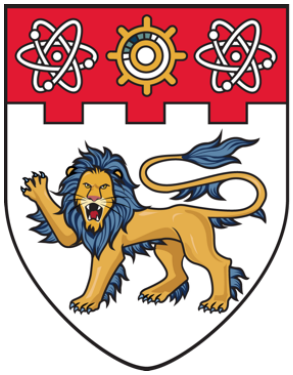

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

AI6121 Computer Vision

Directed Reading and Literature Review on:
Face Recognition

authored by

Ong Jia Hui
JONG119@e.ntu.edu.sg
G1903467L

Abstract

Biometrics is the means of identifying and authenticating a person using a set of recognizable and verifiable human characteristics specific to the person. Facial Recognition (FR) is a subset of biometric solutions, which identifies or verifies a person's identity by analyzing their facial textures and shape.

In this directed reading and literature review, the topic of face recognition will be introduced in the first chapter, followed by the conventional and state-of-the-art approaches of FR techniques. The research paper titled "FaceNet" will be reviewed and reasoned to justify it as one of the most revolutionary papers in modern FR domain.

In the latter parts of this report, numerous open challenges in FR research from conventional computer vision problems to deep learning related issues will be discussed. Finally, future research trends will be listed, where current affairs and market needs are driving factors for these trends.

Contents

1 Introduction **2**
1.1 Background 2
1.2 Motivation 3
1.3 FR Terminology 3
1.4 Components of Modern FR 3

2 Technical Approaches **5**
2.1 Traditional Approaches 5
2.2 SOTA Approaches 6

3 Most Influential Paper **8**

4 Technical Challenges **11**

5 Future Trends **13**

6 Conclusion **15**

References **16**

1. Introduction

1.1 Background

Biometrics is the means of identifying and authenticating a person using a set of recognizable and verifiable human characteristics specific to the person. FR is a subset of biometric solutions, which identifies or verifies a person's identity by analyzing their facial textures and shape. Other forms of biometrics include DNA, fingerprints, iris scans, voice recognition, digitization of veins in the palm, and behavioural measurements [1]. However, as these alternatives are usually more invasive and less pervasive to acquire; contactless, non-intrusive FR promptly becomes prevalent [2].

FR is predominantly performed on two-dimensional (2D) perceptions from a digital image or extracted frames from a video source. However, 2D images suffer from limitations such as the lack of depth information. An example of depth information includes a person's nose bridge height, which can be a discriminating facial feature. Hence, it is also not uncommon for FR systems to utilize three-dimensional (3D) facial imaging through the use of 3D sensors to capture the full face shape. Examples of such sensors include Infrared (IR) cameras or Time of Flight (ToF) camera. There are also applications using both 2D and 3D images for processing, which significantly increase the accuracy of the FR system.

The two general error terms used in FR domain are false positive and false negative rate. A false positive occurs when two images of different individuals have high similarity scores and are considered by the FR software as the same person. On the other hand, a false negative happens when the FR system failed to match images of the same person. The National Institute of Standards and Technology (NIST) began conducting annual Face Recognition Vendor Test (FRVT) since early 2000s. These tests are considered one of the gold standard assessments for FR vendors. In these FRVT report [3], False Non-Match Rate (FNMR) and False Match Rates (FMR) are two main metrics used for ranking the FR algorithms' performance. There are also other metrics studied and used, such as those that measure racial bias in FR [4].

1.2 Motivation

The motivations for FR researches are derived from two major factors; the high adoption rate in various industries as well as the large commercial market revenue.

The application of deep learning and representation learning has driven the great advances in FR technology, which in turn led to the wide adoption in many applications. With state-of-the-art deep FR surpassing human performance, it has become a prevalent tool for automation of video surveillance and security systems. FR have also been incorporated into many government, commercial and banking applications. Automatic face tagging in social media and smartphones' unlocking are some examples of individuals' day-to-day usage of FR.

A market report in June 2019 estimates that the global facial recognition market will generate a total of $7 billion in revenue by 2024 [5]. In similar news, the biometrics-as-a-service (BaaS) market was also forecasted to surpass $10 billion by 2030 [6]. This projected growth and market value spur the demand for higher accuracy in terms of precision and recall.

1.3 FR Terminology

Face recognition is a broad term for two types of authentication methodologies, namely face identification and face verification. Face identification is a form of one-to-many matching (1:N) search for a matching face in a database of multiple faces (also called a gallery). Face enrolment is the process when reference faces are added to the gallery. The most prominent uses of 1:N are in surveillance and crime investigation. On the other hand, face verification is a form of one-to-one matching (1:1), and it is concerned with validating the claimed identity of the probe image with the reference image (known face). Examples of 1:1 include device security and passport verification.

1.4 Components of Modern FR

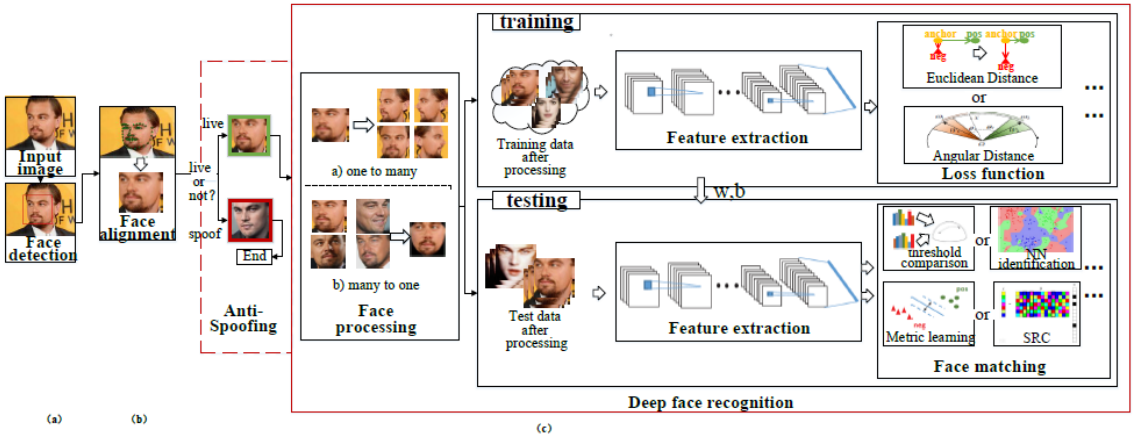

Figure 1.1: Components in a FR system. Source: [7].

As shown in Figure 1.1, a robust FR system comprises of the following six-step process:

1. Face Detection by face detector
2. Face Alignment via landmark locator
3. Anti-Spoofing Analysis such as liveliness check
4. Face Augmentation e.g. one-to-many augmentation or many-to-one normalization
5. Face Extraction of feature representation via neural network
6. Face Matching i.e. similarity score computation

The pipeline begins with the face detection step, which localizes human faces in the image. After which, a landmark detector will perform face alignment in the image. The output of this step will produce faces aligned to normalized canonical coordinates. These faces will then go through anti-spoofing analysis to remove possible spoofed faces. Before passing into the feature extraction neural network, face augmentation generates more images of different facial poses from a single image. The fifth face extraction stage consists of the neural network and loss function design, which learns to output feature representations. Finally, the face matching stage computes the similarity scores using cosine or L2 distances of the feature representations between the probe and the gallery [7].

2. Technical Approaches

2.1 Traditional Approaches

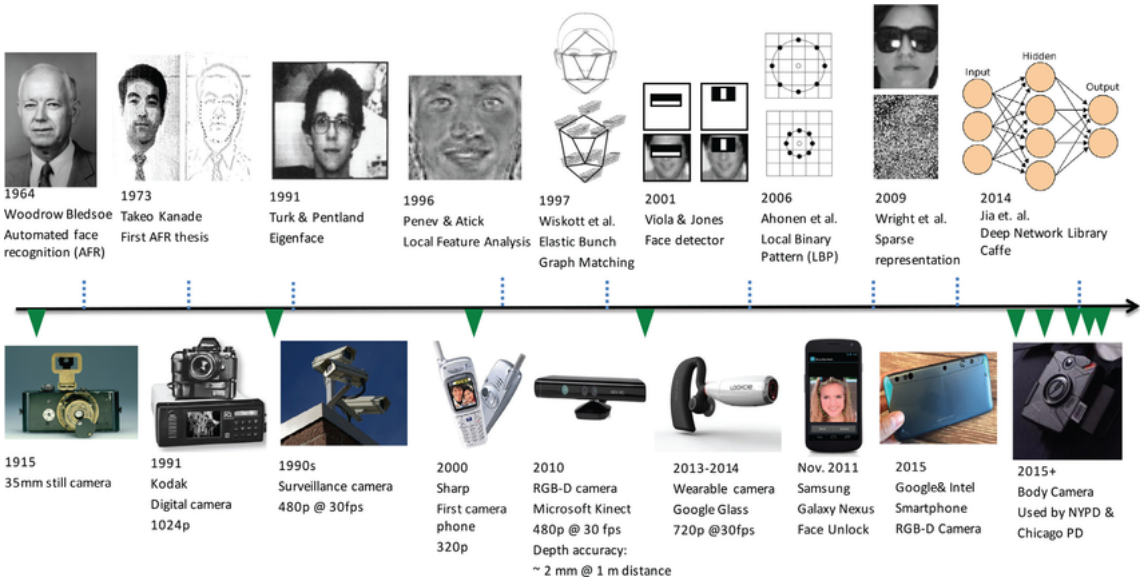

Figure 2.1: Major milestones in history of automated face recognition. Source: [8].

The idea of using a computer to recognize faces first began in the 1960s. Woodrow W. Bledsoe, Helen Chan and Charles Bisson were the pioneer group of researchers working on programming computers to distinguish human faces [9, 10, 11, 12, 13]. Their man-machine project computed normalized distances between human extracted features to generate frontal face representations for comparison [14]. Interestingly, the challenges identified in their early works include issues such as pose, rotation, tilt, scale, lighting, angle, expression and ageing [14, 15]. These are the same challenges faced by modern FR research a demi-decade later.

The proposition of Eigenface representation by Matthew Turk and Alex Pentland in 1991 [16] invoked a significant breakthrough in FR research. They employed the use of Principal Component Analysis (PCA), an unsupervised dimensionality reduction technique to decompose the face images into a much smaller set of significant eigenvectors. It was an efficient approach, but it suffers from unsatisfactory results due to variations in lighting, head size, head orientation [16]. In 1997, another popular dimensional reduction technique called Fisherfaces was introduced. This algorithm uses Linear Discriminant Analysis (LDA) to maximize intra-class from inter-class projections [17]. It is more complex than Eigenface, but has higher resilience towards light variance.

In 2001, the invention of Viola-Jones object detection framework [18] provided near real-time face detection. This improved the performance of FR systems as it is the first step of any FR pipeline (enumerated in Section 1.4). Even though face detection via

Viola-Jones is fast, the major downside of this algorithm is its slow training time.

Since David Lowe published his paper on Scale-invariant Feature Transform (SIFT) in 2004 [19], the technique was widely adopted to the face recognition domain. The main steps of SIFT involve keypoint detection, descriptor establishing, and image feature matching [20]. The SIFT descriptors addressed the problem of matching features with varying scale and rotation. Furthermore, these local features also allowed the identification of faces among clutter and occlusion. Subsequently, there were numerous variants such as PCA-SIFT, GSIFT, CSIFT and SURF developed to improve the robustness of the SIFT algorithm [20].

Following AlexNet's success in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [21], Deep Convolutional Neural Networks (DCNN) trained with massive datasets have performed significantly better in feature extraction than traditional hand-crafted features. The breakthroughs of DeepFace by Yaniv Taigman et al. from Facebook Research using DCNN in 2014 has reshaped the research scene of FR. They achieved an accuracy of 97.35% $\pm$ 0.25% on Labeled Faces in the Wild (LFW) dataset [22], nearly approaching human's standard of 97.53% [23]. In 2015, FaceNet [24] surpassed the state-of-the-art techniques with an accuracy of 99.63% $\pm$ 0.09% on LFW dataset through the novel use of the triplet loss function. However, even with the underlying the great performance of deep learning, deep FR suffers from inherent issues such as dataset biases and weakness to adversarial attacks, which will be further discussed in Chapter 4.

2.2 SOTA Approaches

With deep learning dominating the FR domain, state-of-the-art approaches usually differs in three areas, namely the dataset size, the backbone network architecture and the choice of loss function.

Due to the massive amount of computing power and dataset available to commercial tech giants like Google, Facebook and Baidu, their FR systems have been reportedly trained with hundreds of million faces. In contrast, publicly available dataset typically ranges from thousands to ten million. Examples of public datasets for 2D facial recognition include LFW (2007) [22], MS-Celeb-1M (2016) [25], VGGFace2 (2018) [26], IJB C (2018) [27] and IMDB-Face (2018) [28]. There are also specialized public datasets such as Bosphorus (2008) for 3D facial recognition [29], Youtube Faces Dataset (YTD) (2011) for video FR [30], Cross-Age LFW (CALFW) (2017) [31] and Cross-Pose LFW (CPLFW) [32] (2018) for age and pose variance, Siw (2018) for anti-spoofing research [33] and FairFace (2019) with balanced race composition [34].

These large facial datasets will then be trained in a deep neural network for feature learning. The network architecture used are heavily influenced by the state-of-the-art network architectures of DCNN used in object classification. In terms of variations, DCNN can be broadly categorized into seven different classes namely; spatial exploitation, depth, multi-path, width, feature-map exploitation, channel boosting, and attention-based convolutional neural networks [35]. As shown in Figure 2.2, some of these mainstream architectures include AlexNet (2012) [21], VGGNet (2014) [36], GoogLeNet (2015) [37], ResNet (2016) [38] and Squeeze-and-Excitation Network (SENet) (2017) [39]. Examples

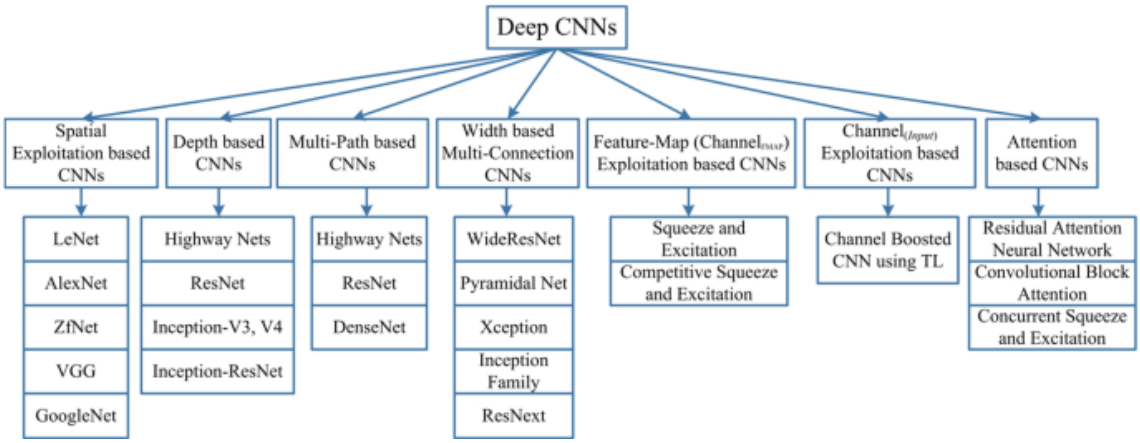

Figure 2.2: Taxonomy of deep CNN architectures. Source: [35].

of the prominent usages of these baseline networks in FR research include AlexNet used in DeepFace (2014) [23], GoogLeNet in FaceNet (2015) [24] and ArcFace (2019) [40] with ResNet and SENet.

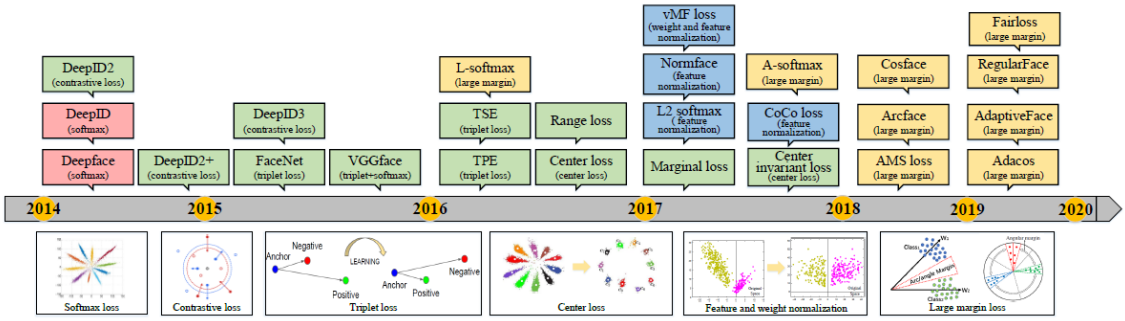

Figure 2.3: Development of loss functions for FR. Source: [7].

In an image classification task, the softmax function often constitutes the last layer of a CNN, used to output the probability distributions of each class. However, in the case of identifying faces, the softmax function is not sufficiently practical for FR as intra-variations in faces could be more significant than inter-differences [7]. Furthermore, with FR often termed as a "zero-shot" or "one-shot" learning problem, facial features learnt need to be highly discriminative in order to correctly distinguish between different identities. As such, there are a large number of works performed in developing novel loss functions for the facial recognition task such as those shown in Figure 2.3. There are three major categories of the loss functions published, namely; euclidean loss, angular or cosine marginal loss and softmax variations. Euclidean loss functions such as the Triplet loss in FaceNet (2015) [24], aim at reducing the intra-variations while increasing the inter-variations of projected distances in the Euclidean space. Moreover, margin-based loss functions such as A-softmax from SphereFace (2017) [41], CosFace (2018) [42] and ArcFace (2019) [40], separate learned features using large angular or cosine distances. Lastly, there are alternate forms of softmax variations like Normface (2017) [43] targeting performance improvement through the use of weights or feature normalization. In summary, state-of-the-art approaches are a combination of available dataset trained or pre-trained on modified backbone networks with the use of appropriate loss function.

3. Most Influential Paper

In this chapter, the paper titled *FaceNet: A Unified Embedding for Face Recognition* by Schroff, F., Kalenichenko, D., Philbin, J. published in 2015 [24] will be analyzed and reviewed. Its contributions will be listed to justify it as being one of the most revolutionary papers in the towards the advancement in deep FR research.

In conventional classification tasks, neural networks typically have a fixed number of classes and use softmax cross-entropy loss for back-propagation. However, in the context of FR, it is often a one-shot learning problem, which makes it infeasible to continuously retrain a giant classifier whenever a new face needs to be added to the gallery.

With this problem in mind, the researchers from Google Inc, Schroff et al. proposed the use of Triplet loss to train deep neural networks that could represent faces using latent variables (embedding) in Euclidean space. The distance between two output embeddings from FaceNet can be regarded as a measure of similarity for face verification, recognition and clustering tasks.

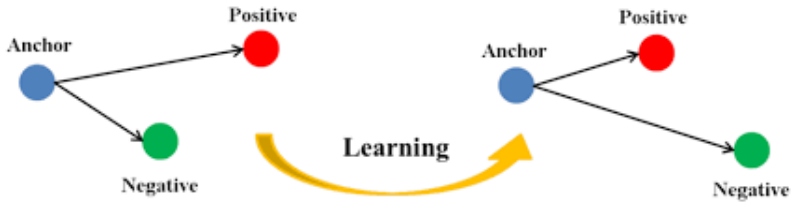

Figure 3.1: The Triplet Learning Problem [24]

The figure 3.2 shows the purpose of training triplets (anchor, positive and negative) together when learning the embeddings. The triplet loss not only minimizes the distance between the anchor image and the positive image of the same identity. It also maximizes the distance between the anchor and the negative image of different labels concurrently.

Using A to depict an anchor image, P as a positive image and N as a negative example of any other person, $f(A), f(P), f(N)$ denote the corresponding output embedding of the images. The network aims to achieve:

$$||f(A) - f(P)||_2^2 + \alpha \leq ||f(A) - f(N)||_2^2$$
$$||f(A) - f(P)||_2^2 - ||f(A) - f(N)||_2^2 + \alpha \leq 0$$

As shown in the equations above, the objective function is to separate the distance between the positive pairs (anchor + positive) and negative pairs (anchor + negative) by a margin (α). The distance margin (α) not only ensures that the model does not merely return a trivial solution ($f(x) = 0$), it also enforces a significant distance between the similar and dissimilar pairs.

$$\mathcal{L}(A, P, N) = \sum_{i}^{M} \left[max\left(||f(A) - f(P)||_2^2 - ||f(A) - f(N)||_2^2 + \alpha, 0 \right) \right] \tag{3.1}$$

As such, the triplet loss to be minimized can be rewritten as equation 3.1. The network will try to learn to reduce the $||f(A) - f(P)||_2^2$ to zero and $||f(A) - f(N)||_2^2$ to be greater than $||f(A) - f(P)||_2^2 + \alpha$.

However, to reap the full benefits of using the triplet loss, there is a need for a good strategy to select the triplets for training. This is because if the triplets are all selected randomly, there is a high chance that there will be too many "hard negatives" or "easy triplets" such that $||f(A) - f(N)||_2^2$ will be always greater than $||f(A) - f(P)||_2^2 + \alpha$. This will cause the loss to easily become zero and it will adversely affect the model's convergence speed.

To solve this constraint, Schroff et al. reported in the same paper of their online negative exemplar mining strategy. The idea is to select "semi-negatives" N that are not too close to the anchor A image than the positive P. The motivation behind this idea is to ensure positive losses. At the same time, by not using "hard negatives" for training can prevent the model from "collapsing" due to bad local minima.

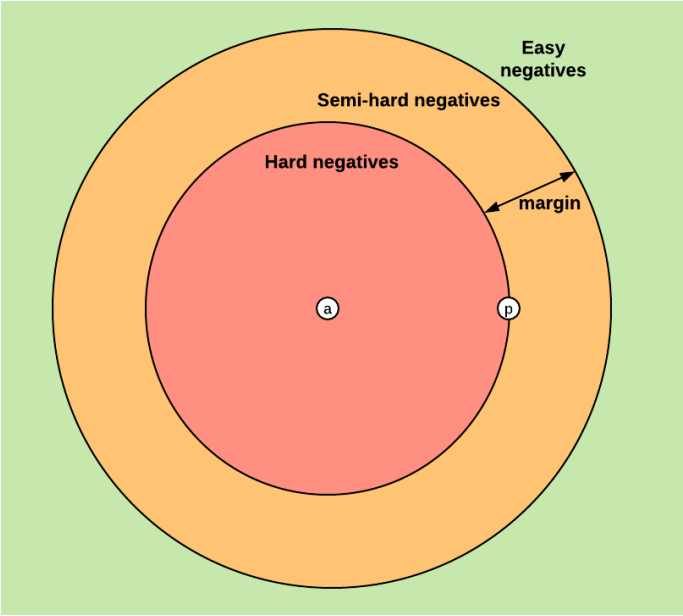

Figure 3.2: Semi-hard negatives versus hard negatives [44]

Figure 3.2 corresponds to the selection of "semi-hard" negatives that are within the range of $||f(A) - f(P)||_2^2 < ||f(A) - f(N)||_2^2 < +\alpha$. These triplets are generated online for each mini-batch of size 1,800. The neural networks were trained with Stochastic Gradient Descent (SGD) using AdaGrad optimizer, and an initial learning rate of 0.05. The margin hyperparameter, α, is set as 0.2, and the two network architectures explored were namely, ZeilerFergus [45] and GoogLeNet style inception models [37].

The trained models were evaluated on the LFW dataset [22] as well as the YTD database [30]. Coupled with an additional face alignment step, FaceNet accomplished a record-breaking classification accuracy of 99.63% ± 0.09% standard error on LFW. Similarly, the model also achieved an accuracy of 95.12% ± 0.39% on YTD, which is 30% better than the former state-of-the-art, DeepId2+'s 93.2% [46].

The innovation in this paper was that of the Triplet loss function. Before FaceNet, multiple separate networks were used for learning features, classifying them and matching of the features' similarities. By enabling an end-to-end method of learning embeddings, it eliminated the need of another CNN bottleneck layer or concatenation of multiple models with PCA. This paper also broke the prevailing belief that the triplet loss is inferior to other classification losses for computer vision tasks [47].

This paper paved the way for applied deep learning in solving a variety of other similarity measurement problems. By posing the problem as a similarity learning problem instead of a classification problem, the authors demonstrated the need for formulating appropriate loss function for the issue at hand. It also drove the research community to develop better loss functions for the FR task. For instance, soon after this 2015 paper, Triplet Probabilistic Embedding (TPE) [48] and Triplet Similarity Embedding (TSE) [49] were inspired to learn a linear projection to construct triplet loss. Current academia state-of-the-arts such as ArcFace (Additive angular margin loss)[40] and SphereFace (A-Softmax)[41] also worked on the betterment of loss functions for deep face recognition.

4. Technical Challenges

Many common challenges in the computer vision field also apply to the FR domain. As delivered by Li Fei-Fei in her lecture on Computer Vision at Stanford University [50], challenges such as viewpoint variation, background clutter, illumination, occlusion, deformation, and scaling are fundamental CV concerns. Aside from these issues, unlike inanimate objects, human faces consist of pose variations [51, 52, 53], expressions [54] or emotions [55]. These facial characteristics have always been active research areas on how to tackle them to improve FR performance [56, 57].

Due to the widespread use of FR for identity and forensic purposes, ensuring robustness towards attacks on the FR system has become one of the greatest challenges in the field. Various forms of attacks ranging from replay or presentation [58] and adversarial attacks [59] pose security concerns to the use of FR applications. Replay and presentation attacks are spoof attacks used to deceive a biometric sensor. These attacks might use a printed photograph, an image or video of a person or wearing a silicone 3D mask. Traditional anti-spoofing solutions of "liveliness" check using image depth and 3D sensors are getting challenged with realistic 3D masks [60]. Hence, anti-spoofing research continues to be a common topic appearing in CVPR conferences [61].

With the rise of deep learning, many of its fundamental research issues are also brought across into the FR domain. First of which is the weakness of neural networks towards adversarial examples. Adversarial examples are intentional feature perturbations designed to cause machine learning models to make a false prediction [62, 63]. Studies on using Generative Adversarial Networks (GANs) to attack FR systems have been ongoing for many years. SOTA FR systems still suffer from the same vulnerability over the years [64, 65]. Consequently, counter detection of adversarial attacks is also extensively researched upon [66]. Recently, benchmarks against adversarial examples such as the TALFW database [67] are getting created to facilitate research on the robustness and defence of deep face recognition.

With deep FR surpassing human performance, deep learning is still often seen as a black-box approach. The challenge to justify AI decision-making is also known as the interpretability problem [68]. Research on Explainable AI (XAI) aims to allow humans to use the knowledge of understanding AI logic to develop better machines. In a recent paper, Philips et al. [69] listed four principles of XAI applied to Biometrics and Facial Forensic algorithms. These are examples of mission-critical applications, which heavily rely on FR systems to deliver fairness, unbiasedness, transparency, and accountability. In 2020, Y.S. Lin et al. also proposed an explainable cosine metric, xCos [70] to measure how deep models determines the similarity score for face verification task.

Training a deep learning model would require a large amount of dataset, which tends to suffer from data imbalance problem. This imbalance problem results in the gender or racial bias concerns raised in machine learning models. The term Cross-Race Effect (CRE) reviewed by SG Young et al. in 2012, is the tendency for perceivers to have more

accurate recognition memory for same-race (SR) faces than for cross-race (CR) faces [71]. This effect is highlighted in the news after MIT Media Lab discovered skin-type and gender biases in many commercial FR vendors [72] in 2018. Nonetheless, in a recent paper titled "Face Recognition: Too Bias, or Not Too Bias?" by Robinson, J. P., et al. (2020) [73] shared compelling insights on the level of biases in state-of-the-art FR systems. After performing human and machine evaluation on a balanced dataset called Balanced Faces in the Wild (BFW) for FR, the results have shown that in many situations, especially when distinguishing with faces of a different demographic group, human fared much worse than machines did. Furthermore, the latest 2020 NIST's FRVT report also showed vanishing small error rates in false positive and false negative rates across demographic groups [74]. This has proved the progress made in recent years by researchers to diminish this challenge.

Aforementioned, the use of FR in mission-critical applications require a constant pursuit of improvement in accuracy and lower false positive and false negative rates. Furthermore, the emerging use of mobile FR applications also demand extreme efficiency in terms of usage and performance [7]. This limits pushing trend remains an open challenge in FR domain to improve algorithms and design further.

Lastly, there are also human rights-related challenges such as privacy and regulatory issues due to the sensitivity of the facial data. Similar to the analogy of password leakage, facial dataset leakage is also of public concern. Numerous privacy-preserving research [75, 76] aims at balancing the trade-off between privacy and security. With calls for FR bans emerging from activists [77], the demand on resolve these open FR challenges becomes increasingly necessary.

5. Future Trends

As discussed in Chapter 4, most of the open challenges will continue to be researched on in the future. This chapter will outline the five most emerging FR research trends of 2020 and beyond.

Figure 5.1: Digitally applied masks with shape variations. Source: [78].

COVID-19 has changed the world's behaviour and appearance. Before the pandemic, banks would usually prohibit customers from entering their branches if they are wearing any form of disguises like sunglasses that would shield their identities. On the contrary, during the containment for the virus outbreak, it is forbidden to enter banks without face masks in 2020. Facial masks worn for protection against contagious viruses are now health necessities, but they pose a new form of challenge to modern FR surveillance and identity applications. In a recent NIST's preliminary study shown the substantial effects of masks on the accuracy of pre-COVID-19 FR algorithms [78]. 1:1 verification tests (Section 1.3) were performed using an unmasked face as the reference against probe faces with digitally applied masks of nine variations (to simulate differences in real-life mask shapes). Figure 5.1 shows an example of the images after applying digital masks. The results revealed that the commercial facial recognition algorithms tested had increased error rates of between 5% to 50%. The NIST report also stated the observation of higher error rates in black masks over light-blue masks. As such, commercial market needs will drive these FR vendors to resolve this mask occlusion research problem.

Fueled by the recent success of Generative Adversarial Networks (GANs) [79], the high realism of deepfakes has a detrimental effect on the credibility of FR systems. For instance, these hyper-realistic artificially modified videos would raise concerns on doctored surveillance camera footage, impairing the use of FR in criminal investigations. This critical risk resulted in the need for inclusion of deepfake detection in the Anti-Spoofing Analysis step (Section 1.4), especially for FR forensic systems. Described as the arms race against deep-

fake trend, it is still an ongoing research problem with limited success [80, 81]. Deepfake detection, along with the various anti-spoofing researches, helps to improve the robustness of a FR system.

Unlike major FR systems that could leverage on large-scale GPU computation and storage capabilities in centralized servers, the use of FR in edge devices such as in mobile phones and tablets have constraints on limited resources. Hence, with the prevalent usage in mobile applications in recent years, the need for high accuracy light-weight deep FR algorithms are on the rise. Some examples of these lightweight algorithms include MobiFace (2019) [82] and AirFace (2019) [83]. In 2019, a Lightweight Face Recognition Challenge was held to evaluate the FR DCNNs based on their model compactness and computation efficiency [84]. Coupled with the need for privacy-preserving FR, federated learning, which is invented by Google researchers in 2016 [85], is a technique of learning a shared machine learning model collaboratively while keeping all the training data on the source device. As such, the trend towards efficient and privacy-enhanced federated learning (PEFL) algorithm designs for mobile FR is evident.

The U.S. Defense Advanced Research Projects Agency (DARPA) initiated the XAI program aimed at researching machine learning techniques that produce more explainable models that are understandable by human users to increase the trust and adaptability of these systems [86]. Google's launch of its own XAI service [87], leads the trend of XAI in commercial markets. Given new laws and regulations on use of FR worldwide, Explainable FR as described in Chapter 4 is undeniably a tradeoff problem that FR researchers have to address.

As mentioned in Section 1.1, commercial facial recognition are predominately performed on 2D facial images. As a result, most FR researches are also focusing on 2D images. This inclination could change as modern sensors become increasingly capable of acquiring both 2D and facial shape (3D) information. Examples of Microsoft and Apple filing patents for their 3D facial recognition software and hardware [88], justify the trend towards 3D FR. When the availability of 3D facial data increases, deep feature extraction on 3D data could potentially lead to significant accuracy improvement due to their lighting and pose invariant properties [89, 90]. Consequently, there would likely be more studies on methods using a hybrid of 2D and 3D data to build a multi-modal FR system [91] as well.

6. Conclusion

In the introduction chapter of this directed reading review, an overview of the topic of facial recognition was given, followed by the motivation behind FR researches as a result of the market widespread adoption. The various FR-related terminologies, especially the differences between face identification and face verification, were also defined and compared. The last section in the chapter gave a brief overview of the components in a modern, robust FR system design.

The chapter on technical approaches dived into the progression of facial recognition. The pioneering works and significant milestones of FR research were documented in the section on traditional approaches. The rationale, pros and cons of each historical works were listed and discussed. The subsequent section on state-of-the-art approaches revealed the trinity in deep FR; namely the dataset size, the backbone network architecture and the choice of loss function. With academia only having access to smaller publicly available datasets than big commercial companies, many accuracies surpassing algorithms were developed through modification of network architectures and loss functions to overcome this constraint.

One of the most influential papers in FR research reviewed in Chapter 3, was developed by the researchers from Google Inc titled "FaceNet: A Unified Embedding for Face Recognition". By posing FR as a similarity learning problem instead of a classification problem, the authors demonstrated the need for formulating appropriate loss function for the issue at hand. The introduction of the Triplet Loss function was inspirational to the development of many state-of-the-art loss functions for deep face recognition.

However, even with its boast of human-surpassing accuracy, FR is not without a fault. The chapter on technical challenges enumerated numerous open challenges of the FR domain. From conventional computer vision problems to deep learning-related issues, many resulted in detrimental impacts on the robustness of FR systems.

In the last chapter, the future trends of FR research were exemplified and reasoned. The five trends described include accuracy improvement for mask occlusion, deepfake and anti-spoofing detection, lightweight DCNN and PEFL algorithms, XAI and 3D face recognition.

References

[1] Biometrics Institute. *Types of Biometrics - Biometrics Institute*. 2018. URL: https://www.biometricsinstitute.org/what-is-biometrics/types-of-biometrics/.

[2] Thales Group. *Facial recognition in 2020 (7 trends to watch)*. Aug. 2020. URL: https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition.

[3] Patrick Grother et al. *Face Recognition Vendor Test (FRVT) Part 2: Identification*. US Department of Commerce, National Institute of Standards and Technology, 2019.

[4] Jacqueline G Cavazos et al. "Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?" In: *arXiv preprint arXiv:1912.07398* (2019).

[5] MarketsandMarkets. *Facial Recognition Market worth $7.0 billion by 2024*. June 2019. URL: https://www.marketsandmarkets.com/PressReleases/facial-recognition.asp.

[6] Luana Pascu. *Global biometrics-as-a service to surpass $10B by 2030, contactless biometrics to top $18B by 2026*. Aug. 2020. URL: https://www.biometricupdate.com/202008/global-biometrics-as-a-service-to-surpass-10b-by-2030-contactless-biometrics-to-top-18b-by-2026.

[7] Mei Wang and Weihong Deng. "Deep Face Recognition: A Survey". In: *CoRR* abs/1804.06655 (2018).

[8] Anil K Jain, Karthik Nandakumar, and Arun Ross. "50 years of biometric research: Accomplishments, challenges, and opportunities". In: *Pattern recognition letters* 79 (2016), pp. 80–105.

[9] Woodrow Wilson Bledsoe. "The model method in facial recognition". In: *Panoramic Research Inc., Palo Alto, CA, Rep. PR1* 15.47 (1966), p. 2.

[10] Woodrow Wilson Bledsoe and Helen Chan. "A man-machine facial recognition system—some preliminary results". In: *Panoramic Research, Inc, Palo Alto, California., Technical Report PRI A* 19 (1965), p. 1965.

[11] W Bledsoe. "Man-machine facial recognition: Report on a large-scale experiment, panoramic research". In: *Inc, Palo Alto, CA* 2 (1966).

[12] WW Bledsoe. "Some results on multicategory pattern recognition". In: *Journal of the ACM (JACM)* 13.2 (1966), pp. 304–316.

[13] Woodrow Wilson Bledsoe. "Semiautomatic facial recognition". In: *Technical report sri project 6693* (1968).

[14] Michael Ballantyne, Robert S Boyer, and Larry Hines. "Woody bledsoe: His life and legacy". In: *AI magazine* 17.1 (1996), pp. 7–7.

[15] Shaun Raviv. *The Secret History of Facial Recognition*. Jan. 2020. URL: https://www.wired.com/story/secret-history-facial-recognition/.

[16] Matthew Turk and Alex Pentland. "Eigenfaces for recognition". In: *Journal of cognitive neuroscience* 3.1 (1991), pp. 71–86.

[17] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman. "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection". In: *IEEE Transactions on pattern analysis and machine intelligence* 19.7 (1997), pp. 711–720.

[18] Paul Viola and Michael Jones. "Rapid object detection using a boosted cascade of simple features". In: *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. Vol. 1. IEEE. 2001, pp. I–I.

[19] David G Lowe. "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60.2 (2004), pp. 91–110.

[20] Jian Wu et al. "A Comparative Study of SIFT and its Variants". In: *Measurement science review* 13.3 (2013), pp. 122–131.

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[22] Gary B Huang et al. "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments". In: 2008.

[23] Yaniv Taigman et al. "Deepface: Closing the gap to human-level performance in face verification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1701–1708.

[24] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 815–823.

[25] Yandong Guo et al. "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition". In: *European conference on computer vision*. Springer. 2016, pp. 87–102.

[26] Qiong Cao et al. "Vggface2: A dataset for recognising faces across pose and age". In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE. 2018, pp. 67–74.

[27] Brianna Maze et al. "Iarpa janus benchmark-c: Face dataset and protocol". In: *2018 International Conference on Biometrics (ICB)*. IEEE. 2018, pp. 158–165.

[28] Fei Wang et al. "The Devil of Face Recognition is in the Noise". In: *arXiv preprint arXiv:1807.11649* (2018).

[29] Arman Savran et al. "Bosphorus database for 3D face analysis". In: *European workshop on biometrics and identity management*. Springer. 2008, pp. 47–56.

[30] Lior Wolf, Tal Hassner, and Itay Maoz. "Face recognition in unconstrained videos with matched background similarity". In: *CVPR 2011*. IEEE. 2011, pp. 529–534.

[31] Tianyue Zheng, Weihong Deng, and Jiani Hu. "Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments". In: *arXiv preprint arXiv:1708.08197* (2017).

[32] Tianyue Zheng and Weihong Deng. "Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments". In: *Beijing University of Posts and Telecommunications, Tech. Rep* 5 (2018).

[33] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. "Learning deep models for face anti-spoofing: Binary or auxiliary supervision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 389–398.

[34] Kimmo Kärkkäinen and Jungseock Joo. "Fairface: Face attribute dataset for balanced race, gender, and age". In: *arXiv preprint arXiv:1908.04913* (2019).

[35] Asifullah Khan et al. "A survey of the recent architectures of deep convolutional neural networks". In: *Artificial Intelligence Review* (2020), pp. 1–62.

[36] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[37] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[38] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[39] Jie Hu, Li Shen, and Gang Sun. "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.

[40] Jiankang Deng et al. "Arcface: Additive angular margin loss for deep face recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4690–4699.

[41] Weiyang Liu et al. "Sphereface: Deep hypersphere embedding for face recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 212–220.

[42] Hao Wang et al. "Cosface: Large margin cosine loss for deep face recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 5265–5274.

[43] Feng Wang et al. "Normface: L2 hypersphere embedding for face verification". In: *Proceedings of the 25th ACM international conference on Multimedia*. 2017, pp. 1041–1049.

[44] Azmath Moosa. *A Comprehensive Guide to Facial Recognition Algorithms - Part 2*. 2018. URL: https://www.baseapp.com/computer-vision/a-comprehensive-guide-to-facial-recognition-algorithms-part-2/.

[45] Matthew D Zeiler and Rob Fergus. "Visualizing and understanding convolutional networks". In: *European conference on computer vision*. Springer. 2014, pp. 818–833.

[46] Yi Sun, Xiaogang Wang, and Xiaoou Tang. "Deeply learned face representations are sparse, selective, and robust". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 2892–2900.

[47] Alexander Hermans, Lucas Beyer, and Bastian Leibe. "In defense of the triplet loss for person re-identification". In: *arXiv preprint arXiv:1703.07737* (2017).

[48] Swami Sankaranarayanan et al. "Triplet probabilistic embedding for face verification and clustering". In: *2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS)*. IEEE. 2016, pp. 1–8.

[49] Swami Sankaranarayanan, Azadeh Alavi, and Rama Chellappa. "Triplet similarity embedding for face verification". In: *arXiv preprint arXiv:1602.03418* (2016).

[50] Fei-Fei Li, Justin Johnson, and Serena Yeung. *CS231n Lecture 2 — Image Classification*. Aug. 2017. URL: https://www.youtube.com/watch?v=0oUX-nOEjG0.

[51] Akshay Asthana et al. "Fully automatic pose-invariant face recognition via 3D pose normalization". In: *2011 International Conference on Computer Vision*. IEEE. 2011, pp. 937–944.

[52] Mostafa Mehdipour Ghazi and Hazim Kemal Ekenel. "A comprehensive analysis of deep learning based representation for face recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2016, pp. 34–41.

[53] Guido Borghi et al. "Face-from-depth for head pose estimation on depth images". In: *IEEE transactions on pattern analysis and machine intelligence* (2018).

[54] Vinay Bettadapura. "Face expression recognition and analysis: the state of the art". In: *arXiv preprint arXiv:1203.6722* (2012).

[55] Wenfeng Chen et al. "Facial expression at retrieval affects recognition of facial identity". In: *Frontiers in psychology* 6 (2015), p. 780.

[56] John Wright et al. "Robust face recognition via sparse representation". In: *IEEE transactions on pattern analysis and machine intelligence* 31.2 (2008), pp. 210–227.

[57] Luan Tran, Xi Yin, and Xiaoming Liu. "Disentangled representation learning gan for pose-invariant face recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1415–1424.

[58] Bowen Zhang, Benedetta Tondi, and Mauro Barni. "Adversarial examples for replay attacks against CNN-based face recognition with anti-spoofing capability". In: *Computer Vision and Image Understanding* (2020), p. 102988.

[59] Yinpeng Dong et al. "Efficient decision-based black-box adversarial attacks on face recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 7714–7722.

[60] Javier Galbally, Sébastien Marcel, and Julian Fierrez. "Biometric antispoofing methods: A survey in face recognition". In: *IEEE Access* 2 (2014), pp. 1530–1552.

[61] Xiao Yang et al. "Face anti-spoofing: Model matters, so does data". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 3507–3516.

[62] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples". In: *arXiv preprint arXiv:1412.6572* (2014).

[63] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. "Adversarial examples in the physical world". In: *arXiv preprint arXiv:1607.02533* (2016).

[64] Mahmood Sharif et al. "Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition". In: *Proceedings of the 2016 acm sigsac conference on computer and communications security*. 2016, pp. 1528–1540.

[65] Qing Song, Yingqi Wu, and Lu Yang. "Attacks on State-of-the-Art Face Recognition using Attentional Adversarial Attack Generative Network". In: *CoRR* abs/1811.12026 (2018). arXiv: 1811.12026. URL: http://arxiv.org/abs/1811.12026.

[66] Fabio Valerio Massoli et al. "Detection of Face Recognition Adversarial Attacks". In: *arXiv preprint arXiv:1912.02918* (2019).

[67] Yaoyao Zhong and Weihong Deng. "Towards Transferable Adversarial Attack against Deep Face Recognition". In: *arXiv preprint arXiv:2004.05790* (2020).

[68] Paul VoosenJul. 6, 2017, and 2:00 Pm. "How AI detectives are cracking open the black box of deep learning". In: *Science — AAAS* (July 2017). URL: https://www.sciencemag.org/news/2017/07/how-ai-detectives-are-cracking-open-black-box-deep-learning.

[69] P Jonathon Phillips and Mark Przybocki. "Four Principles of Explainable AI as Applied to Biometrics and Facial Forensic Algorithms". In: *arXiv preprint arXiv:2002.01014* (2020).

[70] Yu-Sheng Lin et al. "xCos: An Explainable Cosine Metric for Face Verification Task". In: *arXiv preprint arXiv:2003.05383* (2020).

[71] Steven G Young et al. "Perception and motivation in face recognition: A critical review of theories of the cross-race effect". In: *Personality and Social Psychology Review* 16.2 (2012), pp. 116–142.

[72] Joy Buolamwini and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification". In: *Conference on fairness, accountability and transparency*. 2018, pp. 77–91.

[73] Joseph P Robinson et al. "Face recognition: too bias, or not too bias?" In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, pp. 0–1.

[74] Michael McLaughlin and Daniel Castro. *The Critics Were Wrong: NIST Data Shows the Best Facial Recognition Algorithms Are Neither Racist Nor Sexist*. Jan. 2020. URL: https://itif.org/publications/2020/01/27/critics-were-wrong-nist-data-shows-best-facial-recognition-algorithms.

[75] Zekeriya Erkin et al. "Privacy-preserving face recognition". In: *International symposium on privacy enhancing technologies symposium*. Springer. 2009, pp. 235–253.

[76] Kevin W Bowyer. "Face recognition technology: security versus privacy". In: *IEEE Technology and society magazine* 23.1 (2004), pp. 9–19.

[77] Bruce Schneier. "We're banning facial recognition. we're missing the point". In: *The New York Times* (2020).

[78] Mei L Ngan, Patrick J Grother, and Kayee K Hanaoka. "Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms". In: (2020).

[79] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[80] Ekraam Sabir et al. "Recurrent convolutional strategies for face manipulation detection in videos". In: *Interfaces (GUI)* 3.1 (2019).

[81] Yisroel Mirsky and Wenke Lee. "The Creation and Detection of Deepfakes: A Survey". In: *arXiv preprint arXiv:2004.11138* (2020).

[82] Chi Nhan Duong et al. "Mobiface: A lightweight deep learning face recognition on mobile devices". In: *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE. 2019, pp. 1–6.

[83] Xianyang Li et al. "Airface: lightweight and efficient model for face recognition". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2019, pp. 0–0.

[84] Jiankang Deng et al. “Lightweight face recognition challenge”. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2019, pp. 0–0.

[85] Jakub Konečn et al. “Federated learning: Strategies for improving communication efficiency”. In: *arXiv preprint arXiv:1610.05492* (2016).

[86] David Gunning. “Explainable artificial intelligence (xai)”. In: *Defense Advanced Research Projects Agency (DARPA), nd Web* 2 (2017), p. 2.

[87] Google Cloud. *Explainable AI*. 2019. URL: https://cloud.google.com/explainable-ai.

[88] Biometric Update and Chris Burt. *Microsoft patent filing shows new 3D facial biometric technology with multi-spectral camera array*. Dec. 2019. URL: https://www.biometricupdate.com/201912/microsoft-patent-filing-shows-new-3d-facial-biometric-technology-with-multi-spectral-camera-array.

[89] Thibault Napoléon and Ayman Alfalou. “Pose invariant face recognition: 3D model from single photo”. In: *Optics and Lasers in Engineering* 89 (2017), pp. 150–161.

[90] Song Zhou and Sheng Xiao. “3D face recognition: a survey”. In: *Human-centric Computing and Information Sciences* 8.1 (2018), p. 35.

[91] Sima Soltanpour and Qingming Jonathan Wu. “Multimodal 2D–3D face recognition using local descriptors: pyramidal shape map and structural context”. In: *IET Biometrics* 6.1 (2016), pp. 27–35.