SOTA) recognition accuracy and propelled research toward DL for computer vision. Researchers have used a convolutional neural network (CNN) that exhibited strong invariance to face pose, lighting, expression, and other variations to achieve high accuracy. Thus, this research addresses recognition accuracy and investigates the complexity of learning a large number of features, dependency on datasets, protocols addressing application scenarios, and model interpretability. This research also addresses variations encountered owing to cross-posed, aging, and other adversarial conditions.

Since the 1990s, remarkable advances have been made in FD and recognition. This study aims to review the development of learning representations for FR in the past three decades and has resulted in an accuracy increase of 39.85% for labeled faces in the wild (LFW) database from the earlier methods used three decades ago. The remainder of this study is organized as follows. Section 2 describes the initial holistic representation of the learning stage for FR, and section 3 describes the transition to a handcrafted stage. Section 4 presents the shallow learning phase, and section 5 deals with the DL phase and some challenges and current research studies. Finally, section 6 provides the conclusions of this study.

2. Review of Holistic Learning

The earliest holistic stage begins by using eigenfaces, motivated by Sirovich and Kirby [4], to efficiently represent face images using PCA. They then transition to Fisherface algorithms and LDA and later to independent component analysis (ICA), leading to sparse representation-based classification (SRC), a particular case of collaborative representation-based classification (CRC). Later, researchers used distance metric learning with improved class separability, meaning that the holistic stage can assume certain distributions (linear, manifold, and sparse) from which it arrives at a low-dimensional representation. However, these assumptions do not hold firm ground on the variations in facial features.

2.1. Principal component analysis (*PCA*)

Ballantyne et al. [5] mentioned the pioneering work of Woody Bledsoe and his AFR team. They manually classified face images with landmarks (e.g., eye centers and mouth) and saved the metrics in a database. Goldstein et al. [6] enhanced the accuracy by using 21 specific subjective markers on the face. The work of Turk and Pentland [7] in 1991 gave a new direction to using eigenfaces (PCA) to develop the first AFR system. Varying the illumination and pose conditions is a challenging task for this method. It is essential to understand that a particular eigenfeature may not be related to recognition, but to the direction of illumination. Hence, an increase in eigenfeatures does not necessarily lead to better accuracy. PCA can only set apart the linear dependencies in the pixel pair of a facial image.

PCA is a method for expressing data vectors in their principal components (PCs), where the largest variances in the data indicated the direction of the PCs (Fig. 1). PCs capture the most significant data information and correspond to the eigenvectors given by the largest eigenvalues of the autocorrelation matrix of the data vectors. PCA computes the most representational basis for looking at the dataset and generally works as follows. First, it calculates the covariance matrix of the given data points and calculates the eigenvectors and corresponding eigenvalues sorted in decreasing order. Then, the first k eigenvectors are chosen from the n eigenvectors ($k < n$), yielding the novel k dimensions. Thus, the original n higher dimensions were transformed into k fewer dimensions.

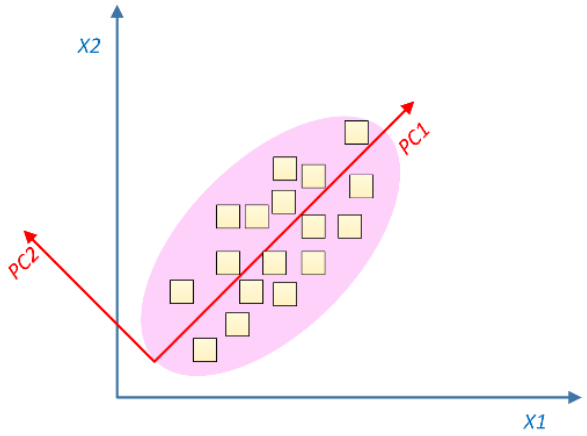


Fig. 1 Original space ($X1$, $X2$) and PCA reduced space ($PC1$, $PC2$)

2.2. Linear discrimination analysis (*LDA*)

LDA constructs a subspace that differentiates between different face images, while Fisher discriminant analysis classifies face images into groups based on their facial features. Zhao et al. [8] used LDA for FR because it encodes discriminatory information. They used PCA to project the face image to a subspace and used an LDA to obtain a linear classifier in the subspace. The pure LDA approach, however, does lead to an overfitting problem and does not perform well for samples from different classes and samples with diverse backgrounds.

2.3. Independent component analysis (*ICA*)

ICA describes a subspace method that transforms data from high to low dimensions. It finds a linear transformation that leads to the minimization of the statistical dependence between its components. However, unlike PCA, it provides an improved probabilistic model, a greater response to high-order statistics, and better reconstruction in noisy environments [9]. A set of statistically independent basis images for a set of face images is found by separating the independent components of the facial images (Fig. 2). Here, let S be a set of statistically independent source images, which is unknown, with X as the source of the face images and A as an unknown combination matrix. W_I is a matrix of learned filters which in turn produces outputs U that are statistically independent. ICA outputs in rows that are $W_I X = U$.

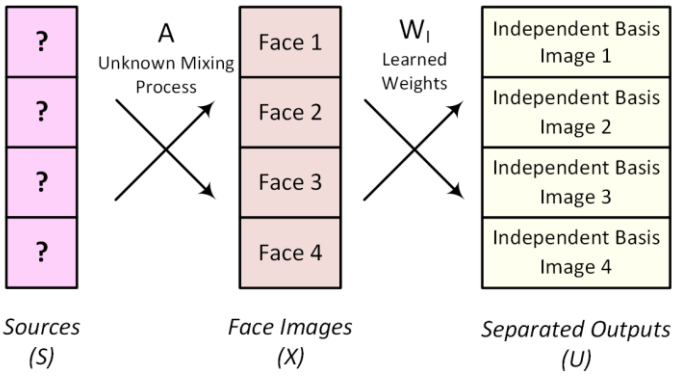


Fig. 2 Image synthesis model

2.4. *Hidden Markov model* (*HMM*)

In a hidden Markov model (HMM), patterns are characterized as parametric random processes. These parameters can be estimated precisely and logically. Samaria et al. [10] used the HMM model to represent the statistics of facial images. They converted a two-dimensional face image to a one-dimensional sequence. As shown in Fig. 3, the face is split into regions (e.g., the forehead, eyes, nose, mouth, and chin). After determining the hidden states (five in the given figure), the HMM is trained to learn the state transitional probability. After training on the output probability, the class was determined. Although HMM has a better detection rate, it also has a higher false-alarm rate.

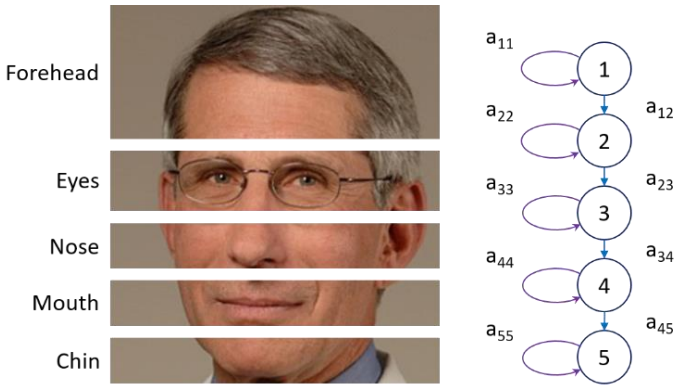


Fig. 3 Five-state HMM

2.5. *Bayesian model*

Schneiderman et al. [11] derived a probabilistic model for FR using local regions, such as the eyes, nose, and mouth. Their statistical model captured the more unique patterns of the human face, such as the intensity patterns around the eye, to represent the local features more uniquely. They also modelled the joint probability of local features and positions, as human faces are easily recognized because of their proper spatial arrangement. They used the Bayesian decision rule, also known as maximum a posteriori (MAP), and calculated a larger probability for a given input image x, namely, $P(\text{face} \mid x)$ or $P(\text{not face} \mid x)$, indicating whether a face was selected. Yang et al. [12] presented two advantages of using a naive Bayes classifier; that is, it provided a better estimation of the subregion conditional density functions and provided an MAP to understand the joint statistics of a local feature and its position.

2.6. *Locality preserving projection* (*LPP*)

He et al. [13] proposed an appearance-based Laplacian method for facial recognition by using locality preserving projections (LPPs) to map facial images into a subspace. Eigenfaces (PCA) preserve the global surface of the face image, whereas the Fisherface algorithm (LDA) preserves discriminating information. The advantage of LPP over PCA and LDA is

that it preserves local features and detects the essential face manifold surface, where the nearest-neighbor graph models this surface. The face images in the lower-dimensional subspaces are called Laplacian faces. Facial recognition was performed in three steps. Laplacian faces were calculated from the given training face image samples, and the test image is then projected onto the Laplacian face subspace. Finally, the nearest-neighbor classifier identifies a new face. As this method considers the face manifold, it considers varying illumination conditions.

2.7. Sparse representation-based classification (SRC) and collaborative representation-based classification (CRC)

SRC and CRC belong to sparse representation-based classifiers. The test input image was a linear connection between the recorded images. The test image can be recognized as the combination coefficients for the target faces, which are larger than the others. In SRC/CRC, the test face images are coded over others with sparsity constraints, such as L_1 minimization. SRC/CRC uses the reconstruction error to determine the face image. In the work of Wright et al. [14], the discriminative property of an SRC model for classification was used, while in the work of Zhang et al. [15] and Zhang et al. [16], it was shown that the good performance of SRC is primarily due to the collaborative representation of the test face image with training samples across different classes.

2.8. Distance metric learning

In distance metric learning, one learns a distance metric for the input space of face images from a given set of similar/dissimilar points in the training face images. Yang et al. [17] categorized the algorithms for distance metric learning into supervised and unsupervised methods. Supervised training face images are placed into pairwise constraints: pairs of same-class data points in the equivalence constraints and those that belong to different classes in equivalence constraints. Supervised learning can be global or local, where global satisfies pairwise constraints simultaneously and local only meets local pairwise constraints. Supervised learning includes supervised global learning, local adaptive supervised learning, neighborhood component analysis, and relevant component analysis (RCA), while unsupervised learning includes linear-like PCA and multidimensional scaling. They also include nonlinear embedding methods such as isometric mapping, linear embedding, and Laplacian eigenmaps.

Jin et al. [18] presented a regularized distance metric learning algorithm that is robust for high-dimensional data. Here, the generalization error of regularized distance metric learning is independent of dimensionality. The algorithm was tested with the baselines of the Euclidean distance metric, Mahalanobis distance metric, large margin nearest neighbor classifier, information-theoretic metric learning, and RCA and was comparable to SOTA approaches for distance learning.

3. Review of Handcrafted Local Feature Learning

To enhance the holistic method, researchers started using handcrafted local features. They used Gabor wavelets, elastic bunch graph matching (EBGM), local binary patterns (LBP), and high dimensional local binary patterns (HD-LBP). These methods did achieve robust performance. However, as the features increased, there was a problem of distinctiveness, and the large size created the problem of non-compactness.

3.1. Gabor wavelet (filter)

Gabor introduced the Gabor wavelet (or Gabor filter) in 1946 as a band-pass filter and has an impulse response given by a Gaussian function, multiplied by a harmonic function. Its resolution is optimal in both the domains of space and frequency. Daugman [19] generalized the 1-D Gabor filters to two-dimensional Gabor filters. Liu et al. [20] described a facial recognition Gabor feature classifier where Gabor wavelets first transform the face images to obtain the augmented Gabor FV and then pass through an enhanced Fisher discrimination model. Their results showed that the classifier can discriminate Gabor features with

low dimensionality and increased discrimination. Barbu [21] proposed a 2-D Gabor filter for human FR. He used 2-D Gabor filter banks, which help extract different orientation and scale features from the input face image, resulting in 3-D face feature vectors. One disadvantage is that Gabor features have high dimensionality and result in redundancy [22]. A hybrid method uses Gabor filters and another technique such as PCA to reduce redundancy. Principal Gabor filters that help reduce redundancy are described in the work of Štruc et al. [23]. Here, they used orthonormal linear combinations and derived a Gabor face representation. However, the tradeoff is that the filters are not optimally localized in the space and frequency domains.

3.2. Local binary pattern (*LBP*)

The human face can be viewed as consisting of micro-patterns and hence can use an LBP as a face descriptor [24-25]. LBP was first proposed for texture description [26], where it was observed that certain LBP are key properties of texture and sometimes represent over 90% of all 3 × 3 patterns present in the textures. After thresholding, a histogram that functions as a texture descriptor can be created (Fig. 4). These patterns have uniform circular structures with few spatial transitions and were used as templates. The LBP operator is only a 3 × 3 neighborhood; therefore, it is difficult to capture the features that are dominant for large-scale structures, with later models using neighborhoods of different sizes to correct this issue. LBP efficiently summarizes the local structures of facial images, where each pixel was compared with its neighboring pixels.

An example is shown in Fig. 5. Here, each pixel is compared with its eight neighbors by subtracting the center pixel value. The encoding process is done in the following steps. Encode a 0 for negative; otherwise, encode a 1. Concatenate all binary values in a clockwise direction. Begin from the top-left neighbor and move clockwise. Convert the binary to a decimal value, the label (LBP codes) for the given pixel. LBP is a non-parametric method that converts the face image into an array of integer labels. Huang et al. [27] surveyed LBP and its variants that offer better performance and improved the robustness of the original LBP. Isnanto et al. [28] used LBP and Haar cascade classifier on low-resolution images for multi-object FR.

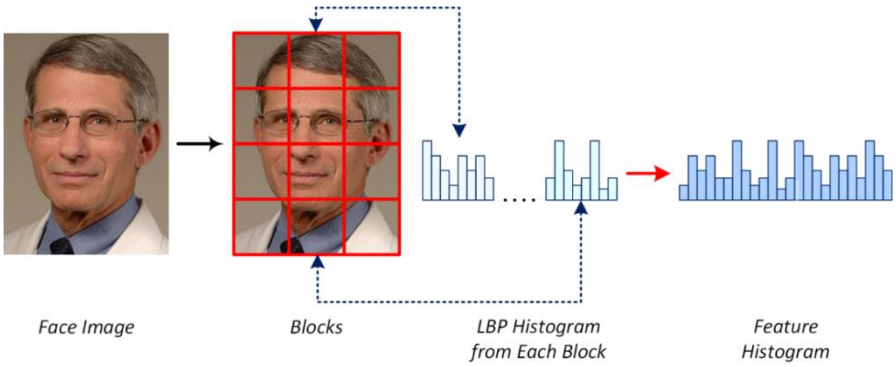


Fig. 4 LBP histogram

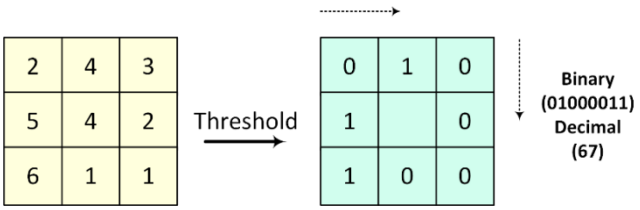


Fig. 5 LBP operator

3.3. Elastic bunch graph matching (*EBGM*)

Bolme [29] described the EBGM FR algorithm. It recognizes new facial images by localizing landmark features and then finds the similarity measure. Facial landmark points were selected manually from a set of model face images with variations. Gabor jets are the names given to the Gabor wavelets extracted from the landmark point and the jets from the model form a face bunch graph. Each node contains a stack of N jets (N = model image). Here, the edge is the distance

between landmark points (Fig. 6). The limitation of the EBGM is that one needs to rely on the model's manual ground truth for landmark selection at the initial recognition stage. Lahasan et al. [30] proposed a method to overcome this shortcoming by posing the EBGM as an optimization problem by using harmony search (HS) to find the optimal facial landmarks using the manual method.

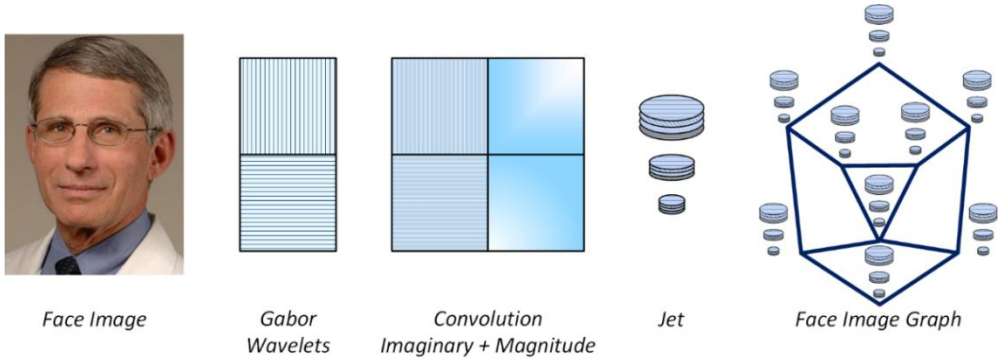


Fig. 6 EBGM process

3.4. Scale-invariant feature transform (*SIFT*)

Scale-invariant feature transform SIFT was proposed by Lowe [32-33]. It creates descriptors that are scale-, rotation-, and translation-invariant and possesses high dimensionality. FR tasks use SIFT features [33-34] to reliably match images. This process includes extracting SIFT keypoints from the face image. How can one find the test image? By finding the matching features. The Euclidean distance was used as the measure; however, a challenge is the reliable extraction of consistent SIFT descriptors. As shown in Fig. 7, the SIFT algorithm has four stages: keypoint detection, keypoint localization, orientation assignment, and keypoint descriptor generation. Keypoint detection uses the difference of the Gaussian (DOG) function to detect feature points, and each keypoint is assigned one or more orientations during the orientation assignment stage. In the last stage, each keypoint is assigned to a vector descriptor. Given that the algorithm is computationally intensive, the actions are performed only at positions that go through the first test. Fig. 8 shows the SIFT features of a 64 × 64 image, its noisy version, and matching features.

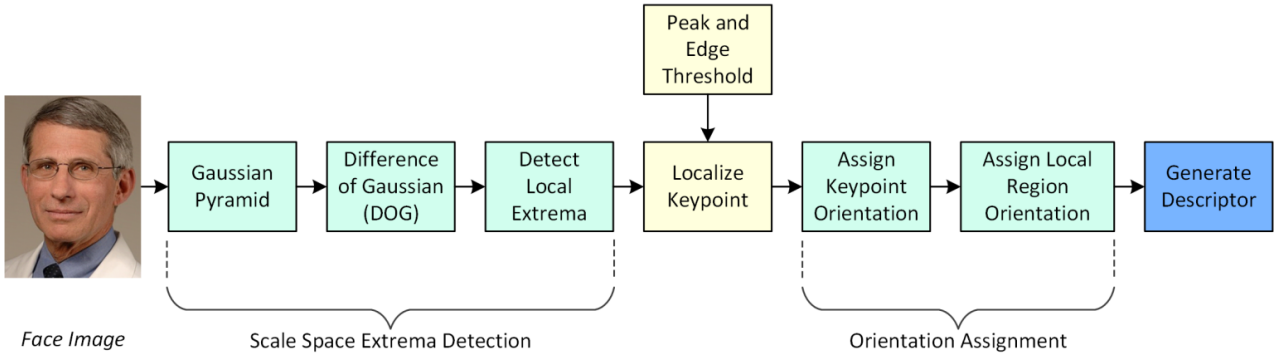
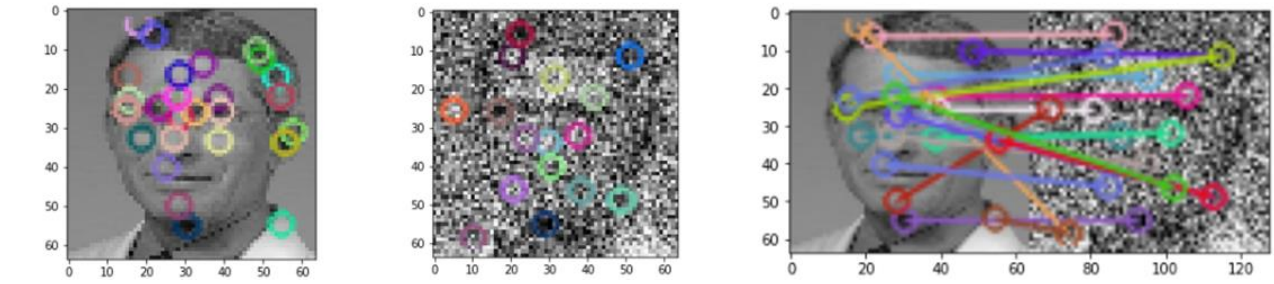


Fig. 7 Stages of the SIFT algorithm



(a) SIFT keypoints of the original image (64 × 64) (b) With added noise (c) SIFT keypoint matching

Fig. 8 Implementation of SIFT

3.5. *Histogram of oriented gradient* (*HOG*)

Dalal et al. [35] developed grids of HOG descriptors, which have the advantage of capturing the gradient (edge) structure, a characteristic of the local shape. The grids count the occurrence of edge orientations in the local neighborhood of the face image. Facial images were split into small and linked regions (cells), and a histogram of the edge orientations was computed for each cell. The histograms were normalized to account for the illumination and combined to form the HOG descriptor. The HOG is invariant to 2D rotation and scaling. Using locally normalized HOG features with an overlapping dense grid yielded better results. Déniz et al. [36] proposed a method for building a robust HOG descriptor by using a regular grid, combining HOG descriptors at different scales, and applying a reduction in linear dimensions.

4. Review of Shallow Learning

The shallow learning-based (LE) local descriptor phase uses local filters to learn distinctiveness and a codebook to achieve compactness. As this was a shallow representation, a one- or two-layer representation, it is not robust to the complex nonlinearity of face images. The method also improves one characteristic, such as pose, light, or expression, but does not address unconstrained changes in the face image in general.

4.1. *Learning-based* (*LE*)

Cao et al. [37] proposed a new LE descriptor that was compact, discriminative, and easy to extract. They list the disadvantages of existing handcrafted methods, as it is challenging to obtain an optimal encoding and unevenly distributed. Their process consisted of extracting face landmarks that aligned nine different parts of the face separately, which were fed into the DOG filter to remove low- and high-frequency illumination variations. Each pixel has a low-level FV encoded by an LE encoder. PCA-reduced histograms were concatenated and then normalized to obtain the LE descriptor, and the similarity of the LE descriptors of the face pair was measured using the L_2 distance norm. The nine component similarity scores were then fed into a pose-adaptive classifier, which resulted in FV.

4.2. *Discriminant face descriptor* (*DFD*)

Lei et al. [38] described a technique for acquiring a DFD. Discriminant local features learn by minimizing the feature differences between the same face images and maximizing those between different face images. The discriminative capability is performed in three steps: learning discriminant image filters, determining the optimal neighborhood sampling, and constructing the dominant patterns. They also used coupled DFDs to view heterogeneous facial data.

4.3. *Feature vector*

Sánchez et al. [39] described the feature vector method for image classification based on the principle of Gaussian mixture distribution. They proposed using the Fisher kernel framework and described their blocks by deviation from a Gaussian mixture distribution with diagonal covariance. Visual vocabulary is a gradient vector for the model parameters. Their method encoded the (probabilistic) count of occurrences and higher-order statistics. The authors listed the advantages of their method as having better results than efficient linear classifiers and compression with a very low loss of accuracy.

4.4. *PCANet*

Chan et al. [40] proposed a baseline model for image classification called PCANet, a precursor to DL models. PCANet consists of cascaded PCA to learn from multistage filter banks, binary hashing, and blockwise histograms and has two variations: RandNet and LDANet. In RandNet, they replaced PCA filters with random filters of the same size at each layer, whereas in LDANet, the supervision of a classification problem was improved by using supervised training. LDA is used to

learn the filters. PCANet eliminated image variability and provided effective accuracy with well-preprocessed images in the datasets. However, PCANet may not sufficiently account for the variability of challenging face images. However, the PCANet is a valuable baseline for studying DL architectures.

5. Review of Deep Learning

The FR landscape saw a fundamental shift with the introduction of AlexNet, which uses DL. DeepFace [41], DeepID [42-43], FaceNet [44], ArcFace [45], and AdaptiveFace [46] have paved the way for an evolution of network architectures, algorithms, and datasets to answer the multi-faceted FR problem. The accuracy results for the LFW database [47] explain the FR development stages. For the holistic stage, the accuracy was 60%, while for handcrafted, it increased to 70%, shallow to 86%, and finally, for DL, especially for DeepFace, it approached human-level performance of 97% for the unconstrained FR.

In the early days of the AFR, the focus was more on developing FD algorithms and less on developing face image datasets. There has been organic growth in the datasets over the past two decades because it has come from the research community in terms of the need for a large number of face images with varying conditions and diversity. Another development has been the challenge to go beyond recognizing faces from laboratory-controlled to unconstrained face images. AFR research has progressed enormously, with some simple datasets achieving 99% accuracy, which has resulted in the development of more complex datasets that can facilitate new directions for FR research.

The number of face images in the datasets and their variations has increased over the years. The past decade with FR research moving toward DL approaches has resulted in the growth of large training datasets required to implement DL algorithms effectively. Taskiran et al. [48] classified face image datasets as image-based or video-based. They may also be 3D or hyperspectral/infrared datasets. Some of the datasets were private, whereas others were public. These datasets are essential for benchmarking new AFR algorithms. A database's choice depends on the given problem that one intends to solve or a property that one wants to test and also depends on the size of the training set required to test the algorithm. Some databases, such as Facebook, Google, CelebFaces+, and VGGFace, were used for training, and others, such as LFW, YTF, and IJB-C, were used for testing.

5.1. Artificial intelligence (AI), machine learning (ML), and deep learning (DL)

John McCarthy, the father of artificial intelligence (AI), coined the term AI in his 1955 proposal for the Dartmouth Conference, USA, in 1956. On a broader scale, AI explores theories and applications to broaden human intelligence and envisions the creation of a future where intelligent machines have human-like perception and cognition. Researchers have made significant progress in understanding and improving learning algorithms; however, the challenge of AI remains [49]. As shown in Fig. 9, DL is a subfield of machine learning (ML), and ML is a subset of the broader field of AI. Some examples of ML problems include classification, clustering, and prediction. Traditional ML techniques are constrained to process data in a basic form and domain experts are required to carefully perform feature extraction [50]. DL is a subset of the ML and learns multiple representations and abstraction levels to understand the data. The raw input was transformed to a higher and more abstract level (Fig. 10). These transformations can help learn complex and intricate functions.

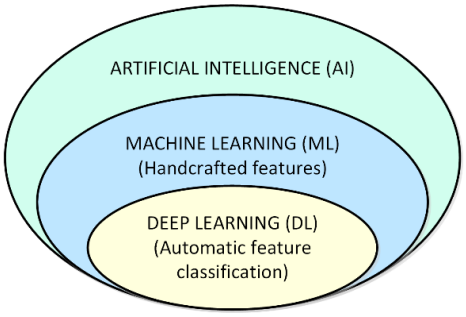
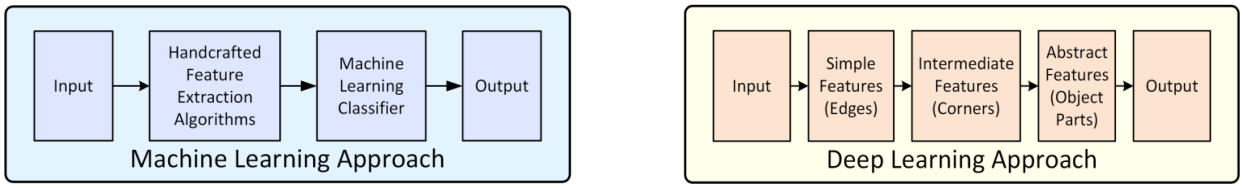


Fig. 9 Relationship of AI, ML, and DL



(a) ML (b) DL

Fig. 10 ML and DL approaches

5.2. *Artificial neural network* (*ANN*)

The unique human brain, especially how neurons interact, has inspired scientists. Artificial neural networks (ANNs) are hardware and software implementations of neural structures in the human brain. The history of neural computing originated with the work of McCulloch and Pitts in 1943. The Warren McCulloch and Walter Pitts model (MCP model, known as the linear threshold gate model) is a binary classifier [51], where the weights were manually adjusted by a human. In the 1950s, Rosenblatt published the Perceptron algorithm, which automatically learns weights without human involvement [52]. This was an enhanced version of the MCP model. The perceptron model adds extra information representing the bias and variable weight values. The 1969 publication by Minsky and Papert [53] weakened neural network research for nearly a decade (1969-1986). They believed that using Perceptrons in practical applications was futile without an adequate basic theory.

In 1979, Fukushima developed a neural network with multiple pooling and convolutional layers called neocognitron, which used a hierarchical and multilayered design that learned how to recognize visual patterns [54]. Rumelhart revived neural network research in 1986 using a backpropagation (BP) algorithm. The neural network iteratively learns weights that are then used to predict class labels. Given sufficient hidden units and sufficient training data multilayers, feedforward networks can closely approximate any function. In 1989, Yann LeCun demonstrated BP at the Bell Labs. He combined CNNs with BP to read handwritten digits. In 1997, long short-term memory for recurrent neural networks (RNNs) was developed by Hochreiter and Schmidhuber, with a gating mechanism to regulate the information to be kept or discarded at each time step.

5.3. *The deep learning phase*

In 2009, Fei-Fei Li launched the challenging benchmark dataset, ImageNet [55]. Between 2011 and 2012, Krizhevsky created AlexNet, a CNN. As shown in Fig. 11, AlexNet has five convolutional layers, followed by max-pooling layers and three fully connected layers. Instead of using *tanh* and *sigmoid* activation functions, he used rectified linear units (ReLUs), which increased the speed and dropout. AlexNet showed that a greater depth resulted in high performance and, despite being computationally expensive, is feasible because of graphics processing units (GPUs). In 2014, DeepFace used neural networks to identify faces from the LFW dataset with 97.35% accuracy, an improvement of 27% over previous efforts [41]. In 2015, the Facenet model, using GoogLeNet-24, achieved 99.63% accuracy for the Google dataset [44]. In 2018, Ring loss model using ResNet-64 achieved 99.5% accuracy for the MS-Celeb dataset [56] and Arcface model using ResNet-100 achieved 99.83% accuracy for the MS-Celeb dataset [45]. In the work of Yan et al. [57], the use of VarGFaceNet resulted in an accuracy of 99.85% for the LFW database.

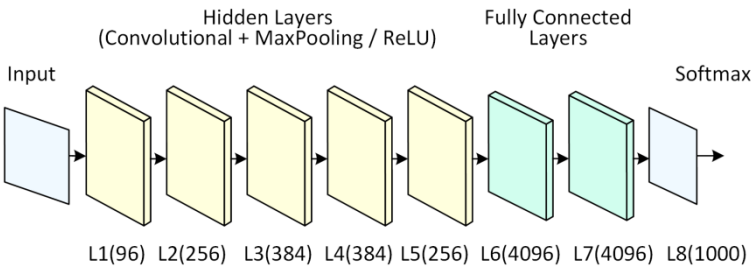


Fig. 11 AlexNet architecture

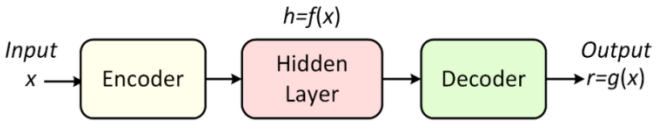


Fig. 12 Autoencoder model

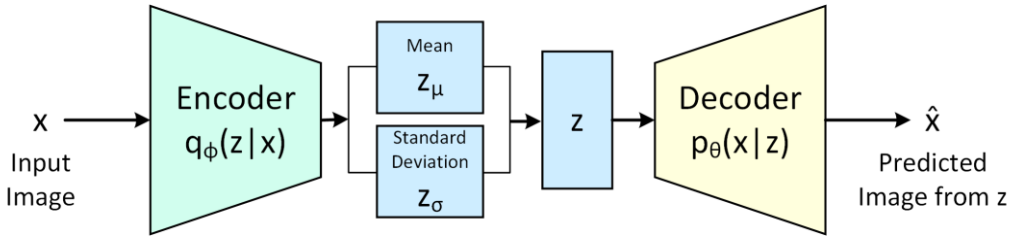


Fig. 13 Variational autoencoder

The evolution of DL is described in detail by Schmidhuber [58]. He explains the hierarchical representation learning for different supervised/reinforcement learning and the various advancements in both feedforward (acyclic) neural networks (FNNs) and recurrent (cyclic) neural networks (RNNs). He also described the evolution of restricted Boltzmann machines (RBMs), as well as the constituents of multilayer learning architectures, such as the deep belief networks (DBNs).

Advances in DL meant working with high dimensional data, which could be reduced to codes of lower dimensionality. In 2006, Hinton and Salakhutdinov [59] trained an "autoencoder" network. Autoencoders [60] are used for dimensionality reduction, denoising, and outlier detection and are made up of three sections, as shown in Fig. 12. The encoder encodes the data to the hidden layer (code) which results in an output $h = f(x)$. The decoder then outputs $r = g(h)$. The training minimizes a mean squared error loss function. Deep autoencoders use numerous internal intermediate representations, and these deep layers help learn more intricate and complex data patterns. Convolutional autoencoder (CAEs) [60] helps integrate the convolutional advantage of a CNN. The encoder is thus made up of convolutional layers and the decoder of deconvolutional layers. Thus, CAEs extract features and gives a feature map containing the image's significant points.

One limitation of an autoencoder is that it has a deterministic latent-space representation. Although the autoencoder learns the input data, it may lack relevant information, which may be due to random encoding in the latent space or empty space. To overcome this, Kingma et al. [61] suggested a variational autoencoder (VAE), as shown in Fig. 13, which uses a probability distribution for latent space code representation. An inference model $q_\phi(z \mid x)$ for VAE is described in [62]. Here, ϕ denotes the variational parameters, optimized for $q_\phi(z \mid x) \approx p_\theta(x \mid z)$. Here, $q_\phi(z \mid x)$ approximates the posterior $p_\theta(z \mid x)$ of the generative model and is optimized using the evidence lower bound (ELBO) [63].

In 2014, Goodfellow et al. [64] introduced generative adversarial networks (GANs) as well as an adversarial network framework. A generative model is matched against a competitor, which they call a discriminative model, and the latter learns to determine whether the query face image is from the model distribution or given data distribution [64]. Both thrive on competition to improve their methods till one cannot be distinguished from the other.

5.4. Some current research in DL for FR

Developing different deep FR methods and their deployment in real-world applications requires a systematic performance evaluation. Iandola et al. [65] provided an evaluation framework for different datasets and SOTA methods. They used the following criteria: data augmentation, network architecture, loss function, training strategy, and model compression. The varied sizes of the datasets, such as CASIA-WebFace, VGG-Face, MS-Celeb-1M, and MegaFace for training and LFW and YTF for testing the models, make comparisons difficult. Here, both the datasets and architectures vary. A critical part of the evaluation is the loss function, which imposes stricter requirements for FR, as it has to discriminate and separate the features from the embedding space. The training strategy also plays an important role in terms of the learning rate and batch size. With

the modern trend of using FR in mobile and embedded devices, they also evaluated SqueezeNet [66] and MobileNet [67], which use compressed models and give better performance. They concluded that the deep ResNet series has advantages over other architectures, and the batch and feature normalization optimizes performance.

Deployment of FR models, especially unconstrained faces on embedded or mobile devices, needs to meet the challenge of recognizing low-resolution faces at a low computational cost. This problem is addressed in the work of Ge et al. [68] by using the selective knowledge distillation approach and calling it the teacher-student model. They used a two-stream CNN, one with high resolution (HR), which collected the essential facial features used to tune the other LR network using regression and classification. Li et al. [69] also take on the challenging task of working with low-resolution unconstrained face images. They explore good-performing models using the SCface [70] and UCCSface [71] datasets. To visually learn the network, they pre-train it with DCGAN [72]. New trends for unconstrained, very low-resolution FR were explored in [73]. They present a classification of very low-resolution FR approaches, characterizing them as heterogeneous or homogeneous based on their belongingness to different or same domains, respectively. The heterogeneous approach can be classified into projection (coupled mapping) and synthesis (super-resolution (SR)) methods. In a homogeneous approach, they discussed lightweight CCNs. They listed the challenges for very low-resolution FR as the availability of datasets for real-world applications, the dearth of discriminative features, discrepancies in the domain, and the efficiency of existing solutions.

One of the challenges in FR is the development of a pipeline that can simultaneously perform FD, alignment, and recognition. Other parameters, such as pose and gender, may also be required in some instances. A CNN pipeline for the different processes is described by Ranjan et al. [74]. They use a deep pyramid single-shot face detector (DPSSD) and a new loss function called crystal loss. They evaluated their end-to-end system on the IARPA Janus Benchmarks IJB-A [75], IJB-B [76], IJB-C [77], and IARPA Janus Challenge Set 5 (CS5) datasets to obtain SOTA performance. They also mentioned that some of the challenges facing current FR systems are dataset bias and domain adaptation.

In mid-March 2020, the World Health Organization (WHO) declared the coronavirus disease 2019 (COVID-19) be a pandemic [78]. DL has been extensively used in the analysis of the COVID-19 pandemic, as elaborated in the work of Heidariet al. [79], for disease prediction, disease monitoring, drug testing, and vaccine development. WHO issued guidelines for wearing a mask to prevent the transmission of the disease. Abboah-Offei et al. [80] provided a detailed analysis of facemasks to control the transmission of respiratory viral infections, and the French government tested AI-based CCTV software to detect whether travelers wore masks or not [81]. The FR research community is engaged in developing systems to monitor the facemasks worn by people. Fig. 14 depicts a block diagram of face mask detection using ML or DL.

Mbunge et al. [82] and Nowrin et al. [83] provide a comprehensive review of ML- and DL-based facemask detection techniques. Most of the facemask detection algorithms are CNN-based. A few are hybrid as they use DL and ML approaches like support vector machine (SVM) and decision tree (DT). CNN-based models include MobileNetV2 [84], ResNet [85], and VGG-16 CNN [86]. MobileNet and ResNet perform better than VGG-16 CNN. MobileNetV2 exhibits better performance because it is a lightweight classifier. SRCNet [87] uses an SR network and a classification network to perform three-class classification with an accuracy of 98.7%. Facemasknet [88], a three-class classifier, has an accuracy of 98.6%. RetinaFacemask [89], which uses both ResNet and MobileNet, incorporates transfer learning to achieve SOTA results. Some challenges for face mask detection are elaborated in the work of Nowrin et al. [83]. These include the availability of benchmarked datasets, variation in mask designs, processing speed for real-time applications, and variations in image resolution and masked face reconstruction.

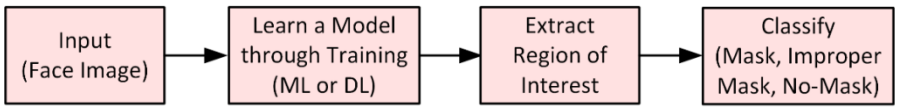


Fig. 14 Face mask detection block diagram

6. Conclusions

This study reviewed the vast literature on the development of different approaches for AFR. Over time, a transition from shallow to modern SOTA methodologies for DL has been observed. Early FR methods used limited images and a laboratory-controlled environment. However, with the advent of DL models, the LFW database achieved 99.85% accuracy. This was possible because of GPUs' massively parallel processing capabilities and large training and testing datasets. The challenges faced by DL models were also examined. As networks deepen, the complexity of the deep convolutional neural network (DCNN) model increases. A deep autoencoder or VAE that preserves some interclass discrimination information and intraclass similarity can feed a DCNN with a lower complexity to reduce the overall DCNN complexity. The performance decreases when the images have low resolution, variations in illumination, and blurry quality. Hence, DL methods must be made more robust under adverse conditions. The advent of new mobile communication technologies presents the challenge of integrating personalized FR applications that can be accessed by mobile users over different clouds and networks.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] G. Guo, et al., "A Survey on Deep Learning Based Face Recognition," Computer Vision and Image Understanding, vol. 189, Article no. 102805, December 2019.
[2] P. Viola, et al., "Rapid Object Detection Using a Boosted Cascade of Simple Features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-9, December 2001.
[3] M. Wang, et al., "Deep Face Recognition: A Survey," https://arxiv.org/pdf/1804.06655v2.pdf, April 2018.
[4] L. Sirovich, et al., "Low-Dimensional Procedure for the Characterization of Human Faces," Journal of the Optical Society of America A, vol. 4, pp. 519-524, March 1987.
[5] M. Ballantyne, et al., "Woody Bledsoe: His Life and Legacy," AI Magazine, vol. 17, no. 1, pp. 7-20, 1996.
[6] A. J. Goldstein, et al., "Man-Machine Interaction in Human-Face Identification," Bell System Technical Journal, vol. 51, no. 2, pp. 399-427, 1972.
[7] M. A. Turk, et al., "Face Recognition Using Eigenfaces," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586-587, January 1991.
[8] W. Zhao, et al., "Subspace Linear Discriminant Analysis for Face Recognition," http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.7.6280&rep=rep1&type=pdf, April 1999.
[9] M. S. Bartlett, et al., "Face Recognition by Independent Component Analysis," IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1450-1464, December 2002.
[10] F. Samaria, et al., "HMM-Based Architecture for Face Identification," Image and Vision Computing, vol. 12, no. 8, pp. 537-543, October 1994.
[11] H. Schneiderman, et al., "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 45-51, June 1998.
[12] M. H. Yang, et al., "Detecting Faces in Images: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, August 2002.
[13] X. He, et al., "Face Recognition Using Laplacianfaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328-340, January 2005.
[14] J. Wright, et al., "Robust Face Recognition via Sparse Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, April 2008.
[15] L. Zhang, et al., "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?" International Conference on Computer Vision, pp. 471-478, November 2011.
[16] L. Zhang, et al., "Collaborative Representation Based Classification for Face Recognition," https://arxiv.org/vc/arxiv/papers/1204/1204.2358v1.pdf, April 2012.
[17] L. Yang, et al., "Distance Metric Learning: A Comprehensive Survey," https://www.cs.cmu.edu/~liuy/frame_survey_v2.pdf, May 2006.

[18] R. Jin, et al., "Regularized Distance Metric Learning: Theory and Algorithm," Advances in Neural Information Processing Systems, vol. 22, pp. 862-870, December 2009.
[19] J. G. Daugman, "Two-Dimensional Spectral Analysis of Cortical Receptive Field Profiles," Vision Research, vol. 20, no. 10, pp. 847-856, January 1980.
[20] C. Liu, et al., "A Gabor Feature Classifier for Face Recognition," 8th IEEE International Conference on Computer Vision, pp. 270-275, July 2001.
[21] T. Barbu, "Gabor Filter-Based Face Recognition Technique," Proceedings of the Romanian Academy, vol. 11, no. 3, pp. 277-283, March 2010.
[22] M. Yang, et al., "Gabor Feature Based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary," European Conference on Computer Vision, pp. 448-461, September 2010.
[23] V. Štruc, et al., "Principal Gabor Filters for Face Recognition," 3rd International Conference on Biometrics: Theory, Applications, and Systems, pp. 1-6, September 2009.
[24] T. Ahonen, et al., "Face Recognition with Local Binary Patterns," European Conference on Computer Vision, pp. 469-481, May 2004.
[25] T. Ahonen, et al., "Face Description with Local Binary Patterns: Application to Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, October 2006.
[26] T. Ojala, et al., "A Comparative Study of Texture Measures with Classification Based on Featured Distributions," Pattern Recognition, vol. 29, no. 1, pp. 51-59, January 1996.
[27] D. Huang, et al., "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, no. 6, pp. 765-781, March 2011.
[28] R. R. Isnanto, et al., "Multi-Object Face Recognition Using Local Binary Pattern Histogram and Haar Cascade Classifier on Low-Resolution Images," International Journal of Engineering and Technology Innovation, vol. 11, no. 1, pp. 45-58, January 2021.
[29] D. S. Bolme, "Elastic Bunch Graph Matching," Master thesis, Department of Computer Science, Colorado State University, CO, 2003.
[30] B. M. Lahasan, et al., "Recognizing Faces Prone to Occlusions and Common Variations Using Optimal Face Subgraphs," Applied Mathematics and Computation, vol. 283, pp. 316-332, June 2016.
[31] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," 7th IEEE International Conference on Computer Vision, pp. 1150-1157, September 1999.
[32] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, November 2004.
[33] J. Luo, et al., "Person-Specific SIFT Features for Face Recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 593-596, April 2007.
[34] C. Geng, et al., "Face Recognition Using SIFT Features," 16th IEEE International Conference on Image Processing, pp. 3313-3316, November 2009.
[35] N. Dalal, et al., "Histograms of Oriented Gradients for Human Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886-893, June 2005.
[36] O. Déniz, et al., "Face Recognition Using Histograms of Oriented Gradients," Pattern Recognition Letters, vol. 32, no. 12, pp. 1598-1603, September 2011.
[37] Z. Cao, et al., "Face Recognition with Learning-Based Descriptor," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2707-2714, June 2010.
[38] Z. Lei, et al., "Learning Discriminant Face Descriptor," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 2, pp. 289-302, June 2013.
[39] J. Sánchez, et al., "Image Classification with the Fisher Vector: Theory and Practice," International Journal of Computer Vision, vol. 105, no. 3, pp. 222-245, December 2013.
[40] T. H. Chan, et al., "PCANet: A Simple Deep Learning Baseline for Image Classification?" IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5017-5032, September 2015.
[41] Y. Taigman, et al., "Deepface: Closing the Gap to Human-Level Performance in Face Verification," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701-1708, June 2014.
[42] Y. Sun, et al., "Deep Learning Face Representation by Joint Identification-Verification," Advances in Neural Information Processing Systems, pp. 1988-1996, December 2014.
[43] Y. Sun, et al., "Deepid3: Face Recognition with Very Deep Neural Networks," https://arxiv.org/pdf/1502.00873.pdf, February 2015.

[44] F. Schroff, et al., "Facenet: A Unified Embedding for Face Recognition and Clustering," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815-823, June 2015.
[45] J. Deng, et al., "Arcface: Additive Angular Margin Loss for Deep Face Recognition," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690-4699, June 2019.
[46] H. Liu, et al., "Adaptiveface: Adaptive Margin and Sampling for Face Recognition," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11947-11956, June 2019.
[47] G. B. Huang, et al., "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition, pp. 1-11, October 2008.
[48] M. Taskiran, et al., "Face Recognition: Past, Present and Future (A Review)," Digital Signal Processing, vol. 106, Article no. 102809, November 2020.
[49] Y. Bengio, Learning Deep Architectures for AI, Hanover: Now Publishers, 2009.
[50] Y. Bengio, et al., "Representation Learning: A Review and New Perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798-1828, March 2013.
[51] S. Hayman, "The McCulloch-Pitts Model," International Joint Conference on Neural Networks, pp. 4438-4439, July 1999.
[52] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," Psychological Review, vol. 65, no. 6, pp. 386-408, November 1958.
[53] M. Minsky, et al., Perceptrons, Cambridge: MIT Press, 1969.
[54] K. Fukushima, "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition," Proceedings of the U.S.-Japan Joint Seminar, pp. 267-285, February 1982.
[55] A. Krizhevsky, et al., "Imagenet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105, December 2012.
[56] Y. Zheng, et al., "Ring Loss: Convex Feature Normalization for Face Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5089-5097, June 2018.
[57] M. Yan, et al., "Vargfacenet: An Efficient Variable Group Convolutional Neural Network for Lightweight Face Recognition," Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 2647-2654, October 2019.
[58] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, vol. 61, pp. 85-117, January 2015.
[59] G. E. Hinton, et al., "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, pp. 504-507, July 2006.
[60] I. Goodfellow, et al., Deep Learning, Cambridge: MIT Press, 2016.
[61] D. P. Kingma, et al., "Auto-Encoding Variational Bayes," https://arxiv.org/pdf/1312.6114v4.pdf, December 2013.
[62] D. P. Kingma, et al., "An Introduction to Variational Autoencoders," https://arxiv.org/pdf/1906.02691v1.pdf, June 2019.
[63] C. Doersch, "Tutorial on Variational Autoencoders," https://arxiv.org/pdf/1606.05908v1.pdf, June 2016.
[64] I. Goodfellow, et al., "Generative Adversarial Nets," Proceedings of the 27th International Conference on Neural Information Processing Systems, pp. 2672-2680, December 2014.
[65] M. You, et al., "Systematic Evaluation of Deep Face Recognition Methods," Neurocomputing, vol. 388, pp. 144-156, May 2020.
[66] F. N. Iandola, et al., "SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5 MB Model Size," https://arxiv.org/pdf/1602.07360.pdf, November 2016.
[67] A. G. Howard, et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision," https://arxiv.org/pdf/1704.04861.pdf, April 2017.
[68] S. Ge, et al., "Low-Resolution Face Recognition in the Wild via Selective Knowledge Distillation," IEEE Transactions on Image Processing, vol. 28, no. 4, pp. 2051-2062, November 2018.
[69] P. Li, et al., "On Low-Resolution Face Recognition in the Wild: Comparisons and New Techniques," IEEE Transactions on Information Forensics and Security, vol. 14, no. 8, pp. 2000-2012, August 2019.
[70] M. Grgic, et al., "SCface-Surveillance Cameras Face Database," Multimedia Tools Applications, vol. 51, no. 3, pp. 863-879, February 2011.
[71] A. Sapkota, et al., "Large Scale Unconstrained Open Set Face Database," IEEE 6th International Conference on Biometrics: Theory, Applications, and Systems, pp. 1-8, September 2013.
[72] A. Radford, et al., "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," https://arxiv.org/pdf/1511.06434v1.pdf, November 2015.
[73] L. S. Luevano, et al., "A Study on the Performance of Unconstrained Very Low Resolution Face Recognition: Analyzing Current Trends and New Research Directions," IEEE Access, vol. 9, pp. 75470-75493, May 2021.

[74] R. Ranjan, et al., "A Fast and Accurate System for Face Detection, Identification, and Verification," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 1, no. 2, pp. 82-96, April 2019.
[75] B. F. Klare, et al., "Pushing the Frontiers of Unconstrained Face Detection and Recognition: IARPA Janus Benchmark A," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1931-1939, June 2015.
[76] C. Whitelam, et al., "IARPA Janus Benchmark-B Face Dataset," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 90-98, July 2017.
[77] B. Maze, et al., "IARPA Janus Benchmark-C: Face Dataset and Protocol," International Conference on Biometrics, pp. 158-165, February 2018.
[78] "WHO Director-General's Opening Remarks at the Media Briefing on COVID-19–11 March 2020," https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020, March 11, 2020.
[79] A. Heidari, et al., "The COVID-19 Epidemic Analysis and Diagnosis Using Deep Learning: A Systematic Literature Review and Future Directions," Computers in Biology and Medicine, vol. 141, Article no. 105141, February 2022.
[80] M. Abboah-Offei, et al., "A Rapid Review of the Use of Face Mask in Preventing the Spread of COVID-19," International Journal of Nursing Studies Advances, vol. 3, Article no. 100013, November 2021.
[81] H. Fouquet, "Paris Tests Face-Mask Recognition Software on Metro Riders," https://www.bloombergquint.com/politics/paris-tests-face-mask-recognition-software-on-metro-riders, May 07, 2020.
[82] E. Mbunge, et al., "Application of Deep Learning and Machine Learning Models to Detect COVID-19 Face Masks–A Review," Sustainable Operations and Computers, vol. 2, pp. 235-245, January 2021.
[83] A. Nowrin, et al., "Comprehensive Review on Facemask Detection Techniques in the Context of Covid-19," IEEE Access, vol. 9, pp. 106839-106864, July 2021.
[84] P. Khandelwal, et al., "Using Computer Vision to Enhance Safety of Workforce in Manufacturing in a Post COVID World," https://arxiv.org/ftp/arxiv/papers/2005/2005.05287.pdf, May 2020.
[85] M. Loey, et al., "Fighting against COVID-19: A Novel Deep Learning Model Based on YOLO-v2 with ResNet-50 for Medical Face Mask Detection," Sustainable Cities and Society, vol. 65, Article no. 102600, February 2021.
[86] S. V. Militante, et al., "Real-Time Facemask Recognition with Alarm System Using Deep Learning," 11th IEEE Control and System Graduate Research Colloquium, pp. 106-110, August 2020.
[87] B. Qin, et al., "Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19," Sensors, vol. 20, no. 18, Article no. 5236, September 2020.
[88] M. Inamdar, et al., "Real-Time Face Mask Identification Using Facemasknet Deep Learning Network," https://ssrn.com/abstract=3663305, July 2020.
[89] M. Jiang, et al., "Retinamask: A Face Mask Detector," https://arxiv.org/pdf/2005.03950v1.pdf, May 2020.