Presentation-Level Privacy Protection Techniques for Automated Face Recognition - A Survey

MD REZWAN HASAN*, School of Engineering, University of Kent, UK
RICHARD GUEST, School of Engineering, University of Kent, UK
FARZIN DERAVI, School of Engineering, University of Kent, UK

The use of Biometric Facial Recognition (FR) Systems have become increasingly widespread, especially since the advent of deep neural network-based architectures (DNNs). Although FR systems provide substantial benefits in terms of security and safety, the use of these systems also raises significant privacy concerns. This paper discusses recent advances in facial identity hiding techniques, focusing on privacy protection approaches that hide or protect facial biometric data before camera devices capture the data. Moreover, we also discuss the state-of-the-art methods used to evaluate such privacy protection techniques. The primary motivation of this survey is to assess the relative performance of facial privacy protection methods and identify open challenges and future work that needs to be considered in this research area.

CCS Concepts: • **Security and privacy → Privacy protections**; *Social aspects of security and privacy*; **Usability in security and privacy**.

Additional Key Words and Phrases: Biometrics, Privacy Protection Techniques, Facial Identity Hiding, Face Detection-Recognition

1 INTRODUCTION

Biometric FR systems have been extensively used in many applications such as surveillance, payment authorisation and automatic border control systems. The fundamental goal of a conventional FR system is to authenticate or identify an individual from a captured image or frames of a video sequence. Early studies from the 1960s and 70s [4, 49] mainly described feature-based approaches to facial recognition systems, but over the past decade, the performance, scale and deployment of FR systems has developed enormously with the help of advanced deep learning techniques [43, 62, 70] as well as innovative cloud storage and processing facilities.

In parallel to the growth of FR technology, there are increasing concerns over the privacy of individual subjects whose data is captured using this technology [47]. The application of FR systems for mass surveillance and identity recognition without explicit consent or awareness of the data subjects are pose significant privacy concerns. Facial images are generally stored in databases whenever photos are shared on online social media platforms or saved on cloud services. There is a potential and growing threat to privacy from the misuse of such accumulated personal data.

It is considered a fundamental human right in many jurisdictions for people to be able to control their biometric data and ensure it is not used without explicit authorisation and consent. The use of FR system-based surveillance cameras to identify demonstrators at public gatherings, for example, can pose a serious threat to the lawful freedom of expression. However, during the protest that took place in Hong Kong [79] is also an example of controlling the freedom of general citizens. Public FR system-based cameras can also cause social harm if they

Authors' addresses: Md Rezwan Hasan, r.hasan@kent.ac.uk, School of Engineering, University of Kent, Canterbury, Kent, UK, CT27NT; Richard Guest, School of Engineering, University of Kent, Canterbury, Kent, UK, r.m.guest@kent.ac.uk; Farzin Deravi, School of Engineering, University of Kent, Canterbury, Kent, UK, f.deravi@kent.ac.uk.

0360-0300/2023/2-ART
https://doi.org/10.1145/3583135

incorrectly recognise subjects or demographic sub-groups. Big Brother Watch [2], a UK based civil liberties group, reported in 2018 that the facial recognition system used by the UK Metropolitan Police was only 2 percent accurate in identifying criminals at a carnival. Such errors may lead to innocent persons or groups of people becoming wrongly implicated in criminal investigation. Several incidents have also taken place in the United States when inaccurate FR technology was deployed by authorities to recognise suspects from security footage [27].

In the last few years, researchers have developed a range of solutions to counter the privacy challenges of FR systems. This paper outlines some of the major solutions proposed for *facial privacy protection (FPP)* against automated face detection and recognition (FD-FR) systems. In this survey, we classify existing FPP methods into two major categories depending on the nature of their implementation. FPP methods that are dependent on the design and mechanism of FR systems or on those people who control the FR systems are categorised under the title of *post-presentation-level FPP* methods. However, there are some methods for FPP where the data subjects themselves are in control of privacy protection, irrespective of the design and functioning of the FR system. In these FPP methods, a person's privacy is protected by intervention prior to image capture, using different privacy protection artefacts (PPA) such as intelligent wearable devices, adversarial patches/stickers, face masks, and makeup. We categorised these methods under the title of *presentation-level FPP* methods. This designation indicates that the PPAs are in use when the data subject is presented to the FR systems. The paper is focused on presentation-level FPP methods, their current limitations and future possibilities. The major contributions of this paper are as follows:

- A comprehensive review of recent and significant studies on facial privacy protection techniques and the categorisation of systems into two implementation methodologies.
- An in-depth assessment of presentation-level FPP techniques.
- A summary of the challenges and limitations of these methods and prospects for future research.

The remainder of the paper is structured as follows:

We briefly present an overview of biometric FR systems in Section 2. In Section 3, the major categories of facial privacy protection techniques are briefly explained. Then we analyse the significant works on presentation-level FPP approaches in Section 4. Finally, Section 5 draws a range of conclusions and suggests possible future research opportunities.

2 OVERVIEW OF FACE DETECTION AND RECOGNITION SYSTEMS

In this section, we briefly discuss how facial detection and recognition systems work and how the mechanism of FD-FR systems has changed over the years from conventional techniques to advanced deep learning-based techniques. Because the main focus of this paper is on facial privacy protection techniques against FD-FR systems, it is useful to have an appreciation of the structure and mechanism of FD-FR systems in that way it can help us to identify the vulnerable parts of FD-FR systems that can be circumvented to hide the facial identity of the users. For any FR system, the facial detection process is the initial step, therefore preventing the detection step or reducing the detection accuracy of an FD system may automatically lead to an unsuccessful facial recognition process. A brief overview of FD methods is discussed in the below section.

2.1 Face Detection Methods

An FD system locates the faces in a photo and provides the coordinates of a bounding box for each face if any are found. FD systems generally initiate its process by looking for eyes, which are one of the easier attributes to identify in a human face, before moving on to detecting brows, mouth, nose, and other key facial features. The detection process varies quite a lot due to the diverse techniques used by different FD algorithms. The progress of FD systems, especially in the security sector, has advanced through the employment of evolutionary

FD algorithms like the *Viola-Jones object detection framework* [67]. The Viola-Jones FD algorithm is dependent on the following three major concepts:

- `Haar-like features` [67] of an image are extracted using an image modelling technique called integral image.
- The `AdaBoost ML` technique is utilised to nominate the essential features to identify the face.
- `Cascade Classifier (multilevel classifier)` is used to remove the insignificant areas of an image promptly.

The FD algorithm based on the *Histogram of Oriented Gradients (HOG)* [10] is another benchmark algorithm in this area. Although the accuracy of the aforementioned algorithms has remained constrained on challenging images that have diverse variation factors such as scale, expression, pose, illumination and occlusion, there are numerous conventional FD methods reported in the literature [78] before the deep learning-based (DL) FD algorithms were introduced. When face datasets like *WIDER FACE* [74] (which has the mentioned challenging factors) were introduced to evaluate the FD algorithms, the accuracy of popular FD systems has decreased significantly. For instance, the accuracy of the Viola-Jones FD algorithm in the WIDER FACE dataset is only 41.20%, and for the HOG FD algorithm, it's also only 39%. However, the advanced development of DL-based algorithms in the computer vision research area has impacted the improvement of FD systems over the last decade.

There are many methods proposed for FD systems using various DL models to date. A comprehensive survey undertaken by Minaee et al. [40] categorised the major works on recent DL-based FD models into the following groups:

- `CNN-Cascade Based` [35].
- `R-CNN Based` [8].
- `Single Shot Detector` [42].
- `Feature Pyramid Network` [36].
- `Other Models` (the FD models which do not fall in the aforementioned category).

Since there are already several surveys reported in the literature on DL-based FD algorithms [40], so in this section, we briefly discuss two of the state-of-the-art (SOTA) FD algorithms. At first, we choose the MTCNN (Multi-Task Cascaded Convolutional Neural Network) FD algorithm [79] because it is mainly used to evaluate the reviewed presentation-level FPP systems in the later section. Then we choose SCRFD (Sample and Computation Redistribution for Efficient Face Detection) FD algorithm because it is one of the best performed FD algorithms reported until May 2021 by Guo et al. [21].

- `MTCNN`: The MTCNN FD algorithm [79] is one of the most popular FD algorithms in this research area. This method is an extended version of the CNN Cascade-based FD algorithm [35], which performs at several resolutions, swiftly discards the background areas of the low-resolution phases, and thoroughly measures the high-resolution phase candidates in the last step. The MTCNN method performs face prediction and landmark location in three phases: P-Net, R-Net and O-Net. The FD accuracy of this algorithm is 85.10% on the SOTA face dataset WIDER FACE.
- `SCRFD`: The SCRFD FD algorithm [21] identified two key aspects of an appropriate FD system: the sampling of training data and distribution techniques of computation. From these identifications, two approaches are proposed. In one of the approaches termed Sample Redistribution (SR), the training data are expanded for the essential phases on the basis of standard dataset information. In another approach termed Computation Redistribution (CR), three crucial parameters of the FD models: the backbone, head and neck, are redistributed for computation on the basis of a precise search technique. This algorithm claimed the highest FD accuracy until published, which is 96.06% on the face dataset WIDER FACE.

2.2 Face Recognition Methods

An FR system analyses the facial information with a database of known faces to identify a match using biometrics to map facial characteristics from an image or video. FR systems are typically comprised of four main steps: Face Detection, Alignment, Representation and Matching. There are numerous methods and algorithms developed over the years for facial recognition systems. The FR methods can be classified into two major categories based on the use of features in FR, before and after the revolution of deep learning (DL) methods: *Conventional FR Methods and Deep Learning-based FR Methods*. This section briefly discusses the two major categories of the FR methods.

2.2.1 Conventional FR Methods: Conventional FR systems have two main stages: training and testing. In the training stage, the input data is pre-processed and then feature extraction methods are used to obtain facial features. After that, the features are stored in a feature template. In the testing stage, the input face data is pre-processed, and the features are extracted in the same manner as the training stage. Extracted features of the testing data are matched with the stored features in the feature template. Therefore, `Fig. 1` shows the major steps of a conventional FR system: input data pre-processing, feature extraction, and feature matching.

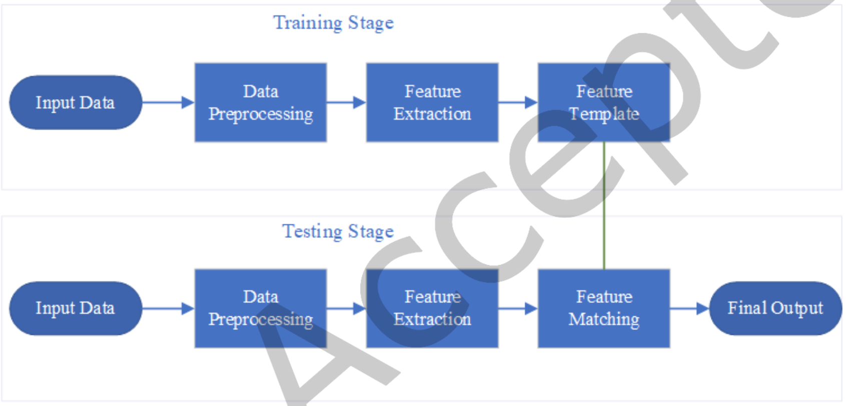


Fig. 1. Block Diagram of a Conventional FR System

In the early stage of the face recognition research, the primary focus was on those techniques which used image processing methods to compare basic features for the topology of facial landmarks. In [22] and [7], the authors tried to detect the location of facial landmarks and to evaluate relative distances and positions among them by using customised contour detectors and edges. In 1993, Brunelli et al. [6] proposed a method that compared gradient images instead of geometrical images, showing improved recognition performance. However, the geometric feature-based methods required less memory and were faster in recognising faces. In [57], the potentiality of employing facial landmarks and their geometry in FR systems was studied comprehensively. When the information of depth was computed in 3D landmarks, then the methods based on geometry became more efficient in 3D FR systems [11, 23]. Later, researchers used the whole face region as input for FR systems, which we call holistic FR methods. Alternatively, the development in computer vision techniques led the researchers to use features that can characterise the image textures at various locations. This kind of feature-based technique is used to match the local features of the images for FR. Furthermore, the feature-based and holistic approaches were combined to improve performance which we call the combined or hybrid methodology for FR systems. These were the conventional methods primarily used in FR systems until the rise of DL-based methods.

2.2.2 Deep Learning-based FR Methods: The use of the convolutional neural network (CNN) and deep learning for facial recognition approaches have significantly impacted performance advancement and growth in the past few years. The performance of FR systems has increased to an extraordinary level [13, 37, 59, 60, 62, 69] with the improvement of highly developed architectures and discriminatory learning techniques. The DL methods have the advantage to be trained with substantial amounts of data to determine an FR system that is stable against different presentations of the training data.

A convolutional neural network mainly comprises three layers: convolutional layers, pooling layers and fully connected (FC) layers. A convolutional layer aims to extract facial features from the given data. The convolutional layer executes the convolution process with a filter kernel and employs a nonlinear transfer operation. The purpose of the pooling layers is to minimise the area of the feature maps by incorporating the results of one layer neuron clusters into a particular neuron in the subsequent layer [20]. The efficiency of FR systems has improved enormously when this kind of CNN-based feature representation procedures are used in the FR systems. DL-based methods also have the same two stages as the conventional methods. In the training stage, the input data is pre-processed either by resizing or changing the alignment or by performing any other necessary steps to create a unified feature map and adapt the input tensor of the neural network architecture such as the number of images, height, width and other parameters. Then, the features are processed in the different layers of the deep CNN architecture. In the testing stage, the input data is similarly pre-processed as the training stage, and then the processed features from the training stage CNN are fetched and compared with a provided feature gallery. Fig. 2 is a generic representation of a deep learning-based face recognition system.

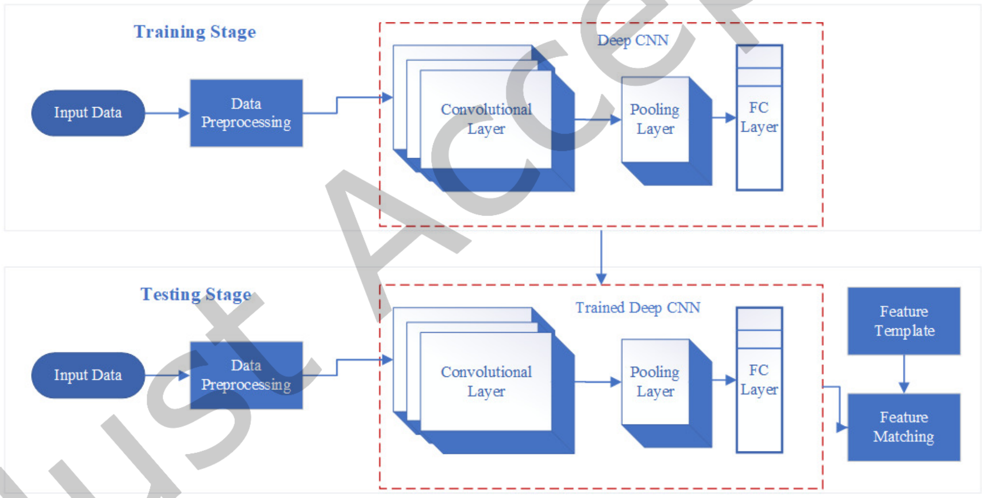


Fig. 2. Block Diagram of a Deep Learning-based FR System

Most of the conventional methods couldn't achieve ground-breaking results due to their low computational capability and moderately small databases available for training at the earlier stage of FR systems. It started changing when one of the first CNN-based methods for the FR system was introduced by Facebook, named DeepFace [62], which used a high power model and obtained an accuracy of 97.35% on the LFW face dataset [28]. At the same time, the DeepID method [59] attained similar results by training several CNNs on patches comprised of different regions, scales and RGB channels. FR systems based on CNN's are motivated mainly by those methods which obtained high accuracy on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [48]. For instance, one of the VGG network [58] versions with 16 layers was applied in [22], and an identical type of network was utilised in [75], which was smaller. GoogleNet style networks [61] and VGG style networks [58] are two various kinds of CNN based networks that were explored in [51], which obtained a high amount of accuracy. In the last few years, Deep Residual Networks (ResNets) [26] have emerged as one of the most favourable choices in several recognition-based works along with FR systems [13, 25, 37]. The primary uniqueness of ResNets is the initiation of a building block that utilises a bypass connection to understand the residual mapping. These kinds of bypass connections let the training of significantly deeper architectures as they assist the flow of info throughout layers. One of the best balances among speed, accuracy and model size was attained with ResNet consisting of 100-layer through a residual block comparable to the proposed method in [73].

3 FACIAL PRIVACY PROTECTION METHODS

Facial privacy protection (FPP) methods attempt to hide the identity of a person to protect their privacy from FR systems. Numerous techniques have been developed to protect the privacy of humans from FR systems. Mainly, these methods try to prevent the detection or identification of person faces either at the presentation level (the image captured by FR system-based devices) or by post-processing the image using algorithms or applications post-capture. Depending on the nature of their implementation, we already mentioned that the FPP methods are categorised into two major classes: *presentation-level FPP* methods and *post-presentation-level FPP* methods. The presentation-level FPP methods can be divided into three more subcategories based on the application of the protection system: *active, passive and other* presentation-level FPP methods. On the other hand, the post-presentation level FPP methods are divided into three subcategories based on the facial identity hiding mechanism of the face recognition process is performed. Those are *image level, feature extraction level and decision level FPP* methods. `Fig.` 3 shows the different categories of FPP methods that we discuss in this paper. These methods are briefly explained in the following sub-section.

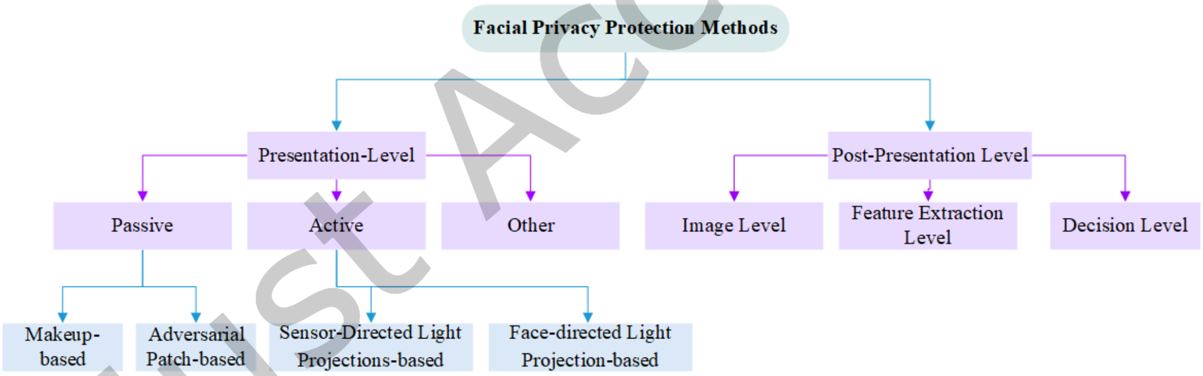


Fig. 3. Different categories of facial privacy protection methods

3.1 Post-Presentation Level Face Privacy Protection (FPP) Systems

In post-presentation level methods, the identifiable facial features of a captured photo or video are protected by using different kinds of image processing-based techniques such as deidentification, anonymisation, obfuscation, blurring, etc. It is already mentioned that the post-presentation level methods are categorised into three more levels. Therefore, the primary characteristics of the three levels are briefly described below:

- `Image-Level FPP Methods`: Image-level methods aim to protect privacy by changing the data visualised using either adversarial attacks or obfuscation or different image processing techniques. Recent works on facial privacy protection techniques are mainly on the privacy issue of facial data visualisation, which can be managed at the image level. Smart privacy protection cameras such as TrustCAM [71], PrivacyCam [7], AnonymousCam [80], De-Identification Camera [41] are some of the examples of this category where privacy is protected at the image level by deploying different image processing techniques on image sensors.

- Feature Extraction Level FPP Methods:Feature extraction level methods intend to protect sensitive information when the data is extracted from the stored feature templates, and it also ensures that the data is not used for any other purposes. These methods can also be categorised into three more categories based on their processing techniques. Those are homomorphic encryption, elimination and transformation-based techniques. The homomorphic encryption-based methods aim to protect data so that only certain predefined tasks like comparing feature templates without decrypting data can be performed in an encrypted domain. This method also works in conjunction between privacy protection and data security. While the elimination-based techniques try to remove the most sensitive features from the original biometric data, the transformation-based methods try to restrain some part of the data by altering the main data into a different form. One of the examples of transformation-based feature extraction-level technique is proposed by Feutry et al. [15], where the facial identity is protected in this method, but face representation can be learned for facial expression detection.
- Decision Level FPP Methods: In this category, privacy protection techniques are implemented during the feature classification or decision-making process. These methods also try to ensure that the purpose of using this data is for the right intentions, such as the PE-MIU method [64]. These methods mainly transform the feature classification process used to find a comparison result in the FR system. The decision level FPP methods can be classified into two types. Those are privacy enhancement at the design level and post-design level. Design-level privacy enhancement methods try to insert an extra layer of privacy at the time of FR system design to ensure that the data is only used for face recognition purposes. On the other hand, there are many FR systems designed without ensuring privacy in the previous decades. So, the Post-design-level privacy enhancement methods try to add an extra layer in the existing FR system design to ensure privacy by following some techniques of the feature extraction-level methods. These approaches were recently introduced in comparison to the other level methods. One example of this method was proposed recently in 2021 [63].

3.2 Presentation-Level Face Privacy Protection (FPP) Systems

In presentation-level methods, the users can protect their identity before their image is captured by any device. It can be achieved by wearing clothes or makeup or attaching adversarial patches in the face, or using some kind of hardware device that obfuscates the users' identifiable features required to perform identification. Fig. 3 and its description show that the presentation-level FPP system categories are passive, active and other methods. The passive methods don't require interactive participation from users or the FPP system to protect their privacy. In contrast, the active methods need at least one activity from the users or the system for privacy protection. Whereas in the other category, we mention those FPP methods found in the random news articles or exhibited in the events as a privacy protection concept but without revealing their evaluation techniques or further information. An overview of the existing presentation-level FPP methods is discussed in the next section.

4 OVERVIEW OF PRESENTATION-LEVEL FACIAL PRIVACY PROTECTION (FPP) SYSTEMS

As already mentioned, presentation-level FPP methods have the advantage of keeping the users' privacy under their control. This review tried to focus explicitly on the characteristics, benefits, and limitations of the existing FPP methods proposed over the last decade. Presentation-level FPP methods use artefacts, which may be called *Privacy Protection Instruments (PPI)*, to prevent the detection of the face and/or the recognition of the correct identity. We specifically concentrated on those methods which have published their evaluation performance and/or generation procedure. Later, we also mention some interesting approaches that did not provide detailed information about their evaluation performance and/or generation procedure. The FPP methods are discussed briefly in this section according to the major categories defined in the previous section.

4.1 Passive FPP Methods

Passive presentation-level FPP methods use artefacts that, once made and applied, do not change with time and may be attached to or worn on different parts of the facial area as a static method to hide the identity against FR systems. Several passive approaches have been proposed to hide or protect a person's identity in the last decade. Some passive methods use an adversarial patch-based approach where small elements (maybe printed paper) are attached to the face, some use other artefacts such as eyeglasses or scarfs, and some use makeup to hide the facial features of a person from FR systems in a real-life scenario. The majority of the FPP methods proposed in recent years are passive. Based on the use of such artefacts to protect against detection and recognition by FD and FR systems, the passive FPP methods can be grouped into two primary levels: makeup-based and adversarial patch-based methods. The relevant published literature for the passive FPP methods is briefly explained and reviewed in the following sub-sections.

4.1.1 Makeup-based FPP Methods: Makeup-based facial privacy protection methods use makeup to modify the anticipated features of the face that are targeted by the FD algorithms to detect a face in an image. For instance, the light areas of the face are transformed to dark and the dark areas into light by using different makeup designs that help break the pattern of facial features required for facial detection. These methods are cheaper to produce due to the availability of low-cost makeup items worldwide. But one of the significant drawbacks of this method is that it can make someone highly noticeable to other people if the makeup is used unnaturally or excessively to make the privacy protection system work. Although the utilisation of makeup-based methods in a real-world scenario is less explored by the researchers, there are some methods reported in the literature.

In 2010, Harvey et al. [24] proposed one of the first concepts called CV-Dazzle (Computer Vision Dazzle Camouflage) to disguise someone from the unauthorised FD systems used mainly in surveillance technologies. The concept and the CV-Dazzle name were inspired by a military camouflage incident in World War I called Dazzle, where ships used in battles were designed following Cubist-Art techniques to separate the perceptible continuousness of the ships and disguise their size and orientation. In the CV-Dazzle approach, to separate the continuousness of the facial features, different makeup designs and hairstyles are used by following the Avant-garde technique [3]. One of the most popular face detection systems of that time, the OpenCV [67] computer vision framework, is used to evaluate the proposed system. OpenCV FD system is built on the Viola-Jones FD algorithm [67], which uses a cascade structure of the Haar-like features.

For the evaluation, five different looks were designed for five individuals by following the above-mentioned techniques to test the system. The results showed almost 99% accuracy in blocking the face. But this concept was developed by targeting only the Viola-Jones Haar Cascade FD system in 2010 when the deep convolutional neural network (DCNN) based FD systems were rarely used. `Fig. 4` is an example of the CV-Dazzle concept design. In May 2021, Yin et al. [76] proposed a facial privacy protection technique based on adversarial makeup (Adv-Makeup) against FR systems. In this approach, a synthetic makeup generation method is developed to attach natural-looking eye shadow over the eye area of face images and later transfer it as physical makeup when the digital makeup achieves the desired success rate. For the makeup generation, generative adversarial network (GAN) [77] based technique is utilised. An attack loss function is proposed for the adversarial makeup by combining the style and content loss functions motivated by the CNN-based image style transfer method [16].

This method is evaluated in both digital and physical scenarios. The digital experiments are performed on the LFW dataset [28] and a high-quality makeup face database, LADN [17]. FR systems used to attack the proposed

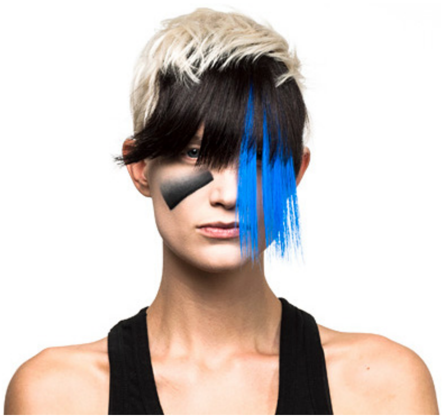


Fig. 4. CV-Dazzle [24]

systems are MobileFace [13] and FaceNet [51] in the digital domain. They compared the results with some of the popular adversarial patch-based techniques, which are Adv-Hat [33], Adv-Glasses [54], and Adv-Patch [5].

For the evaluation, they used the attack success rate (ASR) measure [12] to compare the results with other methods. For the physical scenario, they printed the digitally generated Adv-Makeup on paper and pasted the paper that looks like an eye shadow in the eye region of the face. The physical experiments are performed on two popular online FR systems, which are Face++ [38] and Microsoft Azure [39]. The experimental results in both physical and digital domains outperformed the methods used to compare. The ASR score in Face++ is above 75%, the highest among all other competitor methods in the physical scenario. The `Fig. 5` is an example of the Adv-Makeup effect on a face image where the left side images are without Adv-Makeup, and the right side is with Adv-Makeup.

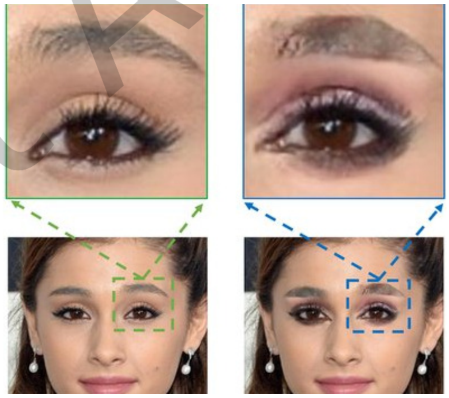


Fig. 5. Example of the Adv-Makeup technique [76]

In September 2021, Guetta et al. [18] proposed another Adv-Makeup-based technique to hide a person's identity from FR systems. In this approach, an adversarial machine learning (AML) based technique is used to apply makeup which looks natural on the face. There are two significant steps in this approach: (1) In the preparation step, this method tries to identify the key areas of the face which has the optimum impact on recognising a face by generating a heatmap using the backward gradient calculation of the triplet loss function proposed by Schroff et al. [51]. After that, makeup is applied digitally on the key areas identified by the heatmap repeatedly till the FR system misidentifies the face of the experimental user. (2) Then, in the final execution step, physical makeup is applied with the help of a professional makeup artist on the user's face by following the final digital makeup image, which successfully hides the face from the FR system.

They created a realistic surveillance environment for the evaluation by setting up cameras in a corridor and using different lighting conditions. They used 20 candidates while maintaining gender equality to evaluate the system. The FaceNet [51] FR system is used for training the data, and the ArcFace [13] FR system is used to evaluate the system. This method achieved 100% success in facial identity hiding for the digital set-up, which means none of the participants was identified correctly. In the physical scenario, the evaluation was performed in three phases. In the first phase, they evaluated without makeup which showed 47.57% accuracy in recognising the face, while in the second phase, random makeup was applied, which showed 33.73% accuracy and in the final phase, where the proposed AML-based method is applied, it performed the best with 1.22% accuracy. That means it achieved 98.78% success in the physical domain to protect the face from the FR system. From `Fig. 6`, we notice that the top image which is without makeup is recognised by the FR system as *Exp subject 15*. In the middle, the AML-based method is applied, and after applying the AML-based makeup, the face is recognised as *Unknown* in the bottom image.

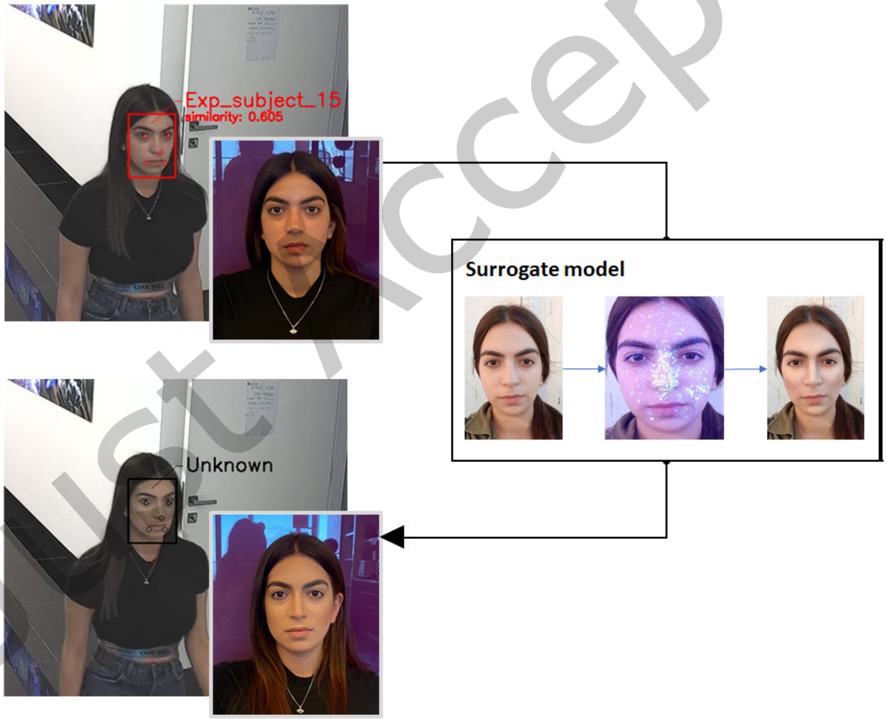


Fig. 6. Example of the AML-based makeup technique [18]

4.1.2 Adversarial Patch-based (Adv-Patch) FPP Methods: An adversarial patch (Adv-Patch) is an approach that is widely used to attack the FD and FR systems for facial identity hiding in both digital and physical domains.

These patches are mainly in printed papers generated by different image pixel perturbation techniques [54]. The patches are attached to different areas of the face to make the face detection system vulnerable. One of the earlier approaches to Adv-Patch based attacks on machine learning (ML) systems in the physical environment is introduced by Kurakin et al. [34] from Google Brain. However, this Adv-Patch based method was only used to attack ML-based object detectors, not for face detectors. But this technique has paved the way for many researchers to develop presentation-level FPP methods based on Adv-Patch. A recent study by Vakhshiteh et al. [66] covered topics on the Adv-Patch based attacks against FR systems but focused more on the techniques implemented in the digital domain. The Adv-Patch based methods are also cheaper to produce like the makeup-based FPP methods because a small printed paper can be used as an Adv-Patch. But this technique also has the issue of noticeability when we use it in a public place. Extensive research has been conducted focusing on the Adv-Patch based attacks on FD and FR systems in the digital domain. Hence, we tried to explore the Adv-Patch based presentation-level FPP methods, which are implemented in the real-world scenario.

Sharif et al. [53] proposed one of the first Adv-Patch based facial identity hiding methods in 2016. In this approach, a systematic methodology is developed to automatically generate a physically achievable Adv-Patch based attack on FR systems by printing eyeglass frames and attaching them to the face. This method is tested both on the white-box scenario (the architecture of the FR system is known by the attacker) and the black-box scenario (the architecture of the FR system is unknown). For the white-box attack, the VGG-Face CNN descriptor [43] is used, and a commercial FR system Face++ [38], is used in the black-box scenario, which has 97.3% accuracy on the SOTA face dataset LFW [28].

This system is evaluated with two sets of experiments. In one experiment, the evaluation was performed under various poses. The users were asked to stand a fixed distance from the camera, keep a neutral expression, and move their heads up-down, left-right, and in a circle. Most of the attempts succeeded to hide the face from FR systems with all video frames incorrectly classified. Even in the worst-case, 81% of video frames were wrongly classified. In the other experiment, they evaluated the effects of alterations to luminance. The results showed that it achieved 96% accuracy to hide the face from the FR system. One of the major drawbacks of this approach is it is tested only in controlled lighting conditions, not in the street, for surveillance purposes.

In April 2019, Thys et al. [65] proposed an identity hiding approach based on Adv-Patch to hide a person from the surveillance technologies. They generated a graphic print (adversarial patch) and attached it to a cardboard plate, and then the cardboard was held by a person to hide from the object detectors. The adversarial patch is generated by following an optimisation process on the image pixels proposed by Chen et al. [9], and the patch is made more robust by performing random transformations like rotating, scaling, adding noise and changing brightness randomly. One of the most popular object detectors, YOLOv2 [46], is used to test the system.

The system is evaluated both on a public dataset of images called Inria [10] and on a physical environment. They have used four kinds of scenarios for adversarial patches to evaluate the system. In the first scenario, named OBJ-CLS (Object Matching-Classification), the motive of the approach is to decrease the total object matching score and classification score. Here, the object matching score and classification score is predicted by the object detector. In the second case, named OBJ, only the object matching score is reduced, and in the CLS case, only the classification score is reduced. Random noise is used in the patch in the fourth case named NOISE. Average Precision (AP) is computed to measure the performance. So, in all four cases, the OBJ case performed the best where AP is reduced to 25.53% from 100% in detecting a person. `Fig. 7` is an example of a person holding a cardboard plate. We can notice from the Figure that the person on the left without holding a patch is successfully identified, and the camera device does not detect the person carrying the cardboard patch on the right side.

Researchers from Huawei Moscow Research Centre have introduced a wearable hat attached with an Adv-Patch to conceal a person's identity from FR systems in August 2019. In this work, they proposed an identity hiding method called AdvHat (Adversarial Hat) [33] which attaches stickers (Adv-Patch) to a hat printed from a colour printer. This method is applied in a real-world environment against one of the most popular FR models, ArcFace

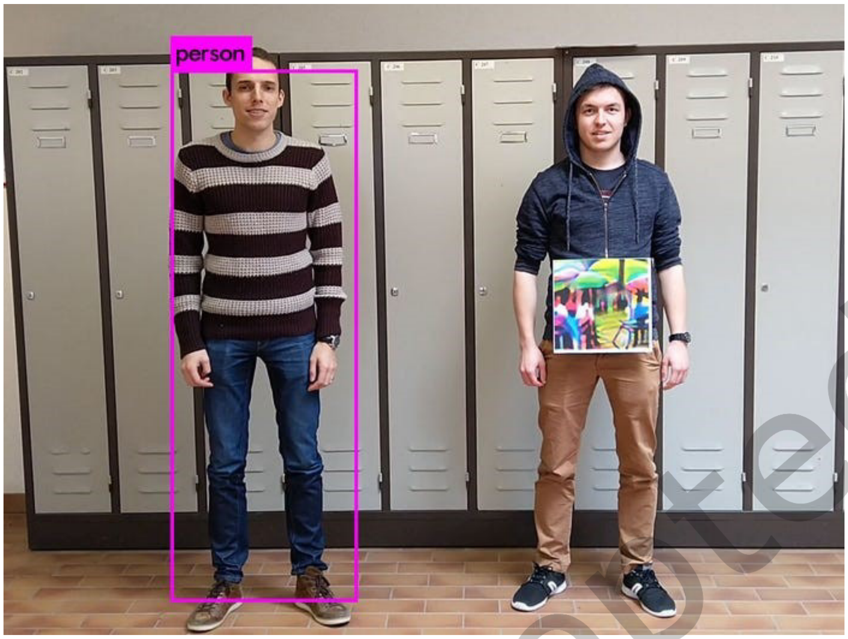


Fig. 7. An adversarial patch hiding persons from person detector [65]

(LResNet100E-IR with ArcFace loss) [13]. To design the Adv-Patch, the iterative Fast Gradient Sign Method (FGSM) that is also called MI-FGSM [14] is used and to fit the patch in a hat, the Off-Plain Sticker Transformation technique and Spatial Transformer Layer (STL) [29] technique is adopted.

The experiments were conducted with both fixed and variable conditions. For the fixed condition, consistent lighting with front-face photos and different lighting conditions with different face rotations were used for variable conditions. To evaluate the system, they calculated three similarity scores: the baseline similarity (similarity between ground-truth score and predicted score without patch), final similarity (similarity between ground-truth score and predicted score with patch) and the difference between these similarities. Here, the authors use the similarity term to explain how similar a face is compared to the training sample before and after attaching an adversarial patch. The experimental results indicated that the proposed system is robust against the ArcFace FR system. After attaching the patch to a hat, the facial recognition rate was reduced by more than 59% in the fixed condition and 43% in the variable condition. In `Fig. 8`, we can notice that the FR accuracy is reduced from 61% to 2% when wearing the hat (bottom-left image) and the pose is straight to the camera, and when the pose is not straight (bottom-right image), then the accuracy is reduced from 54% to 11%.

In October 2019, Kaziakhmedov et al. [32] from the same research institute proposed a similar Adv-Patch based technique to hide the face from FD systems. In this approach, they attached Adv-Patch (stickers with specific patterns similar to QR codes) in the cheek area of the face. They have also used MI-FGSM [14] on the different layers of the well-known FD algorithm MTCNN [79] to generate the patch. This adversarial attack on the face detector is a white-box attack because the architecture of the specific face detector is already known by the attacker. To evaluate the system, two set-ups are used, one with patches on surgical masks and the other set-up patches are directly attached to the cheeks. In each set-up, the probability of misidentification on a video of 1000 frames is calculated. In the mask set-up, the misidentification rate is 95%, and when the patch is directly attached to the face, the misidentification rate goes to almost 98%. Although this method is only applied to the MTCNN

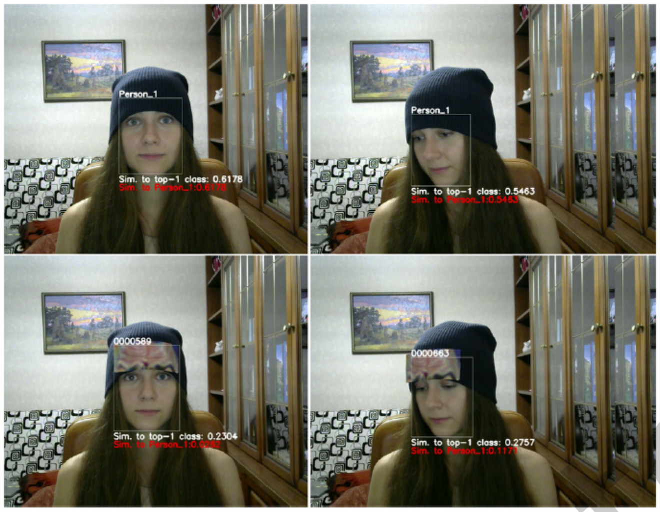


Fig. 8. Example of wearing an Adversarial Hat [33]

face detector system, the performance is very high, and the cost of generating the patch is low. Here, `Fig. 8` shows an example of the proposed method.

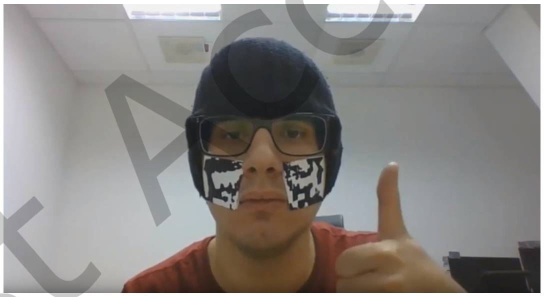


Fig. 9. Example of Adv-Patch on the Cheeks [32]

In April 2020, Pautov et al. [44] proposed a similar identity hiding technique based on Adv-Patch. The robustness of one of the SOTA FR systems, LResNet100E-IR with ArcFace loss [13], is tested in this system by attaching printed Adv-Patch in different locations of the face. The desired face location is selected to design the patch, and a simple chessboard pattern patch is pasted in the selected location. Then, a photo of the face is captured with the chessboard pattern, and the corners of the patch are marked in the image. After that, an Adv-Patch generated by using MI-FGSM [14] is projected on the chessboard pattern of the image. After that, the image with the attached patch is pre-processed using similarity transformation techniques to attack the ArcFace FR system.

For the evaluation, two images from the authors and 200 images from the CASIA-WebFace dataset [75] is used to attack the FR system. The patches were generated for three different locations of the face: nose, forehead and an eyeglass shaped patch for the eye area. But in the physical domain, the Nose area is not attacked due to height adjustment issues of the patch and poor performance in the digital part. The mean similarities and standard errors

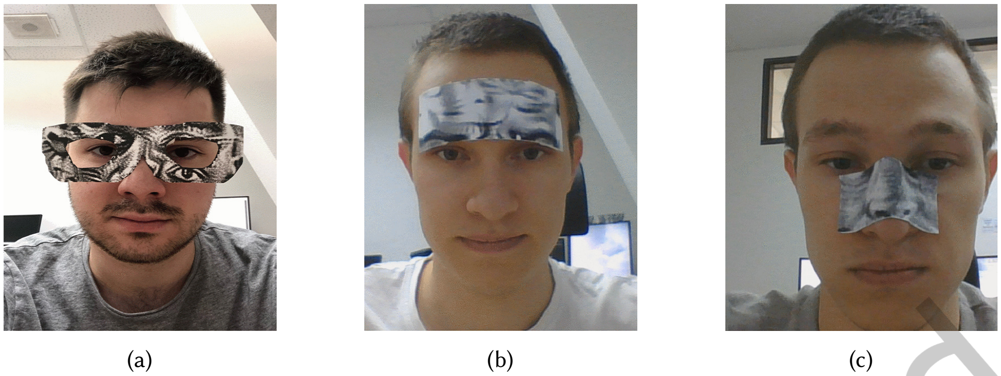


Fig. 10. Example of Adv-Patch on (a) Eyes, (b) Forehead, (c) Nose [44]

of the mean are calculated for training, validation and testing images with ground-truth and anticipated values. The results showed that the Adv-Patch on the forehead and eyeglass area significantly reduces the recognition rate of the ArcFace FR model. The results also indicated that the patches nearer to the eyes performed better than others. `Fig.` 10 shows an example of the patches attached in different face areas.

In August 2021, Shen et al. [56] proposed another Adv-Patch based approach named Face Adversary (FaceAdv) to circumvent FR systems by attaching adversarial stickers on the face. In this approach, the stickers are generated using the Wasserstein GAN with gradient penalty (WGAN-GP) method [19], and the system generates three stickers at a time with different 3D shapes. Then, the FaceAdv system tries to find out the sensitive regions of the face where most FR systems focus for feature extraction using the Guided Grad-CAM method [52]. The efficiency of the system depends heavily on the sticker localisation also. The system finds that areas near the facial organs like the nose, eyes, and mouth are essential for FR decisions by using the method mentioned above. To attach stickers, they choose five sensitive regions of the face (i.e., nasal bone, right and left nasolabial sulcus, right and left superciliary arch). But they attach three stickers in three regions at a time by making ten combinations randomly from the five selected regions. They choose three facial regions to keep a better balance of all factors because attaching five stickers takes a longer time and needs to make the sticker size smaller. This process is performed repeatedly by changing the sticker size in the digital domain to find the best possible size for the lowest FR results.

For the performance evaluation, they used three state-of-the-art FR systems (i.e., ArcFace [13], CosFace [68] and FaceNet [51]) to test the proposed system. They choose the Adversarial Generative Nets (AGNs) technique [54] for comparison since both methods use similar GAN-based approaches to generate and attach stickers. They used the publicly available SOTA face dataset LFW [28] for the digital domain evaluation, and for the real-world experiment, 20 candidates were selected by naming the dataset VoIFace. For the evaluation metrics, the number of frames that are misclassified as the original person is considered the system's success rate. For the VoIFace dataset, the success rates of the FaceAdv method are 77% on ArcFace and 100% on both CosFace and FaceNet in the physical domain. On the other hand, the success rates of AGNs in all three FR systems are much lower, which is around 50-60%. `Fig.` 11 shows an example of the FaceAdv-based sticker attached to the participant's face. The top row on the right side of the image shows that the FR system can recognise any random person other than the participant after applying the sticker. The bottom row shows that the participant is impersonated as the specific person mentioned in the bottom-left image.

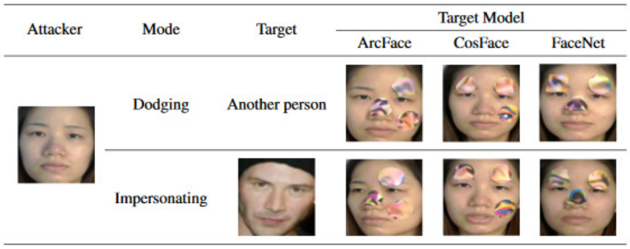


Fig. 11. Example of FaceAdv method-based sticker on the face [56]

4.2 Active FPP Methods

The active presentation-level FPP methods use artefacts that perform some sort of activity like projecting light or signals into the user's face or to the camera sensors to hide the facial identity from FR systems. Some of these methods can activate or deactivate the system while wearing the artefact by the users' choice. These methods include projecting perturbed lights into the face by using different devices such as projectors and LEDs, eyeglasses combined with various devices such as infrared light and LEDs. Some methods are used as wearable devices, such as wearing a cap or headband attached with IR-based light sources. Based on the activity of such artefacts to protect facial identity against FR systems, the active FPP methods can be grouped into two primary levels: *sensor-directed light projections-based* and *face-directed light projection-based* methods. This section briefly discusses the relevant published literature for the active presentation-level FPP methods.

4.2.1 Sensor-Directed Light Projections-based FPP Methods: Sensor-directed light projection-based presentation-level FPP methods intend to project light or noise signals onto the camera sensors by using different kinds of artefacts such as infrared-based LEDs, regular LEDs. The infrared-based LEDs add noise to a captured photo, which helps to distort the key facial features of an image, preventing facial detection. These methods have fewer noticeability issues than makeup-based or adversarial patch-based methods. Because infrared-based lights can be rarely noticed by human eyes, these methods can be a bit more expensive than the mentioned passive methods due to the cost of electronic devices like projectors and LEDs. Since the use of electronic artefacts in the privacy protection research area is relatively newer to other artefacts, we found a limited number of approaches in the literature. Here, we discuss the sensor-directed light projection-based methods below.

In 2012, Yamada et al. [72] developed one of the first facial privacy protection devices, called *Privacy Visor*, to prevent unwanted facial detection by surveillance cameras in a real-world scenario at the National Institute of Informatics Lab of Japan. The Privacy Visor is mainly a set of eyeglasses attached with near-infrared (IR) light-emitting diodes (LEDs). This method is inspired by the technique used to prevent shooting videos in cinema halls where the infrared signal-based method [10] is utilised. Another purpose of this system is to make it unnoticeable to human eyes by using IR signals. However, this system may be considered noticeable now if we compare this with some recently proposed presentation-level FPP systems like Invisible Mask [81].

For the evaluations, a system with 11 near-IR LEDs, plastic frame Googles with polycarbonate lenses and a Lithium-ion battery was constructed. One of the SOTA face detection systems, OpenCV [67], was used to test the system. The OpenCV FD system is built on the Viola-Jones algorithm [67], which uses a cascade structure of the Haar-like features. The signals emitted from the IR LEDs try to manipulate the Haar-like features near the eyes and nose. 10 participants were used to evaluate the system, and the images of the participants were captured from different distances and different angles. The results showed that no faces were detected for all distances and angles when the participants wore the eyeglass (with IR signals turned on). That means the Privacy Visor achieved 100% accuracy by preventing facial detection against the OpenCV FD system. `Fig.` 12 shows examples of the Privacy Visor glasses in use.

Fig. 12. Privacy Visor [72]

In 2018, Perez et al. [45] proposed a facial privacy-enhancing technology named *FacePET*, mainly a pair of eyeglasses. The FacePET system used a similar eyeglass-based technique to the methodology developed by Yamada et al. [72], but this system used simple LED lights instead of near-infrared light in the eyeglass to prevent facial detection. Here, they didn't use infrared light because new generation camera devices have IR filters to block the signals generated by infrared light. This system is also used against the OpenCV's Viola-Jones FD algorithm to manipulate the Haar-like features. The FacePET also included a procedure to provide consent using a Bluetooth Low Energy (BLE) microcontroller to other people who are using third-party devices for capturing pictures. By using the system, the user has the option to enable or disable the visible light when someone asks for permission over Bluetooth to take photos.

A system with 6 LEDs attached to eyeglasses was adopted for the evaluation. The BLE microcontroller was connected to the LEDs using wires. The user of the system controls the microcontroller using a smartphone application to provide permission to others who request to take photos of the user. Sixteen different smartphone models were used to take 16 photos of the user, and only in 2 photos (taken by Samsung Galaxy S7 and OnePlus 6 smartphones), the face is detected by the OpenCV FD system. The accuracy of the FacePET system was reported to be 87.5%. `Fig.` 13 is an example of the wearable FacePET device, where in the left image, LEDs turned off (face detected) and in the right image, LEDs turned on (face not detected).

4.2.2 Face-directed Light Projection-based FPP Methods: Face-directed light projection-based FPP methods aim to project light into the face of the user by using different kinds of devices such as infrared-based LEDs, small projectors. By emitting light, these methods try to force the face landmarking model used by FD systems to generate incorrect facial landmarks that leads to misidentification. These methods also have fewer noticeability issues and are relatively expensive, like the sensor-directed light projections-based methods. Since these methods use similar kinds of artefacts like the sensor-directed methods, the number of approaches found for this category is also limited. Here, we discuss the face-directed light projection-based methods below.

In March 2018, Zhou et al. [81] proposed a physical privacy protection technique on the FR system named *Invisible Mask* to circumvent FR systems without making any noticeable changes in the user's face. Infrared

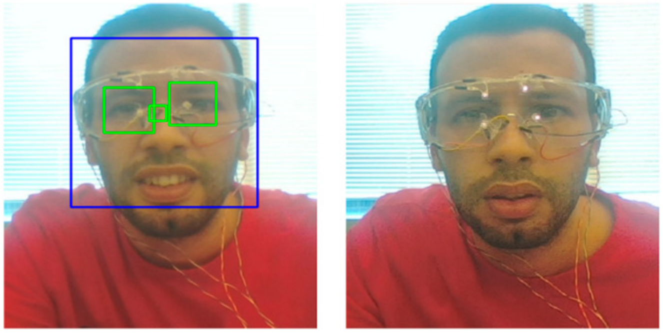
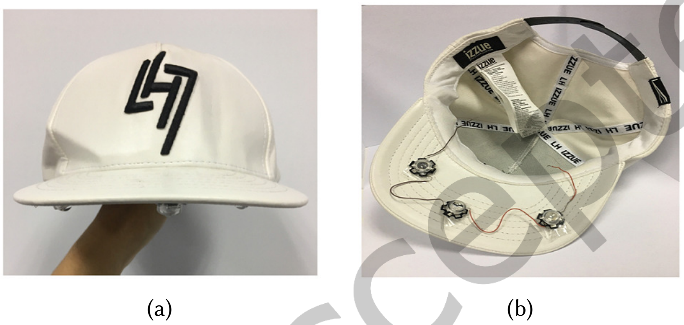


Fig. 13. FacePET [45]



(a)

(b)

Fig. 14. Example of an Invisible Mask [81]

light-based perturbations projected on the face will make it difficult for FR systems to recognise the user's face, but other people will see nothing unusual in the user's face. In this approach, three infrared LEDs, battery and wires are attached to a cap, and the LEDs emit light onto the attacker's face. Human eyes cannot directly notice the infrared light emitted by the infrared LEDs, but the camera sensors can pick the light, and that's why it is also called Invisible Mask.

The FaceNet FR model [51] was used to evaluate this system, and a limited number of participants were employed because of the uncertain health risks associated with the infrared lights. 5 seconds videos of every participant were used, and no frames of the videos were recognised as a face which indicated 100% accuracy in hiding the facial identity. `Fig.` 14 is an example of the Invisible Mask.

In September 2019, Shen et al. [55] proposed a light-based FPP technique called *VLA (Visible Light-based Attack)*. In this approach, an external LCD projector with a lamp is used as a light source to project adversarially perturbed light into the face of the user for identity hiding. This method aimed to make a face invisible to the camera and generate less noticeable light for human eyes so that it can be used in real-world scenarios. To handle both the conditions for adversarial perturbation generation, this method is divided into two parts: perturbation frames and concealing frames and both frames are projected into the face alternatively. The perturbation frame aims to modify the facial features of the attacker, and the concealing frame tries to hide the perturbations for making it invisible to human eyes.

The FaceNet FR system is used to evaluate the proposed approach. They compared the system with a popular adversarial attack generation technique, FGSM. The system is also tested digitally on the LFW large scale dataset, and 9 participants are used for the real-world evaluation. For the performance evaluation, the misclassification rate was calculated where a higher misclassification rate is considered better for the system. For the physical scenario, the misclassification rate of the VLA method is 84.5%, whereas the FGSM is only 31%. In the digital scenario, the FGSM misclassification rate raises to 88.3% but is still less than the proposed system VLA, which is 92.1%. For different head poses and various environmental brightness levels, the misclassification rate is 80%. `Fig.` 15 shows how the VLA system works against an FR system.

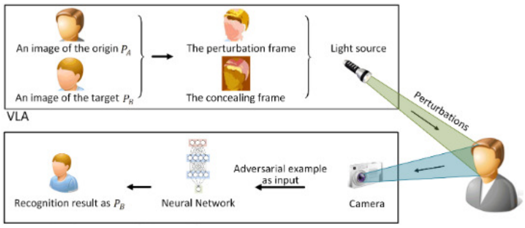

An image of the origin P_A
An image of the target P_B
The perturbation frame
The concealing frame
Light source
VLA
Perturbations
Recognition result as P_B
Neural Network
Adversarial example as input
Camera

Fig. 15. The framework of the VLA system against an FR system [55]

4.3 Other FPP Methods

In this section, we discuss those presentation-level FPP methods where the researchers only proposed concepts or ideas in some cases to protect our facial privacy from unwanted surveillance systems. But they did not provide any detailed information about the proposed approaches' performance results or the procedure to re-generate the systems. However, in some cases, the proposed methodologies are demonstrated in exhibitions or events with only one example data but without publishing the implementation procedures anywhere in the literature. Hence, one instance of such FPP methods cannot be considered robust for facial privacy protection in different challenging scenarios of our life. Here, we discuss some of these methods below.

The concept of Wearable Face Projector, proposed by Liu et al. [30], is different from the techniques mentioned earlier for facial privacy protection. In this approach, a small projector is attached on the top of a headband, and the projector projects a different face image into the user's face. In the Milan Design Week 2017 [8], a group of students from the University of Arts Utrecht, Netherlands, demonstrated some approaches to protect the privacy of ordinary people from the surveillance systems. Jing-Cai Liu from this group presented the concept of Wearable Face Projector at that event. `Fig.` 16 is an example of a Wearable Face Projector.

Weekers et al. [50] from the same group proposed an idea to wear a headscarf designed with several faces. This scarf named Anonymity Scarf can confuse the FD-FR systems to detect the original face. `Fig.` 17 is an example of Anonymity Scarf.

Leeuwenstein et al. [31] from the same group proposed another privacy protection concept at the same event by wearing a transparent face mask designed like a lens. This face mask named by the author as Surveillance Exclusion Face Mask is produced by plastic, and it generates a kind of facial ridges to circumvent FD systems. `Fig.` 18 is an example of wearing the Surveillance Exclusion Face Mask on a human face.

AVG Research Lab presented a novel concept named Invisibility Glasses [1] to protect facial privacy at the Pepcom, Barcelona event in May 2015. The Invisibility Glasses is a pair of eyeglasses designed with retro-reflective components. The components are modelled in such a way that it emits light back to the same direction from where the camera sensor sends light with a flash. The major issue with this technique is that it doesn't perform well when the images are captured without a flashlight as the method depends on reflecting light emitted from flashlights of camera devices. `Fig.` 19 is an example of Invisibility Glasses.

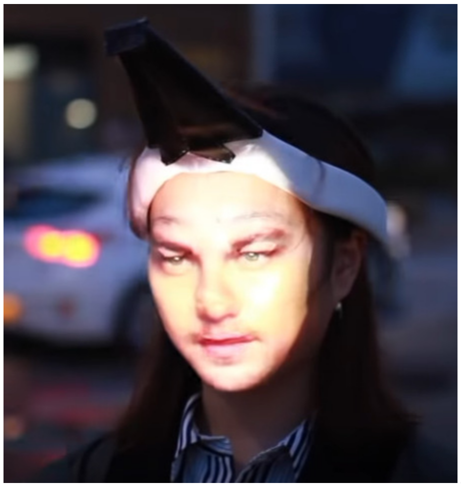


Fig. 16. The Wearable Face Projector [30]

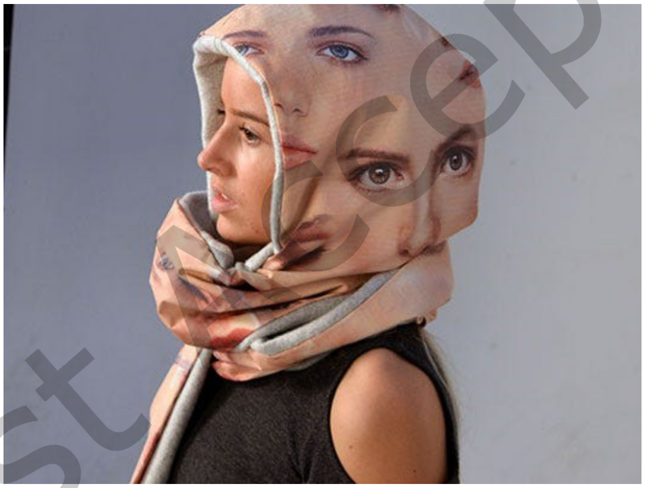


Fig. 17. The Anonymity Scarf [50]

4.4 Summary of Presentation-Level Facial Privacy Protection Approaches

Table 1 provides a general overview of the reported presentation-level FPP methods. The actual performance of these approaches cannot be determined only by the accuracy rates provided for privacy protection because most of the approaches are evaluated on either one or a minimal number of FD and FR systems. In Table 1, the reference of the methods, the institution from where the method is proposed, the name of the proposed system, the face detection or recognition algorithm used to evaluate the system and the claimed success rate in hiding facial identity mentioned in the literature are outlined.

We can notice from Table 1 that there are several approaches where the authors claimed over 99% success rate against the Viola-Jones FD system. However, all those approaches were evaluated against only one FD

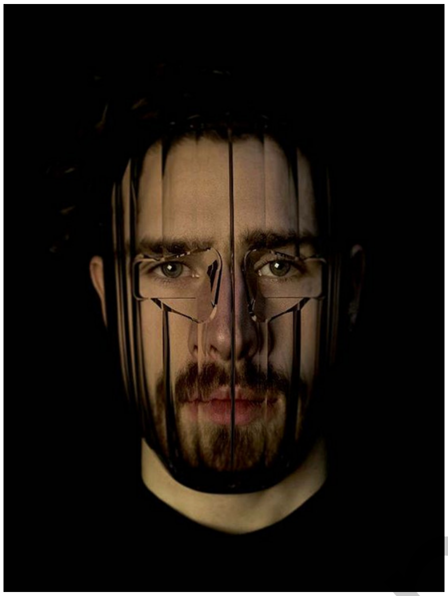


Fig. 18. The Surveillance Exclusion Face Mask [31]

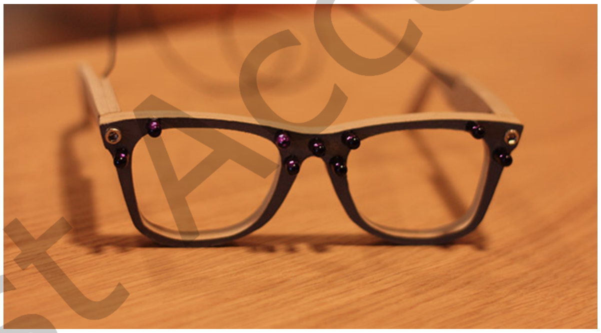


Fig. 19. The Invisibility Glasses by AVG [1]

system, and the architecture of the Viola-Jones FD algorithm is well-known to the researchers, which helps them to attack the system. For the FR systems, although the Invisible Mask technique achieved 100% success against a SOTA FR system FaceNet, this technique is also evaluated against only one FR system, and a minimal number of participants were used. On the other hand, a recently introduced approach, Adv-Makeup, shows that when the method is evaluated against multiple FD and FR systems, we can find mixed performance results, and all the results aren't satisfactory. So, the Adv-Makeup approach also indicates that if the existing methods on presentation-level FPP systems can be evaluated against multiple FD-FR systems and with a large number of participants in real-world scenarios, then there's a considerable chance of exploring the vulnerabilities of existing approaches.

Table 1. General Overview of Presentation-Level Facial Privacy Protection Approaches. *NU- New York University, NII- National Institute of Informatics, CMU- Carnegie Mellon University, FU- Fudan University, CSU- Columbus State University, FacePET- Facial Privacy-Enhancing Technology, OD- Object Detector, AdvHat- Adversarial Hat, BIT- Beijing Institute of Technology, VLA- Visible Light-based Attack, Adv-Patch- Adversarial Patch, FaceAdv-Face Adversary, Adv-Makeup- Adversarial Makeup.

Reference	Institution	Proposed Methods	FD/FR Methods	Success Rate (%)
Harvey et al. 2010 [24]	NU, USA	CV-Dazzle	Viola-Jones FD	99
Yamada et al. 2012 [72]	NII, Japan	Privacy Visor	Viola-Jones FD	100
Sharif et al. 2016 [53]	CMU, USA	Perturbed Eyeglass Frames	Viola-Jones & Face++ FD	96
Zhou et al. 2018 [81]	FU, China	Invisible Mask	FaceNet FR	100
Perez et al. 2018 [45]	CSU, USA	FacePET	Viola-Jones FD	87.5
Thys et al. 2019 [65]	KU Leuven, Belgium	Fooling automated surveillance cameras	YOLOv2 OD	74.47
Komkov et al. 2019 [33][25]	Huawei, Russia	AdvHat	ArcFace FR	59
Shen et al. 2019 [55]	BIT, China	VLA	FaceNet FR	84.5
Kaziakhmedov et al. 2020 [32]	Huawei, Russia	Adv-Patch on Cheeks	MTCNN FD	95
Pautov et al. 2020 [44]	Huawei, Russia	Adv-Patch on Eyeglass	ArcFace FR	70
Shen et al. 2021 [56]	BIT, China	FaceAdv	ArcFace, CosFace & FaceNet FR	77, 100 & 100
Yin et al. 2021 [76]	Tencent, China	Adv-Makeup	MobileFace & FaceNet FR	64 & 33

4.5 Evaluation of Presentation-Level Facial Privacy Protection Methods

Presentation-level FPP systems have numerous key factors to compare and evaluate the systems. These factors can help us find out the significant benefits and drawbacks of the systems, such as which method is easier to deploy, effective in critical conditions and several other key features. In order to evaluate and assess the performance of FPP methods, the key factors can be categorised into the following three levels: low, medium and high. Based on the key factors and their associated levels, we evaluate the FPP methods in the following Table 2. For some cases, we used a range of levels due to their diverse characteristics, such as low to medium or medium to high, etc. The factors are outlined below:

- `Ease of Deployment`: The ease with which the presentation-level FPP system can be deployed and removed is a major consideration for the users and the applications. The ease of deployment can also be classified into the three levels mentioned above but only for this factor; if the low level is considered as Hard level and the high as Easy, then it will be easier to understand, and all the factors can be easier to classify under same levels.
- `Cost`: The cost of developing the FPP system is another key concern because an effective but costly FPP system cannot be affordable to a large group of people.

Table 2. Evaluation of Presentation-Level Facial Privacy Protection Methods. Symbol definition: H - high, M - medium and L - low.

FPP Methods	Ease of Deployment	Cost	Noticeability	Effectiveness
CV-Dazzle	L	L-M	H	L-M
Privacy Visor	H	M-H	H	L-M
Perturbed Eyeglass Frames	H	L	L-M	M-H
Invisible Mask	L-M	M-H	L-M	M-H
FacePET	L	M-H	H	M-H
VLA: Visible Light Attack	L	M-H	L-M	M-H
Fooling automated surveillance cameras	H	L	M-H	M-H
AdvHat: Adversarial Hat	H	L	M-H	M-H
Adv-Patch on Cheeks	H	L	H	M-H
Adv-Patch on Eyeglass	H	L	H	M-H
FaceAdv: Face Adversary	H	L	H	M-H
Adv-Makeup: Adversarial Makeup	L-M	L-M	M-H	M-H

- Noticeability: Some FPP methods use adversarial patches or stickers or wired eyeglasses, and thus it makes the user of the system highly noticeable to other people. It often looks weird because of its size and orientation, so it may not be acceptable to all kinds of users due to its noticeability issue. Therefore, we may need a trade-off between the noticeability and effectiveness factors.
- Effectiveness: Some methods lose their effectiveness at identity hiding in challenging conditions. So, it's another critical concern for a robust FPP system. An FPP system will be acceptable to all users when it is effective against the most challenging environmental scenarios and circumstances.

From the Table 2, we can notice that none of the methods is flawless according to the criteria of the aforementioned factors. However, no method can be perfect for all subjects because every individual may have different requirements to fulfil their criteria for a robust facial privacy protection system. The definition of an ideal facial privacy protection method is a methodology that is very easy to deploy, the cost of developing or using the method is low, the noticeability of the system is also low, and the effectiveness of the method in any kind of challenging conditions is also high. However, we can observe from Table 2 that some presentation-level FPP methods partially fulfil the criteria of the mentioned factors. For example, the adversarial patch-based methods are easy to deploy, the production cost is cheaper, and the effectiveness is also better compared to other categories, but in most cases, it is noticeable to other people while we use it in a public place. So, the adversarial patch-based method can be the system that can be used by a wide range of people if the effectiveness can be enhanced for challenging conditions and if it can be less noticeable to other people while using it in our daily life.

5 CONCLUSION AND DISCUSSION

Privacy protection systems are essential for those people who do not want to share their private data without consent. Hence, individuals would normally have the right to protect or hide their facial data from public surveillance systems where prevailing laws allow it. Methods of identity hiding have been a topic of research by numerous investigators, and it remains important to establish the limits of possibilities through continued research. Such methods provide options for identity hiding and privacy enhancement where and when these are within the law. On other occasions, when revealing identity is required by law, such techniques need the option to be deactivated conveniently or users be clearly warned against their use. Hence, research should be more focused on developing such facial privacy protection methods which have the option to activate or deactivate the system when required by law. This will allow the users of the FPP system to deactivate it when they enter an area which needs mandatory surveillance such as border-controlled areas or banks. In this way, both privacy protection systems and surveillance systems can co-exist. Furthermore, the development of better privacy protection techniques may lead to the development of stronger FR techniques in response. Such technological developments may be inevitable and even good for the development of underlying techniques and legal systems supporting privacy and safety.

Privacy protection is a significant issue when using FR technology given its potential for misuse. As public and private organisations are increasingly relying on FR systems for mass surveillance and public security, the need for stronger measures for protecting privacy has also increased. This paper specifically focused on facial privacy protection techniques at the sensor or presentation level, where the FR systems are confronted in real-world scenarios. The major presentation-level facial privacy protection methods are briefly discussed and their limitations are highlighted.

Different companies or organisations use different FR technologies, and there's no guarantee that one method that performs well to hide the identity in one situation will also perform well on another FR system. So, further research is needed for developing facial privacy protection methods that performs well against a range of FR algorithms.

Another critical concern for these techniques is that most adversarial perturbations like patches, masks, hats are a relatively prominent and noticeable to the human eyes, making their use impractical. There is therefore the need for continued research to develop a physical adversarial perturbation-based interventions that can have a better balance between useability of the system and its effectiveness. Combining multiple approaches, such as adversarial patches with small light projection-based wearable devices, may result in hybrid solutions that are more robust and flexible. It is hoped that this review may contribute to further explorations and innovations in this important area.

ACKNOWLEDGMENTS

This work is supported by the PriMa project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860315.

REFERENCES

[1] 2015. AVG Reveals Invisibility Glasses at Pepcom Barcelona. Available online:. (2015). https://www.avg.com/en/signal/avg-reveals-invisibility-glasses-at-pepcom-barcelona

[2] 2018. "Dangerous and inaccurate" police facial recognition exposed in new Big Brother Watch report — Big Brother Watch. https://bigbrotherwatch.org.uk/2018/05/dangerous-and-inaccurate-police-facial-recognition-exposed-in-new-big-brother-watch-report/

[3] 2021. Avant-garde. https://en.wikipedia.org/w/index.php?title=Avant-garde&oldid=1045792145 Page Version ID: 1045792145.

[4] 2021. History of facial recognition technology. https://en.wikipedia.org/w/index.php?title=Facial_recognition_system&oldid=1038849154, Accessed: 2021-01-10.

[5] Tom B. Brown, Dandelion Mané, Aurko Roy, Martín Abadi, and Justin Gilmer. 2018. Adversarial Patch. *arXiv:1712.09665 [cs]* (May 2018). http://arxiv.org/abs/1712.09665 arXiv: 1712.09665.

[6] R. Brunelli and T. Poggio. 1993. Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 10 (Oct. 1993), 1042–1052. https://doi.org/10.1109/34.254061 Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[7] Ankur Chattopadhyay and T.E. Boult. 2007. PrivacyCam: a Privacy Preserving Camera Using uCLinux on the Blackfin DSP. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 1–8. https://doi.org/10.1109/CVPR.2007.383413 ISSN: 1063-6919.

[8] Dong Chen, Gang Hua, Fang Wen, and Jian Sun. 2016. Supervised Transformer Network for Efficient Face Detection. In *Computer Vision – ECCV 2016 (Lecture Notes in Computer Science)*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Springer International Publishing, Cham, 122–138. https://doi.org/10.1007/978-3-319-46454-1_8

[9] Shang-Tse Chen, Cory Cornelius, Jason Martin, and Duen Horng (Polo) Chau. 2019. ShapeShifter: Robust Physical Adversarial Attack on Faster R-CNN Object Detector. In *Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science)*,

Michele Berlingerio, Francesco Bonchi, Thomas Gärtner, Neil Hurley, and Georgiana Ifrim (Eds.). Springer International Publishing, Cham, 52–68. https://doi.org/10.1007/978-3-030-10925-7_4

[10] N. Dalal and B. Triggs. 2005. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1. 886–893 vol. 1. https://doi.org/10.1109/CVPR.2005.177 ISSN: 1063-6919.

[11] Fahad Daniyal, Prathap Nair, and Andrea Cavallaro. 2009. Compact Signatures for 3D Face Recognition under Varying Expressions. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. 302–307. https://doi.org/10.1109/AVSS.2009.71

[12] Debayan Deb, Jianbang Zhang, and Anil K. Jain. 2019. AdvFaces: Adversarial Face Synthesis. *arXiv:1908.05008 [cs]* (Aug. 2019). http://arxiv.org/abs/1908.05008 arXiv: 1908.05008.

[13] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *arXiv:1801.07698 [cs]* (Feb. 2019). http://arxiv.org/abs/1801.07698 arXiv: 1801.07698.

[14] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. 2018. Boosting Adversarial Attacks with Momentum. IEEE Computer Society, 9185–9193. https://doi.org/10.1109/CVPR.2018.00957

[15] Clément Feutry, Pablo Piantanida, Yoshua Bengio, and Pierre Duhamel. 2018. Learning Anonymized Representations with Adversarial Neural Networks. *arXiv:1802.09386 [cs, stat]* (Feb. 2018). http://arxiv.org/abs/1802.09386 arXiv: 1802.09386.

[16] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2414–2423. https://doi.org/10.1109/CVPR.2016.265 ISSN: 1063-6919.

[17] Qiao Gu, Guanzhi Wang, Mang Tik Chiu, Yu-Wing Tai, and Chi-Keung Tang. 2019. LADN: Local Adversarial Disentangling Network for Facial Makeup and De-Makeup. *arXiv:1904.11272 [cs]* (Aug. 2019). http://arxiv.org/abs/1904.11272 arXiv: 1904.11272.

[18] Nitzan Guetta, Asaf Shabtai, Inderjeet Singh, Satoru Momiyama, and Yuval Elovici. 2021. Dodging Attack Using Carefully Crafted Natural Makeup. *arXiv:2109.06467 [cs]* (Sept. 2021). http://arxiv.org/abs/2109.06467 arXiv: 2109.06467.

[19] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. 2017. Improved training of wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 5769–5779.

[20] Guodong Guo and Na Zhang. 2019. A survey on deep learning based face recognition. *Computer Vision and Image Understanding* 189 (Dec. 2019), 102805. https://doi.org/10.1016/j.cviu.2019.102805

[21] Jia Guo, Jiankang Deng, Alexandros Lattas, and Stefanos Zafeiriou. 2021. Sample and Computation Redistribution for Efficient Face Detection. *arXiv:2105.04714 [cs]* (May 2021). http://arxiv.org/abs/2105.04714 arXiv: 2105.04714.

[22] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. 2016. MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. *arXiv:1607.08221 [cs]* (July 2016). http://arxiv.org/abs/1607.08221 arXiv: 1607.08221.

[23] Shalini Gupta, Mia K. Markey, and Alan C. Bovik. 2010. Anthropometric 3D Face Recognition. *International Journal of Computer Vision* 90, 3 (Dec. 2010), 331–349. https://doi.org/10.1007/s11263-010-0360-8

[24] Adam Harvey. 2010. CV Dazzle: Camouflage from face detection. *Master's thesis* (2010).

[25] Abul Hasnat, Julien Bohne, Jonathan Milgram, Stephane Gentric, and Liming Chen. 2017. DeepVisage: Making Face Recognition Simple Yet With Powerful Generalization Skills. IEEE Computer Society, 1682–1691. https://doi.org/10.1109/ICCVW.2017.197 ISSN: 2473-9944.

[26] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778. https://doi.org/10.1109/CVPR.2016.90 ISSN: 1063-6919.

[27] Kashmir Hill. 2020. Wrongfully Accused by an Algorithm. *The New York Times* (June 2020). https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

[28] Gary B. Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. 2008. Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. https://hal.inria.fr/inria-00321923

[29] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. 2015. Spatial transformer networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15)*. MIT Press, Cambridge, MA, USA, 2017–2025.

[30] Liu Jing. 2017. WEARABLE FACE PROJECTOR. http://jingcailiu.com/wearable-face-projector/

[31] Van Jip. 2017. Surveillance Exclusion Face Mask. http://www.jipvanleeuwenstein.nl/

[32] E. Kaziakhmedov, K. Kireev, G. Melnikov, M. Pautov, and A. Petiushko. 2019. Real-world Attack on MTCNN Face Detection System. In *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*. 0422–0427. https://doi.org/10.1109/SIBIRCON48586.2019.8958122

[33] Stepan Komkov and Aleksandr Petiushko. 2019. AdvHat: Real-world adversarial attack on ArcFace Face ID system. *arXiv:1908.08705 [cs]* (Aug. 2019). http://arxiv.org/abs/1908.08705 arXiv: 1908.08705.

[34] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. 2017. Adversarial examples in the physical world. *ICLR Workshop* (2017). https://arxiv.org/abs/1607.02533

[35] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. 2015. A convolutional neural network cascade for face detection. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5325–5334. https://doi.org/10.1109/CVPR.2015.7299170 ISSN: 1063-6919.

[36] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 936–944. https://doi.org/10.1109/CVPR.2017.106 ISSN: 1063-6919.

[37] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. 2018. SphereFace: Deep Hypersphere Embedding for Face Recognition. *arXiv:1704.08063 [cs]* (Jan. 2018). http://arxiv.org/abs/1704.08063 arXiv: 1704.08063.

[38] MEGVII. 2021. Face++ - (Online) face verification. https://www.faceplusplus.com.cn/

[39] Microsoft. 2021. Online face verification | Microsoft Azure. https://azure.microsoft.com/en-us/

[40] Shervin Minaee, Ping Luo, Zhe Lin, and Kevin Bowyer. 2021. Going Deeper Into Face Detection: A Survey. *arXiv:2103.14983 [cs]* (April 2021). http://arxiv.org/abs/2103.14983 arXiv: 2103.14983.

[41] Mrityunjay and P.J. Narayanan. 2011. The De-Identification Camera. In *2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*. 192–195. https://doi.org/10.1109/NCVPRIPG.2011.48

[42] Mahyar Najibi, Pouya Samangouei, Rama Chellappa, and Larry S. Davis. 2017. SSH: Single Stage Headless Face Detector. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 4885–4894. https://doi.org/10.1109/ICCV.2017.522 ISSN: 2380-7504.

[43] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. 2015. Deep Face Recognition. In *Procedings of the British Machine Vision Conference 2015*. British Machine Vision Association, Swansea, 41.1–41.12. https://doi.org/10.5244/C.29.41

[44] M. Pautov, G. Melnikov, E. Kaziakhmedov, K. Kireev, and A. Petiushko. 2019. On Adversarial Patches: Real-World Attack on ArcFace-100 Face Recognition System. In *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*. 0391–0396. https://doi.org/10.1109/SIBIRCON48586.2019.8958134

[45] Alfredo J. Perez, Sherali Zeadally, Luis Y. Matos Garcia, Jaouad A. Mouloud, and Scott Griffith. 2018. FacePET: Enhancing Bystanders' Facial Privacy with Smart Wearables/Internet of Things. *Electronics* 7, 12 (Dec. 2018), 379. https://doi.org/10.3390/electronics7120379 Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

[46] Joseph Redmon and Ali Farhadi. 2017. YOLO9000: Better, Faster, Stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 6517–6525. https://doi.org/10.1109/CVPR.2017.690 ISSN: 1063-6919.

[47] Antoaneta Roussi. 2020. Resisting the rise of facial recognition. *Nature* 587, 7834 (Nov. 2020), 350–353. https://doi.org/10.1038/d41586-020-03188-2 Number: 7834 Publisher: Nature Publishing Group.

[48] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 3 (Dec. 2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y

[49] Toshiyuki Sakai, Takeo Kanade, Makoto Nagao, and Yu-ichi Ohta. 1973. Picture processing system using a computer complex. *Computer Graphics and Image Processing* 2, 3 (Dec. 1973), 207–215. https://doi.org/10.1016/0146-664X(73)90002-6

[50] Weekers Sanne. 2017. anonymous | Sanne Weekers. http://sanneweekers.nl/big-brother-is-watching-you/

[51] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A Unified Embedding for Face Recognition and Clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), 815–823. https://doi.org/10.1109/CVPR.2015.7298682 arXiv: 1503.03832.

[52] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 618–626. https://doi.org/10.1109/ICCV.2017.74 ISSN: 2380-7504.

[53] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. 2016. 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 (CCS '16)*. Association for Computing Machinery, New York, NY, USA, 1528–1540. https://doi.org/10.1145/2976749.2978392

[54] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. 2019. A General Framework for Adversarial Examples with Objectives. *ACM Transactions on Privacy and Security* 22, 3 (June 2019), 16:1–16:30. https://doi.org/10.1145/3317611

[55] Meng Shen, Zelin Liao, Liehuang Zhu, Ke Xu, and Xiaojiang Du. 2019. VLA: A Practical Visible Light-based Attack on Face Recognition Systems in Physical World. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 103:1–103:19. https://doi.org/10.1145/3351261

[56] Meng Shen, Hao Yu, Liehuang Zhu, Ke Xu, Qi Li, and Jiankun Hu. 2021. Effective and Robust Physical-World Attacks on Deep Learning Face Recognition Systems. *IEEE Transactions on Information Forensics and Security* 16 (2021), 4063–4077. https://doi.org/10.1109/TIFS.2021.3102492 Conference Name: IEEE Transactions on Information Forensics and Security.

[57] J. Shi, A. Samal, and D. Marx. 2006. How effective are landmarks and their geometry for face recognition? *Computer Vision and Image Understanding* 102, 2 (May 2006), 117–133. https://doi.org/10.1016/j.cviu.2005.10.002

[58] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]* (April 2015). http://arxiv.org/abs/1409.1556 arXiv: 1409.1556.

[59] Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang. 2015. DeepID3: Face Recognition with Very Deep Neural Networks. *arXiv:1502.00873 [cs]* (Feb. 2015). http://arxiv.org/abs/1502.00873 arXiv: 1502.00873.

[60] Yi Sun, Xiaogang Wang, and Xiaoou Tang. 2014. Deep Learning Face Representation from Predicting 10,000 Classes. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Columbus, OH, USA, 1891–1898. https://doi.org/10.1109/CVPR.2014.244

[61] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2015. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–9. https://doi.org/10.1109/CVPR.2015.7298594 ISSN: 1063-6919.

[62] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Columbus, OH, USA, 1701–1708. https://doi.org/10.1109/CVPR.2014.220

[63] P. Terhörst. 2021. *Mitigating Soft-Biometric Driven Bias and Privacy Concerns in Face Recognition Systems | Fraunhofer IGD*. Ph. D. Dissertation. Technische Universität Darmtadt. https://www.igd.fraunhofer.de/en/press/news/mitigating-soft-biometric-driven-bias-and-privacy-concerns-face-recognition-systems

[64] Philipp Terhörst, Kevin Riehl, Naser Damer, Peter Rot, Blaz Bortolato, Florian Kirchbuchner, Vitomir Struc, and Arjan Kuijper. 2020. PE-MIU: A Training-Free Privacy-Enhancing Face Recognition Approach Based on Minimum Information Units. *IEEE Access* 8 (2020), 93635–93647. https://doi.org/10.1109/ACCESS.2020.2994960 Conference Name: IEEE Access.

[65] S. Thys, W. V. Ranst, and T. Goedemé. 2019. Fooling Automated Surveillance Cameras: Adversarial Patches to Attack Person Detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 49–55. https://doi.org/10.1109/CVPRW.2019.00012 ISSN: 2160-7516.

[66] Fatemeh Vakhshiteh, Ahmad Nickabadi, and Raghavendra Ramachandra. 2021. Adversarial Attacks against Face Recognition: A Comprehensive Study. *arXiv:2007.11709 [cs, eess]* (Feb. 2021). http://arxiv.org/abs/2007.11709 arXiv: 2007.11709.

[67] Paul Viola and Michael J. Jones. 2004. Robust Real-Time Face Detection. *International Journal of Computer Vision* 57, 2 (May 2004), 137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

[68] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu. 2018. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5265–5274. https://doi.org/10.1109/CVPR.2018.00552 ISSN: 2575-7075.

[69] Mei Wang and Weihong Deng. 2021. Deep Face Recognition: A Survey. *Neurocomputing* 429 (March 2021), 215–244. https://doi.org/10.1016/j.neucom.2020.10.081 arXiv: 1804.06655.

[70] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. 2016. A Discriminative Feature Learning Approach for Deep Face Recognition. In *Computer Vision – ECCV 2016*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Vol. 9911. Springer International Publishing, Cham, 499–515. https://doi.org/10.1007/978-3-319-46478-7_31 Series Title: Lecture Notes in Computer Science.

[71] Thomas Winkler and Bernhard Rinner. 2010. TrustCAM: Security and Privacy-Protection for an Embedded Smart Camera Based on Trusted Computing. In *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. 593–600. https://doi.org/10.1109/AVSS.2010.38

[72] T. Yamada, S. Gohshi, and I. Echizen. 2013. Privacy Visor: Method Based on Light Absorbing and Reflecting Properties for Preventing Face Image Detection. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*. 1572–1577. https://doi.org/10.1109/SMC.2013.271 ISSN: 1062-922X.

[73] Yoshihiro Yamada, Masakazu Iwamura, and Koichi Kise. 2016. Deep Pyramidal Residual Networks with Separated Stochastic Depth. *arXiv e-prints* 1612 (Dec. 2016), arXiv:1612.01230. http://adsabs.harvard.edu/abs/2016arXiv161201230Y

[74] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. 2015. WIDER FACE: A Face Detection Benchmark. *arXiv:1511.06523 [cs]* (Nov. 2015). http://arxiv.org/abs/1511.06523 arXiv: 1511.06523.

[75] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z. Li. 2014. Learning Face Representation from Scratch. *arXiv:1411.7923 [cs]* (Nov. 2014). http://arxiv.org/abs/1411.7923 arXiv: 1411.7923.

[76] Bangjie Yin, Wenxuan Wang, Taiping Yao, Junfeng Guo, Zelun Kong, Shouhong Ding, Jilin Li, and Cong Liu. 2021. Adv-Makeup: A New Imperceptible and Transferable Attack on Face Recognition. (May 2021). https://arxiv.org/abs/2105.03162v1

[77] Mengyao Zhai, Lei Chen, Fred Tung, Jiawei He, Megha Nawhal, and Greg Mori. 2019. Lifelong GAN: Continual Learning for Conditional Image Generation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2759–2768. https://doi.org/10.1109/ICCV.2019.00285 ISSN: 2380-7504.

[78] Cha Zhang and Zhengyou Zhang. 2010. A Survey of Recent Advances in Face Detection. (June 2010). https://www.microsoft.com/en-us/research/publication/a-survey-of-recent-advances-in-face-detection/

[79] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters* 23, 10 (Oct. 2016), 1499–1503. https://doi.org/10.1109/LSP.2016.2603342 Conference Name: IEEE Signal Processing Letters.

[80] Yupeng Zhang, Yuheng Lu, Hajime Nagahara, and Rin-ichiro Taniguchi. 2014. Anonymous Camera for Privacy Protection. In *2014 22nd International Conference on Pattern Recognition*. 4170–4175. https://doi.org/10.1109/ICPR.2014.715 ISSN: 1051-4651.

[81] Zhe Zhou, Di Tang, Xiaofeng Wang, Weili Han, Xiangyu Liu, and Kehuan Zhang. 2018. Invisible Mask: Practical Attacks on Face Recognition with Infrared. *arXiv:1803.04683 [cs]* (March 2018). http://arxiv.org/abs/1803.04683 arXiv: 1803.04683.

Received 9 February 2022; revised 24 Jan 2023; accepted 30 Jan 2023