XAIface

Measuring and Improving Explainability for AI-based Face Recognition

Progress report on influencing factors and models (v1)

Deliverable number: D2.1

Version: 1.0

Acronym of the project: XAIface
Title of the project: Measuring and Improving Explainability for AI-based Face Recognition.
Grant: CHIST-ERA-19-XAI-011
Web site of the project: https://xaiface.eurecom.fr/

Executive Summary

Face recognition has become a fundamental technology in multiple real-world applications using pattern recognition-based tools to offer an increased level of security. Moreover, face recognition solutions increasingly use more sophisticated artificial intelligence (AI) techniques, witnessing significantly increased performance in recent years. However, these performance gains are often associated with higher complexity and harder to understand systems – consequently modern AI-based face recognition systems are sometimes referred to as 'black-box' systems. This raises the risk of not trusting AI-based facial recognition technology if its results cannot be minimally explained, especially when this technology also brings risks in terms of privacy, which is a very sensitive societal issue. Thus, facial recognition explainability has become a pivotal step to understand AI-based facial recognition systems' behaviour and to increase trust on the used technology. Explainability can be the decisive factor to enable the usage of face recognition systems that conform with the European initiative entitled "Artificial Intelligence Act", where most biometric recognition systems are considered "high-risk AI", and the requirements for their authorization will include the need for transparency and information to the user.

The idea underlying AI-based facial recognition explainability is to offer insights into why a specific face probe is matched with one identity instead of another. The explainability process starts by identifying the AI-based face recognition influencing factors, understanding their impact on the overall performance of AI-based facial recognition systems. In this context, the XAIface project aims to contribute to a better understanding and explanation of the working mechanisms of AI-based facial recognition systems, to enhance their effectiveness and the social acceptance of AI-based facial recognition technology.

The main objective of this report is to provide a survey on the influencing factors that impact the recognition performance of AI-based facial recognition systems, in general, and those based on deep learning (DL)-based facial recognition systems, in particular. This report also reviews existing publications that study the impact of the various influencing factors in face recognition performance, and try to model the nature and strength of their effect when compared to each other, as well as any interference between them. Based on the review included in this report, additional studies may be performed to obtain further models considering appropriate criteria, metrics and protocols to more understand the impact of the identified influencing factors on DL-based face recognition performance.

Table of content

List of abbreviations

ADRL	Anomaly-aware Deep Reinforcement Learning
AFRN-GAN	Age Factor Removal Network
AgeDB	Age DataBase
AI	Artificial Intelligence
AR	Augmented reality
AM-Softmax Loss	Additive Margin Softmax Loss
A-Softmax Loss	Angular Softmax Loss
CACD	Cross-Age Celebrity Dataset
CACD-VS	Cross-Age Celebrity Dataset Verification Subset
CAF	Cross-Age Face
CALFW	Cross-Age LFW
CAN	Coupled Auto-encoder Network
CASIA	Chinese Academy of Sciences' Institute of Automation
CelebA	CelebFaces Attributes
CelebFaces	Celebrity Faces
CFEE	Compound Facial Expressions of Emotion
CK	Cohn-Kanade
CNN	Convolutional Neural Network
COCO Loss	COngenerous COsine Loss
CosFace	Cosine Face
CpGAN	Coupled Generative Adversarial Network
COVID	COronaVIrus Disease
CPLFW	Cross-Pose LFW
DAL	Decorrelated Adversarial Learning
DA-GAN	Deep Attention Generative Adversarial Network
DB	Database

DCL	Displacement Consistency Loss
DC-SSDA	Double Channel Stacked Sparse Denoising Auto-encoder
DDAC	Dictionary learning based illicit Drug Abuse face Classification
DDRL	Deep Discriminative Representation Learning
DFM	Dynamic Feature Matching
DFN	Deformable Face Net
DL	Deep Learning
DNA	Deoxyribonucleic Acid
FCN	Fully Convolutional Network
FDDB	Face Detection Data Set and Benchmark
FI-GAN	Feature-Improving Generative Adversarial Network
FRVT	Face Recognition Vendor Test
Fr-DCT	Fractional Discrete Cosine Transform
GAN	Generative Adversarial Network
GANFaces-500k	Generative Adversarial Network for faces
HR	High Resolution
IARPA	Intelligence Advanced Research Projects Activity
ICL	Identity Consistency Loss
ID	IDentity/Identification
IJB	IARPA Janus Benchmark
IMDb-Face	Internet Movie Database for Face
IR	Infrared
JPEG	Joint Photographic Experts Group
JPEG XR	JPEG extended range
KELM	Kernel Extreme Learning Machine
LAG	Large Age-Gap
LFW	Labelled Faces in the Wild
LR	Low-resolution

LWIR	Longwave Infrared
MDS	Multidimensional Scaling
MIT	Massachusetts Institute of Technology
MORPH	Morphology
MS-Celeb-1M	Microsoft Celebrities 1Million
MTCNN	Multi-task Cascaded Convolutional Networks
MWIR	Midwave Infrared
NAN	Neural Aggregation Network
NIR	Near Infrared
NIST	National Institute of Standards and Technology
NR-Network	Noise Resistant-Network
ORE	Other-Race-Effect
OREO	Occlusion-aware face REcOgnition
PaSC	Point-and-Shoot Face Recognition Challenge
PDA	Pyramid Diverse Attention
PIE	Pose, Illumination, and Expression
PTL	Pose-Triplet Loss
R3AN	R3 Adversarial Network
ResNet	Residual Neural Network
ReST	Recursive Spatial Transformer
RL	Reinforcement Learning
RL-RBN	Reinforcement Learning based Race Balance Network
RMFRD	Real-world Masked Face Dataset
SMFRD	Simulated Masked Face Recognition Dataset
SPIHT	Set partitioning In Hierarchical Trees
SRC	Sparse Representation Classification
SWIR	Shortwave Infrared
VDMFP	Video Database of Moving Faces and People

VGG	Visual Geometry Group
XAIface	Explainability of AI-based face analysis
YTC	YouTube Celebrities
YTF	YouTube Face

1. Introduction

Over the past few years, automatic face recognition has attracted remarkable attention due to its increasing performance but has also raised some concerns from industry, research communities, and society in general. This attraction can be demonstrated by the number of developments and contributions coming from multiple backgrounds, including the image processing, computer vision, and pattern recognition communities, among others, all somehow related to face recognition technology [1][2][3][4]. Moreover, multiple face recognition systems have been deployed for day-to-day usage, e.g., Face ID by Apple[1], Rekognition by Amazon[2], DeepFace by Facebook [5], and FaceNet by Google [6]. The main reason behind this growth of face recognition technology are its advantages over other biometric modalities (e.g., fingerprint, iris, etc.) for human identity recognition in a wide range of daily applications, notably in terms of its easy deployment, invasiveness and level of security.

There are two main ways face recognition may be used in practice, namely identification and verification:

- ***Identification***: The purpose is here to answer the question: "Who are you?" by comparing the individual's face with the full set of facial templates stored in the facial recognition system's database. A facial template (henceforth just mentioned as template) is the set of pertinent and unique features to be used for recognition purposes that is extracted from a facial sample. This process of identification is referred to as *one-to-many matching*, which is useful e.g., for law enforcement and forensic investigations.
- ***Verification***: The purpose is here to answer the question: "Are you who you say you are?". Herein, the identity is proved by verifying whether the face presented to the system corresponds to the claimed user. Thus, the input face is used for comparison against a formerly collected template of the user, which is known in biometric terms as *one-to-one matching*; this is useful in applications such as mobile banking, access management, and multi-factor authentication systems to access secured servers.

In the early stages of face recognition technology, several hand-crafted approaches have been developed for face recognition [7][8][9][10][11][12]. With the advent of artificial intelligence (AI) and augmented hardware capacity in recent years, the research community shifted to using AI-based algorithms for face recognition [1][2][6][13][14]. The recent AI-based face recognition solutions typically use deep learning (DL) strategies to extract multiple and powerful face features from the input (facial) data to drive the recognition process ahead, significantly improving the robustness and performance of face recognition technology.

However, notwithstanding the impressive success associated with AI-based face recognition systems, they are still suffering from some drawbacks. On one hand, AI-based facial recognition is used as a powerful tool to offer increased security in daily life. However, this recent demand for increased security may threaten individual privacy, which is a topic of

[1] https://support.apple.com/en-us/HT208109
[2] https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html

intense discussion in human societies, and therefore may raise difficulties to the wide deployment of AI-based facial recognition. For instance, many people do not like having images of their faces recorded and stored in databases, even for anonymous future use, as often people do not trust the entities running facial recognition systems. In several cultures, simply taking facial pictures can be an issue. As a consequence, facial recognition technology needs to be trustworthy to play the key role it has the potential to play in human societies.

On the other hand, like any other technology, face recognition is not perfect. It has been demonstrated that face recognition can exhibit various weaknesses, notably biases, e.g., related to gender, race or age groups. This is particularly critical for AI-based face recognition systems whose performance is largely linked to the training datasets and the deep models used to build the recognition system. For example, in February 2018, Massachusetts Institute of Technology (MIT) researchers discovered that FACE++ tools have a higher error rate when identifying darker-skin women compared to lighter-skin men [15]. In July of the same year, it was reported that the facial recognition systems built by Amazon falsely identified some US Congress members as criminals [16].

These critical issues pose multiple questions about how AI-based face recognition systems are working, and if it merits risking individual privacy in the name of security. Therefore, it becomes a real need to fulfil interpretable reasoning of the AI-based face recognition systems' final decision, in order to understand and explain their behaviour. In this context, the XAIface project aims to better understand the working mechanisms behind AI-based facial recognition systems in terms of recognition performance and explainability to ensure trustworthiness and increase the level of social acceptance of these systems. To accomplish this, a starting task is to identify the influencing factors that impact the AI-based face recognition decision making.

In this context, Section 2 of this report targets reviewing the major influencing factors impacting AI-based face recognition performance. Moreover, since DL solutions have proven their advantage for face detection and recognition, this review specifically addresses the influencing factors in DL-based face recognition. Next, Section 3 will review the main studies in literature on the impact of the various influencing factors on the face recognition performance, notably models expressing the face recognition accuracy as a function of some relevant influencing factor parameters. Following this review, additional studies will be performed to obtain additional models for these impacts, as a fundamental step to fully understand the impact of such factors on the performance of deep face recognition systems and contribute to improving the underlying processes involved. The ultimate goal is enhancing the reputation and trustworthiness of AI-based facial recognition technology in our society.

2. Influencing Factors in AI-based Face Recognition

Understanding and explaining the behaviour of AI-based face recognition systems requires comprehending the influencing factors that impact the final decision of real-world deep face recognition solutions. Generally, the influencing factors can be categorised into two classes, namely intrinsic and extrinsic factors [17]:

- ***Intrinsic Factors***: Include all physical states of the human face impacting the recognition system decision-making process. Examples include facial expressions, ageing, or plastic surgery, among others.
- ***Extrinsic Factors***: Incorporate all factors not directly related to the human face characteristics, but which can influence the facial appearance. Examples include occlusions, changes caused by noise or the spatial resolution or the device used to capture facial images, or illumination and pose variations, but also any algorithmic aspect that can condition the results to be achieved.

As illustrated in *Figure 1*, another type of classification groups the influencing factors into three categories, notably:

- ***Data Quality-related Factors***: Include the extrinsic factors directly linked to the state of the input data (image/video) which may impact the recognition process. Examples include image resolution, illumination variation, face occlusion, noise, etc.
- ***Human-related Factors***: Include all the individual-based variations that change the facial aspect, i.e., the intrinsic factors. These factors may include innate/natural changes, for instance related to ageing or other demographic effects; as well as to facial changes related for instance with face expression variations, or to more permanent changes, such as those resulting from plastic surgery.
- ***Model-related Factors***: Include the extrinsic factors associated with the deep learning features and algorithms used for face recognition. These factors may include the architecture of the deep network, training strategy, loss function, or the strategy for model parameters reduction, among others.

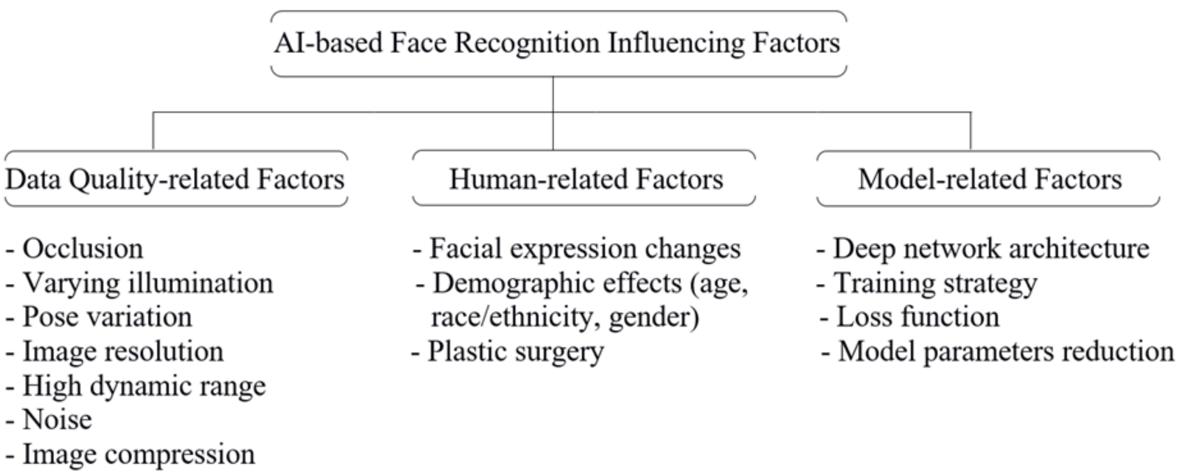


Figure 1: Taxonomy of AI-based face recognition influencing factors.

The objective of the current section is to review the key influencing factors in AI-based face recognition, following the classification into three categories proposed above. Also, the state-of-the-art works which tackle or try to reduce the effect of these factors on face recognition systems are discussed.

2.1. Data Quality-related Influencing Factors

The available literature supports the conclusion that a decrease in data quality can degrade the real-world facial recognition systems performance. In the following, a list of the most critical data quality-related factors is presented and discussed.

2.1.1. Occlusion

In practice, it is not always possible to present fully visible facial images/videos to face recognition systems' sensors. This may be due to the presence of elements such as beard, moustache, sunglasses, scarf, earrings, hairs, etc. (see *Figure 2*), that occlude the appearance of the full face. In addition, contrary to the recognition by humans who are able to deal with significant variations, facial recognition mechanisms tend to perform poorly when some facial parts are missing [18][19]. Thus, the partial occlusion of the human face is one of the most challenging problems for automated face detection and recognition.

Various approaches have been introduced to improve the capacity of face recognition systems to perform adequately in less constrained environments [20][21][22][23][3]. As an example, in [20], a method called dynamic feature matching (DFM) is suggested to deal with partial face recognition. With this purpose in mind, a fully convolutional network (FCN) is combined with a sparse representation classification (SRC) to generate spatial features targeting minimising the intra-variation between the complete facial images and the facial patch images of the individual, which efficiently enhance the recognition rate of the facial patch images. In [22], Xu et al. introduce an occlusion-aware face REcOgnition (OREO) method to improve the robustness of 2D face recognition systems to occlusions. The concept underlying this method includes two stages. First, an attention mechanism is used to extract local identity-related regions; the obtained local features are then merged with the global characteristics to compose a single template. Second, a training strategy and a similarity-based triplet loss function are used to equalise the occluded and non-occluded facial images. In November 2020, the National Institute of Standards and Technology (NIST) introduced a second of a series of reports exploring the performance of face recognition solutions under occluded faces by protective facial masks [3]. This report extended the previous one presented in July of the same year that assesses the impact of masked faces on face recognition solutions submitted before March 2020. The second report appended 65 new face recognition solutions to those tested in the previous report, offering a total of 152 face recognition solutions. The report communicated that the face recognition solutions available before the pandemic had issues with masked faces, while some newer face recognition solutions showed better performance than their predecessors in terms of recognizing masked faces.

Figure 2: Illustration of partially occluded face images from the Face Detection Data Set and Benchmark (FDDB) dataset [24].

2.1.2. Varying Illumination Conditions

Illumination variations can radically reduce the face recognition system's performance, as they can make the face detection and recognition tasks more defying and arduous. As illustrated in *Figure 3*, illumination changes can be caused by the diverse lighting conditions in which images/videos have been acquired, potentially affecting image brightness, contrast, background/foreground light and the presence of shadows.

Illumination variations have been examined in various face recognition research studies [25][26]. For instance, in [25] illumination variations are used as well as pose and background variations to generate synthetic data to be used for training DL-based face recognition systems, in order to reduce the need for real-world training images. The experimental results showed that fusing large-scale real-world data with synthetic data enhances the recognition rate. Moreover, it has been found that training recognition models with only synthetic data with strong variation fulfils well even without dataset adaptation where unrestricted and labelled outside images (i.e. those images do not belong to any of the considered classes, but are mistakenly labelled to one of the existing classes) are used. Moreover, illumination aware facial detection and recognition techniques have been developed [27][28][29]. In [29], the fractional discrete cosine transform (Fr-DCT) approach is proposed to minimise the impact of uncontrolled light sources of multiple directions on the individual's face. Moreover, [28] suggests a new method to enhance the lighting normalisation by developing an underlying reflectance model that characterises the interactions between the lighting source, camera sensor, and skin surface to more effectively represent the formation of facial colour appearance.

Figure 3: The effects of illumination changes on the appearance of a human face, from the CMU Pose, Illumination, and Expression (PIE) dataset [30].

2.1.3. Pose Variation

When a facial image is not frontal, face recognition may still succeed although with a lower accuracy, notably when the face undergoes significant rotations. Notice that under such circumstances the recognition task becomes more complex even for humans. In fact, pose variations that result from image acquisition from different points of view, or correspond to different head movements, modify the spatial relations between the observed facial features and those considered during the user registration. Hence, this can result in insubstantial facial appearance changes, as illustrated in *Figure 4*. In this context, handling pose variation becomes a critical requirement to improve the recognition performance.

Several studies have been carried out to understand how face recognition accuracy is influenced by pose variation [31]. Moreover, multiple pose invariant algorithms and non-frontal face datasets have been developed to deal with pose variations [32][33][34]. In [32], the UHDB31 facial dataset is introduced with pose, illumination, and resolution variations. In addition, three protocols are defined to identify the weaknesses and strengths of 3D, 2D and 2D-2D face recognition under pose, illumination, and resolution variations. In [33], a deformable face net (DFN) algorithm is proposed to handle pose variation for face recognition. This algorithm attempts to learn both face recognition-oriented alignment and identity-preserving features. For this reason, the displacement consistency loss (DCL) is used as a regularisation to impose the learned displacement fields for aligning faces to be locally consistent. Besides, the pose-triplet loss (PTL) and the identity consistency loss (ICL) are used to reduce the intra-class traits variation results for multiple poses. Additionally, they increase the inter classes traits under the same pose.

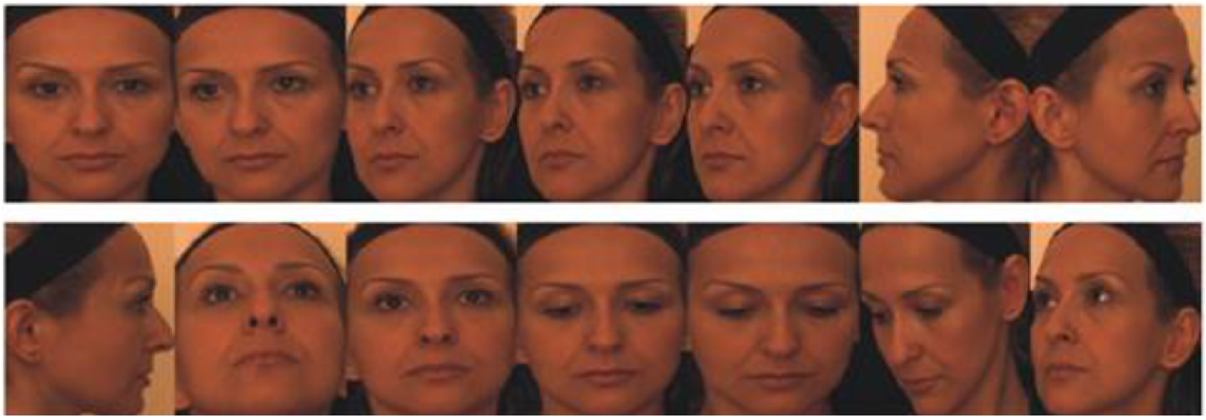


Figure 4: Different facial poses from Bosphorus dataset. The images include 7 yaw face rotations, ranging from -90o to 90o [35] .

2.1.4. Image Resolution

Although modern deep learning face recognition systems report near-perfect performance on well-known datasets, their recognition accuracy can be significantly reduced when low-resolution facial data are used [36][37]. This is often the case when facial data is recorded in environments where the captured faces have limited resolution, such as in surveillance scenarios, as illustrated in *Figure 5*. Moreover, low-resolution images often cannot be used for some pre-processing techniques (e.g., data augmentation) usually considered for training deep face recognition models, which can lead to performance degradation.

Several approaches have been proposed in the literature to tackle the usage of low-resolution images for deep face recognition [38][39][40]. For example, in [39] a discriminative multi-dimensional scaling (MDS) approach is proposed to generate a mapping matrix that projects high resolution and low-resolution images to the same subspace; for example, high-resolution (HR) and low-resolution (LR) images of the same individual are grouped, while the HR and LR images of different individuals are discriminated. In [40], a deep neural network-based identity-preserving end-to-end image-to-image translation method is proposed to super-resolve very LR facial images to their HL counterparts while maintaining the identity-related information. For this purpose, a deep convolutional encoder-decoder network is trained in a multi-scale manner with a combination of facial image reconstruction and an identity-preserving loss.

Figure 5: (a) Example of frame from a surveillance video. (b) Face region with poor resolution.

2.1.5. High Dynamic Range

Facial recognition under uncontrolled lighting conditions may become a hard task. Quite often, the low quality of facial images under uncontrolled lighting conditions reduces the face recognition systems' performance. A common way to mitigate this problem is through facial image pre-processing, which although may alter the resulting images, with a consequent degradation of the facial recognition performance. In this context, high dynamic range (HDR) imaging [41] is introduced as an alternative solution to increase the facial recognition performance when faces are captured under various lighting conditions. The strength of HDR lies in its ability to capture details in high contrast environments, making both dark and bright regions more clearly visible.

HDR image recognition has been largely investigated in the last couple of years [42][43][44]. In [42], the outperformance of HDR imaging over low (or standard) dynamic range (LDR) imaging in the facial recognition process is investigated. For that purpose, the Warwick HDR face dataset is created. After, face recognition is performed by human participants to compare the HDR and LDR imaging performances. The experimental results of this work proved the efficiency of HDR to enhance face recognition accuracy. In [43], P. Korshunov et al. perform a crowdsourcing study to assess the impact of HDR imaging on subjective and objective facial recognition performance. The subjective facial recognition is carried out using the QualityCrowd2[3] open-source framework, while the objective facial recognition is performed using three face recognition algorithms, notably based on Principal Component Analysis (PCA)[45], Linear Discriminant Analysis (LDA)[46], and Local Binary Pattern (LBP)[47]. The experiments are conducted using five facial image datasets created using five tone-mapped solutions to obtain the HDR images. After, the HDR and LDR face recognition performances are compared to conclude that the five tone-mapped versions of HDR affect subjective and objective face recognition differently. Moreover, the impact of using HDR images varies among the three face recognition algorithms. In [44], the benefit of HDR imaging for face recognition is assessed. With this purpose in mind, a novel facial HDR dataset is created under three complex lighting scenarios, namely backlight, left light, and overhead light. Using a speeded-up robust feature (SURF)-based face recognition system, the recognition performance is evaluated under three main conditions, notably including LDR

[3] https://github.com/ldvpublic/QualityCrowd2

images captured with a traditional face recognition system sensor, LDR images pre-processed with conventional methods to minimise the illumination problem, and LDR image obtained from HDR images through conventional HDR tone-mapping techniques. The experimental results show that tone-mapped HDR captured images enhance the recognition performance when used with LDR conventional face recognition systems and help coping with complex lighting scenarios.

2.1.6. Noise

In several face recognition applications, the acquisition of facial images can easily be affected by noise, e.g., salt and pepper noise or Gaussian noise. Additional corruption to the facial images can result from image manipulations including filtering and sparse coding. Additional degradations may result from intentional attacks, sensor errors, or transmission errors [48].

Several works in the literature have been developed seeking to enhance facial recognition performance in the presence of noisy data [49][50]. In [49], Ding *et.al* propose a noise-resistant network (NR-Network) for face recognition under noise. The main idea is to use a multi-input structure in the last fully connected layer of the NR-Network for extracting multi-scale and more distinctive facial features from the input data. In [50], a quantitative analysis of the influence of face image denoising and enhancement methods on the performance of DL-based face recognition is offered. The results of this analysis show that the used image denoising methods enhance the quality of the facial images, thus improving the recognition performance [50].

2.1.7. Image Compression

Most face recognition systems have to work under data size restrictions when storing/transferring biometric samples. For instance, smart cards and mobile applications' databases have a limited capacity to store the reference facial templates. Furthermore, limited-bandwidth transfer mechanisms are impacted by the size of the facial features to transmit. Thus, image compression (e.g., JPEG, JPEG 2000, SPIHT, JPEG XR, etc.) has become an essential post-processing module, frequently employed in multiple facial recognition scenarios [51][52], which means that the face submitted for recognition may suffer from compression artefacts. Despite the popularity of image compression techniques, their use should be acknowledged since, when the original face image is compressed, some digital traces/features of the image may be unintentionally manipulated or erased, which may hamper a reliable face recognition process [53][54].

2.1.8. Acquisition modality

Although the majority of face recognition models have predominantly focused on the use of images captured and represented in the RGB-colour space, considering additional spectra allows for increased robustness, in particular in the presence of different poses, illumination variations, noise, occlusions, as well as increased robustness to presentation attacks (masks, makeup) [191]. One of the most commonly used light spectra for face recognition, in addition to the visible, is the infrared (IR) spectrum, which is briefly described in the following.

The IR spectrum can be broadly divided into two categories: active and passive.
The active IR spectrum consists of near infrared (NIR) and shortwave infrared (SWIR), characterised by reflective material properties, which require an (invisible) IR light to reveal the scene. NIR is placed right next to the red-colour of the visible spectrum with increased wavelength, which is reflected in the visual similarities of VIS and NIR face images. NIR remains invariant to lighting-direction and low-light conditions and hence is mostly employed in monitoring and night-surveillance [195].
SWIR-imagery is, similarly to NIR, visually close to visible imagery. The SWIR-band is significantly larger than the NIR-band and enables sensing in atmospherically challenging conditions such as rain, fog, mist, haze and common urban particulates such as smoke and pollution [196]. In addition, SWIR-sensors are able to capture objects in highly low-light conditions, which makes SWIR suitable for long-range applications (< one kilometre [197]), as well as for identification purposes [198].

Unlike active IR, passive IR does not require illuminating the subject of the acquisition. Beyond wavelength $\lambda = 3\mu m$, the IR band is significantly emissive and passive IR sensors are able to acquire heat radiation emitted from a human face. Thus, cameras are able to capture temperature variations across facial skin tissue, brought to the fore by underlying face vasculature [199]. Sensors able to passively capture temperature variations are called Thermal sensors. Thermal imagery can be acquired in day or night environments [200]. Passive IR consists of Midwave (MWIR) and Longwave (LWIR). MWIR is located between SWIR and LWIR, and has both reflective and emissive properties - allowing for the sensing of different facial-skin-features such as vein patterns [201]. LWIR extends the infrared band up to $\lambda = 14\mu m$ and incorporates exclusively emitted radiation. This shift introduces high intraclass variations. LWIR is visually similar to MWIR with respect to shape and contrast.

2.2. Human-related Influencing Factors

To understand the performance of face recognition systems, human-related factors should also be considered. These factors are directly related to the individual's aspects that impact facial recognition performance. In the following, a list of these influencing factors is reviewed.

2.2.1. Facial Expression Changes

Commonly, human communication includes verbal and non-verbal modes. Non-verbal communication may be carried out via facial expressions. As illustrated in *Figure 6*, humans use different facial expressions to convey their emotions and feelings (e.g., disgust, happiness, surprise, anger, and sadness). Previous research proved that some facial expressions can deceive facial recognition technology and negatively affect its performance [55][56][57]. This is due to the feature variations induced by face muscle contractions, therefore modifying the facial geometry and shape. For example, in [56], the influence of facial expression bias on face recognition technology is assessed. This work addressed two main issues, notably facial expression biases in well-known face recognition datasets and the facial expression influence on face recognition rate. The experiments were conducted with three face recognition architectures, including VGG16 [58], ResNet-50 [59], LResNet100E-IR [60]; four facial datasets, including CFEE [61], CK+ [62], CelebA [63], MS-Celeb-1M [64]; and four different detectors. The experimental results have shown that the most used facial recognition datasets present huge facial expression biases. Additionally, it was demonstrated that the genuine comparison performance is more impacted by facial expression bias than impostor comparison performance. Thus, impersonation attacks via facial expression alterations can be discarded.

Figure 6: Examples of face expression from the FACES dataset with different significant expressions changes [65].

2.2.2. Demographic effects

The performance of automated face recognition systems has long been known to be impacted by demographic cohorts (age, race/ethnicity, and gender) that might lead to potential bias and accuracy variations. Thus, comprehensive studies have been conducted to investigate the impact of demographic cohorts on face recognition systems' performance. For instance, NIST has carried out experiments to quantify the demographic variations in contemporary face recognition algorithms [66]. The main objective of this study is to provide discussions and conclusions concerning the utility and limitations of face recognition solutions toward demographic variations. For this reason, NIST has prepared a report to offer specific knowledge regarding the recognition procedure, notes about where demographic influences may happen, and performance metrics descriptions [66]. Moreover, empirical results and research recommendations have been presented in that report.

In the following, three demographic biases that impact the performance of face recognition systems are addressed, including age, race/ethnicity, and gender.

2.2.2.1. Age

Ageing is one of the key factors that may affect the final decision of facial recognition systems [67]. A person's facial appearance at a given instant is influenced by the interaction of a set of muscles and skin tissues, which are ultimately related to the personal lifestyle, health, and DNA; and those change with human age. When these skin tissues and muscles change over time, they can lead to twisting on the facial surface and affect the face shape, causing significant alterations on the individual's face appearance, as illustrated in *Figure 7*. This means that human face ageing affects the recognition process, notably when a considerable amount of time passes between the acquisition of reference templates and query facial data.

To overcome facial ageing issues, various types of tools have been developed to reduce the effects of age-related influencing factors and boost the recognition algorithms' performance [68][69][70][71][72]. For instance, in [71], a coupled auto-encoder network (CAN) is proposed to address age-invariance on face recognition. In [69], a novel approach to detach age-related characteristics from facial features is proposed. More explicitly, the facial treats are factorized into two separated components, namely identity-dependent and age-dependent components. To accomplish this, the so-called decorrelated adversarial learning (DAL) algorithm is used to reduce the correlation between the paired decomposition traits of age and identity in an adversarial process. The experiments conducted on face ageing datasets (e.g., FG-NET [73], MORPH Al-bum 2 [74], CACD-VS [75], etc.) showed the efficiency of the proposed approach in terms of the facial information preserved in the identity-independent component.

Figure 7: Ageing process over time of the same individual (example from the CALFW dataset [76]*).*

2.2.2.2. Race/Ethnicity

Previous psychological studies have shown that humans tend to recognize faces related to their ethnical group better than other races [77][78]. The elucidation of this phenomenon is that individuals get more experience to distinguish people from a homogeneous population that they subtend daily. This aspect is referred to as the *other-race-effect* (ORE).

Related to this race perceptual prejudice, research has demonstrated that many face recognition systems suffer from critical demographic biases that affect the face recognition decision-making process [66][79][80][81]. For instance, in [82], it has been found out that the facial recognition tool “Rekognition” built by Amazon shows poorer performance when it comes to identifying darker skin women. Additionally, it is proven in [83] that facial recognition algorithms developed by Westerners tend to identify Caucasian faces more correctly than Asian faces, and vice-versa. These prejudices might impact the recognition accuracy for different races but also the security and fairness of facial recognition systems.

2.2.2.3. Gender

Gender bias is considered one of the main sources of errors in automated face recognition systems. This sexual prejudice has been proved in multiple studies revealing that women are more likely to be misidentified through facial recognition than men [84]. For instance, a commercial matcher is used in [85] to explore the performance accuracy variations between women and men. The experimental results showed lower similarity scores for women compared to men. A similar conclusion is presented in the gender section of the NIST report [66], where it was reported that women suffer from higher false-positive rates than men. Another interesting finding is presented in [86], where it was found that the overlapping between genuine and impostor score matching distribution is higher for women than for men, even when using controlled facial images (in terms of facial expressions, makeup, pose variation).

These findings have opened multiple questions about why and how the performance accuracy of face recognition solutions varies between women and men. Thus, various works have been introduced to analyse the gender gap in face recognition systems. For example, in [87], the effect of women under-representation in the training data on the women's performance accuracy was investigated. To accomplish this, the ResNet-50 is evaluated with multiple loss functions using two training datasets. The experimental results showed that (1) balancing both gender cohorts in the training dataset, does not generate balanced accuracies in the test stage. (2) Reducing the gender gap is achieved by training a dataset with more male images, not a gender-balanced dataset.

2.2.3. Artificial beautification

Beautification is the process of making visual alterations to the perceived shape and texture of a human face. Those modifications can therefore compromise the use of face recognition systems in security applications since they distort or modify biometric features. Different types of facial beautification include the use of real-time social media filters, plastic surgery and makeup presence.

2.2.3.1. Make up

Facial cosmetics have the ability to substantially alter the facial appearance. The use of makeup can visually modify the proportions of different face treats such as eyes, lips and cheekbones, having a strong impact on different face recognition systems. The use of make up has been proved as an effective attack for face recognition systems [190]. In this type of attack, a person might apply a high amount of makeup in order to imitate the facial appearance of a target user with the purpose of impersonation.

Related to this makeup presence problematic, in [190] the vulnerability of different open-source face recognition systems such as ArcFace [60] is assessed with regards to different makeup presentation attacks. A makeup attack detection scheme is proposed which compares face depth data with face depth reconstructions obtained from RGB images of potential makeup presentation attacks.

2.2.3.2. Plastic surgery

Facial plastic surgery can be divided into two classes:

- ***Reconstructive Plastic Surgery:*** Rectifies the face features anomalies e.g., cleft lip, palate birthmarks.
- ***Cosmetic Plastic Surgery:*** Enhances the visual appearance of the facial structures and characteristics.

When an individual gets plastic surgery, the facial characteristics are naturally transformed, either locally or globally, to an extent that alters the facial appearance (*Figure 8*). Consequently, these individuals may become unknown to the already existing face recognition systems, including their reference templates. These surgery changes pose a challenge to automatic face recognition technology [88]. Although facial plastic surgery is usually employed for cosmetic and scars treatments to improve the person's appearance, it might also be used by criminals to 'manipulate' their facial identity with the intent to deceive face recognition systems. This increases the challenge to face recognition technology which has not only to preserve the accuracy performance but has also to be robust to the changes produced by facial plastic surgery [89][90].

Figure 8: Illustrations of stars' facial images before (Top row) and after (Bottom row) plastic surgery [89].

2.2.3.3. Social media filters

Face retouching is a widespread application available across a huge spectrum of modern smartphone cameras. In the past years, this type of applications have proved to be more and more relevant as their presence on the top 100 applications on Google Play shows. A social media filter is a photo effect that can be found on those applications and it is applied to images before publishing them. Those effects can range from simple RGB to black-and-white image transformation to different facial shape modifications. Facial social media filters can be divided into four classes:

- ***Colour adjustment filters:*** Change in real-time the captured camera colours. Also modify the image contrast or illumination.
- ***Smoothening filters:*** Smooths and blurrys the face skin, making it look as if a layer of foundation was applied.
- ***Facial features modification filters:*** Enlarge, shrink and sharpen different lines of the face. The most common ones create a smaller nose, a more defined jawline or increase the lips size.
- ***Augmented reality (AR) filters:*** An augmented reality filter is defined as a mask-like filter that incorporates virtual elements to face images in real-time such as long eyelashes, make-up, puppy ears or flower crowns sometimes occluding parts of the face.
- ***Immersive AR face filter:*** Place the users' face into a virtual 3D scene in real time.

Due to this social media face filters trend, some investigations are directing their research towards the creation of facial filters such as skin smothering [188] undetectable for the human eye. The effect of such filters among others, has an impact on face detection and identity recognition, especially when crucial face regions are occluded. In [189] a method to reconstruct the applied manipulation by a filter is developed by implementing a modified version of the U-NET segmentation network. To improve the face recognition system, deep learning algorithms and distance measures are applied to the features extracted using a ResNet-34 network trained to recognize faces.

2.2.4. Lifestyle

The lifestyle of a person, and more specifically the prolonged use of illicit drugs has been proved to alter the texture and geometric features of a person’s face, hence affecting the performance of face recognition systems [194]. Drug abuse can lead to visual face changes such as loss of strength in facial muscles, loss of fat in face, decrease of the skin quality with the appearance of acne and alteration of nose bones. The combination of those effects lead to a highly deformed face presenting a complex challenge to face based authentication systems [193].

To assess the extension of those effects on a face recognition algorithm, [193] analysed using various state-of-art face recognition algorithms the impact of long term illicit drug abuse underlining the importance of the creation of new algorithms able to handle such challenges. In line with this research, [192] proposed a projective Dictionary learning based illicit Drug Abuse face Classification (DDAC) framework to effectively detect and separate normal faces from faces affected by drug abuse. The authors claimed that with this preprocessing step face recognition algorithms can improve their performance on faces affected by drug abuse. In [194], a two-step network was proposed to overcome the challenge presented by those drug affected faces. A scattering transform (ScatNet) based face recognition algorithm is proposed and furthermore optimised with an autoencoder-style mapping function (AutoScat) to learn how to encode the ScatNet representation of a face image in order to reduce the computation time.

2.3. Deep Model-related Influencing Factors

In recent years, the so-called conventional face recognition methods have been overcome by deep learning-based solutions in terms of recognition performance. The shift toward this type of solutions is motivated by their higher accuracy, automatic learning, and significant performance achievements in many computer vision tasks (e.g., image segmentation and classification, object detection and recognition, characters recognition).

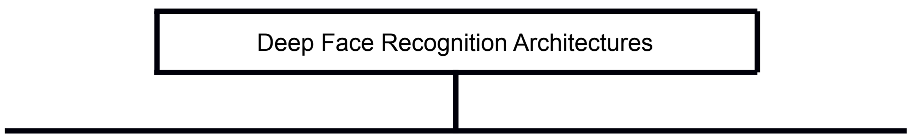
Deep face recognition technology exploits the hierarchical structure of deep learning networks to learn discriminative facial representations from either a face image or a series of facial frames from a video. The learned facial representations are not available for the conventional approaches and can be hard to understand by humans. Consequently, deep learning techniques have boosted face recognition research and the deployment of real-world face recognition applications. Due to the current importance of deep face recognition technology, this section is dedicated to the key factors influencing the recognition performance of this type of technology. Regardless of its high level of success, as demonstrated by the wide dissemination of deep face recognition systems, this technology still presents vulnerabilities, notably related to the large variety of deep learning algorithms' solutions, for instance related to the adopted architectures or training process. In this context, this section reviews the major deep learning technology elements that can impact face recognition performance, namely network architecture, training strategy, loss function, and model parameters reduction.

2.3.1. Deep Network Architecture

The choice of the deep network architecture has received considerable attention in deep face recognition research. In practice, most deep face recognition systems profit from the hierarchical structures of deep learning networks to learn discriminative face representations and adapt to particular application domains. For instance, the performance achieved by convolutional neural networks (CNNs) in various research areas has also proved to be a very effective architecture for face recognition. *Figure 9* presents a list of successful deep learning-based face recognition architectures, categorised into four groups, namely CNN-based, Auto-

Encoder-based, Generative Adversarial Network (GAN)-based, and Deep Reinforcement

Learning (RL)-based architectures.



CNN-Based	Auto-Encoder -Based	GAN-Based	Deep RL-Based
ResNet50 [91]	U-net [92]	R3AN [93]	Fair Loss [13]
SqueezeNet	DC-SSDA [94]	ACSFC [95]	ADRL [96]
ResNet-64 [97]	D2AE [98]	AFRN-GAN [99]	RL-RBN [14]
ResNet-27 [100]	CAN [71]	Age-cGAN [101]	
LightCNN-v29[102]		DA-GAN [101]	
Deep CNN [103]		CpGAN [104]	
MTCNN [105]		FI-GAN [106]	
ReST [107]			
VGG16 [108]			
NAN [109]			
DDRL [110]			
PDA [111]			
Deepface [5]			
DeepID [112]			
VGGFace [113]			
FaceNet [6]			
VGGFace2 [91]			
VGG-16 [58]			
GoogLeNet [114]			
ArcFace [60]			
MagFace [1]			

Figure 9: Overview of deep face recognition architectures.

2.3.2. Training Strategy

Generally, the training strategy comprises multiple hyper-parameters and techniques to adjust these hyperparameters, notably:

- ***Batch Size:*** Defines the number of training samples to propagate through the network before updating the internal model parameters (in one iteration). There are three batch size modes, namely:

- *Batch Mode (Batch Gradient Descent):* The batch size is equal to the total number of samples in the dataset.

- *Mini-batch Mode (Mini-batch Gradient Descent):* The batch size is set to variate among a range of one to the total number of samples in the training dataset.

- *Stochastic Mode (Stochastic Gradient Descent):* The batch size is set to one. Hence, the parameters of the neural network are updated after each sample.

- ***Learning Rate:*** Tuning parameter used to control the adjustment of the neural network weights with respect to the loss function. The lower the learning rate value, the slower the moves toward the minimum of the loss function.
- ***Number of Epochs:*** Defines the number of times the learning algorithm passes through the entire training dataset before that learning algorithm gets updated. In other words, one epoch means that each sample in the training dataset has one opportunity to update the internal model parameters. Generally, the number of epochs is set to be large, e.g., hundreds or thousands, to allow running the learning algorithm to minimise the error between the learning algorithm prediction output and the target true value.
- ***ML Optimization:*** Defines the process of adjusting the hyperparameters of the neural network to reduce the losses by using optimization techniques (henceforth just mentioned as optimizers). Stochastic gradient descent[115], Adam [116], Adagrad [117], and AdaDelta [118] are examples of optimizers used for face recognition.
- ***Activation Functions:*** Includes the set of functions used to help the neural network learn complex patterns in the data by converting the output of the previous neurons into some form that can be taken as the input of the next neurons. ELU, ReLU, Sigmoid, and Softmax are examples of activation functions[4].
- ***Regularisation:*** Includes the set of techniques used to lower the complexity of the model during the training, in order to prevent the overfitting issue and thus improve the accuracy of the model when facing new data. *L1 [119]*, *L2 [120]*, and dropout [121] are the most popular and efficient regularisation techniques.

The appropriate setting of these two hyper-parameters impacts the whole optimization process associated with the training and drives the convergence of the model. Typically, the learning rate is lowered when the training iterations increase to obtain better performance [122][123]. Regarding the batch size, it is well known that a large mini-batch ameliorates the use of the computational resources [124], since quicker updates can be performed by only training with a portion of the dataset. Moreover, using a mini-batch may generate randomness in the training and boost the generalisation performance [125][126].

2.3.3. Loss Function

In the context of deep networks, the loss function plays a fundamental role as it measures the error between the network prediction output and the target true value. If the network's predictions are off, the loss function outputs higher values; if the predictions are good, the

[4] https://ml-cheatsheet.readthedocs.io/en/latest/activation_functions.html#elu

loss function outputs low values. In this context, the amplitude of the error indicates how far the network is from optimal operation.

In the field of deep face recognition, the loss function plays a crucial role in determining the deep face model's performance since it determines what the network must learn. Basically, the loss function is used in the training stage to assist in minimising the facial features' intra-class variation and maximising the facial features' inter-classes variation, consequently leading to a better model in the next training iteration, and consequently to better decisions.

In recent years, it was noticed that the performance accuracy of deep face recognition systems tends to saturate when using a deeper neural architecture [59]. Thus, several studies have been developed to propose novel deep face recognition systems with more suitable loss functions for deep face recognition algorithms [100][127][128][129][130][131][132][133][134]. Generally, the new loss functions are mostly created by modifying the classical loss functions, for instance by adding a penalty term, or a square sum of the distance between the model's predicted values and the true values, to optimise the original loss function (e.g., Softmax loss function, large-margin Softmax, or SphereFace). ArcFace [60] and MagFace [1] are examples of novel and well-performing loss functions used in face recognition. ArcFace incorporates an additive angular margin on the loss function to get highly discriminative features for face recognition while MagFace optimises the facial features with adaptive margin and regularisation based on its magnitude to learn well-structured within-class facial feature distributions. *Figure 10* shows examples of loss functions used in deep face recognition algorithms, including the new algorithms derived from the Softmax loss function.

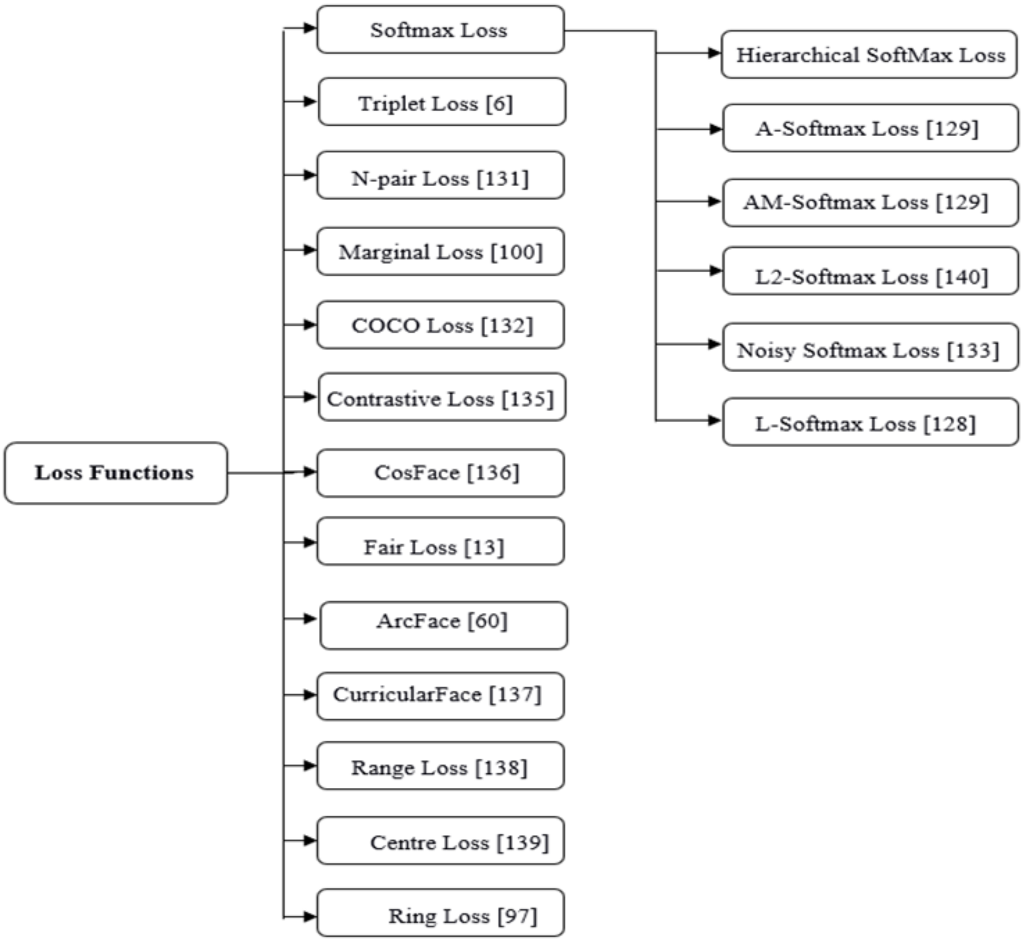


Figure 10: Loss function metrics used in deep face recognition.

2.3.4. Model Parameters Reduction

Face recognition models associated with deep neural networks may include many millions of parameters. This massive number of parameters complicates the deployment of such models, notably on embedded devices and platforms. Thus, the process of reducing the model's parameters (referred to as model compression in some existing work [141][142]) is becoming necessary to minimise the data size associated with the deep learning model. To accomplish this, the convolutional layers may be replaced with smaller size blocks still performing similar tasks, for instance by replacing 3x3 filters with 1x1 filters, or reducing the number of input channels to 3x3 filters, with the ultimate goal of reducing the model's memory footprint by [141][143] storing a lower number of weights and/or adopting a simplified convolutional layer structure. However, since model parameter reduction effectively changes the model to a different, less-ideal, model [142], a deep face recognition performance reduction penalty is often associated with the complexity reduction.

2.4. Summary

In this section, a systematic review of the influencing factors affecting the performance of deep learning-based face recognition systems is presented. The behaviour of face recognition systems is highly related to these influencing factors. Therefore, understanding the role of these factors and their effect on the overall performance of deep learning-based face recognition is an effective tool to explain their decision making, thus increasing the level of trust in face recognition technology.

In addition, the availability and quality of the facial datasets play a pivotal role in face recognition technology development. The impact of facial datasets is mainly linked to their dimensions and proper labelling. Previously, small facial datasets were mostly available. As deep learning-based face detection and recognition solutions have been developed, sizable datasets became essential. Thus, many new datasets have been built considering multiple variations (e.g. variety of racial cohorts, different age ranges, etc.). *Table 1* (resp. *Table 2*) lists some of the recent and most used facial image (resp. videos) datasets.

Table 1: Image datasets for face recognition.

Dataset Name	**Year**	**#Images**	**# Individuals**	**Images Resolution**	**Face Variations**	**Available**
AgeDB [144]	2017	16516	570	-	- Age	Public
UHDB31 [32]	2017	25872	77	2,048×2,448 1,024×1,224 512 ×612 256 ×306 128 ×153	- Pose -Illumination - Resolution	Public
KomNET [145]	2020	39600	-	224 × 224	-	Public
MS-Celeb-1M [64]	2016	10 M	100 k	-	-	Public
CAF [70]	2018	313986	4668	-	- Age - Pose - Race	Private
LFW [146]	2008	13233	5749	250×250	- Pose - Illumination - Focus - Resolution - Expression - Age - Gender - Race - Occlusions - Make-up	Public
IMDb-Face [147]	2018	1.7 M	59 k	-	- Label Noise-controlled	Public

GANFaces-500k [148]	2018	500 k	10 k	108×108	- Pose - Illumination - Expression	Public
GANFaces-5M [148]	2018	5,000 k	10 k	108×108	- Pose -Illumination - Expression	Public
CPLFW [149]	2018	11652	5749	250 ×250	- Pose	Public
CelebFaces [150]	2013	87, 628	5436	-	-	Private
MegaFace [151]	2016	1 M	690 k	-	- Pose - Age	Public
CASIA-WebFace [152]	2014	10575	494414	256×256	- Ages - Expression - Illumination	Public
VGG Face [113]	2015	2.6 M	2622	-	-	Public
VGG Face2 [91]	2018	3.31 M	9131	137×180 (Average resolution)	- Pose - Age - Low Label Noise	Public
RMFRD [153]	2020	95 k	525	250×250	- Masked Face	Public
SMFRD [153]	2020	500 k	10 k	-	- Masked Face	Public
Large Age-Gap (LAG) [154]	2017	3828	1010	-	- Age	Public
CALFW [76]	2017	12000	-	250 ×250	- Age - Gender - Race	Public
FG-NET [73]	2002	1002	82	300 x 400	- Age	Public
CACD-VS [75]	2015	4000	2000	-	- Age	Public
MORPH Album 2 [74]	2009	55000	13000	-	-Gender - Race	Public

Table 2: Video datasets for face recognition.

Dataset Name	**Year**	**#Videos**	**#Individuals**	**#Video Frames**	**Frame Resolution**	**Face Variations**	**Available**
YouTube Celebrities (YTC) [155]	2018	1910	47	7 to 400	20×20	- Pose - Illumination - Expression	Public
YouTube Face (YTF) [156]	2017	3425	1595	48 to 6070	100×100	-	Public

Celebrity-1000 [157]	2014	159726	1000	2405379 (Total)	20×20 (minimum)	- Expression - Pose	Private
VDMFP [158]	2005	958	297	-	-	- Illumination - Gender - Race - Race	-
UMDFaces [159]	2017	22075	3107	2405379 (Total)	-	- Gender - Pose	Public
IARPA Janus Benchmark A (IJB-A) [160]	2015	2042	500	-	-	- Pose	Public
IARPA Janus Benchmark-B (IJB-B) [161]	2017	7011	1845	55026 (Total)	-	-	Public
IARPA Janus Benchmark-C (IJB-C) [162]	2018	11779	3531	117542 (Total)	-	- Race - Occlusion	Public
FaceSurv [163]	2019	460	252	91646 (Total)	-	- Pose - Occlusion - Illumination	Public
PaSC [164]	2013	2802	265	-	-	- Pose	Public

The following section will survey the state-of-the-art mechanisms and models that express and assess the impact of the reviewed influencing factors, notably alone and when they interfere between them.

3. Publications Assessing the Impact of Influencing Factors

Recently, multiple studies assessing the impact of the influencing factors on face recognition systems have been made available in the literature. The main objective of this section is to review some publications addressing these studies to understand the role and impact of the influencing factors on the face recognition systems' decision making, notably by modelling the variation of the face recognition accuracy as a function of relevant parameters related to the influencing factors, e.g. JPEG quality factor or image resolution.

3.1. Overview of Publications

An exhaustive review of the literature addressing the impact of the influencing factors on face recognition systems has been performed. These publications address one or more of the influencing factors surveyed in Section 2 and describe and evaluate the decision making process of both academic and commercial-off-the-shelf (COTS) face recognition systems. Most of these face recognition systems use AI algorithms, with only a few of them relying on hand-crafted algorithms. *Table 3* lists the main characteristics of a selection of the surveyed publications. The first, second, and third columns identify the relevant publications, notably the reference number, publication year and publication journal or conference, respectively. The addressed influencing factors, the evaluated face recognition solutions, and the used facial datasets are presented in the three last columns, respectively.

Table 3: Overview of publications on the impact of relevant influencing factors on face recognition systems performance.

Ref.	**Year**	**Conference/ Journal**	**Influencing Factors**	**Face Recognition Models**	**Datasets**
[165]	2005	- International Conference on Pattern Recognition and Image Analysis	- JPEG Image Compression - JPEG2000 Image Compression	- Principal Component Analysis (PCA) - Independent Component Analysis (ICA) - Linear Discriminant Analysis (LDA)	- FERET
[166]	2006	- International Conference on Control, Automation, Robotics and Vision	- Image Resolution	- Principal Component Analysis (PCA) - Linear Discriminant Analysis (LDA)	- BioID - High-quality FRCG - Low-quality FRCG
[167]	2009	- IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops	- Plastic Surgery	- Principal Component Analysis (PCA) - Fisher Discriminant Analysis (FDA) - Geometric Features (GF) - Local Feature Analysis (LFA) - Local Binary Pattern (LBP) - Neural Network Architecture based 2D - Log Polar Gabor Transform (GNN)	-
[168]	2010	- IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops	- Illumination - Focus Variation	- An Illumination Model - A Focus Model	- CMU-PIE

[169]	2012	- IEEE Transactions on Information Forensics and Security	- Gender - Age - Race	- Cognitec's FaceVACS v8.2 - PittPatt v5.2.2 - Neurotechnology's MegaMatcher v3.1 - Local binary patterns (LBP) - Gabor features - Spectrally Sampled Structural Subspace Features algorithm (4SF@)	-Images from Pinellas County Sheriff's Office
[170]	2015	- IEEE International Conference on Computer Vision Workshop	- Deep Face Model Architecture	- DeepFace - DeepID - WebFace	- LFW
[171]	2016	- International Conference of the Biometrics Special Interest Group	- Image Degradations: . Motion Blur . Noise . Compression . Colour Distortions . Occlusion	- AlexNet - VGG-Face - GoogLeNet	- LFW
[172]	2016	- IEEE Global Conference on Signal and Information Processing	- Limited Training Data - Noisy Labels Data	- CNN	- CASIA WebFace - LFW - WebFace
[173]	2017	- IEEE Conference on Computer Vision and Pattern Recognition Workshops	- Age	- Tow State-of-the-art Commercial-off-the-shelf (COTS) Face Recognition Systems	- PCSO - MSP
[174]	2017	- IET Biometrics	- Illumination - Noise	- GoogLeNet - SqueezeNet - VGG-Face - AlexNet	- VGG-Face - LFW
[147]	2018	- European Conference on Computer Vision	- Dataset Quality (Noisy Labels)	- Attention-56	- MegaFace - MS-Celeb-1M - IMDb-Face
[175]	2018	- International Conference on Electrical, Electronic and Computing Engineering	- Facial expression - Illumination	- Face Recognition Code Based on Histogram of Oriented Gradient (HOG) Features Extraction - Support Vector Machines (SVM) Classifier	-
[176]	2019	- IET Biometrics	- Cross-dataset Training Bias - Dataset's Feature Space Bias	- LBP over landmarks - VGG-Face - ResNet	- FERET - LFW - IJB-A - O2FN - MobBIO - FaceSampler

[177]	2019	- IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops	- Children vs. Adults Bias	- Five Deep Learning Commercial-off-the-shelf (COTS) face recognition systems - Two Government-off-the-shelf (GOTS) Face Recognition Systems - An Open Source Algorithm	- ITWCC−D1 - LFW−D1
[178]	2019	- National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics	- Loss Function	- ResNet - MobileNet	- CASIA-Webface - MS-Celeb-1M - LFW
[66]	2019	- NIST Interagency/Internal Report (NISTIR) – 8280	- Race - Gender - Age	- 106 Face Recognition Algorithms	- Domestic Mugshots - Application Photographs - Visa Photographs - Border Crossing Photographs
[179]	2020	- IEEE Transactions on Technology and Society	- Race	- ArcFace - VGGFace2	- MORPH
[86]	2020	- IEEE Winter Applications of Computer Vision Workshops	- Gender	- ArcFace	- MORPH - Notre Dame - Asian Faces Dataset (AFD)
[67]	2020	- IEEE/CVF Winter Conference on Applications of Computer Vision	- Age	- VGGFace2 (ResNet-50) - FaceNet - ArcFace	- VGGFace2 - MSCeleb-1M - MS-Celeb-1M V2
[142]	2020	- Neurocomputing	- Loss Functions - Model Compression - Network Architecture - Data Augmentation - Deep Face Model Architecture	- VGG-16 - GoogLe-Net - Face- ResNet - ResNet-50	- CASIA-WebFace - UMD-Faces - UMD-CASIA - CFP Frontal-Frontal - LFW - CACD-VS - YTF
[180]	2020	- International Conference of the Biometrics Special Interest Group	- Occlusion by Mask	- ArcFace - SphereFace - A Commercial Face Recognition System	- A New Collected Dataset
[181]	2020	- IEEE/CVF Conference on Computer Vision	- Loss Functions	- ARFace	- MS-Celeb-1M - FG-Net - SCface

		and Pattern Recognition Workshops			- IJB-C - ARFace
[87]	2020	arXiv:2002.02934v2	- Gender Balance in Training Data	- ResNet-50 with three different loss functions . Standard Softmax Loss . Combined Margin Loss . Triplet Loss	- VGGFace2 - MS1MV2 - IJB-B - MORPH - Notre Dame
[182]	2020	- Journal of King Saud University – Computer and Information Sciences	- Deep Face Model Architecture	- AlexNet - GoogleNet - Inception V3 - ResNet50 - SqueezeNet	- FG-NET
[183]	2021	- Computers	- Racial Balance in the Dataset - Deep Face Model Architecture	- ML algorithms: . Support Vector Classifier (SVC) . Linear Discriminant Analysis (LDA) . K-NearestNeighbor (KNN) . Decision Trees(DT) . Logistic Regression (LR) - DL algorithms: . AlexNet . VGG16 . ResNet50	- FERET
[184]	2021	- IEEE Transactions on Biometrics, Behaviour, and Identity Science	- Race	- A2011 - A2015 - A2019 - A2017b	-
[56]	2021	- IEEE International Conference on Image Processing	- Facial Expression	- ResNet-50 - LResNet100E-IR	- CFEE - CK+ - CelebA - MS-Celeb-1M

A subset of the publications presented in *Table 3* are selected for further analysis, according to the following criteria:

- Consider an AI-based face recognition system;
- Propose a relevant study on the impact of one or more influencing factors;
- Published at a top journal or conference.

Table 4 lists the selected subset of publications on the impact of the influencing factors on face recognition systems to be further analysed in this report. As illustrated in the table, the publications are classified according to the influencing factors they address (one or more); some of the previously identified influencing factors are not covered by any of the selected publications.

Table 4: Summary and classification of the selected publications regarding the influencing factors.

		[183]	[169]	[173]	[87]	[147]	[180]	[181]	[174]	[56]	[67]	[66]
Data Quality-related Factors	**Occlusion**						✓		✓			
	Varying Illumination								✓			
	Pose Variation											
	Image Resolutio											
	Noise								✓			
	Availability and Quality of Face Datasets				✓	✓						
	Image Compression								✓			
	High dynamic range											
Human-related Factors	**Facial Expressio Changes**									✓		
	Demographic Effects	✓	✓	✓							✓	✓
	Plastic Surgery											
Deep Model-relat Factors	**Deep Model Architecture**	✓										
	Loss Function							✓				
	Training Strategy											
	Model Parameter Reduction											

In the next sections, the publications listed in *Table 4* are reviewed in more detail. The publications are presented in sequence according to the influencing factors expressed, notably data quality, human, and deep model related influencing factors. For each publication, the analysis will consider: i) objectives; ii) methodology pursued to study and express the effect of the influencing factors; and iii) strengths and weaknesses of the methodology and mechanisms adopted to study the impact of the influencing factors.

3.2. Publications Assessing the Impact of Data Quality-related Influencing Factors

This sub-section provides additional detail about the publications listed in *Table 4* addressing data quality-related influencing factors; each publication is identified by the associated publication.

A. N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, and A. Kuijper, "The Effect of Wearing a Mask on Face Recognition Performance: An Exploratory Study," ***International Conference of the Biometrics Special Interest Group*****, Darmstadt, Germany, September 2020** [180]**.**

- **Influencing Factor(s):** Occlusion by medical mask.
- **Objectives:**
    - To propose a carefully collected masked-face dataset that imitates realistically diverse face capture scenarios, similar to the circumstances in automatic border control, where illumination and background may change.
    - To study the impact of wearing medical masks on the behaviour of one commercial off-the-shelf (COTS) and two academic face recognition systems, e.g., ArcFace [60] and SphereFace [130] in a verification scenario.
- **Methodology Description:** The methodology starts by proposing a new database to study the effect of occlusions by medical masks on the face recognition performance. The facial data are collected in an indoor daylight scenario during three sessions (days), namely ***Session 1***, ***Session 2***, and ***Session 3***. The images of ***Session 1*** are used as reference images (R), while the images of ***Session 2*** and ***Session 3*** are used as probe images (P).

  In addition, two data capturing conditions are considered for each scenario, including the presence/absence of the medical mask and the presence/absence of artificial illumination, which generates multiple subsets for each scenario (see *Table 5*).

Table 5: Overview of the database structure.

Session	***Session 1*****:** References			***Session 2*** and ***3*****:** Probes			
Data split	BLR	M1R	M2R	BLP	M1P	M2P	M12P
Illumination	No	No	Yes	No	No	Yes	Both
Mask	No	Yes	Yes	No	Yes	Yes	Both

BLR: Baseline reference.
M1R: Mask one reference.
M2R: Mask two references.

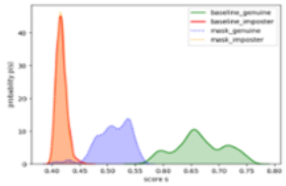
BLP: Baseline probe.
M1P: Mask one probe.
M2P: Mask two probes.
M12P: Data joined from M1P and M2P.

After, four experiments are performed to assess the effect of wearing medical masks on the performance of face recognition pipelines:

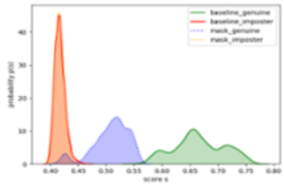
- **Experiment 1:** Measures the recognition performance without a mask (N:N comparison of the subsets BLR and BLP).
- **Experiment 2:** Measures the recognition performance when wearing a mask (N:N comparison of the subsets BLR and M1P).
- **Experiment 3:** Measures the recognition performance in the presence of the mask and electric illumination (N:N comparison of the subsets BLR and M2P).
- **Experiment 4:** Measures the overall recognition performance (N:N comparison of the subsets BLR and M12P).

- **Modelling the Impact of Influencing Factors:** To assess the impact of wearing a mask on the face recognition performance, the genuine and impostor score distribution of Experiment 1 (absence of the mask) and the three remaining experiments (presence of the mask) are compared.

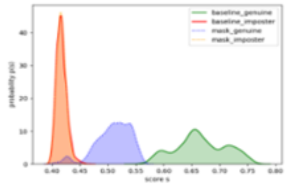
  Figure 11 illustrates the results of the comparison for the three assessed face recognition solutions. For each face recognition solution, the genuine score distribution (blue) is strongly shifted towards the impostor distribution (red) compared to the BLR- BLP setup (green) (see (a), (d) (g)), which leads to a performance degradation. Additionally, this shift becomes stronger when masked images are captured under artificial illumination (see (b), (e) (h)). Moreover, it is found that the impostor score distributions are less influenced by the masked probe images (yellow curves are mostly identical to red curves).



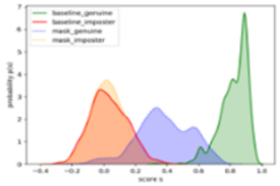
(a) ArcFace: BLR-BLP and BLR-M1P



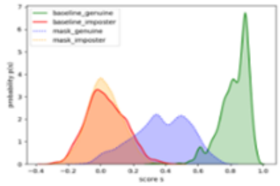
(b) ArcFace: BLR-BLP and BLR-M2P



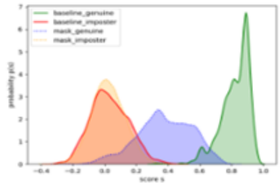
(c) ArcFace: BLR-BLP and BLR-M12P



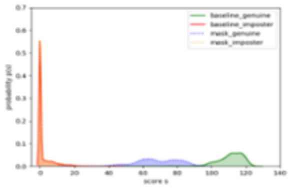
(d) SphereFace: BLR-BLP and BLR-M1P



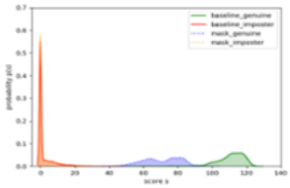
(e) SphereFace: BLR-BLP and BLR-M2P



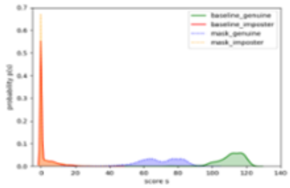
(f) SphereFace: BLR-BLP and BLR-M12P



(g) COTS: BLR-BLP and BLR-M1P



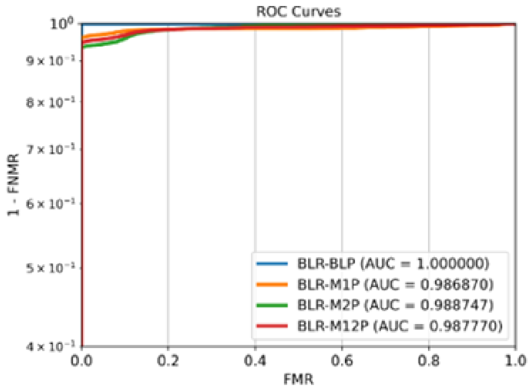
(h) COTS: BLR-BLP and BLR-M2P



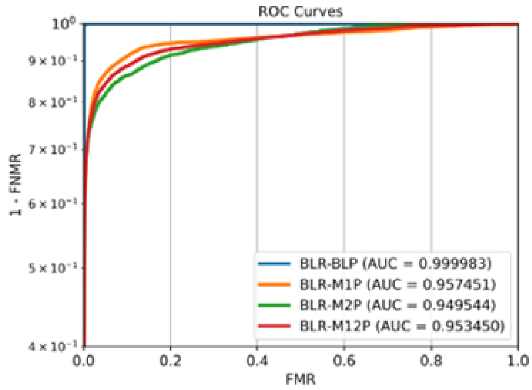
(i) COTS: BLR-BLP and BLR-M12P

Figure 11: Comparison of score (similarity) distributions for the "baseline" BLR-BLP genuine and impostor distributions regarding the distributions including "masked" face probes BLR-M1P (a, d, g); BLR-M2P (b, e, h); and BLR-M12P (c, f, i).

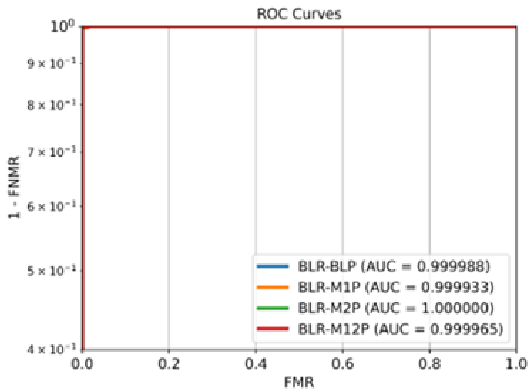
Figure 12 illustrates the face recognition verification performance for the three evaluated face recognition solutions under the four experimental settings, i.e. experiments 1, 2, 3, 4. It may be seen that the COST face recognition solution preserves its verification performance, while the SphereFace and ArcFace performances have a performance degradation when masked facial probes are used (BLR-M1P). Moreover, the degradation becomes stronger when facial data are captured under artificial illumination (BLR-M2P and BLR-M21P).



(a) ArcFace



(b) SphereFace



(c) COTS

Figure 12: Verification performance (as ROC curves) for the three investigated face recognition systems, ArcFace(a); SphereFace (b); and COTS (c).

- **Strengths and Weaknesses:**
  - (+) In addition to the academic face recognition solutions, the publication uses a commercial face recognition pipeline which achieves one of the best performances according to the recent NIST report [124], which increases the relevance of the obtained results.
  - (+) The publication enriches the performance analysis by using multiple verification performance metrics, namely failure to extract rate (FTX), equal error rate (EER), false non-match rate (FNMR), FMR100, FMR1000, ZeroFMR, and receiver operating characteristic (ROC) curves.
  - (-) The suggested dataset is built using day-light images, which might create biases when night-captured images are presented in the identification/verification stage.

B. K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel, "Strengths and Weaknesses of Deep Learning Models for Face Recognition Against Image Degradations" ***IET Biometrics*****, vol. 7, pp. 81–89, October 2017** [174]**.**

- **Influencing Factor(s):** Blur, JPEG compression, noise, image brightness, contrast, and missing pixels.

- **Objectives:** To study the influence of image quality degradation, e.g. blur, JPEG compression, noise, image brightness, contrast and missing pixels, on the performance of four DL-based face recognition systems, namely AlexNet, VGG-Face, GoogLeNet, and SqueezeNet.
- **Methodology Description:** The methodology consists in conducting a rigorous experiment to assess the performance of face recognition including three stages:
    - Train the used deep face recognition solutions from scratch to obtain a fair comparison of their expressivity given the same training dataset, i.e. VGG face.
    - Apply image distortions, notably blur, JPEG compression, noise, image brightness, contrast and missing pixels, with different intensities to the probe images (see *Table 6*).

Table 6: Applied image degradation factors, and associated parameters and intensities.

Blur	Apply Gaussian filters to the probe images with different standard deviations σ, ranging from 2 to 20.
Compression	Encode the probe images with the JPEG coding standard at different quality presets, e.g. of 1, 3, 5, 10, 15, 20, 25, 30, 35 and 40.
Gaussian noise	Add additive Gaussian noise to probe images with a mean of 0 and various standard deviations σ, ranging from 20 to 200.
Salt-and-pepper noise	Truncate all colour components of each probe image pixel to zero with a probability of p/2 and set them to 255 with a probability of p/2 (p between 0.02 and 0.5).
Brightness	Multiply the pixel intensities of all probe images by a brightness factor and clip the resulting pixel values to the valid dynamic range between [0,255].
Contrast	Subtract the central value of the dynamic range from all prob images and multiply the centred images by a contrast factor (between 0.03 and 0.79) and add an offset.
Missing pixels	Remove contiguous pixel areas, e.g. mouth, nose, periocular and eye from the probe images.

- Evaluate the performance of the deep face recognition solutions using the following verification scenario:
    - Consider the facial features output by each evaluated face recognition system as the image descriptor of the input facial image.
    - Conduct one-to-one matching between the facial image descriptors via a cosine-based similarity score using equation *(1)*:

$$g(x_1, x_2, f, T) = \begin{cases} w_1, & \text{if } \delta(f(x_1), f(x_2)) = \delta(y_1, y_2) \geq T \\ w_2, & \text{otherwise} \end{cases} \quad (1)$$

where,

x_1, x_2: Input facial images to be matched (probe and gallery images).

$f(.)$: Selected deep facial recognition model.

y_1, y_2: Descriptors of the images x_1, x_2 respectively.

$\delta(.)$: Cosine similarity.

T: Predefined decision threshold.

- Report the results using labelled faces in the wild (LFW) verification protocol metrics [185], including the mean and standard deviation.
- Conduct a comparative evaluation of the four DL-based face recognition systems.

- **Modelling the Impact of Influencing Factors:** In the following, some of the obtained experimental results are presented, namely the impact of image blurring, JPEG compression, Gaussian noise and salt-and-pepper noise.
  - *Impact of image blurring and JPEG compression*

The left chart in *Figure 13* shows a quick drop of all models' performance with the increase of image blurring intensity which proves the significant influence of image blurring on the recognition performance. Contrarily, the right chart in the same figure shows that the evaluated models are mostly unaffected by the JPEG compression until the intensity of the compression corresponds to very low levels of quality.

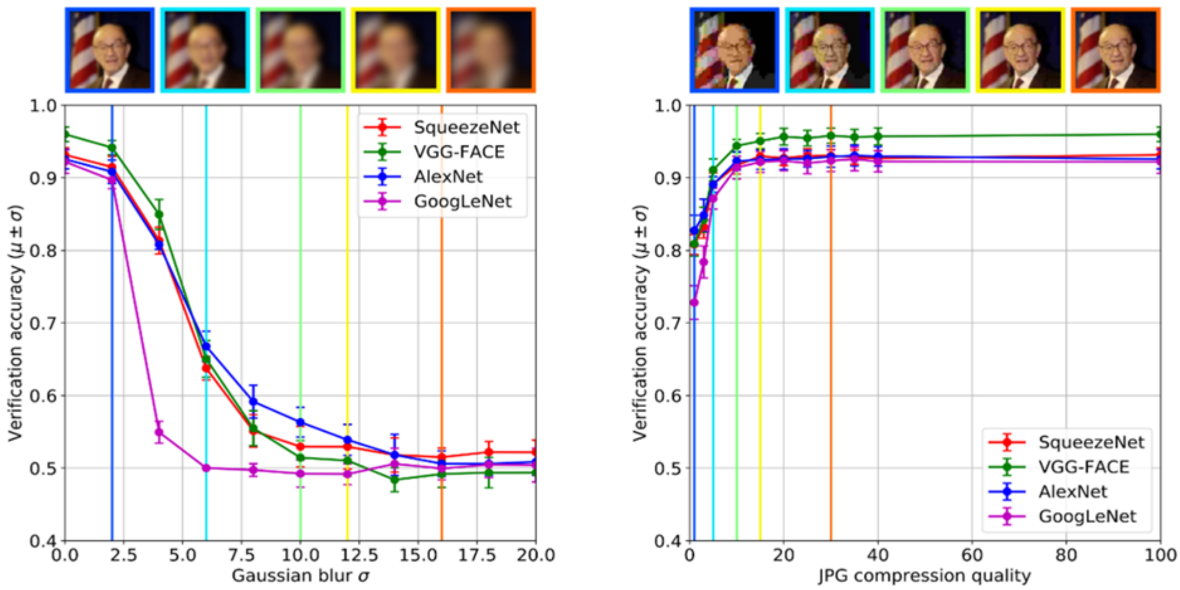


Figure 13: Impact of image blurring (left) and JPEG compression (right) on the recognition performance of the four deep face recognition systems.

- *Impact of Gaussian and Salt-and-pepper noise*

Figure 14 illustrates the verification accuracy of the four recognition systems under Gaussian and Salt-and-pepper noises. The verification accuracy charts show that the evaluated systems behave similarly for the Gaussian noise (left) and Salt-and-pepper noise (right) with a similar impact on the models' recognition performance.

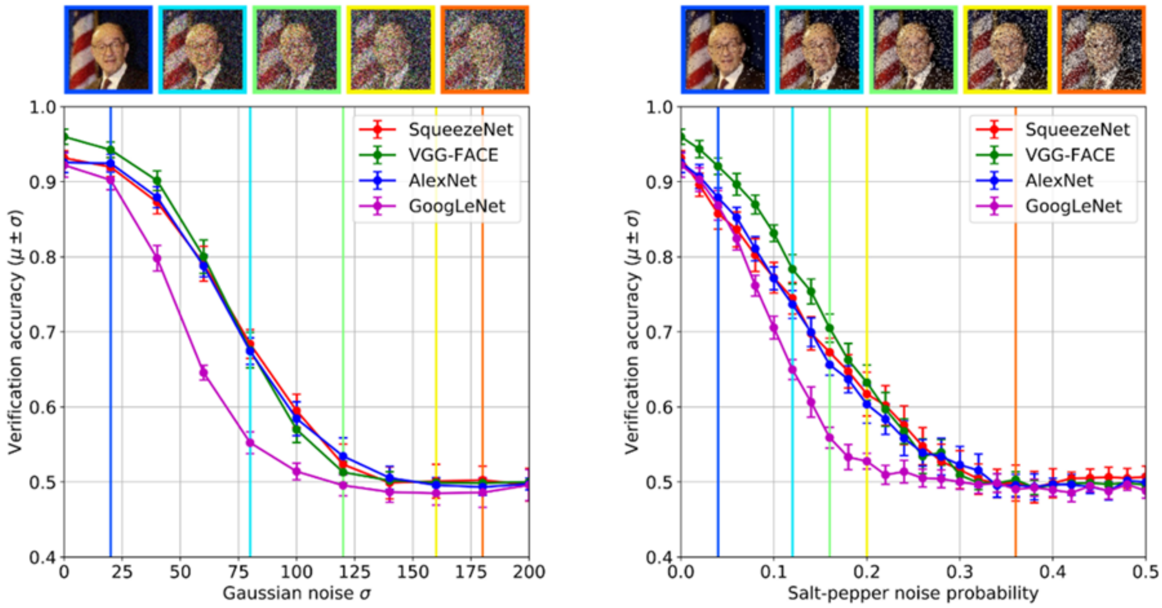


Figure 14: Impact of Gaussian (left) and Salt-and-pepper (right) noise on the recognition performance of the four deep face recognition systems.

- **Strengths and Weaknesses:**
  - (+) The face recognition models are trained from scratch in an unified environment, which leads to a fair comparison among them.
  - (-) The publication evaluates the face recognition systems only under a verification scenario. The face recognition systems need to be evaluated also under the identification scenario to know if they are suitable for both the verification and identification tasks.

C. **V. Albiero, K. Zhang, and K. W. Bowyer, "How Does Gender Balance in Training Data Affect Face Recognition Accuracy?"** ***IEEE International Joint Conference on Biometrics, Houston,*** **TX, USA, October 2020** [87]**.**

- **Influencing Factor(s):** Gender distribution in the training dataset.
- **Objectives:** To study how the gender distribution of the training data impacts the face recognition accuracy and how the female under-representation in the training data degrades the female recognition accuracy in the testing stage.
- **Methodology Description:** The methodology consists in evaluating the performance of the ResNet-50 face recognition system via the following steps:
  - Generate seven subsets from each of the VGGFace2 and MS1MV2 training datasets. The first and second subsets are the initial and gender-balanced (in

terms of subjects and facial images per subject) datasets, respectively, while the remaining subsets are smaller subsets with different ratios between females and males.
    - Train the ResNet-50 model on the various created subsets using three loss functions, notably standard softmax loss, combined margin loss, and triplet loss.
    - Evaluate the face recognition performance of the trained models using three test datasets, notably IJB-B, MORPH, and Notre Dame.
- **Modelling the Impact of Influencing Factors:** In the following, the experimental results of the publication on the gender variation and gender balance in the training data tests are presented.
    - *Gender variation in the training data*

Table 7 presents the gender recognition accuracy of the evaluated face recognition models on the MORPH Caucasian dataset. It is clear from the table that the best accuracy on males (resp. females) is when 100% male (resp. females) training datasets are used. Additionally, it can be seen that the female accuracy overpasses the male accuracy only when 100% female data is used. These results refute the obvious expectation stating that the gender with higher training data gets higher test accuracy.

Table 7: Gender accuracy (%) on the MORPH Caucasian dataset with true accept rate *(TAR) at 0:001% false accept rate (FAR) with different gender balance proportions.*

Loss		VGGFace2					MS1MV2				
		F100	M25F75	M50F50	M75F25	M100	F100	M25F75	M50F50	M75F25	M100
Softmax	Male	56.51	85	86.61	89.55	89.95	76.85	93.08	94.51	95.21	96.09
	Female	80.1	80.24	81.66	68.65	55.24	90.52	89.41	88.74	86.34	81.47
	Avg.	68.31	82.62	84.14	79.1	72.6	83.69	91.25	91.63	90.78	88.78
Combined Margin	Male	0.41	71.25	66.95	69.53	82.98	90.91	98.59	99.24	99.21	99.5
	Female	71.13	71.83	57.66	63.23	52.64	97.71	97.57	97.45	96.51	94.94
	Avg.	35.77	71.54	62.31	66.38	67.81	94.31	98.08	98.35	97.86	97.22
Triplet	Male	39.52	57.23	63.88	68.37	71.36	61.2	80.93	86.83	89.41	90.37
	Female	55.96	56.47	53.68	53.08	35.24	74.61	75.84	77.73	76.87	67.19
	Avg.	47.74	56.85	58.78	60.73	53.3	67.91	78.39	82.28	83.14	78.78

- *Gender balance in the training data*

Table 8 presents the gender recognition accuracy of the evaluated face recognition systems on the MORPH African American dataset. Looking at the female and male accuracies for M50F50 (same ratios of males and females training data are used), it can be noticed that there is a gap between both genders' accuracies indicating that gender balance in the training data does not translate into gender balance in the test accuracy.

*Table 8: Gender accuracy (%) on the MORPH African American dataset with t*rue accept rate *(TAR) at 0:001% false accept rate (FAR) with different gender balance proportions.*

Loss		VGGFace2 F100	M25F75	M50F50	M75F25	M100	MS1MV2 F100	M25F75	M50F50	M75F25	M100
Softmax	Male	55.27	87.55	90.83	89.11	92.92	79.19	96.28	97.58	97.95	98.4
	Female	63.2	77.62	77.61	68.65	58.51	88.7	91.15	89.63	87.79	84.56
	Avg.	59.24	82.59	84.22	78.88	75.72	83.95	93.72	93.61	92.87	91.48
Combined Margin	Male	72.87	76.76	62.7	70.28	84.95	94.79	99.43	99.72	99.78	99.81
	Female	78.2	62.47	45.36	48.97	51.49	98.35	98.85	98.44	98.58	97.88
	Avg.	75.54	69.62	54.03	59.63	68.22	96.57	99.14	99.08	99.18	98.85
Triplet	Male	41.5	58.5	69.09	73.63	76.77	67.4	86.31	91.93	93.32	94.53
	Female	40.38	44.41	49.47	47.28	42.88	71.34	77.51	79.72	79.75	71.66
	Avg.	40.94	51.46	59.28	60.46	59.83	69.37	81.91	85.83	86.54	83.1

- **Strengths and Weaknesses:**
  - (+) Extensive experiments are performed, totaling 42 training conditions.
  - (+) Race and age-balanced training datasets are used to minimise the race and age biases during the experiments.
  - (-) The publication evaluates only one face recognition system.

D. W. Fei et al., "The Devil of Face Recognition is in the Noise," *European Conference on Computer Vision*, Munich, Germany, September 2018 [147]**.**

- **Influencing Factor(s):** Noisy labels in the training dataset, i.e. label errors raising deviations in the original dataset.
- **Objectives:**
  - To understand the source of label noise and assess its influence on the performance of CNN-based face recognition systems using the newly generated free-noise IMDb-Face dataset and two subsets of MS-Celeb-1M and MegaFace datasets.
  - To build a free label noise face recognition dataset for the community.
- **Methodology Description:** The methodology consists in assessing the impact of noisy labels in training datasets by fulfilling the following objectives:
  - Generate free noise label subsets for the MegaFace and MS-Celeb-1M datasets.
  - Build a new large-scale noise-controlled IMDb-Face dataset.
  - Analyse the label noise characteristics of the original and cleaned datasets of MegaFace and MS-Celeb-1M.
  - Assess the association between label flips and outlier noises by injecting noise on the training labels to simulate corruption.
  - Explore techniques to enhance data cleanliness strategies.
- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results of the publication are presented.
  - *Effect of noise on cleaned IMDb-Face dataset*

The proposed free-noise IMDb-Face dataset is assessed under two types of noise, namely:

- *Label flips:* The image has erroneously been given the label of another class within the dataset.
- *Outliers:* The image does not belong to any of the classes under consideration, but mistakenly has one of their labels.

The experiments are performed via the following strategy:

- Infecting the dataset with the label flips and outliers separately by increasing the degree of the noise in the dataset by 10%, 20% and 50%.
- Fixing the size of clean data and diluting it with label flips.

Figure 15 illustrates the effect of the label flips and outliers on the Attention-56 model trained with three loss functions, notably Softmax loss, Centre loss, and A-Softmax loss, on the IMDb-Face dataset. The experimental results showed that:

- Label flips severely deteriorate the performance of the Attention-56 model, more than outliers.
- A-Softmax, which is used to achieve a better performance on a clean dataset, becomes worse than Centre loss and Softmax in the high-noise region.
- Outliers seem to have a less abrupt effect on the recognition performance across all losses.

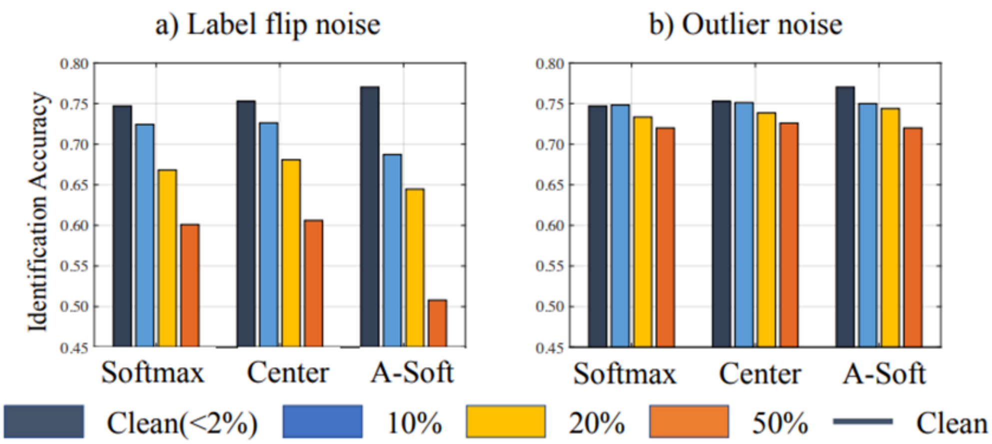


Figure 15: 1:1M rank-1 identification results on MegaFace benchmark: (a) introducing label flips to IMDb-Face; (b) introducing outliers to IMDb-Face.

○ *Comparison of IMDb-Face with other face datasets*

The performance for the IMDb-Face dataset is compared with the performance of other well-established face recognition training datasets, including CelebFaces, CASIA-WebFace, MS-Celeb-1M(v1), and MegaFace. *Table 9* illustrates the comparative performance results with the aforementioned datasets in terms of Rank-1 identification accuracy on the MegaFace dataset benchmark. It is observed that the IMDb-Face dataset is competitive as a training source despite its smaller size, which validates its cleanliness and effectiveness.

Table 9: Recognition performance when using different face datasets for training.

Dataset	#Iden.	#Imgs.	Rank-1 (%)		
			Softmax	Centre Loss	A-Softmax
CelebFaces	10k	0.20M	36.15	42.54	43.72
CASIA-WebFace	10.5k	0.49M	65.17	68.09	70.89
MS-Celeb-1M(V1)	96k	8.6M	71.70	73.82	73.99
MegaFace	670k	4.7M	64.32	64.71	66.95
IMDbFace	59k	1.7M	74.75	79.41	84.06

- **Strengths and Weaknesses:**
  - (+) The publication proposes a large-scale noise-controlled IMDb-Face dataset, which shows competitive performance compared to alternative well-established face recognition training datasets.
  - (-) The publication is based on human annotators to clean large-scale face datasets, which may impact the quality of the cleaned data.

3.3. Publications Assessing the Impact of Human-related Influencing Factors

This subsection will review in more detail the publications listed in *Table 4* addressing human-related influencing factors; each publication will be identified by the associated publication.

A. Deb, L. Best-Rowden, and A. K. Jain, "Face Recognition Performance under Ageing," IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, July 2017 [173]**.**

- **Influencing Factor(s):** Age.
- **Objectives:**
  - Quantify the influence of ageing on the performance of two state-of-the-art commercials off-the-shelf face recognition systems, notably (COTS-A and COTS-B. using two large-scale longitudinal datasets of mugshots (photographic portrait of a person taken when he/she is arrested by police), namely Pinellas County Sheriff's Office (PCSO) and Michigan State Police (MSP).
  - Assess the variation rate of the genuine scores caused by the elapsed time between gallery and query images.
  - Analyse the recognition performance degradation due to cohorts' biases such as ethnicity, gender, and the quality of the query images.

- **Methodology Description:** The methodology consists in analysing the recognition performance degradation due to elapsed time between query and gallery images using two multilevel statistical models:
    - **Level-1 model:** Expresses the genuine scores variations of each subject over time (within-subject change).
    - **Level-2 model:** Describes how genuine score variation differs across subjects (between-subject variation).

  In addition, to get meaningful results, the PCSO and MSP datasets are built following three criteria:

    - The selected subjects should have a sufficient number of images over a minimum of 5 years span (at least 5 for PCSO and 4 for MSP).
    - The acquisition of consecutive images of each subject are separated by at least one month.
    - The youngest subject is at least 18 years old.
- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results obtained with this publication are presented.
    - *Gender and race across age*

Figure 16 shows the population-mean trends for gender and race across age. Two main conclusions can be derived: (1) Differences due to race and gender cohorts are symmetric for COTS-A on PCSO and MSP datasets; (2) Race and gender effects over time are face recognition system-independent when COTS-A and COTS-B are both evaluated on the MSP dataset.

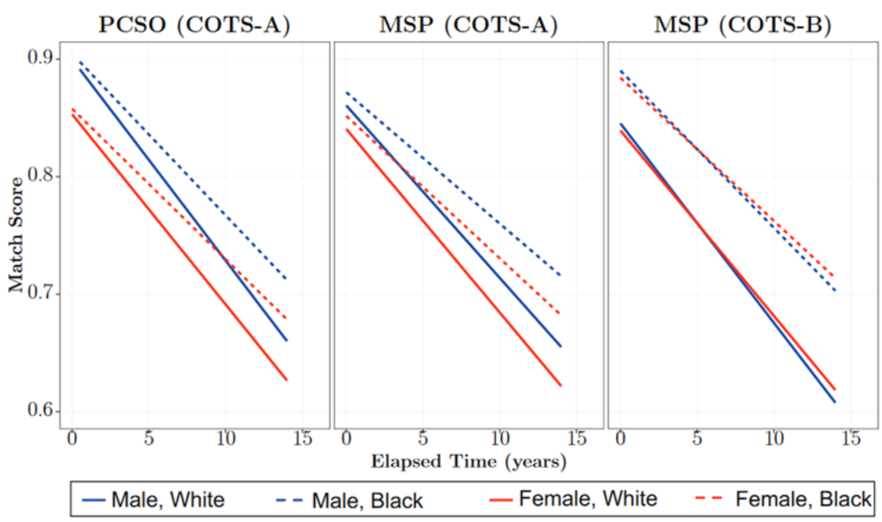


Figure 16: Population-mean trends in COTS-A and COTS-B genuine scores on PCSO dataset and MSP dataset for the four demographic groups in the datasets.

- **Strengths and Weaknesses:**
    - (+) The publication uses two largest longitudinal face datasets (facial images are taken over time), notably, PCSO [186], MSP.

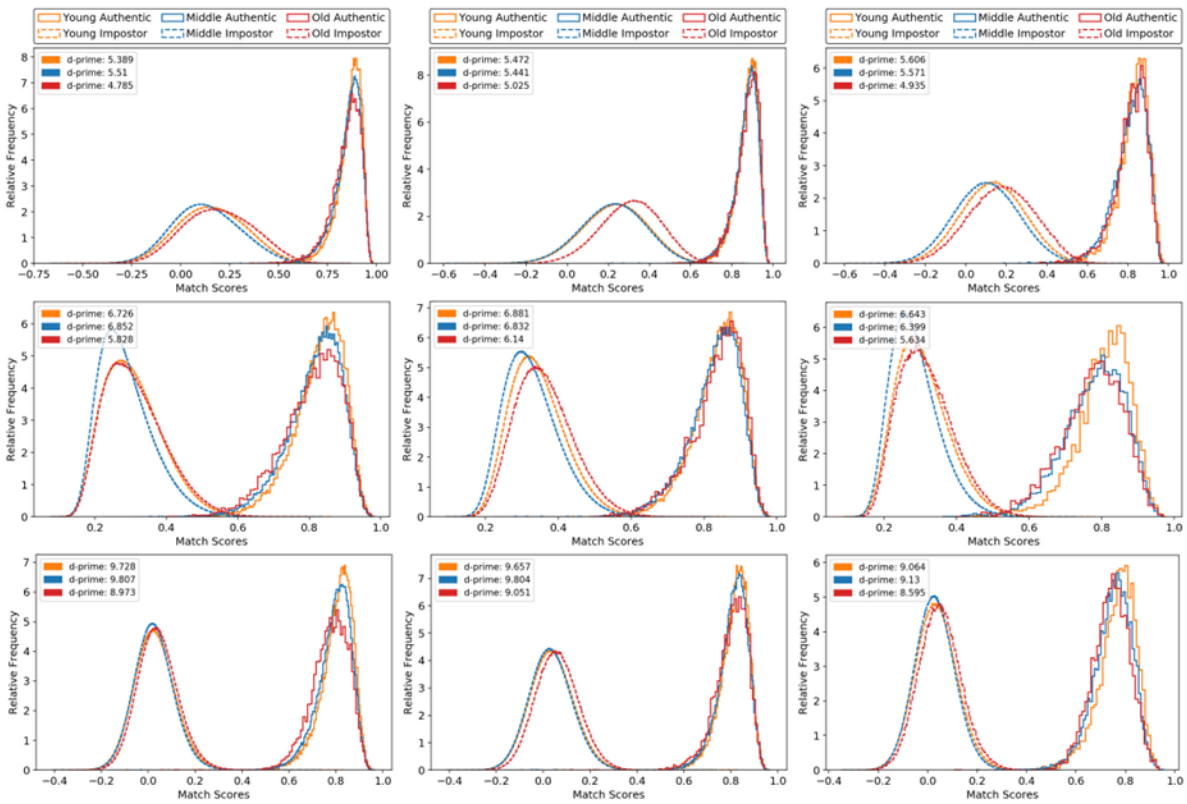
  - (-) The publication evaluates the face recognition systems under a verification scenario only; evaluation for an identification scenario would also be desirable.

B. V. Albiero, K. W. Bowyer, K. Vangara, and M. C. King, "Does Face Recognition Accuracy Get Better with Age? Deep Face Matchers Say No," ***IEEE Winter Conference on Applications of Computer Vision, Snowmass*****, CO, USA, March 2020** [67]**.**

- **Influencing Factor(s):** Age.
- **Objectives:** To investigate the recognition accuracy for three age groups, e.g. 16-29, 30-49, and 50-70, using three pre-trained deep CNN matchers with three different loss functions:
  - VGGFace2 (ResNet-50) trained on the VGGFace2 dataset with standard softmax loss.
  - FaceNet trained on MSCeleb-1M dataset with triplet loss.
  - ArcFace (ResNet-100) trained on MS-Celeb-1M V2 dataset with additive angular margin loss.
- **Methodology Description:** The methodology consists in evaluating and comparing three modern deep CNN face recognition models via the following strategy:
  - Investigate the role of the genuine and impostor score distributions on the accuracy variation across three age groups, e.g. 16-29, 30-49 and 50-70.
  - Assess the impact of elapsed time between genuine pairs.
  - Investigate the role of age groups on the effect of age variation between individuals in an impostor pair.
  - Assess the impact of balanced data in the training stage on the test accuracy.
- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results of the publication are presented.
  - *Impostor and genuine distributions*

  Figure 17 illustrates the impostor and genuine distributions for the three evaluated face recognition systems, e.g. FaceNet (top), VGGFace2 (middle), and ArcFace (bottom). From the nine plots it is possible to observe that the worst impostor distribution is obtained by the old age range followed by the young age range, while the best impostor distribution is obtained by the middle age range.

  On the other hand, it may be observed that the best genuine distribution is obtained by the young age range, which shows a higher peak of high-similarity scores, while the middle and old age ranges obtained the same genuine distribution.



(a) Whole dataset (b) African American males (c) Caucasian males

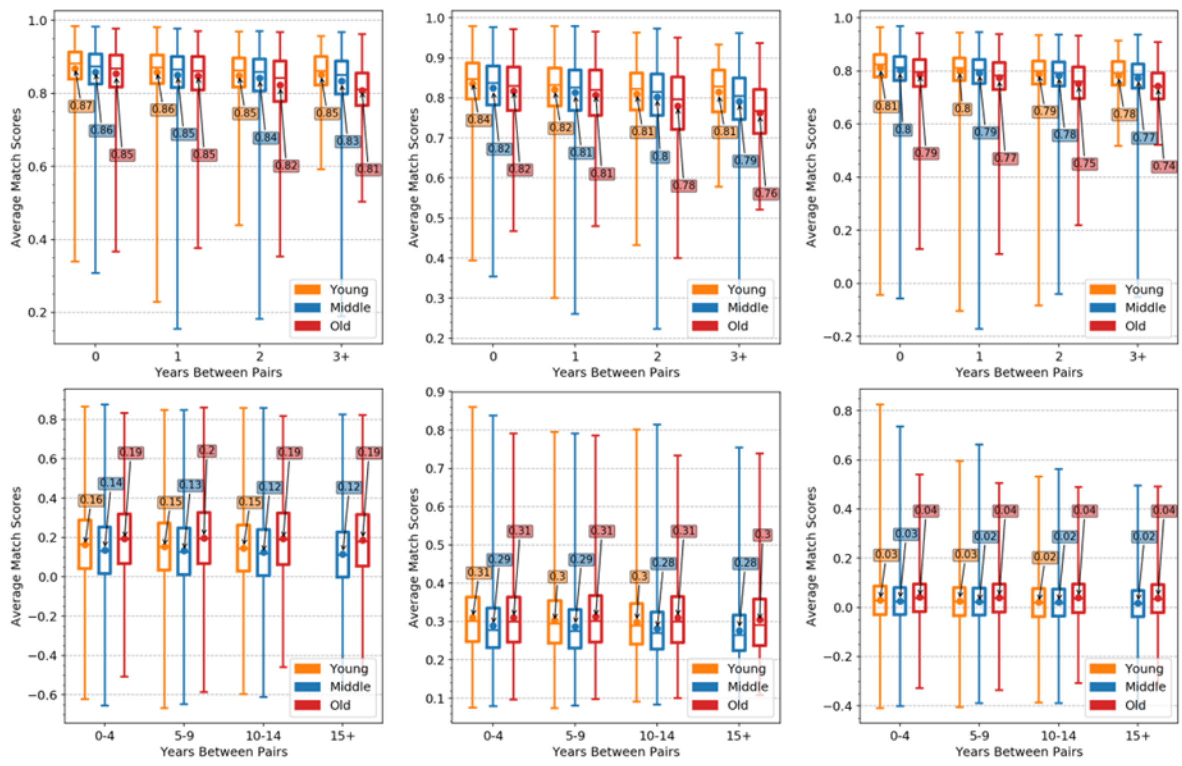
Figure 17: d-prime matching scores distributions for FaceNet (top); VGGFace2 (middle); and ArcFace (bottom).

- *Influence of elapsed time*

The age difference between the individual(s) in two images is represented by the average of the match score for impostor pairs (bottom) and genuine pairs (top) as illustrated in *Figure 18*.

For the genuine score, the results show that the highest similarity score is obtained for the younger age range, while the older age range has the lowest similarity scores.

For the impostor score plots, the results show that the older age range has the worst (highest) average similarity scores, while the middle age range generally has the best (lowest) average similarity scores.



(a) FaceNet (b) VGGFace2 (c) ArcFace

Figure 18: Match scores with increasing elapsed time between authentic (top) and impostors (bottom) pairs for the whole dataset.

- **Strengths and Weaknesses:**
  - (+) The publication reveals and proves a new finding, notably contrasting previous publications , stating that old age groups have lower recognition accuracy values than young age groups.
  - (+) The publication provides a list of convincing arguments regarding the reasons behind the findings of previous relevant publications.
  - (+) The testes are performed with two split datasets, e.g. African-American male and Caucasian male, to reduce racial bias in the age effect analysis.
  - (-) Due to the lower number of female subjects and images in the MORPH dataset (93 subjects and 286 images for African-American, and 34 subjects and 112 images for Caucasians), the subgroup tests are performed only for the two male subsets, which may not be reliable enough to study the age effect across gender.

C. **A. Peña, A. Morales, I. Serna, J. Fierrez, and A. Lapedriza, "Facial Expressions as a Vulnerability in Face Recognition," IEEE International Conference on Image Processing, Anchorage, AK, USA, September 2021** [56].

- **Influencing Factor(s):** Facial expressions.

- **Objectives:** To explore how facial expressions, e.g. happiness, sadness, anger, surprise, disgust, and fear impact the recognition performance and security vulnerability of face recognition technology.
- **Methodology Description:** The methodology consists in performing a series of experiments in controlled scenarios:
  - Using three state-of-the-art face recognition systems, notably VGG16, ResNet-50, and LResNet100E-IR, as features extractors and the Euclidean distance metric as features matcher.
  - Using the COTS Affectiva tool to classify the datasets, notably CFEE, CK+, CelebA, images into seven facial expressions (6 basic emotions plus neutral face).
  - Conducting the experiments inspired by the work [187] to study facial expression bias.
  - Analysing both genuine and impostor scores distributions by comparing each facial expression with the remaining facial expressions.
- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results of the publication are presented.

  Figure 19 illustrates the distributions of the genuine and impostor matching scores (pairs of faces) for the three face recognition systems, using the neutral expression as reference. It is observed that the genuine distributions are different among all the facial expressions for all three face recognition systems. In addition, it may be observed that the genuine distributions are clearly influenced by the facial expressions, notably more than the impostor distributions which barely change across expressions.

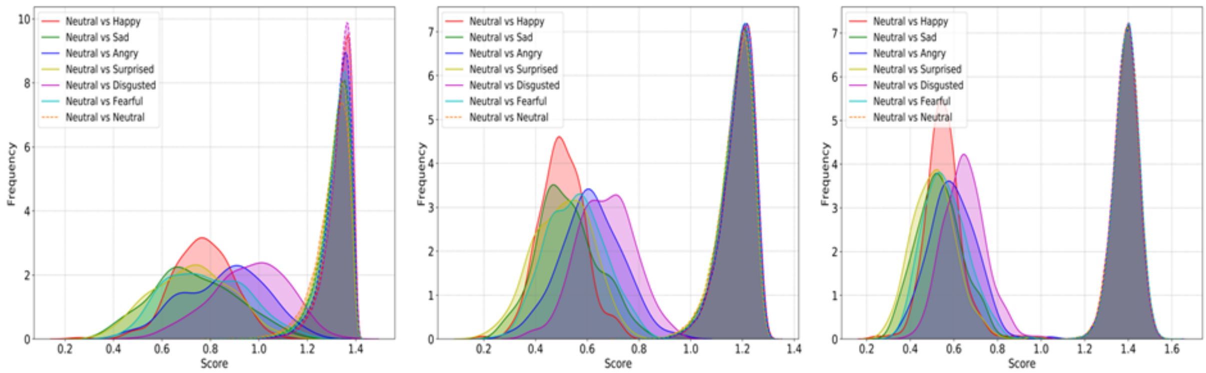


Figure 19: Genuine (continuous line) and impostor (dashed line) matching score distributions by facial expression on the CFEE database for the face matchers: (left) VGG16; (centre) ResNet-50; and (right) LResNet100E-IR.

- **Strengths and Weaknesses:**
  - (+) The publication uses the COTS Affectiva tool as a preprocessing stage to classify the datasets` images into seven facial expressions.
  - (-) The evaluated datasets include a variety of facial expressions, which results in a heterogeneous performance.

D. B. F. Klare, M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain, "Face Recognition Performance: Role of Demographic Information," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1789-1801, October 2012 [169]**.**

- **Influencing Factor(s):** Age, race, and gender
- **Objectives:** To analyse the influence of demographic factors, notably race (Black, White, Hispanic), gender (Female, Male), and age (18 to 30, 30 to 50, 50 to 70), on the recognition performance of six face recognition systems, notably three commercial (COTS-A, COTS-B, and COTS-C), two non-trainable (local binary pattern (LBP) and Gabor) and one trainable (spectrally sampled structural subspace features (4SF)) face recognition systems.
- **Methodology Description:** The methodology consists in performing three separate matching experiments for multiple cohorts, notably male, female, young, middle-aged, old, white, black, and Hispanic, using three separate training and test datasets:
  - **Experiment 1:** Measures the relative performance within the demographic cohort of each demographic factor for each commercial face recognition system. For example, on the gender demographic, this experiment measures the difference in recognition accuracy for commercial face recognition systems on males versus females.
  - **Experiment 2:** Measures the relative performance within the cohort for non-trainable face recognition models.
  - **Experiment 3:** Investigates the influence of the training dataset on the recognition performance. Thus, several versions of the 4SF face recognition models are trained on each demographic cohort. Next, the trained versions are used to separate the testing datasets from each cohort within each particular demographic. For example, this experiment helps to understand how much training exclusively on females improves performance on females and decreases performance on males.
- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results of this publication are presented.
  - *Gender demographic*

    Figure 20 shows that the three commercial face recognition systems perform worse on females than males (subfigures (a-c)). Similarly for the non-trainable models local binary pattern (LBP) and Gabor performed worse on the female cohort (subfigures (d-e)). The same outperformance of the male cohort over the female cohort is obtained with the 4SF face recognition system even when a gender-balanced training dataset is used (subfigure (h)).

  - *Age demographic*

    Figure 21 presents the recognition performance for the six face recognition systems on datasets separated by cohorts within the age demographic. It is observed that the lowest accuracy obtained for these six face recognition systems happens for the subjects in the age range between 18 and 30 (a-f). The COTS-A model performs similarly for both age ranges between 30 and 50 and 50 and 70. However, COTS-B shows higher accuracy for the age range between 30 and 50 than for the age range between 50 and 70, while COTS-C performed slightly better on the age range between 50 and 70 than the age range between 30 and 50.

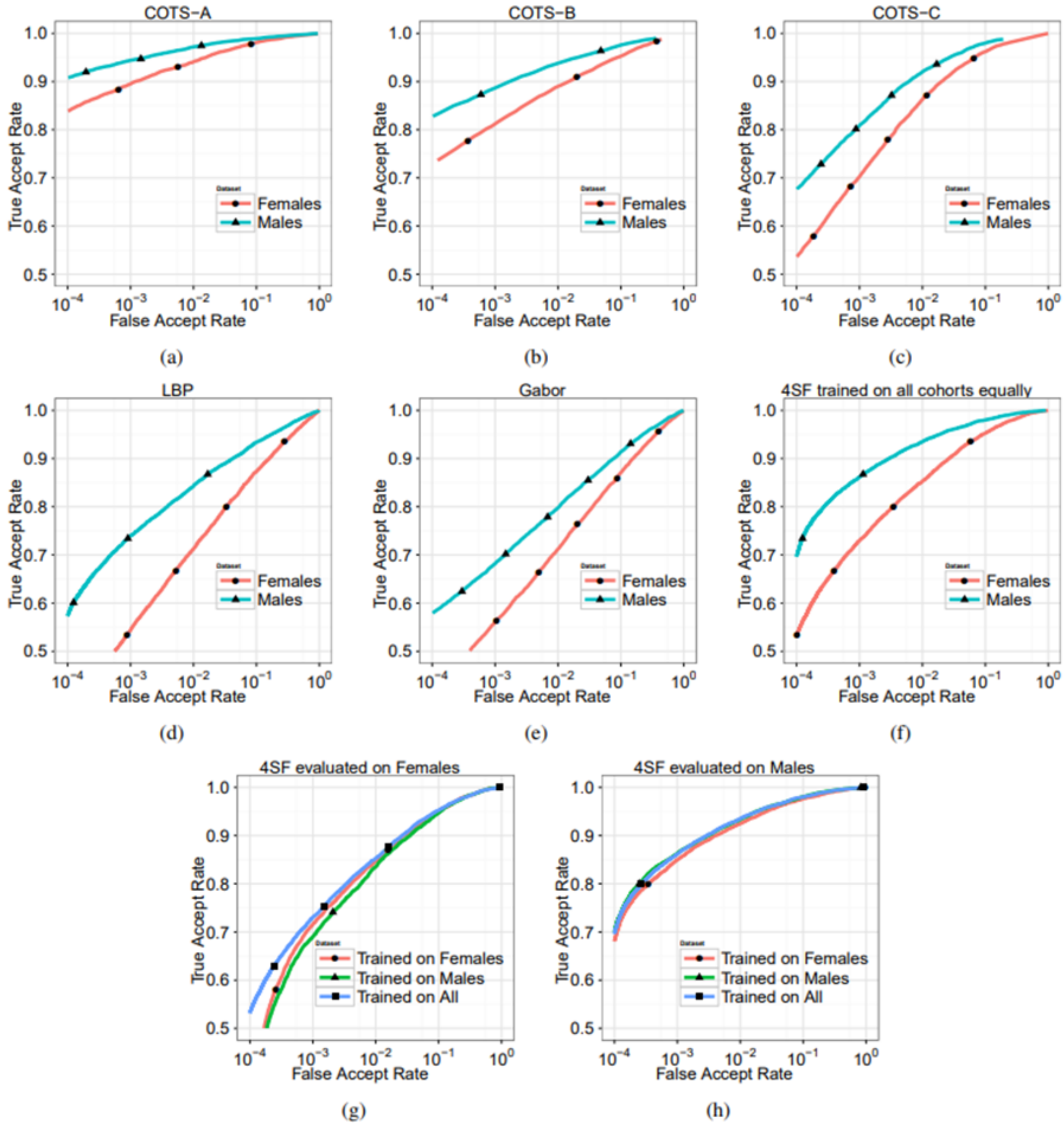


Figure 20: Performance of the six face recognition systems on datasets separated by cohorts within the gender demographic. (a) COTS-A; (b) COTS-B; (c) COTS-C; (d) Local binary patterns (non-trainable); (e) Gabor (non-trainable); (f) 4SF trained on equal number of samples from each gender; (g) 4SF algorithm (trainable) on the females' cohort: (h) 4SF algorithm (trainable) on the males' cohort.

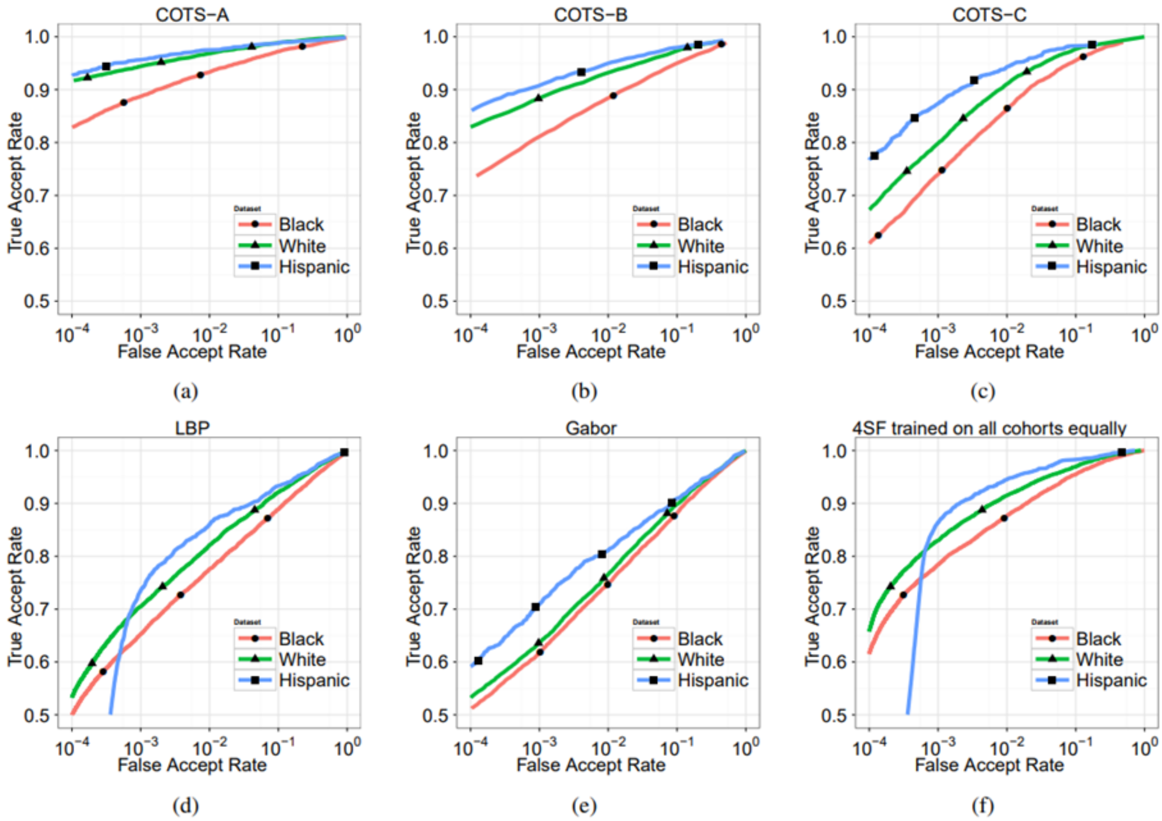


Figure 21: Performance of the six face recognition systems on datasets separated by cohorts within the age demographic. (a) COTS-A; (b) COTS-B; (c) COTS-C; (d) Local binary patterns (non-trainable); (e) Gabor (non-trainable); (f) 4SF trained on equal number of samples from each age; (g) 4SF system (trainable) on the ages 18 to 30 cohort; (h) 4SF algorithm (trainable) on the ages 30 to 50 cohort; (i) 4SF algorithm (trainable) on the ages 50 to 70 cohort.

- **Strengths and Weaknesses:**
  - (+) The publication simultaneously explores the influence of three demographic factors, notably race, gender and age, which helps to compare the interference between them.
  - (+) The publication assesses the impact of the three demographic factors for six face recognition systems, which provides a large-scale performance analysis.
  - (-) The evaluated commercial face recognition models are black boxes, which does not provide any insights into the demographic bias that occurred; instead, they only output a measure of similarity between the compared facial images.

E. P. Grother, M. Ngan, and K. Hanaoka, "Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects," *National Institute of Standards and Technology Interagency Internal Report 8280*, pp. 1-81, December 2019 [66]**.**

- **Influencing Factor(s):** Age, race, and gender
- **Objectives:** To study the utility and limitations of face recognition technology under demographic groups, e.g. age, race, and gender, by analysing the performance accuracy changes across these demographic groups.
- **Methodology Description:** The methodology consists in evaluating 106 pre-trained face recognition systems using four large photograph datasets collected in US governmental applications. The evaluation is performed in one-to-one verification and one-to-many identification scenarios by providing:
  - Details about the recognition process.
  - Notes where demographic effects could occur.
  - Descriptions of specific performance metrics and analyses.
  - Empirical results.
- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results achieved with this publication are presented.

  Figure 22 illustrates the cross-age group false match rate (FMR) for one of the evaluated face recognition systems (imperial_002) for six countries, namely Poland, Mexico, India, Kenya, Nigeria, and China.

  - ***Lower FMR for persons in different groups:*** The heatmaps show that for all the countries and both sexes, the comparison of facial images for different age groups obtains lower (better) FMR than for faces of the same age group.
  - ***Highest FMR in the oldest age group:*** For both sexes from all countries, the comparison of facial images of persons in the 65 years age range group or higher yields the highest FMR.
  - ***High FMR in the youngest age group:*** For both sexes, comparison of facial images of individuals in the age range between 12 and 20 generates the highest FMR.

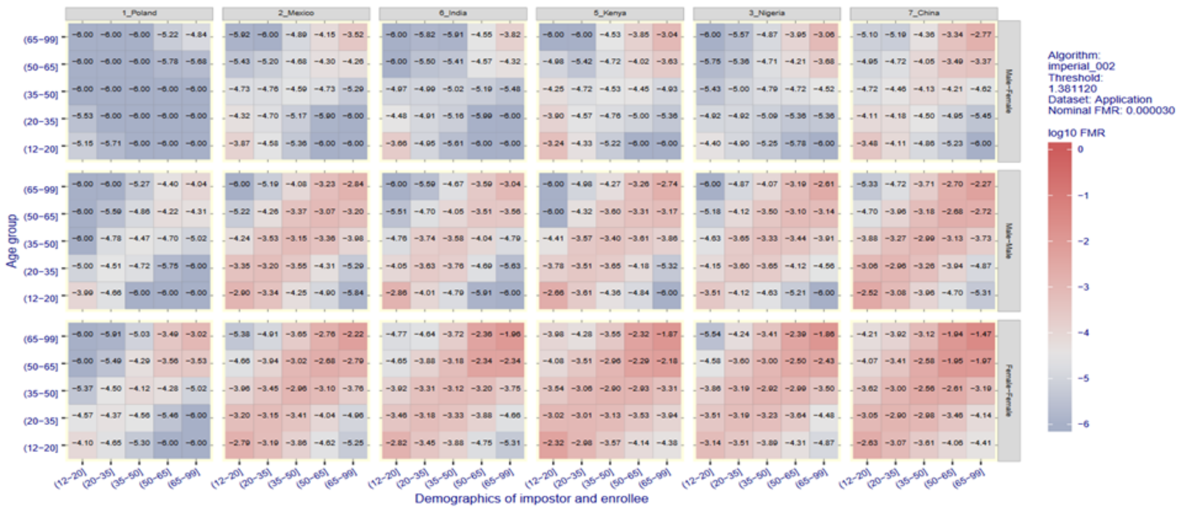


Figure 22: Cross-age false match rates (FMR) for imposters of the same sex from the age groups given on the respective axes. Each cell depicts FMR on a logarithmic scale.

- **Strengths and Weaknesses:**
  - (+) The publication performs face recognition evaluation under one-to-one verification and one-to-many identification scenarios.
  - (+) The publication uses real-life datasets from US governmental applications (e.g. domestic mugshots, application photographs, visa photographs, border crossing photographs), which helps in simulating realistic scenarios.
  - (-) The face recognition systems are tested as submitted by the developers without being refined, adapted or trained in a unified environment; this does not simulate the realistic scenario where the facial recognition system is adapted to the local customer's data.
  - (-) The publication does not make any effort to explain the technical reasons for the obtained results.

3.4. Publications Assessing the Impact of Deep Model-related Influencing Factors

This subsection will review in more detail the publications listed in *Table 4* addressing deep model-related influencing factors; each publication will be identified by the associated publication.

A. **G.-S. J. Hsu, H.-Y. Wu, and M. H. Yap, "A Comprehensive Study on Loss Functions for Cross-Factor Face Recognition," IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, June 2020** [181].

- **Influencing Factor(s):** Loss function, pose variation, age, occlusion, image resolution.
- **Objectives:** To evaluate the state-of-the-art loss functions considered in deep face recognition under four influencing factors, namely pose variation, age, occlusion and image resolution in order to clarify the impact of these factors on loss functions.
- **Methodology Description:** The methodology consists in evaluating and comparing the performance for five loss functions, e.g. Centre Loss, Marginal Loss, Angular Softmax Loss, Large Margin Cosine Loss, and Additive Angular Margin Loss using the same CNN architecture (ResNet-100) via the following process:
  - In the training stage, the ResNet-100 is trained on the MS-Celeb-1M dataset to learn facial features under the supervision of different loss functions.
  - In the test stage, the learned facial features are extracted and compared using the cosine similarity score metric in a verification scenario. In addition, the tests are performed on five different datasets considering the influencing factors:

  - IARPA Janus Benchmark–B (IJB-B) and IARPA Janus Benchmark–C (IJB-C) are used for pose variation.
  - FG-Net Ageing Database (FG-Net) is used for age.
  - AR Face Database (AR Face) is used for occlusion.
  - Surveillance Cameras Face Database (SCface) is used for image resolution.
- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results of the publication are presented.
  - *Pose variation*

Table 10 and *Table 11* illustrate the verification rate in terms of true accept rate (TAR) on the IJB-B and IJB-C datasets, respectively. It may be observed that ArcFace with the Additive Angular Margin loss outperforms the remaining loss functions on both datasets, showing its ability to support facial pose variations. Arcface performance is followed by CosFace with the Large Margin Cosine loss, then SphereFace with the Angular Softmax loss, then Marginal loss and finally Center loss.

Table 10: Verification rate (in % TAR) for state-of-the-art loss functions tested on the IJB-B dataset.

Model	**TAR (%) @FAR**			**AUC (%)**
	0.01%	0.001%	0.0001%	
Centre Loss	88.9	80.2	68.3	98.7
Marginal Loss	90.1	82.5	72.6	98.9
SphereFace	94.3	91.4	81.3	99.6
CosFace	95.9	92.6	89.1	99.4
ArcFace	**97.4**	**94.9**	**92.6 (94.2)**	**99.5**

Table 11: Verification rate (in % TAR) for state-of-the-art loss functions tested on the IJB-C dataset.

Model	**TAR (%) @FAR**			**AUC (%)**
	0.01%	0.001%	0.0001%	
Centre Loss	90.4	83.5	74.1	98.9
Marginal Loss	92.6	87.4	79.9	99.1
SphereFace	96.8	91.7	86.1	99.6
CosFace	96.8	93.3	90.2	99.5

ArcFace	**97.9**	**96.1**	**93.6 (95.6)**	**99.6**

- *Age variation*

  Table 12 shows the cross-age performance on the FG-Net dataset. It may be observed that ArcFace outperforms the remaining loss functions followed by SphereFace, and then CosFace. Additionally, it is found that the cross-age performance is the lowest among the four influencing factors with a clear gap, which may open new future research directions in this field.

Table 12: Verification rate (in % TAR) for state-of-the-art loss functions tested on the FG-Net dataset.

Model	**TAR (%) @FAR**			**AUC (%)**
	0.1%	0.01%	0.001%	
SphereFace	86.1	65.6	43.6	95.1
CosFace	84.1	56.6	33.7	94.2
ArcFace	89.7	71.3	52.3	96.3

- **Strengths and Weaknesses:**
  - (+) The publication evaluates the loss functions under different influencing factors, which provides information on how to better use the available loss functions and better conceive new ones.
  - (+) The loss functions are evaluated in a unified environment (e.g. network architecture settings, training/testing datasets, etc.) to report valid and conclusive assessments.
  - (-) The publication evaluates the loss functions only under four influencing factors, e.g. pose variation, age, occlusion, image resolution, without considering other important factors such as gender and race.

B. J. Coe and M. Atay, "Evaluating Impact of Race in Facial Recognition across Machine Learning and Deep Learning Algorithms," Computers, vol. 10, no. 9, pp. 1-24, September 2021 [183]**.**

- **Influencing Factor(s):** Race.
- **Objectives:** To explore the influence of racial differences on face recognition by pursuing the following strategy:
  - Assess the racial bias across five ML and three DL-based algorithms.
  - Use racially balanced and unbalanced datasets.
  - Analyse the experimental findings and compare the performance, miss rate, and accuracies of the tested systems.
  - Report the systems that most minimise the racial bias.

- **Methodology Description:** The methodology consists in performing two evaluation steps, notably evaluation plan and evaluation procedure:
    - **Evaluation Plan:** Presents the procedure pursued to give an in-depth evaluation by using:
        - Multiple ML and DL systems.
        - Three generated sub-datasets sourced from the FERET dataset that emphasise various races.
        - With 24 subjects, 12 facial images per subject, the sub-datasets are weighted as illustrated in *Table 13*.

Table 13: FERET dataset splits.

	Balanced	Dominant Set 1	Dominant Set 2
Black subjects	12	16	8
White subjects	12	8	16

Since each subject has only 12 samples, three sorts of experiments are performed:

- **Experiment 1:** Takes 11 images for training and 1 image for testing (original split).
- **Experiment 2:** Takes 8 images for training and 1 image for testing (after removing 3 images to eliminate the impact of occlusion, illumination and expression).
- **Experiment 3:** Takes 22 images for training and 1 image for testing (after applying data augmentation on the initial 11 images).
- **Evaluation Procedure:** Regards the global process pursued to evaluate the performance of the tested systems. The process is performed as follows:
    - Select the system.
    - Select the dataset.
    - Measure the metrics.
    - Repeat steps 2 and 3 for each dataset with the selected system.
    - Repeat the entire process for each system.

- **Modelling the Impact of Influencing Factors:** In the following, some of the experimental results from the publication are presented.
    - *Racial bias across ML algorithms on balanced dataset*

Figure 23 illustrates the recognition accuracy of five ML-based face recognition systems on the balanced dataset using experiment 1. It is clear that the support vector classifier (SVC) outperforms the remaining ML-based algorithms. However, the accuracy barely surpassed 80%, which is well below the industry standard and user expectations. Another remarkable finding is the other algorithms` low level performance, particularly the k-nearest neighbours (KNN) algorithm that barely surpasses the 50% performance.

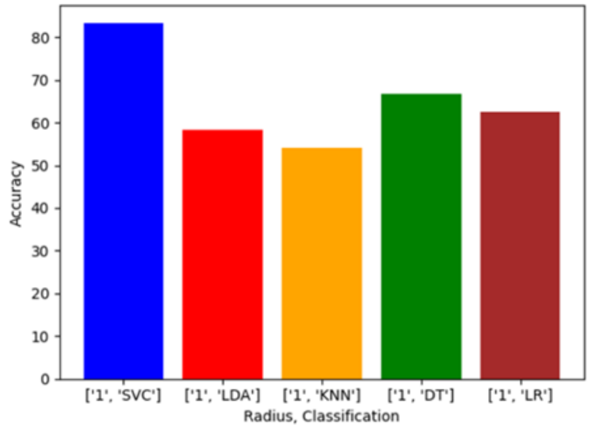


Figure 23: ML recognition accuracies with balanced dataset and 8 training images.

- *Racial bias across ML algorithms on imbalanced dataset*

The same experiment is performed using the two imbalanced datasets (Dominant Set 1 and Dominant Set 2). The results obtained for Dominant Set 1 (see Figure 24(black)) show similar results to *Figure 23* compared to Dominant Set 2 (see *Figure 24* (white)) that yields slightly better performance when the KNN algorithm is applied, demonstrating the importance of the dataset.

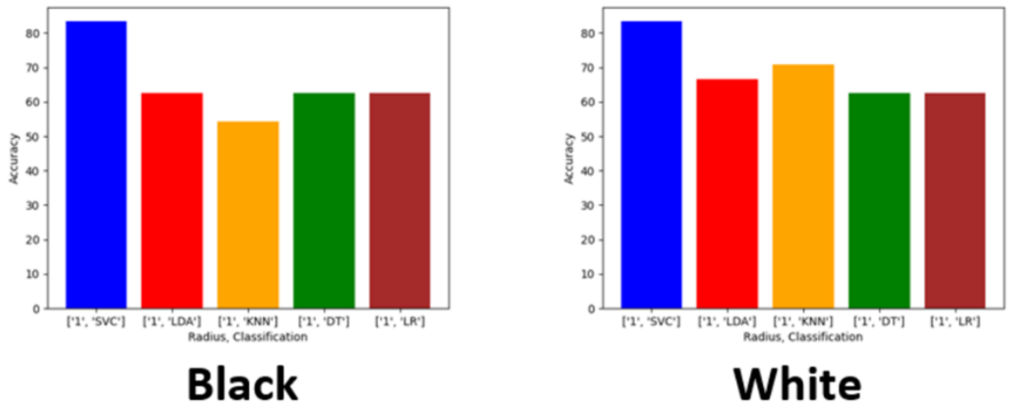


Figure 24: ML accuracies with imbalanced datasets and 8 training images.

- **Strengths and Weaknesses:**
  - (+) The publication evaluates multiple DL and ML face recognition systems.
  - (+) The publication uses various performance metrics such as accuracy, miss rates, precision, recall, and F1.
  - (-) The publication uses only the FERET dataset.
  - (-) The publication conducts the performance evaluation only for white and black cohorts.

3.5. Summary

In Section 3, a review of the state-of-the-art publications assessing the impact of the influencing factors on AI-based face recognition systems is presented. First, a set of publications exploring this research area is introduced. Then a subset of the more relevant publications is identified and classified according to the influencing factors classes. Finally, each of the identified publications is further described and analysed.

As shown in this section, a significant number of publications has already been developed to assess the impact of the influencing factors on facial recognition. However, more attention is given to some influencing factors (e.g. gender, race, age, occlusion, etc.) than others (e.g. image compression, plastic surgery, facial make-up, etc.) that are of fundamental importance and deserve more investigation. Therefore, providing a sufficient and equal level of importance to all the influencing factors might answer some open questions regarding the decision-making process of AI-based face recognition technology.

4. Conclusion

Over recent years, the attempts to understand and explain the facial recognition pipelines behaviour has had significant attention and progress, especially the black box face recognition pipelines based on AI tools. Identifying the reasons responsible for the AI-based face recognition pipelines final decision is very useful for the builders of such pipelines to explain the decision making, improve the pipelines' performance, and ensure their fairness.

Recent AI-based face recognition explainability publications aim to explore the influencing factors and express their effect on the overall performance of AI-based face recognition pipelines. This is particularly relevant for DL-based face recognition systems, which differ from shallow AI algorithms such as decision trees and linear regression that are self-explanatory as the decision boundary used for the recognition can be visualised via the model parameters.

In this context, this report offers a comprehensive overview of the influencing factors impacting the facial recognition process. First, a list of the various influencing factors is identified. More precisely, a novel taxonomy of these influencing factors is introduced, namely, (1) *data quality-related factors*, (2) *human-related factors*, and (3) *deep model-related factors*. Then, a survey of the state-of-the-art publications assessing the impact of the influencing factors on deep face recognition systems is presented. In addition, a subset of the more relevant publications is identified and further analysed.

The survey of the literature publications pointed out a considerable number of challenging influencing factors. The analysis of these publications has shown that despite the number of publications proposed to assess the effect of the influencing factors on deep face recognition, the investigation does not cover well enough all the influencing factors. Most of the publications have predominantly focused on the influence of some specific factors considered as challenging variations such as age, race, illumination and facial expression. Other influencing factors (e.g. image compression, plastic surgery, facial make-up etc.) are less investigated or have not yet been studied. Therefore, more effort needs to be dedicated to evaluate the recognition performance of deep face recognition pipelines under such influencing factors.

To sum up, AI-based face recognition explainability is a wide and yet rather unexplored research area. Therefore, there is still a need to make the explainability mechanisms and models more holistic by assessing the impact of a wide range of influencing factors, which opens new directions for future works. Among the influencing factors that deserve to be assessed, image compression looks rather important since more and more the faces to be recognized are decoded after compression. In fact, compression is typically applied to efficiently transfer the acquired facial images to a distant device for identification or verification. In addition, compression is commonly used to store facial images in limited capacity devices such as smart cards and mobile applications' databases. However, using decompressed facial images for face recognition might lead to recognition performance degradation, depending on the compression level. Thus, studying and analysing the influence of image compression on AI-based face recognition systems seems like an inevitable line of research, especially considering the more recent image compression solutions. e.g. the JPEG XL standard and the emerging JPEG AI solutions.

References

[1] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, “MagFace: A Universal Representation for Face Recognition and Quality Assessment,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, pp. 14225–14234, June 2021.

[2] Y. Shi, X. Yu, K. Sohn, M. Chandraker, and A. K. Jain, “Towards Universal Representation Learning for Deep Face Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 6817-6826, June 2020.

[3] M. Ngan, P. Grother, and K. Hanaoka, “Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face Recognition Accuracy with Face Masks using Post-COVID-19 Algorithms,” NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, pp. 1-104, November 2020.

[4] M. T. H. Fuad et al., “Recent Advances in Deep Learning Techniques for Face Recognition,” IEEE Access, vol. 9, pp. 99112-99142, July 2021.

[5] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-level Performance in Face Verification,” IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, pp. 1701-1708, June 2014.

[6] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 815-823, June 2015.

[7] B. Moghaddam, T. Jebara, and A. Pentland, “Bayesian Face Recognition,” Pattern Recognition, vol. 33, no. 11, pp. 1771-1782, November 2000.

[8] C. Geng and X. Jiang, “SIFT Features for Face Recognition,” 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, China, pp. 598-602, August 2009.

[9] G. Du, F. Su, and A. Cai, “Face Recognition using SURF Features,” Pattern Recognition and Computer Vision, Yichang, China, pp. 7496-749628, October 2009.

[10] S. Chang, D. Xiaoqing, and F. Chi, “Histogram of the Oriented Gradient for Face Recognition,” Tsinghua Science and Technology, vol. 16, no. 2, pp. 216-224, April 2011.

[11] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face Recognition with Local Binary Patterns,” European Conference on Computer Vision, Prague, Czech Republic, pp. 469-481, May 2004.

[12] M. Turk and A. Pentland, “Eigenfaces for Recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, January 1991.

[13] B. Liu et al., “Fair Loss: Margin-Aware Reinforcement Learning for Deep Face Recognition,” IEEE International Conference on Computer Vision, Seoul, Korea, pp. 10052-10061, November 2019.

[14] M. Wang and W. Deng, “Mitigate Bias in Face Recognition using Skewness-Aware Reinforcement Learning,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 9322-9331, June 2020.

[15] . Hardesty, “Study Finds Gender and Skin-type Bias in Commercial Artificial-Intelligence Systems,” February, 2018, Accessed on: October 05, 2021, [Online]. Available:

https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212
[16] J. Murdock, “Amazon Face Recognition Tech Matches 28 Members of Congress with Mugshots,” June, 2018, Accessed on: October 05, 2021. [Online]. Available: https://www.newsweek.com/amazons-face-recognition-tool-matches-28-members-congress-criminal-mugshots-1044850
[17] M. Sharif, F. Naz, M. Yasmin, M. A. Shahid, and A. Rehman, “Face Recognition: A Survey,” Journal of Engineering Science and Technology Review, vol. 10, no. 2, pp. 166–177, June 2017.
[18] U. Ali, Elmahmudi Hassan, “Deep Face Recognition Using Imperfect Facial Data,” Future Generation Computer Systems, vol. 99, pp. 213-225, October 2019.
[19] D. Zeng, R. Veldhuis, and L. Spreeuwers, “A Survey of Face Recognition Techniques Under Occlusion,” arXiv:2006.11366v1, 2020.
[20] H. Lingxiao, L. Haiqing, Z. Qi, and S. Zhenan, “Dynamic Feature Learning for Partial Face Recognition,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 7054-7063, June 2018.
[21] S. Lingxue, G. Dihong, L. Zhifeng, L. Changsong, and L. Wei, “Occlusion Robust Face Recognition Based on Mask Learning with Pairwise Differential Siamese Network,” IEEE/CVF International Conference on Computer Vision, Seoul, Korea, pp. 773-782, November 2019.
[22] X. Xiang, S. Nikolaos, and A. K. Ioannis, “On Improving the Generalization of Face Recognition in the Presence of Occlusions,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, pp. 3470-3480, June 2020.
[23] L. Yong, Z. Jiabei, S. Shiguang, and C. Xilin, “Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism,” IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2439-2450, May 2019.
[24] J. Vidit and L.-M. Erik, “FDDB: A Benchmark for Face Detection in Unconstrained Settings,” University of Massachusetts, Amherst, pp. 1-11, 2010.
[25] A. Kortylewski, A. Schneider, T. Gerig, B. Egger, A. Morel-Forster, and T. Vetter, “Training Deep Face Recognition Systems with Synthetic Data,” arXiv:1802.05891v1, 2018.
[26] D. Lingshuang and H. Haifeng, “Nuclear Norm Based Adapted Occlusion Dictionary Learning for Face Recognition with Occlusion and Illumination Changes,” Neurocomputing, vol. 340, pp. 133-144, May 2019.
[27] Z. Zijian and Y. Min, “Illumination Invariant Face Recognition by Expected Patch Log Likelihood,” IEEE International Symposium on Signal Processing and Information Technology, Louisville, KY USA, pp. 1-4, December 2020.
[28] Z. Wuming, Z. Xi, M. Jean-Marie, and C. Liming, “Improving Shadow Suppression for Illumination Robust Face Recognition,”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 3, pp. 611-624, March 2019.
[29] D. Sahil and P. V. Virendra, “A Novel Approach of Face Recognition Using Optimized Adaptive Illumination-Normalization and KELM,” Arabian Journal for Science Engineering, vol. 45, pp. 9977-9996, May 2020.

[30] S. Terence, B. Simon, and B. Maan, “The CMU Pose, Illumination, and Expression (PIE) Database,” 5th IEEE International Conference on Automatic Face and Gesture Recognition, WA, DC, USA, pp. 53-58, May 2002.

[31] A. R. Shinwari, I. Kosunen, A. A. Alariki, and S. A. A. Naji, “A Comparative Study of Face Recognition under Pose Variation,” 11th International Conference on Computer Modelling and Simulation, North Rockhampton QLD, Australia, pp. 137–141, January 2019.

[32] H. A. Le and I. A. Kakadiaris, “UHDB31: A Dataset for Better Understanding Face Recognition Across Pose and Illumination Variation,” IEEE International Conference on Computer Vision Workshops, Venice, Italy, pp. 2555-2563, October 2017.

[33] M. He, J. Zhang, S. Shan, M. Kan, and X. Chen, “Deformable Face Net for Pose Invariant Face Recognition,” Pattern Recognition, vol. 100, pp. 107-113, April 2020.

[34] J. Zhao et al., “Towards Pose Invariant Face Recognition in the Wild,” IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 2207–2216, June 2018.

[35] A. Savran et al., “Bosphorus Database for 3D Face Analysis,” European Workshop on Biometrics and Identity Management, Roskilde, Denmark, pp. 47-56, May 2008.

[36] L. Pei, P. Loreto, M. Domingo, and F. Patrick, “On Low-Resolution Face Recognition in the Wild: Comparisons and New Techniques,” IEEE Transactions on Information Forensics Security, vol. 14, no. 8, pp. 2000-2012, January 2019.

[37] M. Tomasz, C. Agata, W. Radoslaw, P. Marianna, and D. Adam, “Influence of Low Resolution of Images on Reliability of Face Detection and Recognition,” Multimedia Tools and Applications, vol. 74, pp. 4329-4349, June 2015.

[38] O. A. Aghdam, B. Bozorgtabar, J. Hazım Kemal Ekenel, and E.-P. Thiran, “Exploring Factors for Improving Low Resolution Face Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, pp. 2363-2370, June 2019.

[39] Y. Fuwei, Y. Wenming, G. Riqiang, and L. Qingmin, “Discriminative Multidimensional Scaling for Low-Resolution Face Recognition,” IEEE Signal Processing Letters, vol. 25, no. 3, pp. 388–392, March 2018.

[40] V. Reza Khazaie, N. Bayat, and Y. Mohsenzadeh, “Multi Scale Identity-Preserving Image-to-Image Translation Network for Low-Resolution Face Recognition,” arXiv:2010.12249v2, 2021.

[41] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, “Advanced High Dynamic Range Imaging: Theory and Practice,” A. K. Peters/CRC Press, 1st edition, February 2011.

[42] R. Suma, K. Debattista, D. Watson, E. Blagrove, and A. Chalmers, “Subjective Evaluation of High Dynamic Range Imaging for Face Matching,” IEEE Transactions on Emerging Topics in Computing, vol. 9, pp. 2042-2052, December 2019.

[43] P. Korshunov, M. V. Bernardo, A. M. G. Pinheiro, and T. Ebrahimi, “Impact of Tone-mapping Algorithms on Subjective and Objective Face Recognition in HDR Images,” The Fourth International Workshop on Crowdsourcing for Multimedia, Brisbane Australia, pp. 39-44, October 2015.

[44] E. O. Ge, K. Debattista, R. Muhkerjee, and A. Chalmers, “Exploring Face Recognition under Complex Lighting Conditions with HDR Imaging,” Computer

Graphics and Visual Computing, Bournemouth United Kingdom, pp. 49-56, September 2016.
[45] M. A. Turk and A. P. Pentland, “Face Recognition Using Eigenfaces,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Maui, HI, USA, June 1991.
[46] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 711-720, July 1997.
[47] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 2037-2041, October 2006.
[48] C. Guang Yi, “An Experimental Study for the Effects of Noise on Face Recognition Algorithms Under Varying Illumination,” Multimedia Tools and Applications, vol. 78, pp. 26615-26631, June 2019.
[49] Y. Ding, Y. Cheng, X. Cheng, B. Li, X. You, and X. Yuan, “Noise-Resistant Network: A Deep-Learning Method for Face Recognition under Noise,” EURASIP Journal on Image and Video Processing, vol. 44, pp. 1-14, June 2017.
[50] Z. Jinhua, Q. Xiulian, and S. Shaopei, “Image Processing Effects on the Deep Face Recognition System,” Mathematical Biosciences Engineering, vol. 18, no. 2, pp. 1187-1200, January 2021.
[51] N. Bian, F. Liang, H. Fu, and B. Lei, “A Deep Image Compression Framework for Face Recognition,” 2nd China Symposium on Cognitive Computing and Hybrid Intelligence, Xi'an, Shanxi, China, pp. 99-104, September 2019.
[52] Z. Chen and T. He, “Learning Based Facial Image Compression with Semantic Fidelity Metric,” Neurocomputing, vol. 338, pp. 16-25, April 2019.
[53] C. Rathgeb, K. Bernardo, N. E. Haryanto, and C. Busch, “Effects of Image Compression on Face Image Manipulation Detection: A Case Study on Facial Retouching,” IET Biometrics, vol. 10, no. 3, pp. 342-355, March 2021.
[54] G. W. Quinn and P. J. Grother, “Performance of Face Recognition Algorithms on Compressed Images,” NIST Interagency Report 7830, National Institute of Standards and Technology, pp.1-35, December 2011.
[55] M. Phankokkruad and P. Jaturawat, “Influence of Facial Expression and Viewpoint Variations on Face Recognition Accuracy by Different Face Recognition Algorithms,” 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Kanazawa, Japan, pp. 231–237, June 2017.
[56] A. Peña, A. Morales, I. Serna, J. Fierrez, and A. Lapedriza, “Facial Expressions as a Vulnerability in Face Recognition,” arXiv:2011.08809v2, 2021.
[57] F. A. Pavel and E. Iordănescu, “The influence of facial expressions on recognition performance in facial identity,” Procedia - Social and Behavioral Sciences, vol. 33, pp. 548-552, December 2012.
[58] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv:1409.1556v6, 2015.
[59] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770-778, June 2016.

[60] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 4690–4699, June 2019.
[61] S. Du, Y. Tao, and A. M. Martinez, “Compound Facial Expressions of Emotion,” National Academy of Sciences, vol. 15, pp. 1454-1462, March 2014.
[62] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, SF, CA, USA, August, pp. 94-101, June 2010.
[63] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep Learning Face Attributes in the Wild,” IEEE International Conference on Computer Vision, Santiago, Chile, pp.1-11, December 2014.
[64] G. Yandong, Z. Lei, H. Yuxiao, H. Xiaodong, and G. Jianfeng, “MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition,” European Conference on Computer Vision, Amsterdam, The Netherlands, pp. 87-102, October 2016.
[65] N. C. Ebner, M. Riediger, and U. Lindenberger, “FACES-A database of facial expressions in young, middle-aged, and older women and men: Development and validation,” Behavior Research Methods, vol. 42, pp. 351–362, February 2010.
[66] P. Grother, M. Ngan, and K. Hanaoka, “Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects,” National Institute of Standards and Technology Interagency Internal Report 8280, pp. 1-81, December 2019.
[67] V. Albiero, K. W. Bowyer, K. Vangara, and M. C. King, “Does Face Recognition Accuracy Get Better with Age? Deep Face Matchers Say No,” IEEE Winter Conference on Applications of Computer Vision, Snowmass, CO, USA, pp. 250-258, March 2020.
[68] Z. Huang, J. Zhang, H. Shan, and S. Key, “When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework,” arXiv:2103.01520v2, 2021.
[69] H. Wang, D. Gong, Z. Li, and W. Liu, “Decorrelated Adversarial Learning for Age-Invariant Face Recognition,” IEEE Computer Society Conference on Computer Vision Pattern Recognition, Long Beach, California, USA, pp. 3522-3531, June 2019.
[70] W. Yitong et al., “Orthogonal Deep Features Decomposition for Age-invariant Face Recognition,” European Conference on Computer Vision, Munich, Germany, pp. 738–753, September, 2018.
[71] C. Xu, Q. Liu, and M. Ye, “Age Invariant Face Recognition and Retrieval by Coupled Auto-encoder Networks,” Neurocomputing, vol. 222, pp. 62-71, January 2017.
[72] Y. Wen, Z. Li, and Y. Qiao, “Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 4893-4901, June 2016.
[73] A. Lanitis, C. J. Taylor, and T. F. Cootes, “Toward Automatic Simulation of Ageing Effects on Face Images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442-455, August 2002.
[74] A. W. Rawls and K. Ricanek Jr., “MORPH: Development and Optimization of a Longitudinal Age Progression Database,” European Workshop on Biometrics and Identity Management, Madrid, Spain, pp. 17–24, September 2009.

[75] B.-C. Chen, C.-S. Chen, and W. H. Hsu, “Face Recognition and Retrieval Using Cross-Age Reference Coding with Cross-Age Celebrity Dataset,” IEEE Transactions on Multimedia, vol. 17, no. 6, pp. 804-815, April 2015.

[76] T. Zheng, W. Deng, and J. Hu, “Cross-Age LFW: A Database for Studying Cross-Age Face Recognition in Unconstrained Environments,” arXiv:1708.08197v1, 2017.

[77] R. S. Malpass and J. Kravitz, “Recognition for Faces of Own and Other Race,” Journal of Personality and Social Psychology, vol. 13, no. 4, pp. 330-334, December 1969.

[78] A. J. O’Toole, K. Deffenbacher, H. Abdi, and J. C. Bartlett, “Simulating the ‘Other-race Effect’ as a Problem in Perceptual Learning,” Connection Science, vol. 3, no. 2, pp. 163-178, October 2007.

[79] P. Terhörst et al., “A Comprehensive Study on Face Recognition Biases Beyond Demographics,” arXiv:2103.01592v1, 2021.

[80] C. M. Cook, J. J. Howard, Y. B. Sirotin, J. L. Tipton, and A. R. Vemury, “Demographic Effects in Facial Recognition and Their Dependence on Image Acquisition: An Evaluation of Eleven Commercial Systems,” IEEE Transactions on Biometrics, Behavior and Identity Science, vol. 1, pp. 32–41, February 2019.

[81] K. Krishnapriya, K. Vangara, M. C. King, V. Albiero, and K. Bowyer, “Characterizing the Variability in Face Recognition Accuracy Relative to Race,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, pp. 2278–2285, June 2019.

[82] I. D. Raji and J. Buolamwini, “Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products,” AAAI/ACM Conference on AI, Ethics and Society, Honolulu HI, USA, pp. 429–435, January 2019.

[83] P. J. Phillips, F. Jiang, A. Narvekar, J. Ayyad, and A. J. O’Toole, “Another-race Effect for Face Recognition Algorithms,” ACM Transactions on Applied Perception, vol. 8, no. 2 pp. 1–11, January 2011.

[84] V. Albiero and K. W. Bowyer, “Is Face Recognition Sexist? No, Gendered Hairstyles and Biology Are,” arXiv:2008.06989v1, 2020.

[85] C. M. Cook, J. J. Howard, Y. B. Sirotin, and J. L. Tipton, “Fixed and Varying Effects of Demographic Factors on the Performance of Eleven Commercial Facial Recognition Systems,” IEEE Transactions on Biometrics Behavior and Identity Science, vol. 99, pp. 32-41, February 2019.

[86] V. Albiero, K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer, “Analysis of Gender Inequality in Face Recognition Accuracy,” arXiv:2002.00065v1, 2020.

[87] V. Albiero, K. Zhang, and K. W. Bowyer, “How Does Gender Balance in Training Data Affect Face Recognition Accuracy?,” arXiv:2002.02934v2, 2020.

[88] C. Rathgeb, A. Dantcheva, and C. Busch, “Impact and Detection of Facial Beautification in Face Recognition: An Overview,” EEE Access, vol. 7, pp. 152667–152678, October 2019.

[89] C. Rathgeb, D. Dogan, F. Stockhardt, M. De Marsico, and C. Busch, “Plastic Surgery: An Obstacle for Deep Face Recognition?,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, pp. 806-807, June 2020.

[90] S. Suri, A. Sankaran, M. Vatsa, and R. Singh, “On Matching Faces with Alterations due to Plastic Surgery and Disguise,” 9th IEEE International Conference on Biometrics: Theory, Applications, and Systems, Redondo Beach, CA, USA, pp. 1-7, October 2018.

[91] C. Qiong, S. Li, X. Weidi, M. P. Omkar, and Z. Andrew, “VGGFace2: A Dataset for Recognising Faces across Pose and Age,” 13th IEEE International Conference on Automatic Face & Gesture Recognition, Xi'an, China, pp. 67-74, May 2018.

[92] A. Kantarcı and H. K. Ekenel, “Thermal to Visible Face Recognition Using Deep Autoencoders,” International Conference of the Biometrics Special Interest Group, Darmstadt, Germany, pp. 1-5, September 2019.

[93] K. Chen, Y. Wu, H. Qin, D. Liang, X. Liu, and J. Yan, “R3 Adversarial Network for Cross Model Face Recognition,” IEEE Computer Society Conference Computer Vision Pattern Recognition, Long Beach, CA, USA, pp. 9860-9868, June 2019.

[94] L. Cheng, J. Wang, Y. Gong, and Q. Hou, “Robust Deep Auto-encoder for Occluded Face Recognition,” 23rd ACM International Conference on Multimedia, Brisbane Australia, pp. 1099-1102, October 2015.

[95] R. He, J. Cao, L. Song, Z. Sun, and T. Tan, “Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 5, pp. 1025–1037, December 2020.

[96] Y. Rao, J. Lu, and J. Zhou, “Attention-Aware Deep Reinforcement Learning for Video Face Recognition,” IEEE International Conference on Computer Vision, Venice, Italy, pp. 3951-3960, October 2017.

[97] Y. Zheng, D. K. Pal, and M. Savvides, “Ring loss: Convex Feature Normalization for Face Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 5089-5097, June 2018.

[98] Y. Liu, F. Wei, J. Shao, L. Sheng, J. Yan, and X. Wang, “Exploring Disentangled Feature Representation Beyond Face Identification,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 2080–2089, June 2018.

[99] L. Du, H. Hu, and Y. Wu, “Age Factor Removal Network Based on Transfer Learning and Adversarial Learning for Cross-Age Face Recognition,” IEEE Transactions Circuits and Systems for Video Technology, vol. 30, no. 9, pp. 2830-2842, June 2019.

[100] J. Deng, Y. Zhou, and S. Zafeiriou, “Marginal Loss for Deep Face Recognition,” IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, pp. 60-68, July 2017.

[101] J. Zhao et al., “Dual-Agent GANs for Photorealistic and Identity Preserving Profile Face Synthesis,” Advances in Neural Information Processing Systems, vol. 30, pp. 65–75, December 2017.

[102] J. Chen, J. Chen, Z. Wang, C. Liang, and C.-W. Lin, “Identity-Aware Face Super-Resolution for Low-Resolution Face Recognition,” IEEE Signal Processing Letters, vol. 27, pp. 645-649, April 2020.

[103] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, “Targeting Ultimate Accuracy: Face Recognition via Deep Embedding,” arXiv:1506.07310v4, 2015.

[104] S. M. Iranmanesh, B. Riggan, S. Hu, and N. M. Nasrabadi, “Coupled Generative Adversarial Network for Heterogeneous Face Recognition,” Image and Vision Computing, vol. 94, p. 1-10, February 2020.

[105] Z. Kaipeng, Z. Zhanpeng, L. Zhifeng, and Q. Yu, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Processing Letters, vol. 23, pp. 1499-1503, August 2016

[106] C. Rong, X. Zhang, and Y. Lin, “Feature-Improving Generative Adversarial Network for Face Frontalization,” IEEE Access, vol. 8, pp. 68842-68851, April 2020.

[107] W. Wu, M. Kan, X. Liu, Y. Yang, S. Shan, and X. Chen, “Recursive Spatial Transformer (ReST) for Alignment-Free Face Recognition,” IEEE International Conference on Computer Vision, Venice, Italy, pp. 3772-3780, October 2017.

[108] X. Peng, X. Yu, K. Sohn, D. Metaxas, and M. Chandraker, “Reconstruction-Based Disentanglement for Pose-invariant Face Recognition,” IEEE International Conference on Computer Vision, Venice, Italy, pp. 1623–1632, October 2017.

[109] Y. Jiaolong et al., “Neural Aggregation Network for Video Face Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 4362-4371, July 2017.

[110] J. Yu, D. Ko, H. Moon, and M. Jeon, “Deep Discriminative Representation Learning for Face Verification and Person Re-Identification on Unconstrained Condition,” IEEE International Conference on Image Processing, Athens, Greece pp. 1658–1662, October 2018.

[111] Q. Wang, T. Wu, H. Zheng, and G. Guo, “Hierarchical Pyramid Diverse Attention Networks for Face Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 8326-335, June 2020.

[112] Y. Sun, X. Wang, and X. Tang, “Deep Learning Face Representation from Predicting 10,000 Classes,” IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, pp. 1891-1898, June 2014.

[113] M. P. Omkar, V. Andrea, and Z. Andrew, “Deep Face Recognition,” British Machine Vision Conference, Swansea, UK, pp. 1-12, September 2015.

[114] C. Szegedy et al., “Going Deeper with Convolutions,” IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 1-9, June 2015.

[115] L. Bottou, “Stochastic Gradient Descent Tricks,” Neural Networks: Tricks of the Trade, Springer, Berlin, Heidelberg, pp. 421-436, January 2012.

[116] D. P. Kingma and J. Ba, “Adam: a Method for Stochastic Optimization,” International Conference on Learning Representations, San Diego, CA, USA, December 2014.

[117] J. Duchi, E. Hazan, and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” Journal of Machine Learning Research, vol. 12, pp. 2121−2159, January 2011.

[118] M. D. Zeiler, “ADADELTA: An Adaptive Learning Rate Method,” arXiv:1212.5701v1, 2012.

[119] M. Schmidt, “Least Squares Optimization with L1-Norm Regularization,” Technical Report, CS542B Project Report, January 2005.

[120] C. Cortes, M. Mohri, and A. Rostamizadeh, “L2 Regularization for Learning Kernels,” The Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, Quebec, Canada, pp. 109–116, June 2009.

[121] S. Nitish, H. Geoffrey, K. Alex, S. Ilya, and S. Ruslan, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” Journal of Machine Learning Research, vol. 15, no. 56, pp. 1929–1958, June 2014.

[122] I. Loshchilov and F. Hutter, “SGDR: Stochastic Gradient Descent with Warm Restarts,” arXiv:1608.03983v5, 2017.

[123] L. N. Smith, “Cyclical Learning Rates for Training Neural Networks,” IEEE Winter Conference on Applications of Computer Vision, Santa Rosa, California, USA, pp. 464–472, March 2017.

[124] C. Peng et al., “MegDet: A Large Mini-Batch Object Detector,” IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 6181–6189, June 2018.

[125] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima,” 5th International Conference Learning Representations, Toulon, France, pp. 1-16, April 2017.

[126] D. Masters and C. Luschi, “Revisiting Small Batch Training for Deep Neural Networks,” arXiv:1804.07612v1, 2018.

[127] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, “Large Scale Online Learning of Image Similarity Through Ranking,” Journal of Machine Learning Research, vol. 11, pp. 1109-1135, March 2010.

[128] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-margin Softmax Loss for Convolutional Neural Networks,” 33rd International Conference on International Conference on Machine Learning, New York NY USA, pp. 507-516, June 2016.

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

[130] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “Sphereface: Deep Hypersphere Embedding for Face Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 212-220, July 2017.

[131] K. Sohn, “Improved Deep Metric Learning with Multi-class N-pair Loss Objective,” 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, pp. 1857-1865, December 2016.

[132] Y. Liu, H. Li, and X. Wang, “Learning Deep Features via Congenerous Cosine Loss for Person Recognition,” arXiv:1702.06890v2, 2017.

[133] B. Chen, W. Deng, and J. Du, “Noisy Softmax: Improving the Generalization Ability of DCNN via Postponing the Early Softmax Saturation,” arXiv:1708.03769v1, 2017.

[134] X. Wei, H. Wang, B. Scotney, and H. Wan, “Minimum Margin Loss for Deep Face Recognition,” Pattern Recognition, vol. 97, pp. 107-112, January 2020.

[135] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality Reduction by Learning an Invariant Mapping,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, June 2006.

[136] H. Wang et al., “Cosface: Large Margin Cosine Loss for Deep Face Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 5265-5274, June 2018.

[137] Y. Huang et al., “CurricularFace: Adaptive Curriculum Learning Loss for Deep Face Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, United States, pp. 5901–5910, June 2021.

[138] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, “Range Loss for Deep Face Recognition with Long-tail,” IEEE International Conference on Computer Vision, Venice, Italy, pp. 1-10, October 2017.

[139] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A Discriminative Feature Learning Approach for Deep Face Recognition,” European Conference on Computer Vision, Amsterdam, Netherlands, pp. 499-515, October 2016.

[140] R. Ranjan, C. D. Castillo, and R. Chellappa, “L2-constrained Softmax Loss for Discriminative Face Verification,” arXiv:1703.09507v3, 2017.

[141] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and <0.5MB Model Size,” arXiv:1602.07360v4, 2016.

[142] M. You, X. Han, Y. Xu, and L. Li, “Systematic Evaluation of Deep Face Recognition Methods,” Neurocomputing, vol. 388, pp. 144-156, May 2020.

[143] A. G. Howard et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” arXiv:1704.04861v1, 2017.

[144] M. Stylianos, P. Athanasios, S. Christos, D. Jiankang, K. Irene, and Z. Stefanos, “AgeDB: The First Manually Collected, In-the-Wild Age Database,” IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, pp. 1997-2005, July 2017.

[145] I. N. G. AryaAstawa, I. K. G. D. Putra, M. Sudarma, and R. S. Hartati, “KomNET: Face Image Dataset from Various Media,” Data Brief, vol. 2, pp. 1-5, August 2020.

[146] B. H. Gary, M. Marwan, B. Tamara, and L.-M. Eric, “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, Marseille, France, pp. 1-15, October 2008.

[147] W. Fei et al., “The Devil of Face Recognition Is in the Noise,” European Conference on Computer Vision, Munich, Germany, pp. 780-795, September 2018.

[148] G. Baris, B. Binod, K. Josef, and K. Tae-Kyun, “Semi-supervised Adversarial Learning to Generate Photorealistic Face Images of New Identities from 3D Morphable Model,” European Conference on Computer Vision, Munich, Germany, pp. 217-234, September 2018.

[149] Z. Tianyue and D. Weihong, “Cross-Pose LFW: A Database for Studying Cross-Pose Face Recognition in Unconstrained Environments,” Beijing University of Posts and Telecommunications, Technical Report 18-01, February 2018.

[150] S. Yi, W. Xiaogang, and T. Xiaoou, “Hybrid Deep Learning for Face Verification,” IEEE International Conference on Computer Vision, Sydney, Australia, pp. 1489-1496, December 2013.

[151] K.-S. Ira, S. Steve, M. Daniel, and B. Evan, “The MegaFace Benchmark: 1 Million Faces for Recognition at Scale,” IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 4873-4882, June 2016.

[152] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Learning Face Representation from Scratch,” arXiv:1411.7923v1, 2014.

[153] Z. Wang et al., “Masked Face Recognition Dataset and Application,” arXiv:2003.09093v2, 2020.

[154] B. Simone, “Large Age-Gap Face Verification by Feature Injection in Deep Networks,” Pattern Recognition Letters, vol. 90, no. C, pp. 36-42, April 2017.

[155] Y. Li, W. Zheng, Z. Cui, and T. Zhang, “Face Recognition Based on Recurrent Regression Neural Network,” Neurocomputing, vol. 297, pp. 50-58, July 2018.

[156] L. Wolf, T. Hassner, and I. Maoz, “Face Recognition in Unconstrained Videos with Matched Background Similarity,” Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, pp. 529-534, June 2011.
[157] L. Liu, L. Zhang, H. Liu, and S. Yan, “Toward Large-Population Face Identification in Unconstrained Videos,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 11, pp. 1874-1884, April 2014.
[158] A. J. O’Toole et al., “A Video Database of Moving Faces and People,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 812-816, June 2005.
[159] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, “UMDFaces: An Annotated Face Dataset for Training Deep Networks,” arXiv:1611.01484v2, 2017.
[160] B. F. Klare et al., “Pushing the Frontiers of Unconstrained Face Detection and Recognition: IARPA Janus Benchmark A,” IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 1931-1939, October 2015.
[161] W. Cameron et al., “IARPA Janus Benchmark-B Face Dataset,” IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, pp. 592-600, July 2017.
[162] B. Maze et al., “IARPA Janus Benchmark - C: Face Dataset and Protocol,” International Conference on Biometrics, Queensland, Australia, pp. 58–165, February 2018.
[163] S. Gupta et al., “FaceSurv: A Benchmark Video Dataset for Face Detection and Recognition Across Spectra and Resolutions,” 14th IEEE International Conference on Automatic Face & Gesture Recognition, Lille, France, pp. 1-7, May 2019.
[164] J. R. Beveridge et al., “The Challenge of Face Recognition from Digital Point-and-shoot Cameras,” IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems, Arlington, VA, USA, pp. 1-8, October 2013.
[165] K. Delac, M. Grgic, and S. Grgic, “Effects of JPEG and JPEG2000 Compression on Face Recognition,” International Conference on Pattern Recognition and Image Analysis, Estoril, Portugal, pp. 136–145, June 2005.
[166] B. J. Boom, G. M. Beumer, L. J. Spreeuwers, and R. N. J. Veldhuis, “The Effect of Image Resolution on the Performance of a Face Recognition System,” 9th International Conference on Control, Automation, Robotics and Vision, Singapore, pp. 409-414, December 2006.
[167] R. Singh, M. Vatsa, and A. Noore, “Effect of Plastic Surgery on Face Recognition: A Preliminary Study,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Miami, FL, USA, pp. 72-77, June 2019.
[168] J. R. Beveridge, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, and P. J. Phillips, “Quantifying How Lighting and Focus Affect Face Recognition Performance,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops, San Francisco, CA, USA, pp.74-81, June 2010.
[169] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain, “Face Recognition Performance: Role of Demographic Information,” IEEE Transactions on Information Forensics and Security, vol. 7, no. 6, pp. 1789-1801, October 2012.
[170] G. Hu et al., “When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition,” IEEE International Conference on Computer Vision Workshop, Santiago, Chile, pp. 4321-4329, December 2015.

[171] S. Karahan, M. K. Yildirim, K. Kirtac, F. S. Rende, G. Butun, and H. K. Ekenel, “How Image Degradations Affect Deep CNN-based Face Recognition?,” arXiv:1608.05246v1, 2016.

[172] C. Reale, N. M. Nasrabadi, and R. Chellappa, “An Analysis of the Robustness of Deep Face Recognition Networks to Noisy Training Labels,” IEEE Global Conference on Signal and Information Processing, Washington, DC, USA, pp.1192-1196, December 2016.

[173] D. Deb, L. Best-Rowden, and A. K. Jain, “Face Recognition Performance under Ageing,” IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, pp. 548- 556, July 2017.

[174] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel, “Strengths and Weaknesses of Deep Learning Models for Face Recognition against Image Degradations,” IET Biometrics, vol. 7, pp. 81–89, October 2017.

[175] M. Pavlovic, R. Petrovic, B. Stojanovic, and S. Stankovic, “Facial Expression and Lighting Conditions Influence on Face Recognition Performance,” International Conference on Electrical, Electronic and Computing Engineering, Jawa Timur, Indonesia, pp.1-5, November 2018.

[176] E. López-López, X. M. Pardo, C. V. Regueiro, R. Iglesias, and F. E. Casado, “Dataset Bias Exposed in Face Verification,” IET Biometrics, vol. 8, no. 4, pp. 249-258, February 2019.

[177] N. Srinivas, K. Ricanek, D. Michalski, D. S. Bolme, and M. King, “Face Recognition Algorithm Bias: Performance Differences on Images of Children and Adults,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, pp. 2269-2277, June 2019.

[178] Y. Srivastava, V. Murali, and S. R. Dubey, “A Performance Evaluation of Loss Functions for Deep Face Recognition,” arXiv:1901.05903v2, 2019.

[179] K. S. Krishnapriya, V. Albiero, K. Vangara, M. C. King, and K. W. Bowyer, “Issues Related to Face Recognition Accuracy Varying Based on Race and Skin Tone,” IEEE Transactions on Technology and Society, vol. 1, no. 1, pp. 8-20, February 2020.

[180] N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, and A. Kuijper, “The Effect of Wearing a Mask on Face Recognition Performance: an Exploratory Study,” arXiv:2007.13521v2, 2020.

[181] G.-S. J. Hsu, H.-Y. Wu, and M. H. Yap, “A Comprehensive Study on Loss Functions for Cross-Factor Face Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, pp. 3604-3611, June 2020.

[182] L. Boussaad and A. Boucetta, “Deep-learning Based Descriptors in Application to Ageing Problem in Face Recognition,” Journal of King Saud University - Computer and Information Sciences, In Press, October 2020.

[183] J. Coe and M. Atay, “Evaluating Impact of Race in Facial Recognition across Machine Learning and Deep Learning Algorithms,” Computers, vol. 10, no. 9, pp. 1-24, September 2021.

[184] J. G. Cavazos, P. J. Phillips, C. D. Castillo, and A. J. O’Toole, “Accuracy Comparison Across Face Recognition Algorithms: Where are we on Measuring Race Bias?,” IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, pp. 101-111, September 2020.

[185] G. B. Huang and E. Learned-Miller, "Labeled Faces in the Wild: Updates and New Reporting Procedures," Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.
[186] B. F. Klare, A. Blanton, and B. Klein, "Efficient Face Retrieval Using Synecdoches," International Joint Conference on Biometrics, Clearwater, USA, pp. 1–7, December 2014.
[187] I. Serna, A. Morales, J. Fierrez, M. Cebrian, N. Obradovich, and I. Rahwan, "SensitiveLoss: Improving Accuracy and Fairness of Face Representations with Discrimination-Aware Deep Learning," arXiv:2004.11246v2, 2020.
[188] VELUSAMY, Sudha, et al. FabSoften: Face beautification via dynamic skin smoothing, guided feathering, and texture restoration. En Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020. p. 530-531.
[189] HEDMAN, Pontus, et al. On the Effect of Selfie Beautification Filters on Face Detection and Recognition. arXiv preprint arXiv:2110.08934, 2021.
[190] RATHGEB, Christian, et al. Vulnerability assessment and detection of makeup presentation attacks. En 2020 8th International Workshop on Biometrics and Forensics (IWBF). IEEE, 2020. p. 1-6.
[191] D. Anghelone, C. Chen, A. Ross, A. Dantcheva, "Beyond the Visible: A Survey on Cross-spectral Face Recognition," arXiv preprint arXiv:2201.04435 (2022).
[192] YADAV, Daksha, et al. Effect of illicit drug abuse on face recognition. En 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016. p. 1-7.
[193] RAGHAVENDRA, Ramachandra; RAJA, Kiran B.; BUSCH, Christoph. Impact of drug abuse on face recognition systems: a preliminary study. En Proceedings of the 9th International Conference on Security of Information and Networks. 2016. p. 24-27.
[194] PANDEY, Prateekshit; SINGH, Richa; VATSA, Mayank. Face recognition using scattering wavelet under Illicit Drug Abuse variations. En 2016 International Conference on Biometrics (ICB). IEEE, 2016. p. 1-6.
[195] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4):627–639, 2007.
[196] M.K. Bhowmik, K. Saha, S. Majumder, G. Majumder, A. Saha, A.N. Sarma, D. Bhattacharjee, D. K. Basu, and M. Nasipuri. Thermal infrared face recognition—a biometric identification technique for robust security system. Reviews, refinements and new ideas in face recognition, 7, 2011.
[197] F. Nicolo and N. A. Schmid. Long range cross-spectral face recognition: Matching SWIR against visible light images. IEEE Transactions on Information Forensics and Security, 7(6):1717–1726, 2012.
[198] S. Heist, M. Landmann, M. Steglich, Y. Zhang, P. KAŒhmstedt, and G. Notni. Pattern projection in the short-wave infrared (SWIR): accurate, eye-safe 3D shape measurement. In Dimensional Optical Metrology and Inspection for Practical Applications VIII, volume 10991, pages 99 – 108, 2019.
[199] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos. Physiology-based face recognition in the thermal infrared spectrum. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4):613–626, 2007.

[200] L.B. Wolff, D.A. Socolinsky, and C.K. Eveland. Face recognition in the thermal infrared. In Computer Vision Beyond the Visible Spectrum, pages 167–191. 2005.
[201] T. Bourlai. Mid-wave IR face recognition systems. SPIE Newsroom Magazine-Defense & Security, 2013.