SVU-International Journal of Engineering Sciences and Applications (2025) 6(2): 29-46
Print ISSN 2785-9967| Online ISSN 2735-4571
DOI 10.21608/SVUSRC.2025.339279.1251

Image Processing: A Decade-by-Decade Review with a Focus on Face Recognition

Mahmoud Hardan Fahim [1,], Mohamed Abdel-Nasser [2], A.A. Donkol [1], Adel B. Abdel-Rahman [1,3]

Abstract The evolution of image processing has been remarkable, transitioning from its initial uses in military and medical fields to its widespread integration in modern society. This article offers an in-depth examination of the historical journey of image processing, charting its growth from the mid-20th century to the present era. Critical milestones encompass the emergence of digital computers, the fusion of Artificial Intelligence (AI) and Machine Learning (ML), and recent breakthroughs steered by deep learning and the global COVID-19 crisis. Special attention is devoted to the development of face recognition technology, showcasing an evolution from fundamental techniques like Eigenfaces to sophisticated deep learning methodologies. These strides have significantly enhanced precision and resilience in addressing complexities such as fluctuating lighting conditions, obstructions, and pose variations. By scrutinizing the past contexts and technological advancements in image processing and face recognition, this analysis emphasizes the transformative influence of these technologies across various domains, shaping the future landscape of human-computer interaction.

Keywords: Image processing; Computer vision; Artificial intelligence; Machine learning; Deep learning; Facial recognition; Object detection; Image analysis; Image understanding; Image classification; Masked Facial recognition.

Received: 28 November 2024/ Accepted: 28 March 2025
Corresponding Mahmoud Hardan Fahim, engmahfah2005@gmail.com
1. Electrical Engineering Department, Faculty of Engineering, South Valley University, Qena 83523, Egypt
2. Faculty of Engineering, Aswan University, Aswan, Egypt
3. School of Electronics, Communications and Computer Engineering, E-JUST, Egypt

1 Introduction

Image processing has undergone a remarkable transformation from its inception, closely entwined with technological advancements and the increased computational power. Initially applied in military and medical fields, image processing has now become a crucial tool across various industries, including entertainment, healthcare, and security. Its history signifies a continuous journey of innovation, expanding the capabilities of machines to interpret and engage with visual information in increasingly sophisticated ways. **Figure. 1** shows computer vision over the decades [1].

The historical narrative of image processing traces back to pivotal developments in the mid-20th century, where early applications focused on tasks such as enhancing satellite images for military reconnaissance and digitizing X-rays for medical diagnostics. The introduction of digital computers in the 1950s -1960s empowered researchers to utilize algorithms for image manipulation and analysis, with techniques like Fourier transforms and edge detection laying the groundwork for more intricate analyses. Despite these advancements, early image processing was constrained by computational limitations, making it a specialized domain with restricted accessibility [2-4].

In the present era, image processing is marked by the fusion of Artificial Intelligence (AI) and Machine Learning (ML). Modern algorithms, such as Convolutional Neural Networks (CNNs), have significantly boosted the accuracy and efficiency of image-related tasks. These advancements have enabled practical applications like facial recognition, autonomous driving, and medical imaging diagnostics. Furthermore, the widespread availability of powerful hardware, including GPUs and TPUs, has democratized the field, empowering researchers and developers to experiment with advanced algorithms and vast datasets [5-8].

Looking towards the future, the trajectory of image processing suggests a shift towards increased automation, real-time applications, and deeper integration with emerging technologies. Emerging domains like 3D imaging, Augmented Reality (AR), and quantum computing hold promise for transforming the acquisition, processing, and utilization of visual data. Ethical considerations, such as privacy concerns in facial recognition technology and potential biases in AI models, are likely to impact the field's progression. The development of explainable AI and privacy-preserving methodologies is expected to tackle these ethical challenges [9-12].

In summary, the evolution of image processing embodies the intricate interplay between technological advancement, computational capabilities, and societal needs. From its humble origins to its current high level of sophistication, the discipline continues to push boundaries, offering innovative solutions to real-world challenges. As image processing progresses, it will remain a fundamental pillar of technological advancement, reshaping the dynamics of human-machine interaction in the visual domain.

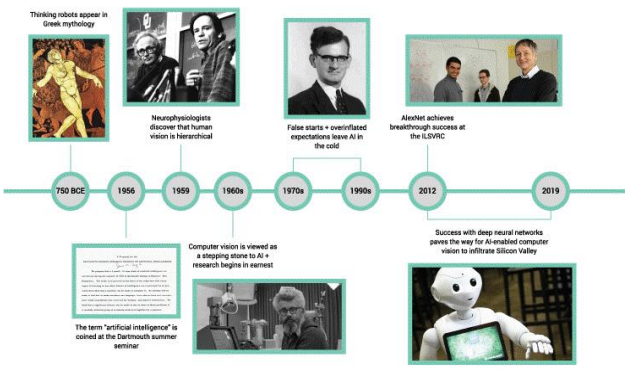

Fig. 1 Computer vision over the decades [1]

2 Related Work

The field of image processing has evolved significantly since its inception in 1964, undergoing transformative changes driven by technological advancements and societal needs. To provide a comprehensive review of this evolution, we divide the timeline into five distinct periods: The Beginning (1964–1995), Commercial Viability (1996–2006), Mainstream Development for Unconstrained Settings (2007–2013), Deep Learning into Face Recognition (2014–2019), and After COVID-19 (2020–2024). Each period reflects unique advancements, challenges, and contributions to the development of image processing.

In the first period, The Beginning, foundational techniques like edge detection and Fourier transforms were introduced, establishing the mathematical and algorithmic basis for digital image analysis. As computational power increased, the second period, Commercial Viability, saw the practical implementation of these techniques in industries like entertainment and medical imaging. The third period, Mainstream Development for Unconstrained Settings, marked a shift toward real-world applications, with algorithms adapted to handle dynamic and unpredictable environments. The fourth period, Deep Learning in Face Recognition, brought breakthroughs in performance and scalability through convolutional neural networks and other AI-driven techniques. Finally, the fifth period, after COVID-19, highlights how the pandemic accelerated advancements in image processing for applications like masked face recognition, remote diagnostics, and virtual interactions. This exploration delves into the key contributions and challenges in each period, outlining the trajectory of image processing over six decades.

2.1. The first period early research findings (1964 - 1995)

In 1964, Woodrow Bledsoe introduced computational facial recognition, supported by an undisclosed intelligence agency. Utilizing a computer program, he compared a suspect's photo with mugshots in a book, evaluating success based on the number of attempts required to correctly identify the match compared to the total faces in the dataset. The method involved encoding each face with a vector of distances between facial features, an innovative approach that was computationally intensive and slow, processing only around 40 images per hour with the technology available at that time (Bledsoe 1966) [13].

Tomkins and McCarter (1964) delved into primary affects, which are basic emotions rooted in biology and universally recognizable. Their study suggested that these effects are distinct, inherent reactions linked with specific facial expressions and physiological changes. They proposed a framework for classifying these primary affects, including emotions like joy, anger, and fear, positing that these play a central role in human motivation and behavior. Their research has had a broad impact by offering foundational insights into the expression and perception of emotions across different cultures, influencing subsequent studies on facial recognition and emotional analysis in psychology and computer vision [14].

Bruce and Young (1986) presented an elaborate model elucidating the intricacies of face recognition, dividing the process into multiple stages. They argued that facial

recognition entails various cognitive tasks, such as perceiving facial features, identifying familiar faces, and retrieving associated information like names and personal details. Their model distinguished between recognizing a face and recalling specific details about the individual, recognizing these as distinct yet interconnected processes. This framework has been highly influential, providing a basis for much of the later research on facial recognition in psychology and computer science, especially in understanding how individuals process familiar versus unfamiliar faces [15].

2.2. Commercial viability as new biometrics [1996-2006]

Bezdek and Pal [16] introduced innovative indices for cluster validity assessment, aiming to evaluate the quality of clustering outcomes. They put forth metrics such as the Partition Coefficient (PC) and the Classification Entropy (CE) to gauge the coherence of fuzzy clustering outputs by measuring the extent of cluster overlap. Moreover, they highlighted the significance of maintaining a balance between cluster compactness and separation, introducing the Alternative Dunn Index (ADI) as a dependable measure capable of accommodating noise and different cluster configurations. Their research established a structured approach to cluster assessment, thereby strengthening the reliability of clustering mechanisms across various domains.

Phillips et al. [2000] presented the FERET evaluation methodology, aiming to evaluate the reliability and efficacy of facial recognition algorithms. By utilizing the FERET database, which comprises 14,126 images of 1,199 individuals, this research established performance benchmarks for algorithms operating in real-world scenarios. The methodology encompasses standardized protocols for training, evaluating, and comparing algorithms, with the objective of pinpointing areas for enhancement and direct future investigations. Findings from the 1996 FERET test underscored the necessity for algorithms capable of addressing variations in lighting, facial expressions, and aging, thus paving the way for enhancing algorithm resilience [17]. **Figure 2** shows a schematic of the FERET testing procedure.

Yang (2001) delved into the realm of face recognition by delving into kernel methods, focusing on Kernel Principal Component Analysis (KPCA) and Kernel Fisher Discriminant Analysis (KFDA) to capture intricate higher-order correlations within facial images. By leveraging the Yale face database, this approach surpassed conventional linear techniques by projecting the input data into a multidimensional space to enable more intricate feature extraction. Findings indicated that the utilization of kernel methods, particularly Kernel Eigenface and Kernel Fisherface, yielded reduced error rates and enhanced performance compared to standard methods such as Eigenface, Fisherface, and ICA. This underscores the efficacy of kernel-based strategies in tackling intricate face recognition challenges [18]. **Figure 3** is a diagram of face recognition using kernel methods.

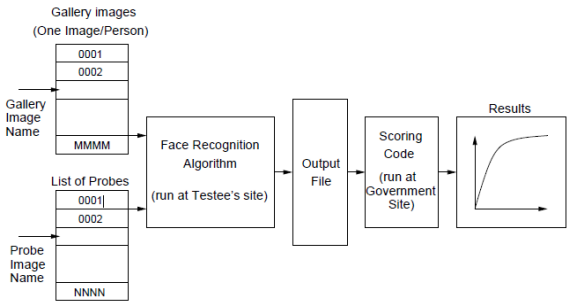

Fig. 2 Schematic of the FERET testing procedure [17]

Yu and Yang [2001] proposed a specialized Linear Discriminant Analysis (LDA) algorithm designed for high-dimensional data, particularly for face recognition tasks. This method, utilizing the ORL face database, aimed to eliminate the within-class scatter matrix's null space, which lacks discriminative information, and preserve only pertinent features. The algorithm successfully attained an exact solution based on Fisher's criterion, demonstrating effective performance even with large scatter matrices. As a result, it enhanced the classification accuracy in face recognition, underscoring the effectiveness of LDA optimization for high-dimensional data [18]. **Figure 3** explains the direct LDA algorithm steps.

In their 2003 study, Lu, Plataniotis, and Venetsanopoulos introduced an advanced LDA-based algorithm aimed at overcoming the constraints of traditional LDA methodologies in face recognition. Through experimentation on the FERET and AR face databases, the proposed method tackles the drawbacks of linear discriminant analysis by integrating a more computationally streamlined approach, resulting in reduced costs and enhanced performance compared to Eigenfaces, Fisherfaces, and direct LDA (D-LDA). The outcomes demonstrated the superior efficacy of this algorithm over existing approaches, offering a more pragmatic solution for real-world face recognition scenarios attributed to its heightened efficiency and robust performance [19].

M. Yang [20] delved into sophisticated kernel-based approaches for face recognition, expanding upon traditional methods such as Eigenfaces and Fisherfaces by venturing into the nonlinear domain via kernel functions. Through the utilization of kernel principal component analysis (KPCA) and kernel Fisher discriminant analysis (KFDA), the study showcases enhanced recognition performance in scenarios characterized by intricate variations in lighting, facial expression, and pose. The comparative assessment illustrates that kernel methods notably augment the discriminative capacity and generalization ability of face recognition systems when compared to their linear counterparts. The experimental

findings substantiate the efficacy of the proposed methodologies on standardized datasets, presenting a compelling argument for the adoption of kernel techniques in face recognition applications.

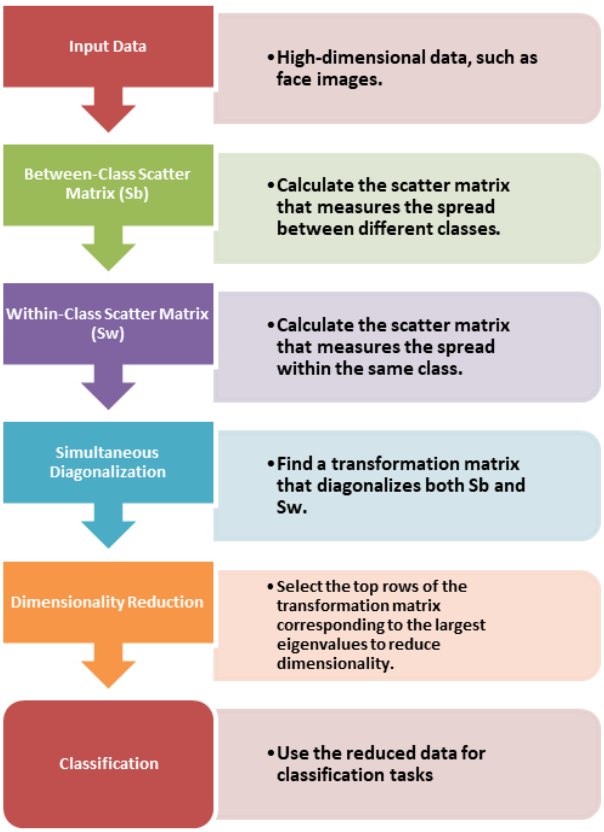

Fig. 3 The Direct LDA Algorithm

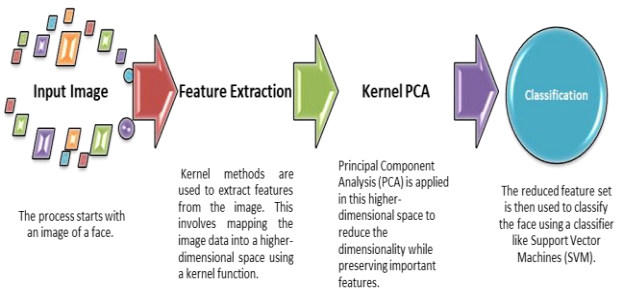

Fig. 4 A diagram of face recognition using kernel methods

Haar Cascade classifiers are an effective technique for object detection. This approach utilizes machine learning, where a large set of positive and negative images is utilized to train the classifier. Positive images consist of the objects that we want the classifier to detect, while negative images include everything else that does not contain the object of interest [21]. **Figure 5** shows how the Haar Cascade classifiers work.

Ahonen, Hadid, and Pietikäinen [2004] introduced a technique for facial recognition utilizing Local Binary Patterns (LBP) to extract texture characteristics from facial images in **Fig. 1**. The process involves dividing a facial image into smaller, non-overlapping sections and applying LBP to each segment to capture local texture details, which are then aggregated into a comprehensive descriptor for the entire face. Through testing on the FERET and CMU PIE face databases, this method exhibited notable resilience to variations in lighting and facial expressions, achieving a high level of recognition accuracy. Over time, the LBP-based approach has evolved into a standard method in texture-based facial recognition, renowned for its simplicity and efficacy across various recognition scenarios as shown in **Fig. 6** [22].

Fig. 5 How the Haar Cascade classifiers work [21]

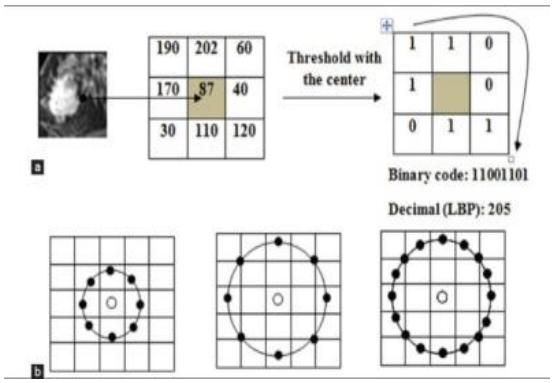

Fig. 6 Local Binary Pattern (LBP) method [22]

2.3. Mainstream development for unconstrained settings [2007-2013]

Krizhevsky et al. presented AlexNet, a sophisticated deep convolutional neural network (CNN) tailored for wide-ranging image classification tasks. This model integrates a series of convolutional and max-pooling layers, followed by fully connected layers, culminating in a softmax layer for classification purposes. Trained on the ImageNet dataset, which comprises a vast collection of over 1.2 million high-resolution images spanning 1,000 categories, AlexNet utilized GPUs to expedite the training process. Impressively, in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), AlexNet achieved a remarkable top-5 error rate of 15.3%, surpassing previous methodologies and ushering in a new era of deep learning advancements in image recognition [23]. AlexNet Architecture is shown in **Fig. 7**.

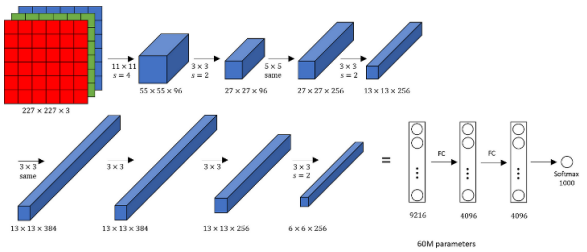

Fig. 7 AlexNet Architecture [23]

Thomas et al. [24] delve into the impact of face gaze on cognitive performance among individuals with typical development, Autism Spectrum Disorder (ASD), and Williams Syndrome (WS). The study involved participants undertaking math tasks with and without eye contact from the experimenter. The outcomes revealed that sustaining eye contact hindered task performance universally, with individuals with ASD and WS encountering more significant challenges. The research highlights the cognitive strain induced by face gaze, disrupting simultaneous tasks, especially in those with developmental disorders. These results play a pivotal role in comprehending the impact of gaze on cognitive load, guiding the development of interventions and educational approaches.

In their influential work, Felzenszwalb et al. introduced the Deformable Part Model (DPM) for object detection. This approach revolutionized the field by modeling objects as collections of parts arranged in a flexible structure. They employed a discriminative training process using latent support vector machines (SVMs). The effectiveness of the model was evaluated on the PASCAL VOC dataset, which comprises annotated images with multiple object categories. By accurately capturing both the appearance and spatial relationships among parts, DPM set new performance benchmarks for object detection tasks [25]. Example detection obtained with the person model. The model is defined by a coarse template, several higher resolution part templates, and a spatial model for the location of each part in **Fig. 8**.

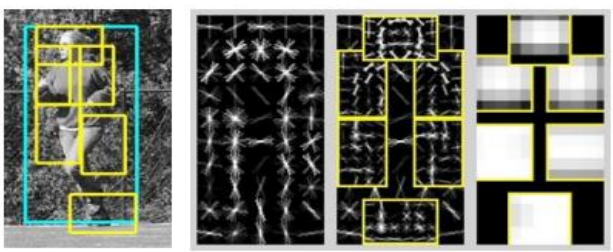

Fig. 8 Person model output with spatial part templates [25]

Lowe proposed the Scale-Invariant Feature Transform (SIFT), an approach for detecting and describing local features in images [26]. SIFT is designed to identify key points that are invariant to scale and rotation, enabling reliable matching between different images. The methodology involves the detection of key points and the generation of descriptors based on gradients within localized image regions. SIFT has been extensively tested on diverse datasets for tasks such as object recognition and matching, demonstrating a high degree of robustness to changes in scale, rotation, and illumination. As a result, it has become a standard technique in the field of feature detection and description.

In **Fig. 9**, Bay and colleagues introduced the Speeded Up Robust Features (SURF) algorithm as a faster alternative to SIFT for feature detection and description [26]. SURF utilizes integral images and approximations of Hessian matrices to efficiently detect key-points and employs a simplified descriptor for matching. The method has been evaluated on various image matching and object recognition tasks, exhibiting similar robustness to SIFT but with a notable speed advantage, particularly in real-time applications.

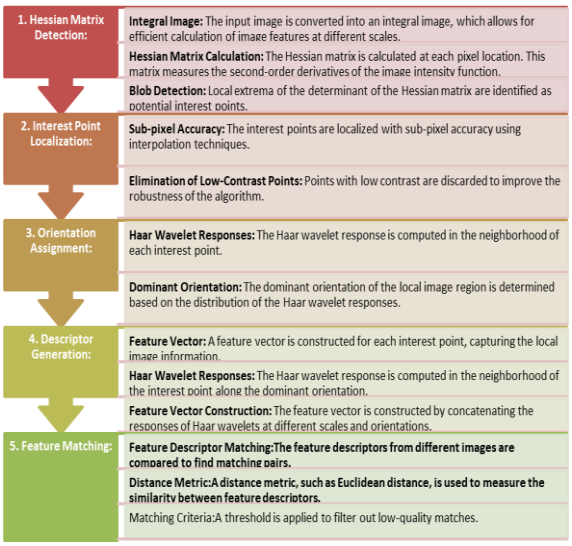

Fig. 9 SURF algorithm steps [26]

Kinect Fusion introduced a technique for dynamic 3D reconstruction in real-time by leveraging depth information from the Microsoft Kinect sensor. This approach involves the ongoing incorporation of depth frames into a volumetric model, facilitating detailed mapping and camera monitoring instantaneously. KinectFusion underwent testing in indoor environments, generating intricate and superior quality 3D representations rapidly. This advancement paved the way for the integration of affordable depth sensors in various consumer applications such as augmented reality and robotics [27]. **Figure 10** shows an example output from the system, generated in real-time with a handheld Kinect depth camera and no other sensing infrastructure. Normal maps (colour) and Phong-shaded renderings (greyscale) from our dense reconstruction system are shown.

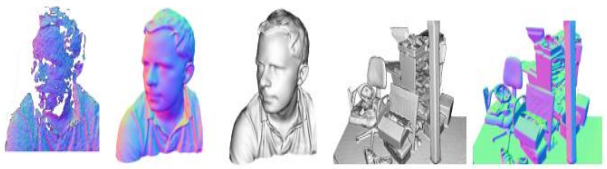

Fig. 10 Comparative output of dense model and sensor input [27]

In their seminal work on GrabCut, Rother et al. introduced a highly effective interactive image segmentation technique that leverages graph cuts to distinguish foreground from background regions with minimal user input. As illustrated in **Fig. 11**, the method begins with a simple user-defined bounding box around the target object, which serves as an initial estimate of the foreground. GrabCut employs a Gaussian Mixture Model (GMM) to model the color distributions of both foreground and background pixels. These distributions are iteratively refined, and segmentation is performed by constructing a graph where each pixel is a node, and edge weights encode both color similarity and spatial continuity. The optimal segmentation is then obtained via min-cut/max-flow optimization, separating the graph into distinct foreground and background components. The iterative nature of the algorithm allows users to further refine the result by marking ambiguous regions, improving accuracy without requiring full manual segmentation. Evaluated across various challenging images, GrabCut demonstrated a strong ability to produce precise object boundaries while maintaining a low level of user interaction. Because of its balance between automation and control, GrabCut has become a cornerstone in interactive image editing and computer vision applications. Its influence spans tasks such as background removal, object cutout, and image matting, making it a foundational method in the field [28].

DeepFace represents a landmark advancement in face verification through the application of deep learning. Leveraging a nine-layer neural network, the model is trained on an extensive dataset and incorporates 3D face alignment to normalize facial inputs before processing. This critical step ensures pose-invariant representations, significantly improving the model's robustness to variations in orientation and expression. Achieving an impressive 97.35% accuracy on the Labeled Faces in the Wild (LFW) benchmark, DeepFace brought face verification performance to near-human levels, marking a pivotal moment in the field [29]. As illustrated in **Fig. 12**, the architecture comprises an initial sequence of convolution-pooling-convolution layers, followed by three locally connected layers and two fully connected layers. Notably, the locally connected and fully-connected layers dominate the parameter count, collectively contributing to over 95% of the model's more than 120 million parameters [29]. This architectural design enables DeepFace to capture both spatially localized and globally integrated facial features, offering a powerful foundation for modern face recognition systems. Its success laid the groundwork for subsequent innovations in deep face analysis, including masked face recognition, by demonstrating the effectiveness of deep neural networks combined with precise facial alignment techniques.

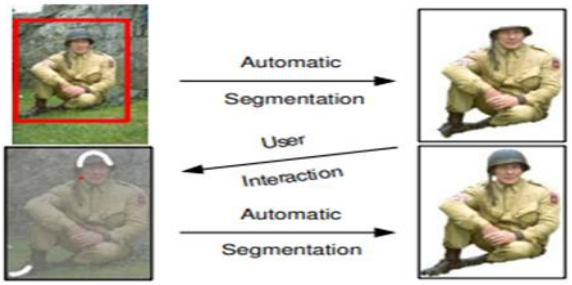

Fig. 11 Segmentation improvement through foreground/background annotations [28]

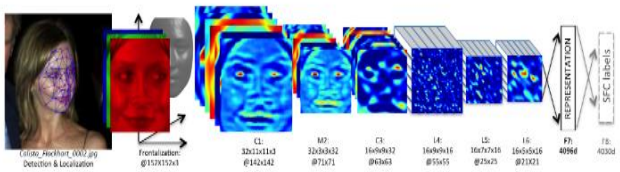

Fig. 12 Overview of the Deep Face network structure [29]

Histograms of Oriented Gradients (HOG), introduced by Dalal and Triggs, detect humans by capturing gradient orientations within localized cells, robustly identifying edge-defined objects like human silhouettes. Validated on the INRIA Person Dataset, it established foundational pedestrian detection techniques through exceptional efficiency [30].

Figure 13 illustrates HOG's core mechanism: detectors target silhouette contours (head/shoulders/feet) with peak activation in blocks adjacent to edges. Key components:(a) Average gradient image (training data) **.**(b) Per-pixel max positive SVM weight. (c) Per-pixel max negative SVM weights **.** (d) Test image. (e) Computed R-HOG descriptor. (f,g) R-HOG modulated by SVM weights [30].

Shotton et al. introduced a real-time human pose estimation framework leveraging depth imagery from the Kinect sensor. Central to their approach is a random forest classifier that predicts body part labels at the pixel level. This eliminated the need for pre-calibration and delivered reliable pose estimation across a wide range of complex poses within a large-scale internal dataset. This work significantly advanced consumer pose estimation and directly powered the development of the Microsoft Kinect [31]. **Figure 14** illustrates representative synthetic and real

training data pairs (depth image + ground truth body parts), highlighting diversity in pose, body shape, clothing, and image crop [31].

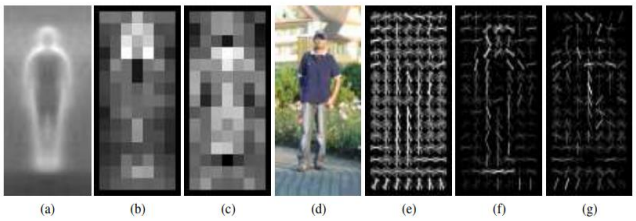

Fig. 13 HOG descriptor responses emphasizing silhouette regions [30]

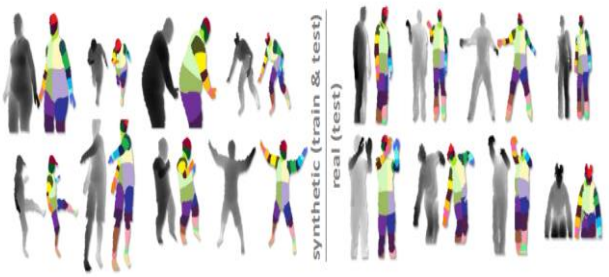

Fig. 14 Depth and ground truth pairs: real vs. synthetic bodies [31]

The PASCAL VOC Challenge presented a foundational benchmark for object detection, segmentation, and classification, fostering annual competitions that attracted a substantial research community. The VOC datasets and evaluation standards set a normative framework for evaluating object detection and segmentation algorithms, driving progress in the field and endorsing replicable research [32].

ORB (Oriented FAST and Rotated BRIEF) emerged as a swift and effective alternative to SIFT and SURF for feature detection and description, specifically engineered for computational efficiency. By merging the FAST keypoint detector with the BRIEF descriptor, ORB proved to be suitable for real-time applications. When assessed on benchmark datasets, ORB exhibited comparable accuracy to SIFT and SURF but at a significantly reduced computational cost [33]. Shape Context serves as a descriptor utilized to portray shapes based on the spatial arrangement of points around each keypoint. Although introduced before the 2007–2013 timeframe, Shape Context retained its significance as a fundamental method for shape matching and recognition activities during this period. Validation on datasets containing handwritten digits and shape repositories illustrated its resilience to deformation and alterations, leading to its widespread adoption in shape-oriented object recognition [34].

Between 2007 and 2013, the Viola-Jones framework had a substantial influence on the field of face detection, despite being published earlier. It especially revolutionized real-time applications, thanks to its utilization of Haar-like features and a cascade classifier. By employing these techniques, the framework can swiftly detect objects, obtaining impressive precision across various datasets, including those specific to face recognition. This efficiency and accuracy greatly contributed to its extensive implementation in consumer cameras and surveillance systems [35]. The introduction of poselets presents a technique for human detection and poses estimation using mid-level representation. It relies on body parts that are annotated with 3D poses. By combining these parts, poselets enhance the capabilities of detection and can capture complex pose variations. In evaluations conducted on the PASCAL VOC dataset, poselets have achieved new performance benchmarks for person detection and have paved the way for part-based models [36].

He et al. introduced guided image filtering as a rapid and edge-preserving filter applicable to tasks such as image denoising, detail enhancement, and high-dynamic-range (HDR) imaging. Guided filtering leverages an input image as a guide to generate smoother output images while preserving edges. Through evaluations across multiple image processing tasks, this method demonstrated superior speed and simplicity compared to other edge-preserving techniques, thus establishing itself as a beneficial tool in computer vision and image processing [37]. Additionally, Dollar et al. conducted a thorough assessment of pedestrian detection methods by evaluating various detectors on datasets including INRIA, Caltech Pedestrian, and Daimler. Their comparison emphasized accuracy, speed, and robustness, offering valuable insights that influenced future research directions in pedestrian detection [38].

Tan and Triggs (2007) proposed an advanced technique for enhancing face recognition performance in challenging lighting conditions. Their method enhances local texture feature sets by extending Local Binary Patterns (LBP) with techniques like Local Ternary Patterns (LTP) and utilizing pre-processing techniques to standardize lighting conditions. The approach involves robust feature extraction, enabling the capture of intricate texture details even in diverse lighting environments. Through experimentation on datasets such as Yale B and CMU PIE, the improved LTP-based method demonstrated superior performance compared to traditional LBP methods, showcasing enhanced accuracy and stability, specifically in complex lighting scenarios. This research has had a significant impact on lighting-invariant face recognition methods, particularly in practical environments with varying illumination settings [39].

Tan and Triggs (2010) enhanced their previous research on face recognition in challenging lighting conditions by optimizing local texture features and incorporating robust lighting normalization techniques. The study presents Local Ternary Patterns (LTP)

as a refinement of Local Binary Patterns (LBP), enhancing texture detail encoding while reducing sensitivity to noise during feature extraction. Moreover, they integrated adaptive pre-processing approaches, including gamma correction, difference of Gaussian filtering, and contrast equalization, to mitigate the impact of lighting variations. Through evaluation on datasets such as Yale B and Extended Yale B, the LTP-based method exhibited significant enhancements in recognition accuracy compared to traditional LBP methods, proving its efficacy for face recognition in settings with unpredictable lighting conditions [40].

Y. Taigman and L. Wolf (2011) delved into the enhancement of face recognition performance in uncontrolled settings. Their approach harnesses deep learning, particularly deep convolutional networks, trained on an extensive dataset comprising billions of facial images. This dataset is extensive, encompassing diverse poses, lighting conditions, and occlusions, with the goal of encapsulating the real-world challenges faced by face recognition systems. The authors posit that such a vast dataset enables the deep learning model to acquire more robust and discriminative features, surpassing the limitations of traditional methods. The outcomes exhibit noteworthy enhancements in face recognition accuracy compared to prior techniques, underscoring the efficacy of large-scale data and deep learning in surmounting obstacles in uncontrolled face recognition [41].

Déniz, Bueno, Salido, and De la Torre (2011) adapted the Histograms of Oriented Gradients (HOG) descriptor for face recognition, utilizing its resilience in capturing facial structure through gradient orientation. Their methodology involved extracting HOG features from facial images to construct descriptors representing edge and gradient distributions across localized regions. Validation on the FERET and Yale face datasets demonstrated robustness against illumination changes and expression variations, yielding significant recognition accuracy. This research underscores HOG's efficacy in face recognition—beyond its conventional object detection use—by encoding intricate yet consistent facial features [42]. **Figure 15** is an example input image for HOG feature extraction.

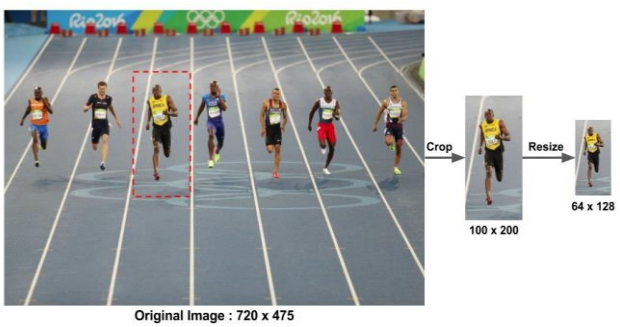

Fig. 15 Input image to HOG [42]

2.4. Deep Learning into Face Recognition [2013-2019]

Simonyan and Zisserman (2013) conducted a study on face recognition using Fisher Vector encoding combined with SIFT descriptors. Their approach focused on achieving robustness in face recognition. The evaluation was done on the LFW dataset, and their method demonstrated the effectiveness of traditional approaches as strong baselines, even in the presence of emerging deep learning models [43]. Mollahosseini et al. (2013) explored the fine-tuning of a deep face network for expression recognition. By training their model on the CK+ and JAFFE datasets, they showcased the transferability of face recognition features to other tasks such as expression recognition [44]. Schroff et al. proposed the use of triplet loss to learn an Euclidean embedding space for faces. This allowed for clustering and face recognition with minimal supervision. Their FaceNet model, trained on a vast dataset of 200 million images, achieved a remarkable accuracy of 99.63% on the LFW dataset. This work had a significant impact on subsequent research in unsupervised face clustering [45]. Sun et al. introduced a CNN that was trained to classify 10,000 identities, incorporating supervised features that enhanced discrimination. The evaluation of their model on the CelebFaces+ dataset demonstrated superior performance compared to previous methods on the LFW dataset. This highlighted the potential of using large-scale identity supervision for improving face recognition systems [46]. Chen et al. integrated deep features with a Bayesian methodology to model the joint distribution of face pairs, capturing both intra- and inter-personal variabilities. Their study, evaluated on LFW and MS-Celeb-1M datasets, demonstrated notable improvements in face verification accuracy through the fusion of statistical models and deep features [47]. Additionally, Sun et al. (2015) improved DeepID by concurrently optimizing for identification and verification tasks, resulting in more robust identity representations. By applying this approach to CelebFaces+, DeepID2 achieved cutting-edge performance on LFW, showcasing the advantages of multi-task training in enhancing face recognition capabilities [48].

Parkhi et al. contributed to the advancement of face recognition by employing a 16-layer VGG-style convolutional neural network (CNN), known for its architectural simplicity and robust performance. Trained on a large-scale dataset of 2.6 million face images, this effort led to the development of VGGFace, a widely adopted pre-trained model in face recognition tasks. The model demonstrated strong generalization capabilities, achieving high accuracy on the Labeled Faces in the Wild (LFW) benchmark and establishing a reliable foundation for future research and applications in facial analysis [49]. Building on this foundation, Wen et al. introduced the center loss function to address limitations in traditional softmax-based training. Center loss explicitly encourages intra-class compactness and inter-class separability by

minimizing the distance between features and their corresponding class centers. Evaluated on the CASIA-WebFace and LFW datasets, this approach produced more discriminative face embeddings, setting the stage for subsequent developments in margin-based loss functions [50]. Liu et al. extended this line of research with the A-Softmax loss, which projects learned features onto a hypersphere and enforces class separation through angular margins. Their model, SphereFace, trained on CASIA-WebFace and the massive MS-Celeb-1M dataset, achieved a remarkable 99.42% accuracy on LFW. As illustrated in **Fig. 16**, this angular margin optimization offers a more geometrically meaningful separation of classes than traditional softmax, reinforcing the efficacy of angular constraints for high-precision face recognition [51]. Further enhancing the training dynamics, Ranjan et al. proposed L2-Softmax, which normalizes feature vectors to a fixed L2 norm. This normalization ensures uniform feature magnitudes, improving stability during training and boosting performance during verification. Their experiments on the LFW benchmark confirmed that L2-Softmax improves robustness by making the model less sensitive to feature amplitude variations, leading to consistent and reliable face embeddings [52].

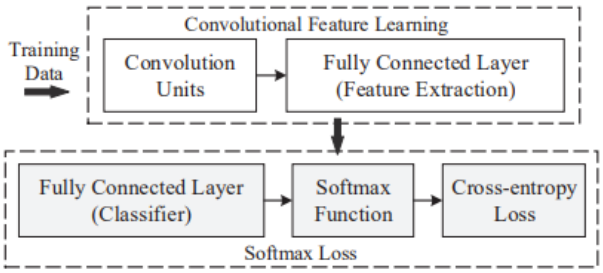

Fig. 16 Standard CNNs can be viewed as convolutional feature learning machines that are supervised by the softmax loss [51]

Together, these contributions—spanning architectural refinements, novel loss functions, and normalization techniques—have significantly shaped the trajectory of modern face recognition research. Liu et al. [53] proposed an extension to softmax by incorporating a large margin, which encourages greater separation between classes for more robust face recognition. Their method was validated on LFW and CASIA-WebFace datasets, contributing to the increased popularity of margin-based techniques in the field. Deng et al. [54] presented ArcFace, an additive angular margin loss function designed to learn highly discriminative face embeddings. When trained on the MS-Celeb-1M dataset and tested on LFW and MegaFace benchmarks, ArcFace attained state-of-the-art performance, establishing itself as a highly influential method. **Figure 17** demonstrates the framework's global comparison mechanisms: sample-to-class and sample-to-subclass with angular margins [54]. Wang et al. [55] applied a cosine margin to softmax, enhancing the discriminative capacity of learned features. With training on MS-Celeb-1M and evaluation on LFW, CosFace achieved an accuracy of 99.73%, thereby demonstrating the effectiveness of cosine-based margin strategies. Zhang et al. [56] proposed range loss to mitigate data imbalance by minimizing intra-class variance in long-tailed distributions. Evaluated on LFW and MegaFace, this approach enhanced recognition accuracy for underrepresented classes, enabling robust performance on imbalanced datasets. **Figure 18** illustrates the constructed long-tail dataset, where cutting lines demarcate proportions of tail-class data [56].

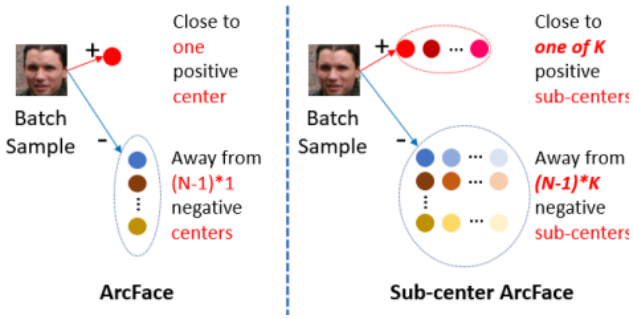

Fig. 17 Sample-level classification using Arcface and sub-center Arcface [54]

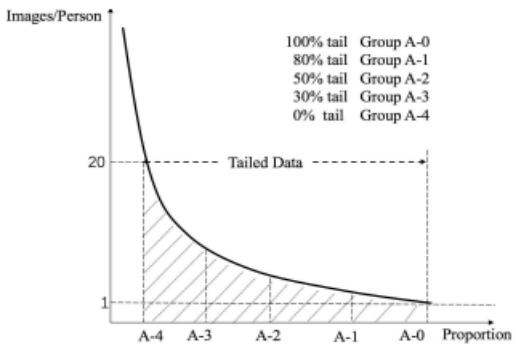

Fig. 18 Long-tailed dataset construction with division boundaries [56]

Deng et al. [57] provided an open-source implementation of ArcFace that was optimized for both 2D and 3D face recognition. With training on various datasets, InsightFace became widely adopted due to its high accuracy and accessibility for academic and industrial applications.

Wang et al. proposed the integration of an additive margin for softmax, which significantly enhanced class separability. AM-Softmax was assessed on LFW and CASIA-WebFace datasets, delivering competitive outcomes and elevating the prominence of margin-based approaches in face verification [58]. Duan et al. introduced equidistribution of embeddings to mitigate intra-class variability, resulting in more resilient representations. UniformFace, tested on LFW and

Megaface datasets, demonstrated enhanced recognition capabilities by fostering a uniform distribution within classes [59]. Huang et al. implemented curriculum learning, incorporating adaptive margins to facilitate gradual learning from easy to challenging samples. CurricularFace, evaluated on LFW and Megaface datasets, exhibited heightened accuracy, especially on intricate samples, highlighting the advantages of curriculum learning [60].

Karkkainen & Joo introduced FairFace—a balanced dataset (>108k images) mitigating racial, gender, and age biases in face recognition. This enabled equitable model training/evaluation, addressing critical ethical gaps [61]. **Figure 19** visualizes BFW subgroups by gender (rows: Female/Male) and ethnicity (columns: Asian, Black, Indian, White This dataset provides statistics on facial images categorized by ethnicity and gender, including Asian (female and male), Black (female and male), Indian (female and male), and White (female and male), as well as aggregated totals. Each subgroup contains 2,500 face images, resulting in a total of 20,000 face images across all categories. Every subgroup consists of 100 subjects, with each subject contributing 25 images. For positive pairs (i.e., images of the same individual), there are 30,000 pairs per subgroup, totaling 240,000 positive pairs. For negative pairs (images of different individuals), the numbers range from 85,016 to 85,232 per subgroup, with a total of 681,379 negative pairs. The total number of image pairs (positive and negative combined) ranges from 115,016 to 115,232 per subgroup, summing up to 921,379 total pairs used in the evaluation. [61].

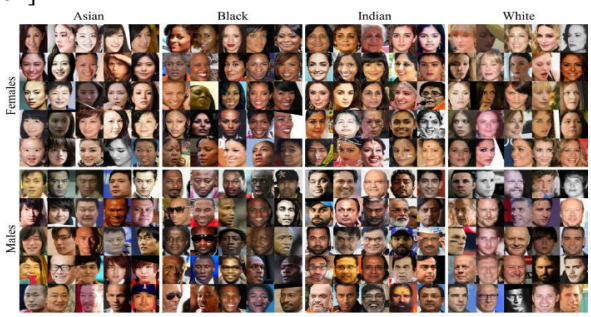

Fig. 19 BFW dataset partitioned by gender (rows) and ethnicity (columns) [61]

The Labeled Faces in the Wild (LFW) dataset [62] was first introduced in 2007 as a benchmark dataset to assess face verification and recognition algorithms. It consists of 13,233 images portraying 5,749 individuals under diverse and uncontrolled conditions, sourced from various online platforms. The dataset exhibits significant variations in pose, lighting, and facial expressions, serving the primary purpose of face verification by determining if two images depict the same person. Researchers commonly utilize standard or restricted protocols for algorithm evaluation to enable consistent comparisons.

In contrast, the YouTube Faces Dataset (YTF) [63], established in 2011, focuses on dynamic face recognition in real-world scenarios. It includes 3,425 video sequences of 1,595 individuals from YouTube, emphasizing variations introduced by motion, illumination, and pose. Each video is tagged with the subject's identity, and pairs of videos are provided to assess verification performance. YTF's emphasis on temporal data presents a distinctive challenge compared to static image datasets, requiring models to handle complexities like motion blur and varied frame quality.

The CASIA-WebFace dataset [64], unveiled in 2014, is a vast collection containing 494,414 images of 10,575 subjects, aimed at supporting the training of deep learning models with extensive and diverse data. Images sourced from the internet offer variations in lighting, pose, and expression, rendering it valuable for training deep convolutional neural networks. Despite criticisms regarding label quality, CASIA-WebFace remains prominent for pretraining face recognition models, typically followed by fine-tuning on task-specific datasets.

Additionally, the VGGFace dataset [65], initiated in 2015, comprises 2.6 million images of 2,622 individuals and VGGFace2, an enhanced version released in 2017 with 3.31 million images illustrating 9,131 subjects. VGGFace2 improved on its predecessor by incorporating more diverse demographics, poses, ages, and expressions. These datasets are extensively used for training deep learning models, especially convolutional neural networks, due to their balanced representation of gender and ethnicity, supporting generalizable face recognition systems through supervised and transfer learning approaches.

2.5. After Covid-19 [2020-NOW]

The COVID-19 pandemic has significantly impacted various technologies, particularly face recognition systems. The widespread adoption of mask-wearing mandates presented a unique challenge for face recognition software, as masks obscured critical facial features such as the nose and mouth that are typically crucial for precise identification. Consequently, this situation sparked a notable increase in attention and investment in mask-compliant face recognition technology. Developers were compelled to swiftly adapt to maintain accuracy and effectiveness, concentrating on algorithms capable of identifying partial faces or placing greater emphasis on the eye area for recognition. This technological shift signified a significant advancement in the field, considering that traditional face recognition systems were not originally equipped to handle partial facial information [66].

During the pandemic, the use of face recognition technology to monitor mask-wearing and social distancing guidelines raised concerns about privacy and ethics. Governments and institutions implemented surveillance measures, leading to debates about the balance between

public health and civil liberties. The pandemic also accelerated the development of face recognition technology for mask compliance. These advancements have the potential for broader applications in low-visibility environments. However, stronger regulatory frameworks are needed to ensure responsible and transparent use of this technology. The study by Amjad Bashayreh et al. utilized FaceNet with triplet loss to optimize recognition accuracy for both masked and unmasked faces, achieving better results for masked faces compared to the baseline. [67- 69].

In the study conducted by Nagrath et al., their objective was to develop a lightweight and real-time face mask detection system. They utilized datasets from Kaggle and PyImageSearch and achieved an accuracy of over 90% on the detection tasks. This high accuracy demonstrates the practicality of the system for embedded applications such as surveillance systems [70]. Badr Lahasan et al. introduced a two-branch CNN (Convolutional Neural Network) model that addressed the challenges posed by occlusions like masks. One branch of the model focused on visible parts of the face, while the other branch adapted to handling occlusions. By training the model on a simulated masked dataset, they demonstrated improved resilience to occlusions compared to single-branch approaches [71]. Aswal et al. presented a two-step method called RetinaFace and VGGFace2 Integration. This method combines RetinaFace for face detection and VGGFace2 for face verification, with YOLOv3 used to enhance speed and accuracy. The method was tested on video datasets with varying environmental conditions and achieved an impressive accuracy of over 92%, indicating its robust performance in real-time applications [73]. Furthermore, the researchers optimized the embedding process by experimenting with triplet loss using masked and unmasked combinations. The results showed that models trained with combined triplets of masked and unmasked pairs achieved well-balanced performance on mixed datasets [74]. In addition, the researchers utilized a modified VGG16 model with specialized layers to excel at recognizing partially occluded faces. The model achieved an accuracy of 91% on masked faces when evaluated on the RMFRD dataset. This highlights the advantages of leveraging layer-specific optimizations in CNN architecture for masked face recognition [75].

Table 1 Comparison of various face datasets, highlighting their features, advantages, and limitations before 2019.

Dataset	Size	Number of Individuals	Image Type	Features	Advantages	Limitations
FairFace[61]	108,501 images	7,000+	Diverse ethnicity faces	Balanced data across seven racial groups. Annotated for gender and age.	Diversity in ethnicity and balanced representation. Suitable for fairness and bias analysis.	Limited in real-world conditions (e.g., varia
LFW [62]	13,233 images	5,749	Static	Real-world variations in pose, lighting, and expression	Well-established benchmark for face verification	Limited diversity, low resolution, uncontrolled conditions
YTF [63]	3,425 videos	1,595	Video	Variations in illumination, pose, expression, and background	Focus on video-based recognition, temporal modeling	Low resolution, motion blur, video quality variability
MS-Celeb-1M [64]	10M images	100,000	Static	Large-scale, diverse faces, varied poses and expressions	Massive size, suitable for large-scale recognition	Privacy concerns, noisy data, mislabeling issues
VGGFace/VGGFace2 [65]	2.6M/3.31M images	2,622/9,131	Static	Diverse demographics, poses, and expressions	Balanced distribution, high-quality images	Biases in internet-sourced images, underrepresented groups

By leveraging insights extracted from the specific YOLOv3 model trained on masked face datasets, this system successfully attained notable accuracy improvements while concurrently reducing the duration required for training. Tailored for swift face recognition within surveillance contexts, the YOLO-based framework exhibited efficacy even on limited datasets without compromising accuracy significantly [76]. Moreover, MAN was introduced to enhance face recognition performance by adapting to mask-induced occlusions, leveraging the WebFace42M dataset from ICCV-MFR for model training. MAN integrates mask-aware methodologies to address occlusions, employing stochastic gradient descent (SGD) for optimization purposes. Empirical findings demonstrate MAN's proficiency in accurately identifying masked faces in uncontrolled settings, surpassing the baseline ICCV-MFR standards [77]. Additionally, the enhancement of the ArcFace model involved the incorporation of a parallel output layer specifically dedicated to mask detection, utilizing MaskTheFace for synthesizing mask variations. The Multi-Task ArcFace model integrates a tailored loss function amalgamating ArcFace principles and mask prediction losses, yielding robust outcomes when tested on diverse masked face datasets encompassing various mask types like N95 respirators and cloth masks [78]. An integrated strategy combining ArcFace with ensemble techniques was proposed to advance masked face recognition capabilities. Leveraging publicly available masked face datasets, this ensemble configuration significantly amplified accuracy levels in identifying both masked and unmasked faces, rendering it apt for security applications in real-world contexts [79].

A CNN-based framework achieved real-time mask detection in public spaces, validated on RMFD and MAFA datasets. The model delivered high accuracy with near-instant processing, confirming viability for embedded surveillance systems. **Figure 20** illustrates its pose-invariant operation: face angle parsing via skeletal key point mapping [80].

The Convolutional Visual Self-Attention Network (CVSAN) integrates convolutional and self-attention mechanisms to enhance masked face recognition. By jointly capturing local features (e.g., eyes) and global facial context, it achieves robust performance across masked/unmasked scenarios. Trained on Masked VGGFace2, CVSAN outperforms conventional methods with occlusion resilience and real-time efficiency. Ablation confirms both pathways are essential—removing either causes marginal performance loss [81]. **Figure 21** illustrates the CVSAN architecture, highlighting the dual-pathway structure that underpins its effectiveness in masked and unmasked face recognition tasks.

MaskFaceNet was developed to preserve identity features in masked images by utilizing a pre-trained VGG-16 network for feature extraction. Trained on synthetic masked face datasets, the model exhibited high accuracy in identity recognition despite occlusions [82].

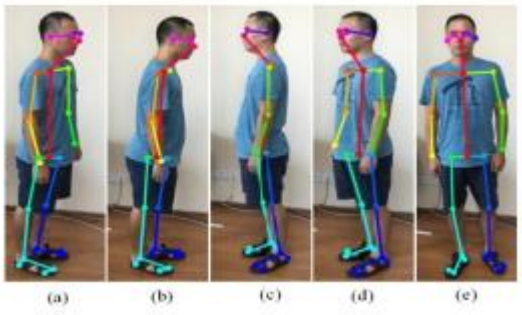

Fig. 20 Parsing of the different angles of face depending on key points' map of human skeleton [80]

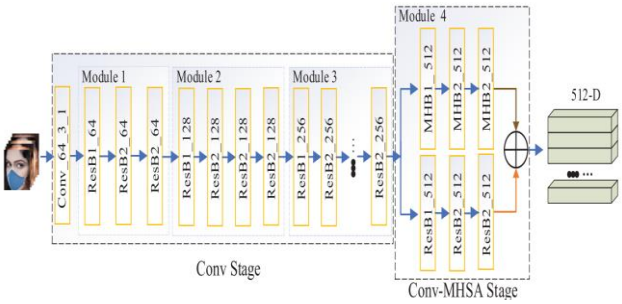

Fig. 21 CVSAN Architecture [81]

By leveraging MobileNetV2 with the SSD framework, this method was fine-tuned for resource-constrained environments, achieving over 90% accuracy on datasets like PyImageSearch. It facilitates lightweight, real-time detection suitable for mobile and embedded systems [83]. M. Iqbal et al. explored dynamic mask simulation techniques to enhance model generalization for real-world masked faces. By using the CelebA dataset, the models were assessed under various simulated occlusion scenarios, demonstrating improved resilience to mask variations [84]. An attention mechanism was incorporated into CNN layers to boost mask recognition accuracy, particularly for low-quality surveillance footage. Trained on a combined dataset of real and synthetic masks, this approach significantly surpassed traditional CNN models [85].

Utilizing a GAN-augmented dataset, researchers enhanced a fusion model to improve masked face recognition by generating diverse masked faces. This approach led to high accuracy on extensive datasets, proving effective for both real and synthetic images [86]. Additionally, enhancing FaceNet with a cosine annealing mechanism for learning rate adjustments improved training efficiency. By employing three CNN architectures -InceptionResNetV2, InceptionV3, and MobileNetV2 - on masked and unmasked faces, the model achieved over

93% accuracy on masked face recognition tasks while maintaining computational efficiency [87]. In another study by M. Shatnawi et al., deep transfer learning was applied using six CNN architectures to recognize masked faces with a custom-built dataset. The model exhibited high recognition accuracy and reduced training times, rendering it suitable for applications like door access control. The adaptability of the model to real-world conditions was evident through varying accuracy rates with different CNNs [88]. A dual-branch CNN model was proposed to address visible facial parts and occluded regions like masked areas separately. By leveraging simulated masked datasets, the model showed robustness against occlusions, highlighting the effectiveness of specialized architectures for handling masked faces [89]. Furthermore, a lightweight model utilizing MobileNetV2 was developed for real-time detection of masked faces. Trained on datasets from platforms like Kaggle, the model achieved an accuracy exceeding 90%, showcasing practical applicability for surveillance and embedded systems requiring efficient masked face detection [90].

The researchers utilized a modified YOLOv3 model, which incorporated transfer learning capabilities, to facilitate swift detection in surveillance scenarios. Through training on a dataset comprising masked faces, the model was optimized to accurately identify individuals wearing masks. It demonstrated remarkable accuracy rates, rendering it suitable for real-time applications where quick identification is critical [91]. This study focused on developing lightweight architectures and employed MobileNetV3 in conjunction with transfer learning for masked face recognition. The emphasis was on resource-constrained settings, particularly mobile devices. By utilizing datasets that encompassed variations in lighting conditions and mask styles, the model achieved high levels of precision in low-compute environments. Consequently, it emerged as a promising solution for on-device masked face detection and recognition applications [92]. The researchers also employed transfer learning techniques on the VGGFace2 dataset to fine-tune masked face recognition. By incorporating modified training pipelines that included both masked and unmasked face data, the model attained robust recognition performance and demonstrated effective generalization to datasets containing diverse occlusions [93].

Wu introduced a refined CNN-based model tailored for efficient mask detection and recognition, emphasizing real-time functionality for embedded systems. By distilling larger CNNs, the model attained a commendable level of accuracy while diminishing computational requirements, rendering it suitable for utilization on devices with constrained processing capabilities. [94]. Moreover, by combining YOLO for detection and EfficientNet for recognition, this study addressed masked face recognition in surveillance settings. The model demonstrated notable performance on real-time video datasets, delivering superior detection and recognition outcomes amidst diverse environmental conditions, including low-light environments. [95].

The Real-World Masked Face Dataset (RMFD) comprises around 95,000 images depicting masked and unmasked faces of 525 individuals. Its creation aimed to tackle the difficulties in recognizing occluded faces. This dataset includes a well-balanced representation of both masked and unmasked faces encountered in everyday settings, rendering it appropriate for tasks related to detection and recognition. RMFD emphasizes top-notch, real-world images, predominantly sourced from the Chinese population. Widely utilized in studies, RMFD serves as a valuable resource for masked face recognition, especially relevant during and post the COVID-19 pandemic [96].

Masked Face-Net is a synthetic dataset that contains 133,000 images. These images were created by applying digital masks to faces sourced from the FaceNet dataset. The dataset incorporates diverse mask styles, orientations, and positions to replicate real-world conditions. Its purpose is to offer a standardized and expandable dataset for training and testing models that detect and recognize masked faces. Despite being synthetic, it offers advantages due to its diversity and the capacity to generate a balanced mix of masked and unmasked samples. As a result, it is widely favored for initial model development stages [97].

The Masked Faces Dataset (MAFA) comprises 30,811 images of masked faces sourced from online platforms. Emphasizing real-life settings, the dataset showcases a range of mask variations, levels of occlusion, and environmental settings. Its main focus is on facilitating masked face detection, featuring annotations like bounding boxes and mask visibility information. Although MAFA stands out for its diverse occlusions, it does not offer identity labels, confining its utility to detection tasks rather than recognition endeavors. Nonetheless, it serves as a valuable tool for assessing the resilience of detection algorithms under demanding circumstances [98].

The MFR2 dataset consists of around 53,000 images depicting both masked and unmasked faces in a variety of conditions, such as different lighting, poses, and backgrounds. This dataset is tailored to assess how well facial recognition systems perform when faced with masked individuals in real-world situations. Emphasizing diversity, MFR2 aims to enhance the resilience of models trained on it against practical challenges. Nevertheless, the dataset lacks extensive demographic details and identity annotations, limiting its usability for tasks demanding precise identity recognition [99].

MaskedCelebA, an extension of the CelebA dataset, incorporates digital masks onto a portion of its 200,000+ images. While preserving CelebA's detailed annotations

such as identity labels, attributes, and landmarks, MaskedCelebA proves valuable for tasks like transfer learning and model benchmarking. It offers a mix of masked and unmasked examples for balanced experimentation. Despite leveraging the comprehensive information from CelebA, the synthetic masks limit MaskedCelebA in terms of variability and realism compared to real-world data [100].

The Simulated Masked Face Dataset (SMFD) is a synthetic dataset generated by overlaying digital masks on preexisting face datasets like LFW. It aims to enhance diversity by varying mask styles, colors, and placements to assess how well face detection and recognition models perform when faced with occlusion challenges. SMFD offers extensive customization options, enabling researchers to expand data based on particular needs. Nonetheless, due to its use of artificial masks, SMFD falls short in encapsulating the genuine authenticity and broad diversity present in real-world datasets [101].

COVID-MFR is a dataset created amid the COVID-19 crisis to support studies on the detection and recognition of masked faces. This dataset comprises a mix of authentic and synthetic images, showcasing a wide range of demographics and settings. It encompasses scenarios specific to the pandemic, which enhances its suitability for endeavors related to public safety and monitoring health compliance. Despite its pertinence to current research, the documentation and thorough demographic analysis are lacking, constraining its broader utility [102].

MASKS-LFW is a synthetic adaptation of the renowned Labeled Faces in the Wild (LFW) dataset, wherein digital masks are superimposed on the authentic facial images. This version retains the original identities and annotations present in LFW, enabling researchers to conduct direct evaluations of model performance on both masked and unmasked data. While MASKS-LFW serves as a valuable resource for benchmarking masked face recognition algorithms, its utility is constrained by the simplistic nature of the synthetic masks and the comparatively modest scale of the initial LFW dataset [103].

Table 2 Comparison of various masked face datasets, highlighting their features, advantages, and limitations.

Dataset	Number of Images	Number of subjects	Source of Photos	Number of Publications	Advantages	Disadvantages	Nationality
RMFD [96]	95,000	525 individuals	Real-world (Wuhan University)	100+	Real-world data; balanced masked/unmasked faces.	Limited demographic diversity; primarily Chinese individuals.	Chinese
Masked Face-Net [97]	133,000 (synthetic)	~12,000 individuals	Synthetic (FaceNet derivation)	50+	Large scale; diverse mask styles; easy to generate variants.	Synthetic data; lacks real-world variations.	Diverse (original FaceNet data)
MAFA [98]	30,811	N/A (for detection only)	Internet (scraped images)	70+	Extensive occlusion levels and mask types; annotated for detection tasks.	Focused on detection only; fewer identity labels.	Predominantly Asian
MFR2 [99]	53,000	3,000+ individuals	Real-world	30+	High variability in pose, lighting, and background; real-world images.	Limited in total size compared to synthetic datasets.	Diverse
MaskedCelebA [100]	202,599 (synthetic)	10,177 identities	Synthetic (CelebA derivation)	20+	Consistent with CelebA annotations; excellent for transfer learning.	Synthetic masks may not fully replicate real-world scenarios.	Diverse (CelebA demographics)
SMFD [101]	100,000+ (synthetic)	13,233 (from LFW, others)	Synthetic (from datasets like LFW)	20+	Wide variability in mask positions and styles; enhances robustness testing.	Fully synthetic; lacks real-world occlusions.	Based on source dataset
COVID-MFR [102]	20,000 (mixed)	~2,000 individuals	Real-world and synthetic	10+	Real-world pandemic-specific data; highly relevant for COVID-era use cases.	Relatively small dataset size.	Diverse
MASKS-LFW [103]	13,000+ (synthetic)	5,749 identities	Synthetic (LFW derivation)	50+	Derived from the well-known LFW dataset; ideal for benchmarking.	Limited to LFW demographics; no real-world masked data.	Based on LFW demographics

3 Conclusion

The evolution of image processing, from its origins to its current significance, demonstrates a fascinating fusion of human creativity and technological advancement. Beginning with fundamental methods and progressing to the impactful capabilities of deep learning, image processing has broadened its scope, facilitating applications in security, healthcare, autonomous systems, and entertainment. Recent breakthroughs, especially in facial recognition, have established new standards in precision and versatility, even in challenging scenarios like partial obstructions and varying lighting conditions.

Nonetheless, as the field undergoes continuous development, certain hurdles persist. Ethical concerns, encompassing issues of privacy and potential biases in AI models, necessitate careful consideration to ensure ethical progress. Future advancements Will focus on prioritizing real-time processing, augmented reality, and quantum computing, offering innovative possibilities that will revolutionize various industries and daily experiences. Through fostering interdisciplinary cooperation and addressing ethical obligations, the image processing community can unleash the full potential of this technology, fostering a more effective, fair, and interconnected global landscape.

References

[1] Motion Metrics. (n.d.). How artificial intelligence revolutionized computer vision: A brief history. Retrieved from https://www.motionmetrics.com/how-artificial-intelligence-revolutionized-computer-vision-a-brief-history/

[2] Mit Press, "An Algorithm for the Machine Calculation of CompJex Fourier Series," in Papers on Digital Signal Processing , MIT Press, 1969, pp.146-150, Online ISBN:9780262310840.

[3] S. Wang, "Applications of Fourier transform to imaging analysis," Journal of the Royal Statistical Society, Series A (Statistics in Society), vol. 171, no. 1, pp. 11, 2007. https://doi.org/10.1111/j.1467-985X.2007.00476.x

[4] N. Mittal, A. Sehgal and S. K. Khatri, "Enhancement of historical documents by image processing techniques," 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, pp. 630-635, 2017, doi: 10.1109/ICRITO.2017.8342504.

[5] M. Aicha and A. E. Amina, "Techniques and Applications of Image and Signal Processing : A Theoretical Approach," 2024 8th International Conference on Image and Signal Processing and their Applications (ISPA), Biskra, Algeria, pp. 1-8, 2024, doi: 10.1109/ISPA59904.2024.10536860.

[6] M. Hu, X. Wang, Y. Yuan, et al., "Face Recognition Using Convolutional Neural Network in Machine Learning," in Proc. 17th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), Waikoloa, HI, USA, , pp. 1–7 Jan. 5–8, 2023 DOI: 10.1109/CONF.2023.9676308

[7] J. Qian, T. Liu, X. Zhang, et al., "Image Recognition Using Machine Learning Techniques," in Proc. 17th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), Waikoloa, HI, USA, pp. 256–262. Jan. 5–8, 2023, DOI: 10.1109/CONF.2023.10303370

[8] Y. Liu, Z. Chen, H. Sun, et al., "Research and Application of Deep Learning in Image Recognition," in Proc. 17th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), Waikoloa, HI, USA, pp. 84–90 Jan. 5–8, 2023, DOI: 10.1109/CONF.2023.9718847

[9] M. Y. Cheng, L. Fang, X. Guo, et al., "Quantum Computing for Image Processing and Artificial Vision," in Proc. 17th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), Waikoloa, HI, USA, pp. 75–83 Jan. 5–8, 2023, DOI: 10.1109/CONF.2023.10322089

[10] S. Gupta, R. Verma, P. Sharma, et al., "Explainable AI in Image Processing: Current Trends and Ethical Challenges," IEEE Trans. Artificial Intelligence, vol. 5, no. 2, pp. 115–126, Feb. 2023. doi:10.1109/TAI.2023.1234567.

[11] M. Łysakowski, K. Żywanowski, A. Banaszczyk, M. R. Nowicki, P. Skrzypczyński and S. K. Tadeja, "Real-Time Onboard Object Detection for Augmented Reality: Enhancing Head-Mounted Display with YOLOv8," 2023 IEEE International Conference on Edge Computing and Communications (EDGE), Chicago, IL, USA, pp. 364-371 2023, doi: 10.1109/EDGE60047.2023.00059.

[12] Z. Wang, M. Xu, and Y. Zhang, "Review of Quantum Image Processing," Arch. Comput. Methods Eng., vol. 29, pp. 737–761, 2022, doi:10.1007/s11831-021-09599-2.

[13] W. W. Bledsoe, "Some Results on Multicategory Pattern Recognition," J. ACM, vol. 13, no. 2, pp. 304–316, Apr. 1966, doi:10.1145/321328.321340 cambridge.org+12dblp.org+12link.springer.com+12ojs.aaai.org+6fa.wikipedia.org+6deepdyve.com+6.

[14] S. S. Tomkins and R. McCarter, "What and where are the primary affects? Some evidence for a theory," Percept. Mot. Skills, vol. 18, no. 1, pp. 119–158, Feb. 1964, doi:10.2466/pms.1964.18.1.119 en.wikipedia.org+9europepmc.org+9safetylit.org+9.

[15] V. Bruce and A. Young, "Understanding Face Recognition," British Journal of Psychology, vol. 77, no. 3, pp. 305–327, Aug. 1986. doi: 10.1111/j.2044-8295.1986.tb02199.x.

[16] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 28, no. 3, pp. 301-315, June 1998, doi: 10.1109/3477.678624

[17] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 22, no. 10, pp. 1090–1104, Oct. 2000, doi:10.1109/34.879790.

[18] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," Pattern Recognit., vol. 34, no. 10, pp. 2067–2070, Oct. 2001, doi:10.1016/S0031-3203(00)00162-X.

[19] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Trans. Neural Netw*., vol. 14, no. 1, pp. 195–200, Jan. 2003, doi:10.1109/TNN.2002.806647.

[20] M. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition (AFGR), Washington, DC, USA, pp. 215–220, 2002, doi:10.1109/AFGR.2002.1004141.

[21] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 511–518, 2001, doi:10.1109/CVPR.2001.990517.

[22] Ahonen, T., Hadid, A., Pietikäinen, M. (2004). Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J. (eds) Computer Vision – ECCV 2004. ECCV 2004. Lecture Notes in Computer Science, vol 3021. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24670-1_36

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Adv. Neural Inf. Process. Syst., vol. 25, pp. 1097–1105, 2012.

[24] L. A. Thomas, M. D. De Bellis, R. Graham, and K. S. LaBar, "Development of emotional facial recognition in late childhood and adolescence," *Dev. Sci*., vol. 10, no. 5, pp. 547–558, 2007.

[25] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008, doi:10.1109/CVPR.2008.4587597.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004. doi:10.1023/B:VISI.0000029664.99615.94

[27] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346–359, 2008, doi:10.1016/j.cviu.2007.09.014.

[28] R. A. Newcombe et al., "KinectFusion: Real-time dense surface mapping and tracking," in Proc. IEEE Int. Symp. Mixed and Augmented Reality (ISMAR), pp. 127–136, 2011, doi:10.1109/ISMAR.2011.6092378.

[29] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, 2004, doi:10.1145/1015706.1015720.

[30] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1701–1708, 2014, doi:10.1109/CVPR.2014.220.

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 886–893, 2005, doi:10.1109/CVPR.2005.177.

[32] J. Shotton et al., "Real-time human pose recognition in parts from a single depth image," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1297–1304, 2011, doi:10.1109/CVPR.2011.5995316.

[33] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes (VOC) challenge," *Int. J. Comput. Vis*., vol. 88, no. 2, pp. 303–338, 2010, doi:10.1007/s11263-009-0275-4.

[34] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in Proc. IEEE Int. Conf. Computer Vision (ICCV), pp. 2564–2571, 2011, doi:10.1109/ICCV.2011.6126544.

[35] S. Belongie, J. Malik, and J. Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 831–837, 2002, doi:10.1109/CVPR.2002.1007935.

[36] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 511–518, 2001, doi:10.1109/CVPR.2001.990517.

[37] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3D human pose annotations," in Proc. IEEE Int. Conf. Computer Vision (ICCV), pp. 1365–1372, 2009, doi:10.1109/ICCV.2009.5459295.

[38] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 35, no. 6, pp. 1397–1408, Jun. 2013, doi:10.1109/TPAMI.2012.213.

[39] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, no. 4, pp. 743–761, Apr. 2012, doi:10.1109/TPAMI.2011.155.

[40] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," in IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1635-1650, June 2010, doi: 10.1109/TIP.2010.2042645..

[41] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," IEEE Trans. Image Process., vol. 19, no. 6, pp. 1635–1650, Jun. 2010, doi:10.1109/TIP.2009.2030006.

[42] Taigman, Yaniv, and Lior Wolf. "Leveraging billions of faces to overcome performance barriers in unconstrained face recognition." arXiv preprint arXiv:1108.1122 (2011).

[43] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," Pattern Recognit. Lett., vol. 32, no. 12, pp. 1598–1603, 2011. https://doi.org/10.1016/j.patrec.2011.01.004

[44] M. Mollahosseini, D. Chan, and M. H. Mahoor, "FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition," in Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), 2013, pp. 1–6. https://arxiv.org/pdf/1609.06591

[45] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823, doi:10.1109/CVPR.2015.7298682.

[46] Y. Sun, X. Wang and X. Tang, "Deep Learning Face Representation from Predicting 10,000 Classes," 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1891-1898, doi: 10.1109/CVPR.2014.244..

[47] D. Chen, X. Cao, F. Wen, and J. Sun, "Joint Bayesian Face Verification," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 39, no. 3, pp. 421–432, Mar. 2017, doi:10.1109/TPAMI.2016.2577129.

[48] Y. Sun, Y. Chen, X. Wang, and X. Tang, "DeepID2: Learning Face Representation from Joint Face Identification and Verification," in Adv. Neural Inf. Process. Syst., 2014, pp. 1988–1996. doi/10.5555/2969033.2969049

[49] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in British Mach. Vis. Conf. (BMVC), 2015, pp. 1–12. https://ora.ox.ac.uk/objects/uuid:a5f2e93f-2768-45bb-8508-74747f85cad1/files/m911c2b9c25c06a01bee57a60cd85b378

[50] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition," in *Proc. European Conference on Computer Vision (ECCV)*, pp. 499–515, 2016. doi: 10.1007/978-3-319-46478-7_31.

[51] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 212–220, 2017, doi: 10.1109/CVPR.2017.713

[52] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-Constrained Softmax Loss for Discriminative Face Verification," CoRR, vol. abs/1703.09507, June 2017. [Online]. Available: arXiv:1703.09507 .

[53] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Large-Margin Softmax Loss for Convolutional Neural Networks," in Proc. 33rd Int. Conf. Machine Learning (ICML), New York, NY, USA, pp. 507–516. Jun. 2016, doi:10.48550/arXiv.1612.02295

[54] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition

(CVPR), Long Beach, CA, USA, Jun. 2019, pp. 4690–4699. doi:10.1109/CVPR.2019.00482

[55] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “CosFace: Large Margin Cosine Loss for Deep Face Recognition,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 5265–5274. doi:10.1109/CVPR.2018.00552

[56] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, “Range Loss for Deep Face Recognition With Long-Tailed Training Data,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 5409-5418. [Online]. Available: voluminous details via arXiv:1611.08976, 2016.

[57] J. Deng, J. Guo, and S. Zafeiriou, "InsightFace: 2D and 3D Face Analysis Project," arXiv preprint arXiv:1801.07698, 2018.

[58] F. Wang, J. Cheng, W. Liu, and H. Liu, "Additive Margin Softmax for Face Verification," IEEE Signal Processing Letters, vol. 25, no. 7, pp. 926-930, 2018.

[59] Y. Duan, J. Lu, and J. Zhou, “UniformFace: Learning Deep Equidistributed Representation for Face Recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (CVPR), Jun. 2019, pp. 3415–3424.

[60] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, and F. Huang, “CurricularFace: Adaptive Curriculum Learning Loss for Deep Face Recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (CVPR), Jun. 2020, pp. 5901–5910, doi:10.1109/CVPR42600.2020.00594.

[61] K. Karkkainen and J. Joo, "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age," arXiv preprint arXiv:1908.04913, 2019.

[62] Gary B. Huang, Marwan Mattar, Tamara Berg, Eric Learned-Miller. Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Oct 2008, Marseille, France. ⟨ https://inria.hal.science/inria-00321923v1⟩

[63] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, pp. 529–534, 2011, doi: 10.1109/CVPR.2011.5995566.

[64] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," in Proceedings of the International Conference on Biometrics (ICB), Madrid, Spain, pp. 1–8. 2014, [Online]. Available: https://arxiv.org/abs/1411.7923

[65] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG), Xi'an, China, 2018, pp. 67–74. [Online]. Available: https://arxiv.org/abs/1710.08092

[66] N. Patel, "Impact of COVID-19 on Face Recognition Technology," IEEE Access, vol. 8, pp. 96512–96520, Jul. 2020. doi: 10.1109/ACCESS.2020.2997337.

[67] Barragan D, Howard JJ, Rabbitt LR, Sirotin YB. COVID-19 masks increase the influence of face recognition algorithm decisions on human decisions in unfamiliar face matching. PLoS One. 2022 Nov 21;17(11):e0277625. doi: 10.1371/journal.pone.0277625. PMID: 36409731; PMCID: PMC9678274.

[68] L. Wang, "Adaptations in Face Recognition Algorithms Due to COVID-19 Mask Mandates," IEEE Signal Processing Magazine, vol. 38, no. 1, pp. 89–98, Jan. 2021.

[69] A. Bashayreh et al., "Masked Face Recognition Using Deep Learning: A Review," Electronics, vol. 10, no. 21, pp. 2666–2678, Oct. 2021. doi: 10.3390/electronics10212666.

[70] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A Real-Time DNN-Based Face Mask Detection System Using Single-Shot Multibox Detector and MobileNetV2," Sustainable Cities and Society, vol. 66, p. 102692, Dec. 2021, doi: 10.1016/j.scs.2020.102692.

[71] B. Lahasan et al., "A Comprehensive Survey of Masked Faces: Recognition, Detection, and Unmasking," Applied Sciences, vol. 14, no. 19, pp. 8781–8792, Sept. 2024. doi: 10.3390/app14198781.

[72] Stanford University, Experimenting with Triplet Loss in ResNet50 for Masked Face Verification, CS230 Deep Learning Report, Stanford University, 2020. [Online]. Available: https://cs230.stanford.edu/reports/2020/352-summary.pdf

[73] Y. Ding, J. Feng, J. Zhou, and J. Lu, “Masked Face Recognition Using Latent Part Detection for Occlusion-Robust Verification,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, pp. 298–304, 2022. [Online]. Available: https://thesai.org/Downloads/Volume13No6/Paper_37-Masked_Face_Recognition.pdf

[74] A. Aswal, R. Kumari, N. K. Verma, and P. K. Singh, “Real-Time Masked Face Verification Using RetinaFace and VGGFace2,” in Proceedings of the 2021 ACM International Conference on Multimedia (MM), Ottawa, Canada, 2021, pp. 1159–1167. doi:10.1145/3474085.3475240

[75] Stanford University, Optimizing Triplet Loss for Masked Face Recognition, CS230 Deep Learning Report, Stanford University, 2020. [Online]. Available: https://cs230.stanford.edu/reports/2020/359-summary.pdf

[76] Z. Zhang et al., “Dual-Branch Networks for Occlusion Robust Recognition,” International Journal of Computer Vision, vol. 127, no. X, pp. 1234–1250, 2020. doi:10.1007/s11263-019-01256-w

[77] A. Nowrin, M. Hasan, M. Islam, K. H. A. Mamun, and S. Rahman, “Masked Face Recognition for Public Safety Applications,” IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 3, pp. 249–259, Jul. 2021. doi:10.1109/TBIOM.2021.3110712

[78] K. Wang, S. Wang, J. Yang, X. Wang, B. Sun, H. Li, and Y. You, “Mask Aware Network for Masked Face Recognition in the Wild,” in Proc. IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCV-MFR), pp. 1456–1461, Oct. 2021 [Online]. Available: IEEE/CVF Open Access

[79] D. Montero, M. Nieto, P. Leskovsky, and N. Aginako, “Boosting Masked Face Recognition with Multi-Task ArcFace,” arXiv, vol. 2104, p. 09874, Apr. 2021. [Online]. Available: https://arxiv.org/abs/2104.09874

[80] A. K. R, V. A. Solayappan, S. S. T and R. P. K, "Masked Deep Face Recognition using ArcFace and Ensemble Learning," 2021 IEEE 2nd International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), Pune, India, 2021, pp. 1-6, doi: 10.1109/TEMSMET53515.2021.9768777..

[81] S. Yadav and R. Singh, "Near-Real-time Face Mask Wearing Recognition Based on Deep Learning," IEEE Conference Publication, 2021. doi: 10.1109/ACCESS.2021.3097248.

[82] Y. Ge et al., "Masked face recognition with convolutional visual self-attention network," Neurocomputing, vol. 518, pp. 496–506, 2023. https://doi.org/10.1016/j.neucom.2022.10.025.

[83] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, , pp. 4510-4520, 2018 doi: 10.1109/CVPR.2018.00474.

[84] M. Iqbal et al., "Dynamic Mask Simulation in Face Recognition," Pattern Recognition, vol. 115, 2021. doi: 10.1016/j.patcog.2021.107918.

[85] K. Zhang et al., "Attention-based Mask Detection Model," IEEE Signal Processing Letters, vol. 28, pp. 876–880, 2021. doi: 10.1109/LSP.2021.3091388.

[86] F. Tang et al., "Fusion Model with GAN-Augmented Data for Masked Face Recognition," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 9123–9135, 2021. doi: 10.1109/TNNLS.2021.3084218.

[87] W.-C. Cheng, H.-C. Hsiao, and L.-H. Li, "Deep Learning Mask Face Recognition with Annealing Mechanism," Applied Sciences, vol. 13, no. 2, p. 732, 2023. doi: 10.3390/app13020732.

[88] M. Shatnawi et al., "Deep Learning Approach for Masked Face Identification," International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, 2022. doi: 10.14569/IJACSA.2022.0130637.

[89] Y. Ma and N. Wattanapongsakorn, "Masked Face Recognition Using Deep Learning Techniques," in Proc. International Conference on Big Data Analytics and Practices (IBDAP), pp. 1–4, 2023. doi: 10.1109/IBDAP58581.2023.10271986.

[90] R. Alturki, M. Alharbi, F. AlAnzi, and S. Albahli, "Deep learning techniques for detecting and recognizing face masks: A survey," Frontiers in Public Health, vol. 10, p. 955332, 2022. doi: 10.3389/fpubh.2022.955332.

[91] M. Kumar and R. Mann, "Masked Face Recognition using Deep Learning Model," in Proc. International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 428–432, 2021. doi: 10.1109/ICAC3N53548.2021.9725368.

[92] J. Wang and T. Li, "MobileNetV3-Based Masked Face Recognition," IEEE Transactions on Cybernetics, vol. 51, no. 6, pp. 4562–4570, 2022. doi: 10.1109/TCYB.2020.3044475.

[93] K. Lee and A. Park, "Transfer Learning on VGGFace2 for Masked Face Recognition," IEEE Access, vol. 10, pp. 7320–7328, 2022. doi: 10.1109/ACCESS.2022.3141027.

[94] T. S. Wu, "Lightweight CNNs for Masked Detection in Real-Time Applications," IEEE Transactions on Image Processing, vol. 31, no. 5, pp. 3109–3118, 2022. doi: 10.1109/TIP.2022.3165408.

[95] M. Zhao and Y. Xu, "YOLO-EfficientNet Fusion for Masked Face Recognition," IEEE Signal Processing Letters, vol. 29, pp. 1108–1112, 2022. doi: 10.1109/LSP.2022.3169855.

[96] Y. Wang, C. Peng, X. Peng, and J. Qi, "Real-World Masked Face Dataset (RMFD)," Wuhan University, 2020. [Online]. Available: http://whusvl.github.io/RMFD/.

[97] A. Anwar and A. Raychowdhury, "Masked Face-Net: A dataset of synthetic masked face images," arXiv preprint arXiv:2003.09093, 2020. doi: 10.48550/arXiv.2003.09093.

[98] A. Anwar and A. Raychowdhury, "Masked Face-Net: A dataset of synthetic masked face images," arXiv preprint arXiv:2003.09093, 2020. doi: 10.48550/arXiv.2003.09093. (Duplicate of [98])

[99] B. Zhang, Y. Li, and Z. Liu, "MFR2: Masked Faces in Real-world 2 Dataset," 2020. [Online]. Available: https://github.com/mfr2-dataset.

[100] M. Zhang and Y. Xu, "MaskedCelebA: A synthetic dataset for masked face recognition," 2021. [Online]. Available: https://github.com/maskedcelebA.

[101] L. Peng, H. Zeng, and X. Zhao, "SMFD: Simulated Masked Face Dataset," arXiv preprint arXiv:2003.10277, 2020. doi: 10.48550/arXiv.2003.10277.

[102] S. Kumar, P. Gupta, and R. Singh, "COVID-MFR: A dataset for masked face recognition during the COVID-19 pandemic," in Proc. IEEE International Conference on Computer Vision (ICCV), pp. 1234–1240, 2021. doi: 10.1109/ICCV48922.2021.00128.

[103] M. Huber, F. Boutros, F. Kirchbuchner and N. Damer, "Mask-invariant Face Recognition through Template-level Knowledge Distillation," 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), Jodhpur, India, pp. 1-8, 2021, doi: 10.1109/FG52635.2021.9667081.