APSIPA Transactions on Signal and Information Processing, 2023, 12, e25

Overview Paper

Digital and Physical Face Attacks: Reviewing and One Step Further

Chenqi Kong[1], Shiqi Wang[1] and Haoliang Li[2*]

[1]*Department of Computer Science, City University of Hong Kong, Hong Kong SAR, China*

[2]*Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China*

ABSTRACT

With the rapid progress over the past five years, face authentication has become the most pervasive biometric recognition method. Thanks to the high-accuracy recognition performance and user-friendly usage, automatic face recognition (AFR) has exploded into a plethora of practical applications over device unlocking, checking-in, and financial payment. In spite of the tremendous success of face authentication, a variety of face presentation attacks (FPA), such as print attacks, replay attacks, and 3D mask attacks, have raised pressing mistrust concerns. Even worse, as attack techniques are getting more and more powerful and smart, FPA is becoming increasingly realistic and advanced. Besides physical face attacks, face videos/images are vulnerable to a wide variety of digital attack techniques launched by malicious hackers, causing potential menace to the public at large. Due to the unrestricted access to enormous digital face images/videos and disclosed easy-to-use face manipulation tools circulating on the internet, non-expert attackers without any prior professional skills are able to readily create sophisticated fake faces, leading to numerous dangerous applications such as financial fraud, impersonation, and identity theft. Nowadays, face information has become the dominant biometric trait of a person and unique non-verbal but powerful FaceID. How to safeguard personal face information against

*Corresponding author: Haoliang Li, haoliang.li@cityu.edu.hk.

Received 29 September 2022; Revised 24 December 2022
ISSN 2048-7703; DOI 10.1561/116.00000069

both physical and digital attacks is of great importance. This survey aims to build the integrity of face forensics by providing thorough analyses of existing literature and highlighting the issues requiring further attention. In this paper, we first comprehensively survey both physical and digital face attack types and datasets. Then, we review the latest and most advanced progress on existing counter-attack methodologies and highlight their current limits. Moreover, we outline possible future research directions for existing and upcoming challenges in the face forensics community. Finally, the necessity of joint physical and digital face attack detection has been discussed, which has never been studied in previous surveys.

Keywords: Face attacks, digital face attack, physical face attack, face forensics

1 Introduction

Significant progress on face recognition techniques has been made since the advent of Apple's highly touted FaceID and the follow-up face authentication works. Face recognition systems have pervaded into billions of people's daily lives over various applications such as device unlocking, log-in, and e-banking. Consequently, face information nowadays has become the dominant biometric trait of a person, a unique FaceID, and a vehicle itself of non-verbal but powerful messages [201]. With the rapid proliferation of face multimedia content circulating on social media platforms, unrestricted access to digital media content has posed high level of risks over privacy leakage, identity theft, and financial fraud. In spite of the achieved tremendous success on face authentications, potential malicious face attacks, including digital and physical attacks, have raised pressing security concerns to the public at large.

As shown in Figure 1, digital face attacks can be basically classified into four categories: (1) identity swap; (2) face reenactment; (3) attribute manipulation; and (4) entire face synthesis [43]. Identity swap [1, 19, 60] is actually not a new problem. The first ever work on identity swap dates back to 1860, where Abraham Lincoln's head is stitched up with the body of southern politician John Calhoun [41]. Heading to the era of artificial intelligence and deep learning, deepfake techniques, employing powerful various generative models, are able to create sophisticated fake faces with the target identity. Face reenactment (*a.k.a.* expression edition) [195, 196] aims to transfer the source person's facial expression to the target one. Face2Face [195] and NeuralTextures [196] are two of the most prominent facial expression editing techniques. Moreover, attribute manipulation [38, 69] empowered by numerous image translation methods [79, 250, 251] attempts to edit face attributes such as hair, glasses,

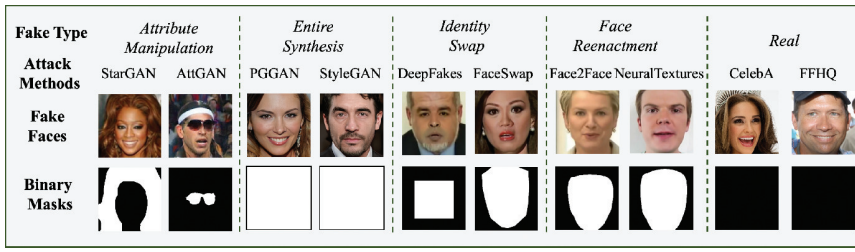

Figure 1: Four digital face attack types and corresponding forgery regions. White pixels indicate the forged region.

and skin color in face images/videos. Thanks to the recent advances of various generative models [61, 93, 94, 199, 200], entire face synthesis [87, 88] can generate face pictures whose identity does not exist with a high level of realism. Generally speaking, attribute manipulation and entire face synthesis techniques tend to bring positive impacts to human lives, while identity swap and face reenactment could easily cause disconcerting security problems. For this reason, this survey mainly focuses on identity swap and face reenactment.

Based on the attack techniques and intents, physical face attacks (*a.k.a.* Face presentation attacks (FPA)) can be broadly categorized into two classifications: impersonation and obfuscation. As shown in Figure 3(a)–(d), impersonation attacks include typical print attack, replay attack, and 3D mask attack, where attackers impersonate the target identity by covering the whole face region to fool face recognition systems. Generally speaking, 2D print and replay attacks can be easily launched by non-expert persons. In turn, 3D mask attacks, including silicon masks, resin masks, plastic masks, and mannequins, always demand advanced fabrication systems to capture the target person's 3D facial information, which requires great efforts and costs [157]. On the other hand, more advanced FPA types have been subsequently proposed, such as makeup attack, tattoo attack, funny glasses attack, and wig attack, as shown in Figure 3(e)–(h). We cast these attacks as the obfuscation attack, where attackers partially obfuscate the face region to hide the attacker's identity. Compared with the impersonation FPA, the latter is more realistic and challenging to detect.

Malicious hackers can handily download the media content circulating on the internet and launch two types of face attack: physical and digital face attacks. Figure 2 illustrates the general pipeline. On the side of physical attacks, numerous face presentation attacks, such as 2D print/replay attack, 3D mask attack and makeup attack, can be easily launched to hack the face authentication systems over various application scenarios. On the other side, the attacker can also employ off-the-shelf APPs (e.g., ZAO [4], Facebrity [2], and Reface [3]) or disclosed face manipulation algorithms (e.g., Deepfakes [60]) to edit face

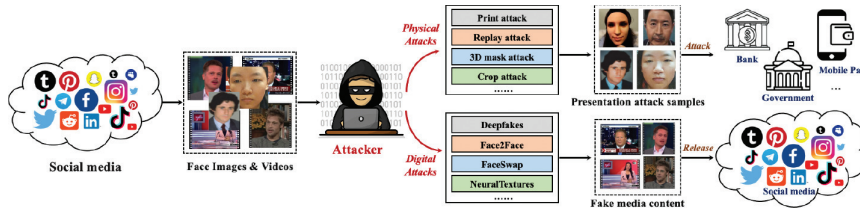

Figure 2: Overview of physical and digital face attacks.

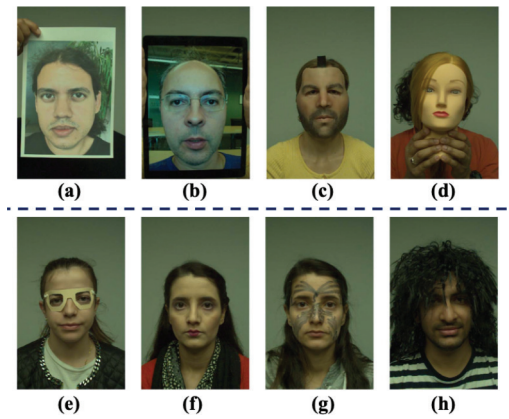

Figure 3: Typical face presentation attack (FPA) examples Heusch *et al.* [71]. The top row shows the impersonation attacks: (a) print, (b) replay, (c) 3D mask, and (d) mannequin. The bottom row presents the obfuscation attacks: (e) glasses, (f) makeup, (g) tattoo, and (h) wig.

content fueled with targeted disinformation or misinformation. The generated fake content can be released or disseminated to the social network platforms, causing detrimental mistrust issues. Even worse, as the attack methodologies are getting increasingly advanced, the produced fake faces are becoming more and more sophisticated. Powerful as the attack technique is, there is a thin line between bonafide and fake faces that can be hardly distinguished by human naked eyes, and it is easy to cross over. To that end, the abuse of either physical or digital face attacks will certainly lead to the tendency of reducing the trust of digital media content and raising tangible concerns, in the long run.

To counter various malicious physical face attacks and safeguard face recognition systems, numerous traditional methods have been first proposed. These methods mine informative artifacts via extracting hand-crafted features such as histograms, gradients, and texture [21, 52, 96, 137, 157, 159]. With the advent of deep learning, the accuracy of learning-based PAD methodologies [23, 82, 108, 130, 179, 227, 230, 241] significantly boosts. To overcome the overfitting

problem of the data-driven models, some methods seek to employ auxiliary modality information, such as remote physiological signals (rPPG) [115, 123, 130, 229], pseudo depth maps [15, 130, 211, 233, 234, 241], Near-infrared (NIR) maps [124, 129, 189, 230], and Thermal maps [138, 175]. On the other hand, great efforts have been dedicated to tackling the digital face attack problems over the past five years. Most existing face forgery detection algorithms are AI- or deep-learning based [39, 97, 139, 142, 165, 192, 243]. Apart from some models focusing on distinguishing input face images/videos between real and fake, some recent works [43, 76, 97, 206, 224] propose to localize forged regions for fake face appearances. Unsurprisingly, most learning-based detectors suffer significant performance drops when deployed to unforeseen datasets or attack types. As such, [66, 114, 126, 135, 183, 244] designed more generalized face forgery detection models that mine more inherent clues of fake media and mitigate the severe domain gaps. Figure 4 shows the explosive increase of published literature numbers for both digital and physical face attack detection in recent years. It can be seen that significant efforts have been devoted to the face forensics community, making it an active research area.

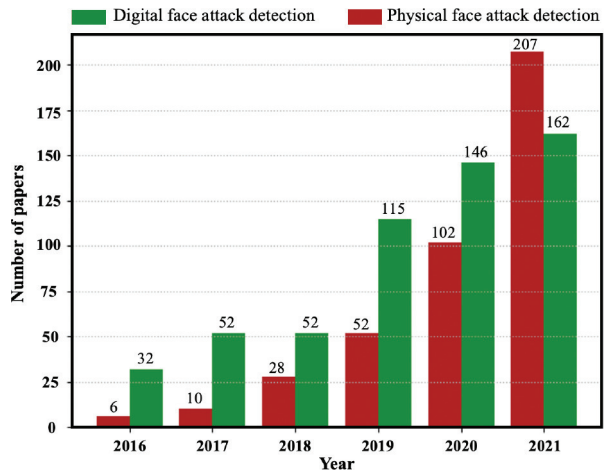

Figure 4: The growing trend in the number of papers in the digital face attack detection and physical face attack detection fields [86, 231].

Compared with previous surveys, we note that this survey is unique and superior in the following three aspects:

- We, for the first time to the best of our knowledge, aggregate and examine the literature on both digital and physical face attacks into one survey. We outline that the unified face attack detection would be a promising research area in the face forensics community.

- As shown in Table 1, this survey provides, by far, the latest and most comprehensive overview of the face forensics literature (>250 research papers) over attack types, datasets, and detection methodologies.

- This survey poses severe security, privacy, and explainability issues that have been largely understudied in the existing literature. Based on the outlined issues, we further suggest future research orientations to facilitate the development of this community.

Table 1: Comparisons with prior survey papers.

Prior Surveys	Timelines	Ref. Scale	#Dataset	Physical Attack	Digital Attack	Unified Attack
Souza *et al.* [185]	2018	98	9	✓	-	-
Raheem *et al.* [168]	2019	90	14	✓	-	-
Pereira *et al.* [161]	2019	57	7	✓	-	-
Jia *et al.* [80]	2020	74	10	✓	-	-
Safaa El-Din *et al.* [174]	2020	127	8	✓	-	-
Kotwal *et al.* [102]	2020	42	12	✓	-	-
Yu *et al.* [231]	2022	252	36	✓	-	-
Nguyen *et al.* [151]	2019	106	3	-	✓	-
Verdoliva [201]	2020	274	9	-	✓	-
Lyu [136]	2020	34	5	-	✓	-
Tolosana *et al.* [197]	2020	200	7	-	✓	-
Mirsky and Lee [143]	2021	192	3	-	✓	-
Ours	2022	253	32	✓	✓	✓

This survey starts with reviewing the physical face attacks and digital face attacks in Sections 2 and 3, respectively. We briefly discuss the importance of the problem and concretely review related literature regarding attack types, datasets, and detection methodologies. Then we analyze the existing security issues and suggest possible future research directions to facilitate the development of the face forensics community. Section 4 innovatively investigates the unifying detection works against face spoofing and face forgery. We also thoroughly analyze and discuss the motivations, benefits, and future research of the unified face attack defense. Finally, we draw the conclusion in Section 5.

2 Physical Face Attacks

As automatic face recognition (AFR) systems have been prevalently deployed in a wide variety of applications, face presentation attack detection (PAD) has attracted extensive attention from both industry and academia. It is of utmost necessity to safeguard AFR against malicious physical face attacks. In

this section, we provide a comprehensive review on the literature of physical face attacks. The main literature structure is illustrated in Figure 5.

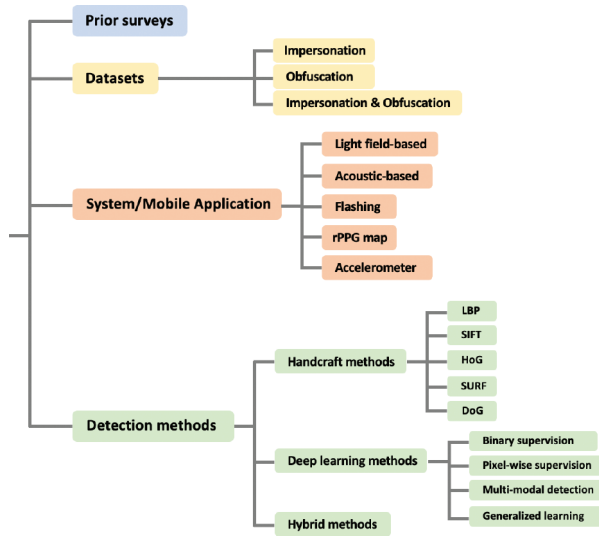

Figure 5: Tree diagram of physical face attack paper structure.

2.1 Importance of the Problem

Physical face attacks, also known as face presentation attacks (FPAs), can be deployed either as an obfuscation attack or as an impersonation attack, where the former attempts to hide one's identity and the latter aims at impersonating the target person. Past decades have witnessed the rapid proliferation of face authentication systems, and they have exploded into various practical applications ranging from log-in, financial payment, check-in, etc. FPA is getting increasingly notorious because it can easily bypass the face authentication system. For example, Apple's FaceID was hacked by a 3D mask FPA [5] in 2017 and caused disturbing security concerns. With the rapid development of attack methods and fabrication techniques in this era, FPAs tend to be more and more sophisticated and challenging. As such, it is of great importance to design highly accurate and secure presentation attack detection (PAD) models to safeguard face recognition systems against FPAs. To better illustrate how current face recognition systems assemble with face attack detection models, we present the relationship between face recognition and face digital/physical attack detection in Figure 6. Generally speaking, there are two popular schemes for deployed face recognition systems: (a) parallel scheme and (b) serial scheme. For the parallel scheme, the face

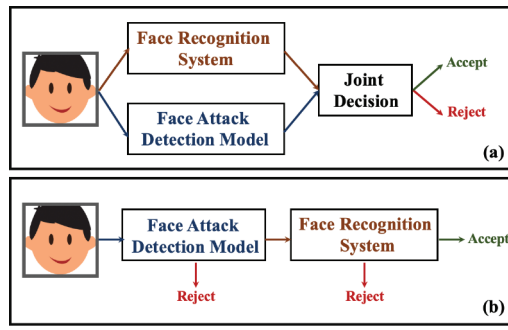

Figure 6: Illustration of the relationship between face recognition and face digital/physical attack detection. (a) parallel scheme and (b) serial scheme.

recognition system and face attack detection model are employed to jointly decide the authorization of the input face. In turn, the serial scheme first feeds the input face into the face attack detection model. Once the attack is detected, the input face will be rejected immediately. Otherwise, the face will be delivered to the face recognition system for final decision-making.

2.2 Face Presentation Attacks and Datasets

Face presentation attacks can be generally divided into two categories: obfuscation attacks and impersonation attacks. As shown in Figure 3(e)–(h), obfuscation attacks such as glasses, makeup, tattoo, and wig attempt to hide someone's identity. On the other hand, impersonation attacks (Figure 3(a)–(d)) attempt to mimic the target person's identity by copying the target person's face to specific mediums, such as paper, screen, and 3D mask. Over the past fifteen years, substantial efforts have been devoted to building face anti-spoofing (FAS) datasets for facilitating the algorithm design of presentation attack detection. In Table 2, we comprehensively summarize the existing face presentation attack databases in terms of modality, quantity, spoof medium, and acquisition device. To fit the uncontrollable environmental variables (e.g., illumination, scene, acquisition device, spoof medium, etc.) in practical scenarios, face anti-spoofing (FAS) databases tend to be increasingly diverse. On the other hand, scale is another pivotal factor of FAS databases, as most deep learning-based methods demand large-scale training data to guarantee high-level PAD performance when deployed in real-world applications. Besides the RGB vision modality, more modalities such as depth map, near-infrared (NIR), thermal map, flashing, and acoustic have been gradually incorporated in lastly released databases. The additional modalities can serve as auxiliary information to improve the generalization capability of PAD models. Due to the two-player nature between FPA and PAD, novel presentation attacks

Table 2: Summary of face presentation attack databases.

Database	Release year	Modalities	#Images or Videos (Live, Spoof)	Spoof medium	Acquisition device
ZJU EyeBlink [156]	2007	RGB	(80, 100)	High-quality photo	Webcam (320×240)
NUAA [193]	2010	RGB	(5105, 7509)	A4 paper	Webcam (640×480)
IDIAP Print Attack [13]	2011	RGB	(200, 200)	A4 paper	MacBook Webcam (320×240)
CASIA FASD [242]	2012	RGB	(200, 450)	iPad 1 (1024×768) Printed photo	Sony NEX-5 (1280×720) USB camera (640×480) Webcam (640×480)
IDIAP Replay Attack [35]	2012	RGB	(200, 1000)	iPad 1 (1024×768) iPhone 3GS (480×320)	MacBook Webcam (320×240) Cannon PowerShot SX 150 IS (1280×720)
3DMAD [148]	2013	RGB, Depth	(51100, 25500)	3D Mask	Microsoft Kinect for Xbox 360 (640×480)
MSU-MFSD [212]	2015	RGB	(110, 330)	iPad Air (2048×1536) iPhone 5s (1136×640) A3 paper	Nexus 5 (720×480) MacBook (640×480) Canon 550D (1920×1088) iPhone 5s (1920×1080)
MSU-RAFS [158]	2015	RGB	(55, 110)	MacBook (1280×800)	Nexus 5 (1920×1080) iPhone 6 (1920×1080)
IDIAP Multi-spectral-Spoof [36]	2016	RGB, Near-Infrared	(1689, 3024)	A4 paper	u-Eye camera (1280×1024)
MSU-USSA [157]	2016	RGB	(1140, 9120)	MacBook (2880×1800) Nexus 5 (1920×1080) Tablet (1920×1200) 11×8.5 in. paper	Nexus 5 (3264×2448) Cameras used to capture celebrity photos
HKBU-MARs [128]	2016	RGB	(504, 504)	3D Mask	Logitech C920, industrial camera, Canon EOS M3, Nexus 5, iPhone 6 Samsung S7, Sony Tablet S
OULU-NPU [22]	2017	RGB	(1980, 3960)	A3 paper Dell UltraSharp 1905FP Display (1280×1024) MacBook 2015 (2560×1600)	Samsung Galaxy S6 edge HTC Desire EYE, OPPO N3 MEIZU X5, ASUS Zenfore Selfie Sony XPERIA C5 Ultra Dual

Table 2: Continued.

Database	Release year	Modalities	#Images or Videos (Live, Spoof)	Spoof medium	Acquisition device
SiW (Spoofing in the Wild) [130]	2018	RGB	(1320, 3300)	Samsung Galaxy S8 iPhone 7, iPad Pro PC (Asus MB168B) screen	Canon EOS T6 Logistech C920 webcam
ROSE-YOUTU [108]	2018	RGB	4225	A4 paper Lenovo LCD (4096×2160) Mac screen (2560×1600)	Hasee smartphone (640×480) Huawei Smartphone (640×480) iPad 4 (640×480) iPhone 5s (1280×720) ZTE smartphone (1280×720)
IDIAP-CSMAD [18]	2018	RGB, near-infrared (NIR), Thermal from long-wave infrared (LWIR)	(87, 159)	3D Mask	Realsense SR300, Compact Pro
3DMA [214]	2019	RGB, NIR	(536, 384)	3D Mask	R0710A binocular camera (640×480)
CUHK MMLab CelebA-Spoof [240]	2020	RGB	625,537	A4 paper Face mask, PC	24 sensors with 4 types (PC, Camera, Tablet, Phone)
CASIA-SURF 3D Mask [233]	2020	RGB	(288, 864)	3D Mask	Apple, Huawei, Samsung iPhone11, iPhoneX, MI10, P40, S20, Vivo, HJIM
CASIA-SURF 3D HiFi Mask [125]	2020	RGB	(13650, 40950)	3D Mask	
Ambient-Flash [46]	2021	RGB, additional light flashing	(7503, 7503)	Printed paper Digital screen	Logitech C920 HD webcam LenovoT430u laptop webcam (640×480), MotoG4
Echo-Spoof [99]	2022	RGB, Acoustic	(82,850, 166,666)	A4 paper iPad Pro (2388×1668) iPad Air 3 (2224×1668)	Samsung Edge Note (2560×1440) Samsung Galaxy S9 (3264×2448) Samsung Galaxy S21(4216×2371) Xiaomi Redmi7 (3264×2448)

with higher quality will be constantly proposed with the development of smart attack algorithms and advanced fabrication techniques. As such, it is unsurprising that attack types tend to be more and more diverse in newly published databases.

2.3 Overview of Presentation Attack Detection Methodologies

2.3.1 Face Liveness Detection Systems and Mobile Applications

Face liveness detection techniques have been widely deployed in real-world applications. With various sensors (e.g., RGB camera, speaker, microphone, accelerometer, etc.) assembled, most devices such as smartphones are able to take advantage of multi-modality information captured by different sensors to conduct more accurate and generalized PAD. Thanks to the pervasive availability of speakers and microphones on mobile devices, acoustic signals have been demonstrated to be effective in capturing biometric information from users for various mobile-oriented applications. Recently, great efforts [30, 99, 217, 246] have been devoted to devising acoustic-based face liveness detection frameworks to perform more reliable PAD in practical scenarios. EchoPrint [246], for the first time, incorporated both RGB vision modality and acoustic modality to conduct the user authentication. However, face anti-spoofing has been largely ignored in EchoPrint. Follow-up works such as Echoface [30] achieved more than 96% accuracy by using acoustic signals solely. Rface [217] demonstrated that radio frequency signals could identify both 2D print/replay and 3D mask attacks with high-level accuracy. Moreover, EchoFAS [99], designed a more advanced signal configuration and aggregated CNN and vision transformer to achieve outstanding PAD performance.

Besides acoustic signal, some recent works [26, 49, 194] proposed to use the flash to conduct a secure face liveness detection, based on the theory that the reflection characteristics of bonafide and PAs are distinguishable. FaceRevelio [50] demonstrated that varying illumination could enable reconstructing the 3D face surface of the input, thereby achieving robust and accurate face liveness detection. Given the fact that dual-pixel sensors have been widely built in mobile devices, Wu *et al.* [213] proposed to capture dual-pixel images to reconstruct the depth map and subsequently distinguish the bonafide from PAs. Chen *et al.* [33] used an RGB camera to conduct PAD by comparing the rPPG maps of face and fingertip videos, which should be highly consistent if they are captured from a live person. Apart from the sensors mentioned above, the accelerometer and gyroscope have also been incorporated into face liveness detection. FaceLive [117] employed accelerometer and gyroscope to measure the movement data, and the head pose video was meanwhile recorded by the built-in camera. Then the consistency between the two modality data would be used to discriminate the bonafide from attacks.

2.3.2 Presentation Attack Detection Methodologies

In this survey, we classify PAD methods into three categories: traditional handcraft methods, deep learning methods, and hybrid methods. Traditional handcraft methods attempt to extract hand-crafted features such as local binary pattern (LBP) [11], scale-invariant feature transform (SIFT) [157], histograms of oriented gradients (HoG) [42], speeded-up robust features (SURF) [21], and difference of gaussian (DoG) [193] to perform face liveness detection. As illustrated in Figure 7(a), deep learning methods, empowered by effective neural network architectures, aim at directly extracting deep features from input face images/videos for PAD. In turn, hybrid methods assemble handcrafted feature extraction and deep feature extraction modules into one framework for final decision-making.

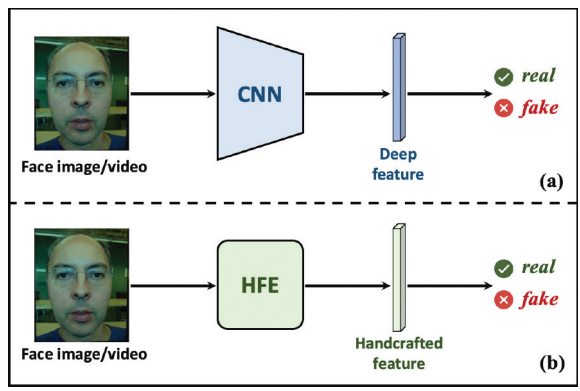

Figure 7: (a) Deep learning methods for FAS. (b) Traditional handcrafted methods for FAS. HFE indicates a variety of hand-crafted feature extraction algorithms.

Traditional Handcraft Methods. We illustrate the general pipeline of handcraft method in Figure 7(b), where HFE indicates a variety of hand-crafted feature extractors such as LBP [11], SIFT [157], HoG [42], SURF [21], and DoG [193]. LBP [11] was taken as a local texture descriptor that assigned a binary label to each pixel, and the binary number was determined by the values of the central pixel and its neighbor pixels. SIFT [157] had been widely employed to capture image representations in many computer vision tasks as it was invariant to various image distortions, such as rotation, scale, translation, etc. Besides, HoG [42] showed great superiority in capturing representative features of images compared with previous edge and gradient based descriptors. SURF [21] was a fast and efficient scale and rotation invariant descriptor. It could effectively speed-up the computation and extract robust scale-independent features to perform accurate face liveness detection. Moreover, the DoG [193]

filter could effectively remove the noise in the high-frequency domain, hence empowering high-performance face anti-spoofing.

Learning-based Methods. Thanks to the advent of deep learning, enormous progress has been achieved in this research field. Early deep learning attempts on FAS date back to 2014, where Yang *et al.* [220] first proposed to design a convolutional neural network with some data pre-processing, such as spatial and temporal augmentations, to achieve an outstanding FAS performance. Presentation attack detection can be regarded as a binary classification problem. Lucena *et al.* [134] found that pretraining the VGG16 [184] on ImageNet [172] and transferring the learned knowledge to FAS could effectively save computational resources and avoid the overfitting problem. With the rapid progress in network architectures, more advanced networks such as the siamese network [67] and transformer [59] have been applied to the FAS task. Moreover, Chen *et al.* [29] designed a two-stream framework complementarily combining RGB feature and multi-scale retinex (MSR) feature via an attention-based fusion module and achieved outstanding generalization capability. Deb and Jain [44] demonstrated that local face patches could effectively reflect the inherent cues for more generalized detection. Similarly, Wang *et al.* [202] designed PatchNet to mine informative local cues and proposed asymmetric margin-based classification loss and self-supervised similarity loss to regularize the patch embedding space. On the other hand, PAD against video replay attacks plays a critical role in securing automatic face recognition systems. As such, some methods proposed to employ LSTM [55, 218] and RNN [146] to detect the temporal consistency. Yang *et al.* [221] exploited a novel spatial-temporal network to capture subtle evidence in both spatial and temporal domains.

Generally speaking, binary supervision can easily cause severe overfitting problems (*i.e.*, lack generalization capability to unseen environments). To mitigate the domain gap between training and testing data, many methods seek to use auxiliary supervision in the training phase, such as the binary mask [57, 73, 131, 132, 188, 232] and depth map [15, 24, 56, 124, 130, 152, 160, 182, 211, 230, 234]. Binary mask-based methods assigned 0/1 to each pixel in bonafide/fake regions. The idea of the depth map is based on the fact that live faces contain rich 3D facial structures while 2D FPA can barely reflect depth information. Typically, Sun *et al.* [188] demonstrated that the local label supervision scheme, including local depth map and local binary label supervisions, is superior to the global binary supervision for FAS. George and Marcel [57] conducted pixel-wise binary supervision at the feature level, thereby achieving a more accurate and robust detection performance. Liu *et al.* [131] designed a deep tree learning scheme with binary map supervision for zero-shot face anti-spoofing. On the other hand, depth map supervision

has been widely used in FAS since it can reflect rich intrinsic spoofing cues for 2D PAD. Liu *et al.* [130] designed a CNN-RNN framework and simultaneously estimated depth and rPPG maps for FAS at the video level. Yu *et al.* [234] proposed central difference convolutional operators and extended this work by further incorporating central difference pooling [233] to estimate the depth maps for PAD.

Moreover, since the reflection characterastics of PAs and bonafide are discriminative, some works [92, 226, 239] seek to use auxiliary geometric information such as reflection maps to conduct generalized face anti-spoofing. Benefiting from the advances of FAS systems, abundant auxiliary modality information is available in practical applications. For this reason, many methods propose to conduct robust PAD via multi-modal fusion. Besides RGB space, some detectors [20, 21, 103] demonstrated that HSV and YCbCr could provide informative clues. Recently, researchers found that near-infrared (NIR) modality [84, 103, 124, 129, 152, 182, 230, 238] contains abundant discriminative and generalized information than RGB and depth data since NIR measures the amount of heat radiated from a live face. Specifically, Liu *et al.* [129] proposed a multi-modal two-stage cascade framework that fused three modalities of RGB, depth map, and NIR to perform PAD. Liu *et al.* [124] proposed a modality translation-based FAS method that translated the RGB face image into more generalized NIR image, thereby achieving an excellent generalization capability. Besides NIR modality, rPPG signals have also been exploited in PAD since rPPG signals can reflect periodic heart rhythms of input faces. Some models [115, 123, 130, 225, 229] attempt to incorporate rPPG modality to mine inherent face spoofing cues and conduct more robust FAS.

To further improve the generalization capability of PAD, researchers recently turned to domain generalization and domain adaptation algorithms that have been demonstrated effective in a wide variety of tasks, including computer vision, natural language processing, and multi-modality problems. Domain adaptation aims at learning a model on source domain data that can adapt well to target domains with different data distributions. Recently, various domain adaptation-based methods have been proposed for FAS [83, 108, 145, 204, 205, 207, 247]. Li *et al.* [108], for the first time, used the knowledge of domain adaptation to tackle the FAS problem. The authors proposed to minimize the Maximum Mean Discrepancy (MMD) to align the distributions of training and test datasets in high-dimension feature space. Wang *et al.* [205] designed an unsupervised adversarial domain adaptation framework to learn domain-invariant features for robust FAS. To overcome the problem that the target domain data is always unavailable in the training stage, Wang *et al.* [207] designed a meta-learning based model that could adapt better to target domains. On the other hand, domain generalization aims at learning a robust model on source domains that can generalize well to unforeseen target

domains. Due to the uncontrollable environmental variables (e.g., illumination, capture device, and attack types), the trained PAD models tend to easily suffer significant performance drops in practical applications. For this reason, extensive efforts [17, 27, 58, 82, 109, 122, 127, 153, 162, 177, 203, 215] have been devoted to domain generalization-based FAS methods in the recent few years.

Typically, Jia *et al.* [82] proposed a single-side domain generalization model to obtain compact and generalized features on the real side. Similarly, George and Marcel [58] designed a multi-channel CNN and used a one-class classifier to learn compact embedding for the bonafide class. Besides, Shao *et al.* [177] proposed to use adversarial learning to align the feature distributions between source and target domains. Wang *et al.* [203] used disentangled representation learning to disentangle spoofing-related features from subject-related features and achieved outstanding generalization performance. Chen *et al.* [27] designed a two-branch framework to capture camera-invariant features for robust PAD. Last but not least, more effective learning schemes such as zero-shot learning [166], meta learning [24, 34, 131, 163, 179, 207], knowledge distillation [114, 121], and progressive transfer learning [167] have been deployed to PAD and achieved promising generalization capability.

Hybrid Methods. Hybrid methods aim at taking advantage of discriminative handcrafted features and powerful learning-based models for PAD. These methods can be typically divided into three categories: (1) Extracting handcraft features first and then feeding them forward to neural networks [90, 110, 112, 228]; (2) Using deep models to extract deep features first and subsequently extracting handcrafted features from deep features [14, 111, 178]; (3) Handcraft features and deep features are fused together for final classification [51, 169, 170, 181, 228]. To be more specific, Li *et al.* [111] demonstrated that motion blurs could reflect informative clues for discriminating the bonafide from replay attacks. They first extracted the motion blur indicator for each input video and then applied 1D CNN to extract deep features. Besides, TransRPPG [228] designed a novel transformer-based framework for 3D mask PAD. TransRPPG first extracted the rPPG map from an input video and then designed a two-branch ViT to extract rPPG and environmental features for final decision-making. Feng *et al.* [51] proposed a neural network-based method by synchronously aggregating the image quality and motion cues from input face videos to conduct FAS.

For the second category, Shao *et al.* [178] proposed deep dynamic textures for 3D mask FAS by using pre-trained VGG Net to extract deep features and then applying optical flow [16] to form the deep dynamic texture features. Li *et al.* [110] proposed to fine-tune the VGG-face model to extract deep features and then used PCA [6] to overcome overfitting problems. Moreover, Rehman *et al.* [170] combined deep features and HOG maps of input face images to

perform PAD. Rehman *et al.* [169] designed a FAS framework that enhanced discriminative features by aggregating deep features and LBP texture maps of input faces. Overall, although some hybrid methods can benefit from the advantages of handcrafted features and deep features, they still have obvious drawbacks, such as needing expert prior knowledge for handcrafted feature extraction. Thus, they cannot guarantee a global optimum for FAS.

2.4 Counter-Forensics Issues

The advent of effective detection tools always comes with more powerful attack methods since attacks and defenses are in an arms race. Skilled attackers are able to launch adversarial attacks to bypass PAD models. Herein, we generally categorize them into two types: physical adversarial attacks and digital adversarial attacks.

As shown in Figure 8, physical adversarial attacks, including adversarial hat [95], adversarial glasses [180], adversarial makeup [223], and adversarial sticker [63], are capable of hacking face recognition systems under the black-box or white-box setting. Even worse, these attack methods can easily bypass most existing PAD models since the live persons are indeed actively present in front of the devices. Thus, how to design more generalized FAS models to counter the evasion of PAD remains an open problem.

Figure 8: Physical adversarial face attack samples: adv. hat [95], adv. examples [180], adv. makeup [223], and adv. sticker [63].

Besides physical adversarial attacks, digital adversarial attacks also pose a significant challenge in this research community. As illustrated in Figure 9, although the existing spoofing detection model can effectively distinguish bonafide from PAs, the expert attacker can launch adversarial attacks on input media content to fool the detection model. With the rapid developments of adversarial attack algorithms, more and more attack methods [9, 10, 235] on spoof faces have been recently proposed to deceive existing FAS models, among which [235] can even achieve a 100% attack success rate. There is no doubt that these attack methods will cause severe security concerns and hazardous crises. Therefore, it is non-trivial to develop a more secure and robust FAS model to counteract the menace of these face adversarial attack techniques.

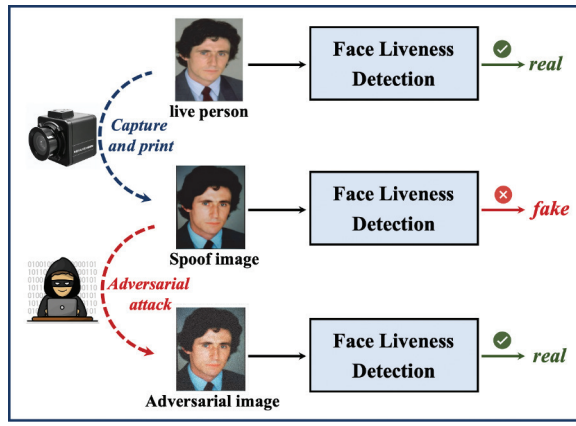

Figure 9: The spoofing detection model can effectively discriminate bonafide from presentation attacks. However, expert attacker is able to apply adversarial attack technologies on spoof images to nullify the detection model.

2.5 Future Research Directions

To date, there are still many open issues that need to be properly addressed in the FAS research field. On the one hand, industry is now somewhat ahead of academia. For instance, Apple FaceID takes advantage of the aggregation of three modules: a dot projector, a flood illuminator, and an infrared camera to capture the 2D infrared face image and reconstruct the 3D facial structure. However, in the research community, most existing FAS databases are somewhat outdated. More advanced PA databases are expected in the future. On the other hand, generalized PAD is a long-standing challenge in this research area. Mining inherent spoof clues and designing more effective networks are necessary to empower the generalization capability. In addition, the problem of privacy leakage during the face recognition process has raised pressing concerns. Proposing privacy-reserved PAD methodologies is also of great importance to address the concerns and secure the user privacy.

2.6 Discussion

In this section, we comprehensively reviewed the existing literature on PAD in terms of face spoofing datasets, PAD systems, and PAD methodologies. We further analyzed the existing security issues and main risks of adversarial attacks. Moreover, we outlined promising future research areas in physical face attacks to facilitate the development of both industry and academia.

3 Digital Face Attacks

The digital face attack on media content are actually not a new problem. The first ever attempt at face identity swap dates back to 1860, where Abraham Lincoln's head is stitched up with the body of southern politician John Calhoun [41]. Figure 12 depicts the tree diagram of literature structure on digital face forensics. Previous works can be basically classified into five categories: surveys, dataset papers, attack methodologies, detection methods, and other works. Numerous existing survey papers have reviewed prior literature on face forgery attacks and detection methodologies. However, these surveys are somewhat outdated and uninspiring to neither industry nor academia. Herein, we comprehensively review the forgery generation methodologies, deepfake datasets, existing attack detection models, and counter-forensics works. Moreover, we thoroughly analyze existing issues needed to be properly addressed and propose possible future research directions.

3.1 Importance of the Problem

In recent years, falsified media content has become a vital problem on social media platforms. Faces play a central role in human communication, as a person's face can emphasize a message or even convey a message in its own right [53]. However, due to the unrestricted access to enormous face media content on the network, face forgery attacks aim at manipulating pristine face images/videos have posed pressing security risks to the public at large. The situation gets even worse with the advent of AI and deep learning. The fake faces generated by deep learning methods are referred to as Deepfakes in face forensics community. Empowered by Deepfakes, the quality-level and fidelity-level of fake multimedia content have been improved so rapidly that human eyes can hardly identify the authentication. Due to the zero-barrier accessibility of the high-performance face attack resources on some open platforms (e.g., Github), non-expert persons without any prior professional knowledge can readily use disclosed face forgery algorithms or APIs to create sophisticated fake content for either entertaining or malicious purposes. In this vein, these techniques could be easily fueled with targeted disinformation or misinformation and cause harmful consequences over fraud, impersonation, and rumor. For this reason, it is urgent to propose effective and robust face forgery detection methods to build digital media integrity and safeguard social platforms from face forgery attacks.

3.2 Digital Face Attack Methodologies

Digital face forgery can be generally classified into four categories: identity swap, face reenactment, attribute manipulation, and entire synthesis. Figure 1 summarizes digital face attack types and the corresponding forgery regions,

where white pixels indicate the forged regions. Note that the manipulation regions of identity swap and face reenactment are provided by the official FF++ [171] dataset. Some works also define the manipulation region as the absolute difference between the pristine images and the corresponding forged ones. Generally speaking, Attribute manipulation and entire face synthesis techniques tend to bring positive impacts to human lives, while identity swap and face reenactment could cause disconcerting security problems [86]. For this reason, this survey mainly focuses on identity swap and facial reenactment (*a.k.a.* expression edition).

3.2.1 Identity Swap

The first ever work on identity swap dates back to 1860, where Abraham Lincoln's head is stitched up with the body of southern politician John Calhoun [41]. A typical face swap pipeline is shown in Figure 10, which can be generally divided into four components: face alignment, face warping, face replacement, and post-processing. Heading to the era of deep learning, numerous learning-based face swap frameworks have been designed that extensively boost the quality of generated fake faces. Figure 11 illustrates a typical fake face generation pipeline. In the training phase, two face auto-encoders that share one identical encoder are trained for the source person and the target person. The encoder learns the shared information from input faces, while two decoders are responsible for capturing the specific information for the two identities [98]. As such, in the inference stage, the source face is firstly fed forward to the encoder and subsequently passed to the target person's decoder to produce the forged face. With the advent of powerful generative models, such as autoregressive models [199, 200], generative adversarial networks (GANs) [61], and variational autoencoders (VAEs) [93, 94], generative networks have become the main-stream deepfake generation architectures. Over the past three years, numerous terrific deepfake generation methodologies have been proposed to generate forgery face images with high-level quality and realism [31, 54, 113, 133, 154, 216, 253].

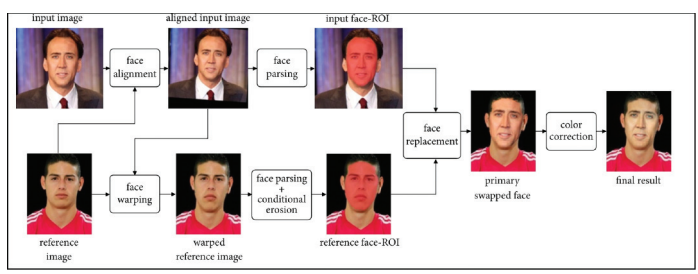

Figure 10: Typical pipeline for face swapping [28].

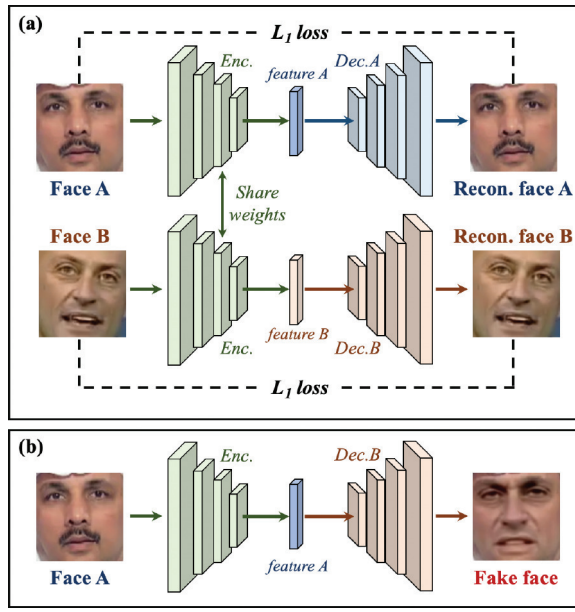

Figure 11: Deepfake content creation pipeline. (a) Training phase: two face auto-encoders with one identical encoder and two specific decoders are trained under the supervision of face reconstruction loss and (b) Inference phase: feed forward the source face to the encoder and employ the target person's decoder to produce the fake face.

3.2.2 Face Reenactment

Face reenactment is also known as the expression edition. Face2Face [195] is one of the typical expression edition methods. It proposed a real-time face reenactment method that could reenact the target video sequence of photo-realistic quality by using a three-step solution. Follow-up works such as A2V [191] designed a cross-modal framework that was able to generate high-quality mouth texture with accurate lip sync. It employed a recurrent neural network to learn the mapping from audio features to mouth shapes, thereby could synthesize realistic speech videos. Tripathy *et al.* [198] achieved facial expression transfer with a single source and target face images by using GANs. To further improve the fidelity and quality levels of synthesized videos, many powerful frameworks had been designed in recent two years. For example, Ha *et al.* [64] designed three components: image attention block, target feature alignment, and landmark transformer to fix the identity mismatch issue between the target identity and the driver identity. Recently, 3DMM [209] attempted to use a single source image and a driving video to synthesize the speech video. Hyun *et al.* [78] further improved the quality of face reenactment in terms of

appearance consistency and motion coherency in videos. The majority of the following works focused on making generated videos look more natural and realistic. Zhang *et al.* [236] demonstrated that audio not only had a high correlation with lip motion but also had a low correlation with head movement and eye-blinking. Moreover, Zhang *et al.* [236] further learned to render the head pose and eye-blinking in the synthesized videos to make them more natural.

3.3 Digital Fake Datasets

Based on the release time, we summarize the existing face forgery datasets in Table 3. According to the level of scale, quality, fidelity, and the real-world application scenarios, we generally divide these datasets into three generations. 1^{st} generation: UADFV [222], DF-TIMIT [100], and FaceForensics++ [171]; 2^{nd} generation: DFD [47], DFDC [47], and Celeb-DF [120]; 3^{rd} generation: DF-Forensics-1.0 [85], ForgeryNet [68], FFIW [248], KoDF [105], and FakeAVCeleb [89]. The datasets are elaborated one-by-one as follows:

- **UADFV** [222] consists of 49 real videos and 49 fake videos, with 17.3k frames manipulated. The deepfake videos are produced by using generative neural networks and post-processing algorithms.

- **DF-TIMIT** [100] contains 320 pristine videos and 640 deepfake videos generated by faceswap-GAN [1] with 32 subjects. Fake videos are equally split into high-quality(HQ) and low-quality(LQ) subsets, corresponding to the face regions with different resolutions: 128×128 and 64×64. Compared with UADFV, DF-TIMIT has a higher diversity and a larger scale.

- **FaceForensics++** [171] is one of the most pervasive digital face attack datasets in the community. It composites of two forgery types: identity swap and facial reenactment, with each one containing one traditional and one deep learning-based attack, resulting in four automated face manipulation methods: Deepfakes, Face2face, FaceSwap, and NeuralTextures. FaceForensics++ covers three quality levels, and each level contains 1,000 pristine videos and 4,000 manipulated videos.

- **DFD** [48]. Deepfake detection dataset (DFD) was released by Google/Jigsaw in 2019. It consists of 363 real videos and 3,068 deepfake videos with 28 consented subjects in various practical scenes.

- **DFDC** [47] includes 1,131 real videos and 4,113 fake videos, which are generated by two face-swap algorithms. DFDC is of high diversity in terms of scenes and actor characteristics.

- **Celeb-DF** [120] has a higher level of quality and fidelity compared with the datasets released earlier. It collects 590 real celebrity videos from YouTube and generates 5,639 fake videos based on real videos.
- **DF-Forensics-1.0** [85] is a large-scale deepfake dataset that contains 50,000 real videos and 10,000 forged videos generated by an end-to-end automatic face swapping model. The dataset shoot videos from 100 paid actors of various ages, skin colors, nationalities, and genders. To better imitate real-world scenarios and produce more challenging videos, extensive perturbations are applied in this dataset.
- **ForgeryNet** [68] builds an extremely large forgery dataset with both image- and video-level labels. ForgeryNet provides 221,247 videos shot from more than 5400 subjects, and the fake videos are generated by 15 different manipulation algorithms. It synchronously facilitates the development of four vital tasks in digital face forensics: image forgery classification, spatial forgery localization, video forgery classification, and temporal forgery localization.
- **FFIW** [248] constructs a large-scale and high-quality deepfake dataset by designing a novel domain-adversarial quality assessment framework. Meanwhile, it proposes a novel algorithm to tackle the multi-person problem in face forgery detection. FFIW contains 10,000 real videos and 10,000 fake videos, with an average of more than three faces in each frame.
- **KoDF** [105] is a large-scale collection of deepfake and genuine videos on 403 Korean subjects. It contains 175,776 fake videos generated by six synthetic methods.
- **FakeAVCeleb** [89] fills the gap that existing deepfake datasets either contain deepfake videos or deepfake audios. FakeAVCeleb contains 19,500 fake videos, with both videos and audios manipulated.

3.4 Overview of Face Forgery Detection Methodologies

Face forgery detection methodologies can be generally classified into two categories: frame-level (image-level) detection and video-level detection. The former focuses on mining key spatial or frequency information to discriminate real faces from fake ones. On the other hand, the video-level detection methods can utilize the temporal-inconsistency features to distinguish the input video clip between real and fake. We elaborate on the details of both frame-level and video-level detection as follows.

Frame-level detection. As illustrated in Figure 12, we summarize frame-level detection methodologies as the following five types: (1) DNN-based

Table 3: Summary of digital face forgery attack databases.

Database	Release year	#Videos (Real, Fake)	#Synthetic Methods	#Face Per-frame	Subjects	Deepfake Audio	Finegrained Labeling
UADFV [222]	2018. 11	(49, 49)	1	1	49	N	N
DF-TIMIT [100]	2018. 12	(320, 640)	2	1	43	N	N
FaceForensics++ [171]	2019. 01	(1,000, 4,000)	4	1	1000	N	N
DFD [48]	2019. 09	(363, 3,068)	5	1	28	N	N
DFDC [47]	2019. 10	(1,131, 4,113)	2	1̃	960	N	N
Celeb-DF [120]	2019. 11	(590, 5,639)	1	1	59	N	N
DF-Forensics-1.0 [85]	2020. 05	(50,000, 10,000)	1	1	100	N	N
ForgeryNet [68]	2021. 03	(99,630, 121,617)	15	1	5400+	N	Y
FFIW [248]	2021. 03	(10,000, 10,000)	3	3.15	-	N	N
KoDF [105]	2021. 08	(62,166, 175,776)	6	1	403	N	N
FakeAVCeleb [89]	2022. 12	(500, 19,500)	4	1	500	Y	Y

detection: DNN-based detection methods are data-driven methods, including convolutional neural networks (CNN) [7, 39, 150, 192], recurrent neural networks (RNN) [173], and vision transformer (ViT) [70]. Afchar *et al.* [7] designed MesoNet and MesoInception4 to detect Deepfake and Face2Face videos automatically. Besides, some generic networks such as Xception Net [39], Efficient Net [192], and Capsule Net [150] have been demonstrated effective on deepfake detection tasks. Follow-up architectures such as RNN [173] and ViT [70] have been employed to further improve the forgery detection accuracy. Tremendous progress has demonstrated that DNN-based methods are able to achieve promising detection methods. However, they are vulnerable to adversarial attacks and tend to suffer severe overfitting problems. (2) Hand-crafted features such as the LBP map [210], color component [107], and DCT map [165] have been taken as informative indicators for Deepfake detection; (3) Spatial-based models are the most common forgery detection methods.

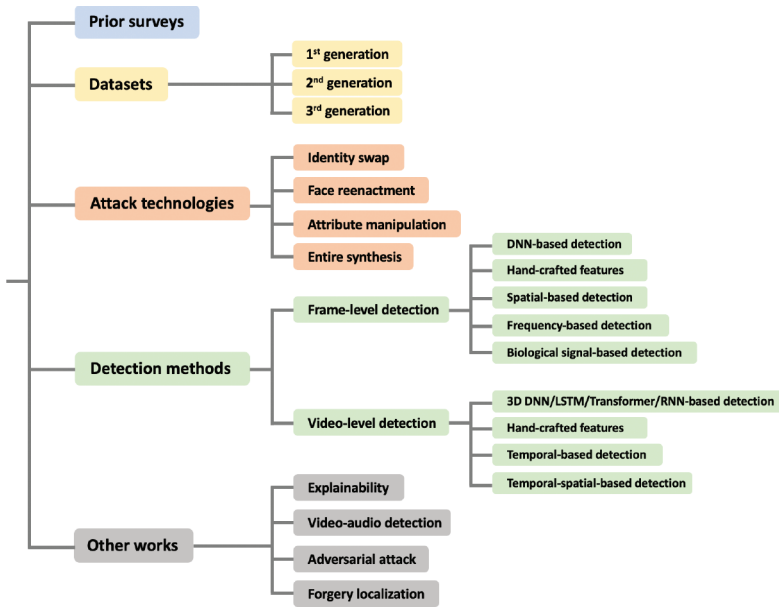

Figure 12: Tree diagram of digital face attack paper structure.

A wide variety of spatial-based features, such as local relations [32], pixel region relations [176], and context discrepancies [155], have demonstrated their effectiveness on different datasets. Besides, some prior arts [104, 116, 243] proposed to use attention mechanisms and multi-instance learning to capture informative clues in the spatial domain. (4) Frequency-based detection: F3Net [142] mined rich artifacts in the frequency domain and performed robust and accurate face manipulation detection. Miao *et al.* [142] extended this idea by

designing hierarchical frequency-assisted interactive networks to conduct more robust detection. Li *et al.* proposed a method that extracts frequency-aware discriminative features supervised by single-center loss. (5) Biological signal-based detection: some remote photoplethysmography (PPG) methods have been proposed to expose manipulation in synthesized videos. We illustrate classical face manipulation detection methods in Figure 13. The basic idea of these methods is grounded on the fact that fake videos cannot replicate the biological signal of synthesized faces. In this vein, DeepRhythm [164] utilized dual-spatial-temporal attention to capture normal heartbeat rhythms and detect deepfake videos. Similarly, [40] extracted ppg maps and computed spatial coherence and temporal consistency to identify the authentication of input videos.

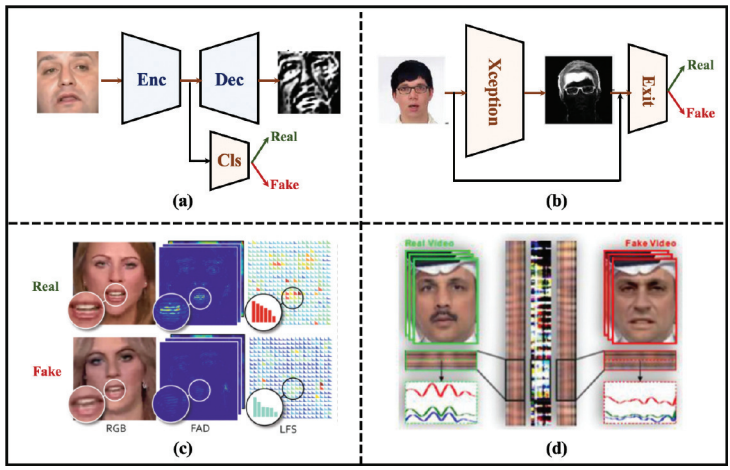

Figure 13: Illustration of classical face digital attack detection methods. (a) hand-crafted features; (b) spatial-based detection; (c) frequency-based detection; and (d) biological signal-based detection.

Due to the two-player nature between face forgery and forgery detection, attack techniques are getting smarter and smarter. Previous detection methodologies can achieve outstanding detection performance under intra-settings while they are struggling in detecting unforeseen deepfake attacks or datasets. It is of great significance to mitigate these domain gaps and propose more robust and generalized detection models. As shown in Figure 14, there are two general steps for generating manipulated faces. Given two input faces, **Step 1** applies face manipulation algorithms to alter the face content, and **Step 2** conducts various post processes like blending, color correction, and other post-processing. Inspired by the fake face generation pipeline, face X-ray [114] focused on the blending process, which employs ground-truth boundary

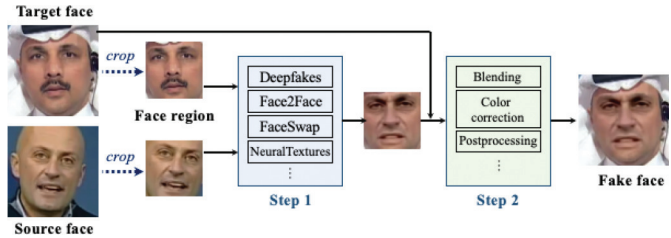

Figure 14: The overview of face manipulation pipeline can be generally regarded as a two-step process. **Step 1** aims at applying various algorithms to modify the face content. **Step 2** conducts the blending, color correction, and postprocessing processes.

maps and binary labels to jointly supervise the training process. Luo *et al.* [135] found that CNN-based models tend to overfit to training data. So they proposed to use more intrinsic high-frequency noise features to conduct generalized face forgery detection. SPSL [126] observed that the up-sampling operation is common in most manipulation techniques, and this operation introduces unique forgery traces in the frequency domain. Shiohara and Yamasaki [183] further extended this idea by incorporating more common forgery artifacts such as landmark mismatch, blending boundary, color mismatch, and frequency inconsistency to further improve the generalization capability. On the other hand, Zhao *et al.* [244] hypothesized that images' distinct source features could be preserved in manipulated faces. Based on this assumption, they proposed to measure the consistencies of image patches and achieved promising performance. Cao *et al.* [25] built an end-to-end reconstruction architecture to learn the optimal forgery patterns. Zhu *et al.* [252] found that decomposing an image into several constituent elements and utilizing direct light and identity texture can remarkably extract subtle forgery patterns. Moreover, lots of powerful learning regularities such as meta learning [186], few-shot learning [101], contrastive learning [187], and neural coverage [208] have been demonstrated effective for general forgery detection.

Besides forgery detection, forgery localization is another vital task in the face forensics community. Localizing manipulated regions of forgery faces can not only provide solid evidence for final decision-making but also unveil the potential intents of attackers. Some methods such as multi-task [149], DFFD [43], Detect and Locate [97, 140], and Fakelocator [76] have been recently proposed. They can accurately localize forgery regions and identify the authentication of input faces.

Video-level Detection. Most video-level detection methodologies capture temporal inconsistency in fake videos and combine spatial artifacts to jointly conduct the final decision-making. Generic neural networks such as 3DCNN [237], LSTM [12, 72], RNN [37, 173], and ViT [91] have achieved impressive

detection performance. Some methodologies focus on extracting hand-crafted features like eye-blinking [118], head pose [222], face warping [119], and lip movement [219]. Other models [62, 74, 75, 190] attempt to combine both spatial and temporal artifacts in manipulated videos and perform a more accurate deepfake detection. To defend against unforeseen attacks and datasets, more generalized and robust detectors have been designed in recent three years. DeepRhythm [164] demonstrated that the rppg maps could reflect heartbeat rhythms, which can be further taken as a reliable and robust indicator for video-level deepfake detection. Masi *et al.* [139] proposed a two-branch framework to capture the intrinsic low-level artifacts while suppressing the high-level semantic information in input videos. Haliassos *et al.* [65] only exploited real talking faces to conduct a more robust and generalized detection in a self-supervision manner. Lipforensics [66] focused on the irregularities in mouth movement, which are common in most manipulated videos. Besides, Temporal Coherence [245] proposed an end-to-end framework combining a fully connected convolution network and a temporal transformer network for extracting the temporal features and long-term temporal coherence. Moreover, some multi-modal methodologies [98, 144, 249] jointly used visual and audio information to achieve a variety of deepfake tasks.

3.5 Counter-Forensics Issues

Although existing detectors have shown effectiveness and robustness on various face forgery datasets, they also stimulate the births of more powerful attacks. Attacks and defenses are in an arms race of such typical two-player games. Thus, it is unsurprising that adversarial attacks have recently fueled the face forensics community. Most face forgery detectors are vulnerable to both black-box and white-box adversarial attacks. Neekhara *et al.* [147] launched adversarial attacks on deepfake detectors in a black-box setting. They demonstrated that the designed universal adversarial perturbations could be flexibly deployed on face images and bypass forgery detectors. Hussain *et al.* [77] proposed that adversarial perturbations could fool DNN-based detectors and the produced adversarial videos were robust to video and image compression. Jia *et al.* [81] proposed a meta-learning framework to generate more imperceptible adversarial samples by injecting adversarial perturbations into the frequency domain. Adversarial attacks have posed pressing new challenges for both industry and academia. They demand more powerful and robust face forgery detectors to properly counteract the potential risks caused by counter-forensics issues.

3.6 Future Research Directions

Extraordinary success in deepfake attacks and digital face forensics has been achieved in the last few years. Nonetheless, there are still lots of issues that

need to be addressed. Although accurate and secure, most of the deepfake detectors lack explainability and interpretability, thus limiting their reliability when deployed in practical scenarios. More explainability-related works are expected in the future to better interpret why the decision is made by the defense system, and then the decision can be adjusted accordingly. Besides accurately detecting forgery faces, localizing forgery regions is another vital task in this community. Forgery localization is able to provide evidence for detecting deepfakes and unveil attackers' intents. However, this task has been largely understudied so far. On the other hand, due to the two-player nature between face forgery and forgery detection, attack algorithms will be more and more powerful, and the generated fake faces will get increasingly realistic. This research field calls for more robust detection methods to counteract the menace of unforeseen advanced attack methods and address the generalization issues. Moreover, deepfake videos in the wild always involve both visual and audio manipulation to make the fake videos look more realistic. As shown in Table 3, only FakeAVCeleb [89] considers deepfake audio (*a.k.a.* audio manipulation). To facilitate accurate deepfake detection in the wild, more visual-audio joint deepfake datasets and multi-modal detectors are expected in future research works.

3.7 Discussion

In this section, we comprehensively reviewed the existing digital face forgery literature over several important tasks, including face forgery generation, deepfake datasets, and face forgery detection methodologies. We thoroughly analyzed the potential risks and dangerous consequences of digital face attacks and adversarial attacks. Besides, we outlined the existing and upcoming challenges in the face forensics community and suggested possible future research directions for industry and academia.

4 Unifying Security Efforts Against Physical and Digital Face Attacks

4.1 Importance of the Problem

Automated face recognition (AFR) systems have been pervasively deployed to billions of human beings all around the world for various applications. It is reported that the market of AFR will reach USD 3.35B by 2024 [8]. However, as shown in Figure 15, AFRs are vulnerable to both physical and digital face attacks. Malicious attackers can readily launch various physical attacks on the image/video capture stage or hack the device with digital attacks. Most current defense methods are only capable of detecting either physical or digital

attacks, thereby requiring the attack type be known as a prior. Generalizing to unknown attack types has been remained as an open issue in this research community. Therefore, it is of utmost importance to propose generalized and unified detectors for safeguarding AFR systems from various malicious attacks. On the other hand, presentation attack detection and digital face forgery detection are two highly related tasks. Training a unified detector can be cast as a multi-task learning problem. Yu *et al.* [225] demonstrated that the generalization capacities of models could be obviously improved via the joint training scheme compared with single-task learning. Therefore, it is of much necessary to devote more efforts to unifying security for AFR against physical and digital face attacks, which have been barely studied in the existing literature.

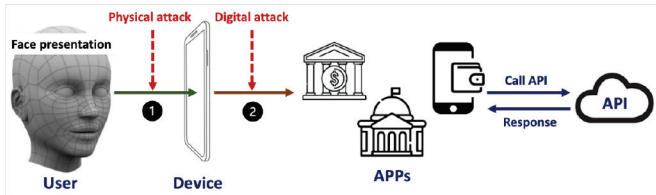

Figure 15: Overview of AFR process. Step 182 and step 183 are vulnerable to physical and digital face attacks, respectively.

4.2 Overview of Joint Face Spoofing and Forgery Detection Methodologies

Joint detection is a brand new task that requires more attention in this community. DFFD [43] is the first attempt at unifying the detection over four digital face attacks, including identity swap, face reenactment, attribute manipulation, and entire face synthesis. On the other hand, Li *et al.* [106] demonstrated that face liveness verification systems are vulnerable to not only presentation attacks but also digital face attacks (Deepfake). Inspired by these two arts, follow-up works [45, 141, 225] attempt to propose unified detection to counteract physical face spoofing and digital face forgery. Mehta *et al.* [141] proposed to use the cross asymmetric loss function to supervise the training process and achieved promising attack detection performance in three scenarios: ubiquitous environment, individual databases, and cross-attack/cross-database. Deb *et al.* [45] cast the task of unified detection of digital and physical face attacks as a multi-task problem and achieved a more generalized defense for automatic face recognition. Yu *et al.* [225] firstly built a benchmark for joint face spoofing and forgery detection. Then they proposed a novel multi-modal framework that combined rPPG facial signals and RGB face images and achieved the best detection performance. Yu *et al.*

[225] demonstrated that joint training can greatly boost the generalization capability as spoofing detection and forgery detection are two highly related tasks.

4.3 Future Research Directions

By far, only few efforts have been dedicated to this unified detection task. Although the benchmark on joint physical and digital face attack detection has been built in [225], it only considered video-level detection. It is necessary to benchmark the joint detection at the image-level because image-level attacks are prominent in many real-world scenarios. Besides, standard protocols for this task should be properly built in future works to facilitate the development of new models. Apart from benchmarks and protocols, more generalized features and intrinsic clues between these two highly-related tasks are expected to be extracted to further improve the generalization capability. Last but not least, the interpretability for why the generalization capability of joint detectors boosts compared with the single-task learning scheme is still vague. More explainability and interpretability works are expected in the future.

4.4 Discussion

We innovatively analyzed and discussed the pivotal joint detection problem in this section, which, to the best of our knowledge, was never mentioned in the existing surveys. The importance of this problem had been firstly clarified to attract more attention to this research field. Then, we reviewed early attempts on this joint face anti-spoofing and forgery detection task. We analyzed the main drawbacks of these works and suggested promising areas for future research. As clearly stated before, the joint detection task is largely understudied so far. To fill this gap in the face forensics community, more efforts regarding generalized unifying fake face detection are expected in future research works.

5 Conclusion

Securing face data circulating on the internet and face recognition systems deployed in real-world applications is becoming a significant necessity to the public at large. Over the past decades, we have witnessed tremendous progress in both face attacks and face forensics. For sure, attack and safeguard are two players in a competitive arms race, and both of them are becoming more and more mature. Generally speaking, attack samples tend to be increasingly sophisticated and realistic, which demands powerful detection tools to counteract the pressing menace. It also requires industry and academia to design

robust models to defend against various unforeseen attacks. In this survey, we have provided a comprehensive overview and concrete discussions on the literature on both physical and digital face attacks. For each respective topic, we have provided a clear problem definition and analyzed the importance of the problem. On the other hand, the taxonomy of various attack methodologies and associated databases have been listed. We presented numerous attack detectors and analyzed their technique soundness, and also pointed out the main drawbacks of existing works. More importantly, future research directions have been highlighted in this survey for addressing unsolved problems that remained in the face forensics community. One step further, at the end of the survey, we extensively surveyed and analyzed the research works on joint face spoofing and forgery detection and concluded with suggestions for future research directions. We hope this survey can help facilitate the development of the face forensics community and attract more attention to contribute to face security.

Acknowledgements

This work is supported by the Research Grant Council (RGC) of Hong Kong through Early Career Scheme (ECS) under the Grant 21200522 and Sichuan Science and Technology Program 2022NSFSC0551.

References

[1] [EB/OL], "Deepfakes faceswap," 2019, https://github.com/deepfakes/faceswap.
[2] [EB/OL], "Facebrity, Apple App Store.," 2022, https://apps.apple.com/us/app/facebrity-face-swap-morph-app/id1449734851.
[3] [EB/OL], "Reface, Apple App Store.," 2022, https://apps.apple.com/us/app/reface-face-swap-videos/id1488782587.
[4] [EB/OL], "Zao, Apple App Store.," 2022, https://apps.apple.com/cn/app/zao/id1465199127.
[5] "3D mask attack for Apple FaceID," 2017, https://www.theverge.com/2017/11/13/16642690/bkav-iphone-x-faceid-mask.
[6] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 2010, 433–59.
[7] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: A compact facial video forgery detection network," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, IEEE, 2018, 1–7.
[8] "AFR market," 2021, https://bwnews.pr/2OqY0nD.

[9] A. Agarwal, A. Sehwag, R. Singh, and M. Vatsa, "Deceiving face presentation attack detection via image transforms," in *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, IEEE, 2019, 373–82.

[10] A. Agarwal, A. Sehwag, M. Vatsa, and R. Singh, "Deceiving the protector: Fooling face presentation attack detection algorithms," in *2019 International Conference on Biometrics (ICB)*, IEEE, 2019, 1–6.

[11] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2006, 2037–41.

[12] I. Amerini and R. Caldelli, "Exploiting prediction error inconsistencies through LSTM-based classifiers to detect deepfake videos," in *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security*, 2020, 97–102.

[13] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *2011 International Joint Conference on Biometrics (IJCB)*, IEEE, 2011, 1–7.

[14] M. Asim, Z. Ming, and M. Y. Javed, "CNN based spatio-temporal feature extraction for face anti-spoofing," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, IEEE, 2017, 234–8.

[15] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face anti-spoofing using patch and depth-based CNNs," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2017, 319–28.

[16] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, 12(1), 1994, 43–77.

[17] Y. Baweja, P. Oza, P. Perera, and V. M. Patel, "Anomaly detection-based unknown face presentation attack detection," in *2020 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2020, 1–9.

[18] S. Bhattacharjee, A. Mohammadi, and S. Marcel, "Spoofing deep face recognition with custom silicone masks," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, 2018, 1–7.

[19] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: Automatically replacing faces in photographs," in *ACM SIGGRAPH 2008 Papers*, 2008, 1–8.

[20] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2015, 2636–40.

[21] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face antispoofing using speeded-up robust features and fisher vector encoding," *IEEE Signal Processing Letters*, 24(2), 2016, 141–5.

[22] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "Oulu-npu: A mobile face presentation attack database with real-world variations," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, IEEE, 2017, 612–8.

[23] R. Cai, H. Li, S. Wang, C. Chen, and A. C. Kot, "DRL-fas: a novel framework based on deep reinforcement learning for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, 16, 2020, 937–51.

[24] R. Cai, Z. Li, R. Wan, H. Li, Y. Hu, and A. C. Kot, "Learning Meta Pattern for Face Anti-Spoofing," *IEEE Transactions on Information Forensics and Security*, 2022.

[25] J. Cao, C. Ma, T. Yao, S. Chen, S. Ding, and X. Yang, "End-to-End Reconstruction-Classification Learning for Face Forgery Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 4113–22.

[26] P. P. Chan, W. Liu, D. Chen, D. S. Yeung, F. Zhang, X. Wang, and C.-C. Hsu, "Face liveness detection using a flash against 2D spoofing attack," *IEEE Transactions on Information Forensics and Security*, 13(2), 2017, 521–34.

[27] B. Chen, W. Yang, H. Li, S. Wang, and S. Kwong, "Camera invariant feature learning for generalized face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, 16, 2021, 2477–92.

[28] D. Chen, Q. Chen, J. Wu, X. Yu, and J. Tong, "Face swapping: realistic image synthesis based on facial landmarks alignment," *Mathematical Problems in Engineering*, 2019, 2019.

[29] H. Chen, G. Hu, Z. Lei, Y. Chen, N. M. Robertson, and S. Z. Li, "Attention-based two-stream convolutional networks for face spoofing detection," *IEEE Transactions on Information Forensics and Security*, 15, 2019, 578–93.

[30] H. Chen, W. Wang, J. Zhang, and Q. Zhang, "Echoface: Acoustic sensor-based media attack detection for face authentication," *IEEE Internet of Things Journal*, 7(3), 2019, 2152–9.

[31] R. Chen, X. Chen, B. Ni, and Y. Ge, "Simswap: An efficient framework for high fidelity face swapping," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, 2003–11.

[32] S. Chen, T. Yao, Y. Chen, S. Ding, J. Li, and R. Ji, "Local relation learning for face forgery detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 2, 2021, 1081–8.

[33] Y. Chen, J. Sun, X. Jin, T. Li, R. Zhang, and Y. Zhang, "Your face your heart: Secure mobile face authentication with photoplethysmograms," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, IEEE, 2017, 1–9.

[34] Z. Chen, T. Yao, K. Sheng, S. Ding, Y. Tai, J. Li, F. Huang, and X. Jin, "Generalizable representation learning for mixture domain face anti-spoofing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 2, 2021, 1132–9.

[35] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)*, IEEE, 2012, 1–7.

[36] I. Chingovska, N. Erdogmus, A. Anjos, and S. Marcel, "Face recognition systems under spoofing attacks," in *Face Recognition Across the Imaging Spectrum*, Springer, 2016, 165–94.

[37] A. Chintha, B. Thai, S. J. Sohrawardi, K. Bhatt, A. Hickerson, M. Wright, and R. Ptucha, "Recurrent convolutional structures for audio spoof and video deepfake detection," *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 2020, 1024–37.

[38] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 8789–97.

[39] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 1251–8.

[40] U. A. Ciftci, I. Demir, and L. Yin, "Fakecatcher: Detection of synthetic portrait videos using biological signals," *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[41] Daily Mail, "[EB/OL]," 2022, https://www.dailymail.co.uk/news/article-2107109/Iconic-Abraham-Lincoln-portrait-revealed-TWO-pictures-stitched-together.html.

[42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1, Ieee, 2005, 886–93.

[43] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, "On the detection of digital face manipulation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern recognition*, 2020, 5781–90.

[44] D. Deb and A. K. Jain, "Look locally infer globally: A generalizable face anti-spoofing approach," *IEEE Transactions on Information Forensics and Security*, 16, 2020, 1143–57.

[45] D. Deb, X. Liu, and A. K. Jain, "Unified detection of digital and physical face attacks," *arXiv preprint arXiv:2104.02156*, 2021.

[46] J. M. Di Martino, Q. Qiu, and G. Sapiro, "Rethinking Shape From Shading for Spoofing Detection," *IEEE Transactions on Image Processing*, 30, 2020, 1086–99.

[47] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (dfdc) preview dataset," *arXiv preprint arXiv:1910.08854*, 2019.

[48] N. Dufour and A. Gully, "Contributing data to deepfake detection research," *Google AI Blog*, 1(3), 2019.

[49] A. F. Ebihara, K. Sakurai, and H. Imaoka, "Efficient Face Spoofing Detection with Flash," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4), 2021, 535–49.

[50] H. Farrukh, R. M. Aburas, S. Cao, and H. Wang, "FaceRevelio: a face liveness detection system for smartphones with a single front camera," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, 1–13.

[51] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung, "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *Journal of Visual Communication and Image Representation*, 38, 2016, 451–60.

[52] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "LBP-TOP based countermeasure against face spoofing attacks," in *Asian Conference on Computer Vision*, Springer, 2012, 121–32.

[53] C. Frith, "Role of facial expressions in social interactions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 2009, 3453–8.

[54] G. Gao, H. Huang, C. Fu, Z. Li, and R. He, "Information bottleneck disentanglement for identity swapping," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 3404–13.

[55] H. Ge, X. Tu, W. Ai, Y. Luo, Z. Ma, and M. Xie, "Face Anti-Spoofing by the Enhancement of Temporal Motion," in *2020 2nd International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)*, IEEE, 2020, 106–11.

[56] A. George and S. Marcel, "Cross modal focal loss for rgbd face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 7882–91.

[57] A. George and S. Marcel, "Deep pixel-wise binary supervision for face presentation attack detection," in *2019 International Conference on Biometrics (ICB)*, IEEE, 2019, 1–8.

[58] A. George and S. Marcel, "Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks," *IEEE Transactions on Information Forensics and Security*, 16, 2020, 361–75.

[59] A. George and S. Marcel, "On the effectiveness of vision transformers for zero-shot face anti-spoofing," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2021, 1–8.

[60] "Github deepfake faceswap," 2018 https://github.com/deepfakes/faceswap.

[61] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative Adversarial Nets," in *NIPS*, 2014.

[62] Z. Gu, Y. Chen, T. Yao, S. Ding, J. Li, F. Huang, and L. Ma, "Spatiotemporal inconsistency learning for deepfake video detection," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, 3473–81.

[63] Y. Guo, X. Wei, G. Wang, and B. Zhang, "Meaningful adversarial stickers for face recognition in physical world," *arXiv preprint arXiv:2104.06728*, 2021.

[64] S. Ha, M. Kersner, B. Kim, S. Seo, and D. Kim, "Marionette: Few-shot face reenactment preserving identity of unseen targets," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 07, 2020, 10893–900.

[65] A. Haliassos, R. Mira, S. Petridis, and M. Pantic, "Leveraging Real Talking Faces via Self-Supervision for Robust Forgery Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 14950–62.

[66] A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic, "Lips don't lie: A generalisable and robust approach to face forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 5039–49.

[67] H. Hao, M. Pei, and M. Zhao, "Face liveness detection based on client identity using siamese network," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Springer, 2019, 172–80.

[68] Y. He, B. Gan, S. Chen, Y. Zhou, G. Yin, L. Song, L. Sheng, J. Shao, and Z. Liu, "Forgerynet: A versatile benchmark for comprehensive forgery analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 4360–9.

[69] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "Attgan: Facial attribute editing by only changing what you want," *IEEE Transactions on Image Processing*, 28(11), 2019, 5464–78.

[70] Y.-J. Heo, Y.-J. Choi, Y.-W. Lee, and B.-G. Kim, "Deepfake detection scheme based on vision transformer and distillation," *arXiv preprint arXiv:2104.01353*, 2021.

[71] G. Heusch, A. George, D. Geissbühler, Z. Mostaani, and S. Marcel, "Deep models and shortwave infrared information to detect face presentation attacks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4), 2020, 399–409.

[72] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, 9(8), 1997, 1735–80.

[73] M. S. Hossain, L. Rupty, K. Roy, M. Hasan, S. Sengupta, and N. Mohammed, "A-DeepPixBis: Attentional angular margin for face anti-spoofing," in *2020 Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2020, 1–8.

[74] J. Hu, X. Liao, J. Liang, W. Zhou, and Z. Qin, "FInfer: Frame Inference-based Deepfake Detection for High-Visual-Quality Videos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 1–9.

[75] J. Hu, X. Liao, W. Wang, and Z. Qin, "Detecting compressed deepfake videos in social networks using frame-temporality two-stream convolutional network," *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3), 2021, 1089–102.

[76] Y. Huang, F. Juefei-Xu, Q. Guo, Y. Liu, and G. Pu, "FakeLocator: Robust localization of GAN-based face manipulations," *IEEE Transactions on Information Forensics and Security*, 2022.

[77] S. Hussain, P. Neekhara, M. Jere, F. Koushanfar, and J. McAuley, "Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, 3348–57.

[78] S. Hyun, J. Kim, and J.-P. Heo, "Self-supervised video gans: Learning for appearance consistency and motion coherency," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 10826–35.

[79] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 1125–34.

[80] S. Jia, G. Guo, and Z. Xu, "A survey on 3D mask presentation attack detection and countermeasures," *Pattern recognition*, 98, 2020, 107032.

[81] S. Jia, C. Ma, T. Yao, B. Yin, S. Ding, and X. Yang, "Exploring Frequency Adversarial Attacks for Face Forgery Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 4103–12.

[82] Y. Jia, J. Zhang, S. Shan, and X. Chen, "Single-side domain generalization for face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 8484–93.

[83] Y. Jia, J. Zhang, S. Shan, and X. Chen, "Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing," *Pattern Recognition*, 115, 2021, 107888.

[84] F. Jiang, P. Liu, X. Shao, and X. Zhou, "Face anti-spoofing with generated near-infrared images," *Multimedia Tools and Applications*, 79(29), 2020, 21299–323.

[85] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 2889–98.

[86] F. Juefei-Xu, R. Wang, Y. Huang, Q. Guo, L. Ma, and Y. Liu, "Countering malicious deepfakes: Survey, battleground, and horizon," *International Journal of Computer Vision*, 2022, 1–57.

[87] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," in *International Conference on Learning Representations*, 2018.

[88] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 4401–10.

[89] H. Khalid, S. Tariq, M. Kim, and S. S. Woo, "FakeAVCeleb: a novel audio-video multimodal deepfake dataset," *arXiv preprint arXiv:2108.05080*, 2021.

[90] M. Khammari, "Robust face anti-spoofing using CNN with LBP and WLD," *IET Image Processing*, 13(11), 2019, 1880–4.

[91] S. A. Khan and H. Dai, "Video transformer for deepfake detection with incremental learning," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, 1821–8.

[92] T. Kim, Y. Kim, I. Kim, and D. Kim, "Basn: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, 0–0.

[93] D. P. Kingma, M. Welling, *et al.*, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning*, 12(4), 2019, 307–92.

[94] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[95] S. Komkov and A. Petiushko, "Advhat: Real-world adversarial attack on arcface face id system," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021, 819–26.

[96] J. Komulainen, A. Hadid, and M. Pietikäinen, "Context based face anti-spoofing," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, IEEE, 2013, 1–8.

[97] C. Kong, B. Chen, H. Li, S. Wang, A. Rocha, and S. Kwong, "Detect and Locate: Exposing Face Manipulation by Semantic-and Noise-Level Telltales," *IEEE Transactions on Information Forensics and Security*, 17, 2022, 1741–56.

[98] C. Kong, B. Chen, W. Yang, H. Li, P. Chen, and S. Wang, "Appearance matters, so does audio: Revealing the hidden face via cross-modality transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, 32(1), 2021, 423–36.

[99] C. Kong, K. Zheng, S. Wang, A. Rocha, and H. Li, "Beyond the Pixel World: A Novel Acoustic-based Face Anti-Spoofing System for Smartphones," *IEEE Transactions on Information Forensics and Security*, 2022.

[100] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," *arXiv preprint arXiv:1812.08685*, 2018.

[101] P. Korshunov and S. Marcel, "Improving generalization of deepfake detection with data farming and few-shot learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2022.

[102] K. Kotwal, S. Bhattacharjee, and S. Marcel, "Multispectral deep embeddings as a countermeasure to custom silicone mask presentation attacks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(4), 2019, 238–51.

[103] H. Kuang, R. Ji, H. Liu, S. Zhang, X. Sun, F. Huang, and B. Zhang, "Multi-modal multi-layer fusion network with average binary center loss for face anti-spoofing," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, 48–56.

[104] A. Kumar, A. Bhavsar, and R. Verma, "Detecting deepfakes with metric learning," in *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, IEEE, 2020, 1–6.

[105] P. Kwon, J. You, G. Nam, S. Park, and G. Chae, "Kodf: A large-scale korean deepfake detection dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 10744–53.

[106] C. Li, L. Wang, S. Ji, X. Zhang, Z. Xi, S. Guo, and T. Wang, "Seeing is living? rethinking the security of facial liveness verification in the deepfake era," *CoRR abs/2202.10673*, 2022.

[107] H. Li, B. Li, S. Tan, and J. Huang, "Identification of deep network generated images using disparities in color components," *Signal Processing*, 174, 2020, 107616.

[108] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, 13(7), 2018, 1794–809.

[109] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 5400–9.

[110] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An original face anti-spoofing approach using partial convolutional neural network," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, 2016, 1–6.

[111] L. Li, Z. Xia, A. Hadid, X. Jiang, H. Zhang, and X. Feng, "Replayed video attack detection based on motion blur analysis," *IEEE Transactions on Information Forensics and Security*, 14(9), 2019, 2246–61.

[112] L. Li, Z. Xia, X. Jiang, Y. Ma, F. Roli, and X. Feng, "3D face mask presentation attack detection based on intrinsic image analysis," *Iet Biometrics*, 9(3), 2020, 100–8.

[113] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "Advancing high fidelity identity swapping for forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 5074–83.

[114] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, "Face x-ray for more general face forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 5001–10.

[115] X. Li, J. Komulainen, G. Zhao, P.-C. Yuen, and M. Pietikäinen, "Generalized face anti-spoofing by detecting pulse from face videos," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, IEEE, 2016, 4244–9.

[116] X. Li, Y. Lang, Y. Chen, X. Mao, Y. He, S. Wang, H. Xue, and Q. Lu, "Sharp multiple instance learning for deepfake video detection," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, 1864–72.

[117] Y. Li, Y. Li, Q. Yan, H. Kong, and R. H. Deng, "Seeing your face is not enough: An inertial sensor-based liveness detection for face authentication," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, 1558–69.

[118] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," in *2018 IEEE International workshop on information forensics and security (WIFS)*, IEEE, 2018, 1–7.

[119] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," *arXiv preprint arXiv:1811.00656*, 2018.

[120] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 3207–16.

[121] Z. Li, R. Cai, H. Li, K.-Y. Lam, Y. Hu, and A. C. Kot, "One-Class Knowledge Distillation for Face Presentation Attack Detection," *IEEE Transactions on Information Forensics and Security*, 2022.

[122] Z. Li, H. Li, K.-Y. Lam, and A. C. Kot, "Unseen face presentation attack detection with hypersphere loss," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, 2852–6.

[123] B. Lin, X. Li, Z. Yu, and G. Zhao, "Face liveness detection by rppg features and contextual patch-based cnn," in *Proceedings of the 2019 3rd international Conference on Biometric Engineering and Applications*, 2019, 61–8.

[124] A. Liu, Z. Tan, J. Wan, Y. Liang, Z. Lei, G. Guo, and S. Z. Li, "Face Anti-Spoofing via Adversarial Cross-Modality Translation," *IEEE Transactions on Information Forensics and Security*, 16, 2021, 2759–72.

[125] A. Liu, C. Zhao, Z. Yu, J. Wan, A. Su, X. Liu, Z. Tan, S. Escalera, J. Xing, Y. Liang, *et al.*, "Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection," *IEEE Transactions on Information Forensics and Security*, 17, 2022, 2497–507.

[126] H. Liu, X. Li, W. Zhou, Y. Chen, Y. He, H. Xue, W. Zhang, and N. Yu, "Spatial-phase shallow learning: Rethinking face forgery detection in frequency domain," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 772–81.

[127] S. Liu, K.-Y. Zhang, T. Yao, M. Bi, S. Ding, J. Li, F. Huang, and L. Ma, "Adaptive normalized representation learning for generalizable face anti-spoofing," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, 1469–77.

[128] S. Liu, P. C. Yuen, S. Zhang, and G. Zhao, "3D mask face anti-spoofing with remote photoplethysmography," in *European Conference on Computer Vision*, Springer, 2016, 85–100.

[129] W. Liu, X. Wei, T. Lei, X. Wang, H. Meng, and A. K. Nandi, "Data Fusion based Two-stage Cascade Framework for Multi-Modality Face Anti-Spoofing," *IEEE Transactions on Cognitive and Developmental Systems*, 2021.

[130] Y. Liu, A. Jourabloo, and X. 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, 389–98.

[131] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 4680–9.

[132] Y. Liu, J. Stehouwer, and X. Liu, "On disentangling spoof trace for generic face anti-spoofing," in *European Conference on Computer Vision*, Springer, 2020, 406–22.

[133] Y. Lu, J. Chai, and X. Cao, "Live speech portraits: real-time photorealistic talking-head animation," *ACM Transactions on Graphics (TOG)*, 40(6), 2021, 1–17.

[134] O. Lucena, A. Junior, V. Moia, R. Souza, E. Valle, and R. Lotufo, "Transfer learning using convolutional neural networks for face anti-spoofing," in *International conference Image Analysis and Recognition*, Springer, 2017, 27–34.

[135] Y. Luo, Y. Zhang, J. Yan, and W. Liu, "Generalizing face forgery detection with high-frequency features," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 16317–26.

[136] S. Lyu, "Deepfake detection: Current challenges and next steps," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, IEEE, 2020, 1–6.

[137] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *2011 international Joint Conference on Biometrics (IJCB)*, IEEE, 2011, 1–7.

[138] K. Mallat and J.-L. Dugelay, "Indirect synthetic attack on thermal face biometric systems via visible-to-thermal spectrum conversion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 1435–43.

[139] I. Masi, A. Killekar, R. M. Mascarenhas, S. P. Gurudatt, and W. AbdAlmageed, "Two-branch recurrent network for isolating deepfakes in videos," in *European Conference on Computer Vision*, Springer, 2020, 667–84.

[140] G. Mazaheri and A. K. Roy-Chowdhury, "Detection and localization of facial expression manipulations," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, 1035–45.

[141] S. Mehta, A. Uberoi, A. Agarwal, M. Vatsa, and R. Singh, "Crafting a panoptic face presentation attack detector," in *2019 International Conference on Biometrics (ICB)*, IEEE, 2019, 1–6.

[142] C. Miao, Z. Tan, Q. Chu, N. Yu, and G. Guo, "Hierarchical Frequency-Assisted Interactive Networks for Face Manipulation Detection," *IEEE Transactions on Information Forensics and Security*, 2022.

[143] Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," *ACM Computing Surveys (CSUR)*, 54(1), 2021, 1–41.

[144] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "Emotions don't lie: An audio-visual deepfake detection method using affective cues," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, 2823–32.

[145] A. Mohammadi, S. Bhattacharjee, and S. Marcel, "Domain adaptation for generalization of face presentation attack detection in mobile settengs with minimal information," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, 1001–5.

[146] U. Muhammad, T. Holmberg, W. C. de Melo, and A. Hadid, "Face Anti-Spoofing via Sample Learning Based Recurrent Neural Network (RNN).," in *BMVC*, 2019, 113.

[147] P. Neekhara, B. Dolhansky, J. Bitton, and C. C. Ferrer, "Adversarial threats to deepfake detection: A practical perspective," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 923–32.

[148] E. Nesli and S. Marcel, "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect," in *IEEE 6th International Conference on Biometrics: Theory, Applications and Systems (BTAS'13)*, 2013, 1–8.

[149] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," in *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, 2019, 1–8.

[150] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using capsule networks to detect forged images and videos," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, 2307–11.

[151] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep learning for deepfakes creation and detection," *arXiv preprint arXiv:1909.11573*, 1, 2019, 2.

[152] O. Nikisins, A. George, and S. Marcel, "Domain adaptation in multi-channel autoencoder based features for robust face anti-spoofing," in *2019 International Conference on Biometrics (ICB)*, IEEE, 2019, 1–8.

[153] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel, "On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing," in *2018 International Conference on Biometrics (ICB)*, IEEE, 2018, 75–81.

[154] Y. Nirkin, Y. Keller, and T. Hassner, "Fsgan: Subject agnostic face swapping and reenactment," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, 7184–93.

[155] Y. Nirkin, L. Wolf, Y. Keller, and T. Hassner, "DeepFake detection based on discrepancies between faces and their context," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[156] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *2007 IEEE 11th International Conference on Computer Vision*, IEEE, 2007, 1–8.

[157] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Transactions on Information Forensics and Security*, 11(10), 2016, 2268–83.

[158] K. Patel, H. Han, A. K. Jain, and G. Ott, "Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks," in *2015 International Conference on Biometrics (ICB)*, IEEE, 2015, 98–105.

[159] B. Peixoto, C. Michelassi, and A. Rocha, "Face liveness detection under bad illumination conditions," in *2011 18th IEEE International Conference on Image Processing*, IEEE, 2011, 3557–60.

[160] D. Peng, J. Xiao, R. Zhu, and G. Gao, "Ts-Fen: Probing feature selection strategy for face anti-spoofing," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, 2942–6.

[161] L. A. Pereira, A. Pinto, F. A. Andaló, A. M. Ferreira, B. Lavi, A. Soriano-Vargas, M. V. Cirne, and A. Rocha, "The Rise of Data-Driven Models in Presentation Attack Detection," in *Deep Biometrics*, Springer, 2020, 289–311.

[162] D. Pérez-Cabo, D. Jiménez-Cabello, A. Costa-Pazo, and R. J. López-Sastre, "Deep anomaly detection for generalized face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, 0–0.

[163] D. Pérez-Cabo, D. Jiménez-Cabello, A. Costa-Pazo, and R. J. López-Sastre, "Learning to Learn Face-PAD: A lifelong learning approach," in *2020 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2020, 1–9.

[164] H. Qi, Q. Guo, F. Juefei-Xu, X. Xie, L. Ma, W. Feng, Y. Liu, and J. Zhao, "Deeprhythm: Exposing deepfakes with attentional visual heartbeat rhythms," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, 4318–27.

[165] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in *European conference on computer vision*, Springer, 2020, 86–103.

[166] Y. Qin, C. Zhao, X. Zhu, Z. Wang, Z. Yu, T. Fu, F. Zhou, J. Shi, and Z. Lei, "Learning meta model for zero-and few-shot face anti-spoofing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 07, 2020, 11916–23.

[167] R. Quan, Y. Wu, X. Yu, and Y. Yang, "Progressive transfer learning for face anti-spoofing," *IEEE Transactions on Image Processing*, 30, 2021, 3946–55.

[168] E. A. Raheem, S. M. S. Ahmad, and W. A. W. Adnan, "Insight on face liveness detection: A systematic literature review.," *International Journal of Electrical & Computer Engineering (2088-8708)*, 9(6), 2019.

[169] Y. A. U. Rehman, L.-M. Po, and J. Komulainen, "Enhancing deep discriminative feature maps via perturbation for face presentation attack detection," *Image and Vision Computing*, 94, 2020, 103858.

[170] Y. A. U. Rehman, L.-M. Po, M. Liu, Z. Zou, and W. Ou, "Perturbing convolutional feature maps with histogram of oriented gradients for face liveness detection," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019)*, Springer, 2019, 3–13.

[171] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, 1–11.

[172] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, 115(3), 2015, 211–52.

[173] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," *Interfaces (GUI)*, 3(1), 2019, 80–7.

[174] Y. Safaa El-Din, M. N. Moustafa, and H. Mahdi, "Deep convolutional neural networks for face and iris presentation attack detection: Survey and case study," *IET Biometrics*, 9(5), 2020, 179–93.

[175] J. Seo and I.-J. Chung, "Face liveness detection using thermal face-CNN with external knowledge," *Symmetry*, 11(3), 2019, 360.

[176] Z. Shang, H. Xie, Z. Zha, L. Yu, Y. Li, and Y. Zhang, "PRRNet: Pixel-Region relation network for face forgery detection," *Pattern Recognition*, 116, 2021, 107950.

[177] R. Shao, X. Lan, J. Li, and P. C. Yuen, "Multi-adversarial discriminative deep domain generalization for face presentation attack detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 10023–31.

[178] R. Shao, X. Lan, and P. C. Yuen, "Joint discriminative learning of deep dynamic textures for 3D mask face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, 14(4), 2018, 923–38.

[179] R. Shao, X. Lan, and P. C. Yuen, "Regularized fine-grained meta face anti-spoofing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 07, 2020, 11974–81.

[180] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, "A general framework for adversarial examples with objectives," *ACM Transactions on Privacy and Security (TOPS)*, 22(3), 2019, 1–30.

[181] O. Sharifi, "Score-Level-based Face Anti-Spoofing System Using Handcrafted and Deep Learned Characteristics," *International Journal of Image, Graphics and Signal Processing*, 10(2), 2019, 15.

[182] T. Shen, Y. Huang, and Z. Tong, "Facebagnet: Bag-of-local-features model for multi-modal face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, 0–0.

[183] K. Shiohara and T. Yamasaki, "Detecting Deepfakes with Self-Blended Images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 18720–9.

[184] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[185] L. Souza, L. Oliveira, M. Pamplona, and J. Papa, "How far did we get in face spoofing detection?" *Engineering Applications of Artificial Intelligence*, 72, 2018, 368–81.

[186] K. Sun, H. Liu, Q. Ye, Y. Gao, J. Liu, L. Shao, and R. Ji, "Domain general face forgery detection by learning to weight," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 3, 2021, 2638–46.

[187] K. Sun, T. Yao, S. Chen, S. Ding, J. Li, and R. Ji, "Dual contrastive learning for general face forgery detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, No. 2, 2022, 2316–24.

[188] W. Sun, Y. Song, C. Chen, J. Huang, and A. C. Kot, "Face spoofing detection based on local ternary label supervision in fully convolutional networks," *IEEE Transactions on Information Forensics and Security*, 15, 2020, 3181–96.

[189] X. Sun, L. Huang, and C. Liu, "Context based face spoofing detection using active near-infrared images," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, IEEE, 2016, 4262–7.

[190] Z. Sun, Y. Han, Z. Hua, N. Ruan, and W. Jia, "Improving the efficiency and robustness of deepfakes detection through precise geometric features," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 3609–18.

[191] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing obama: learning lip sync from audio," *ACM Transactions on Graphics (ToG)*, 36(4), 2017, 1–13.

[192] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, PMLR, 2019, 6105–14.

[193] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *European Conference on Computer Vision*, Springer, 2010, 504–17.

[194] D. Tang, Z. Zhou, Y. Zhang, and K. Zhang, "Face flashing: a secure liveness detection protocol based on light reflections," *arXiv preprint arXiv:1801.01949*, 2018.

[195] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 2387–95.

[196] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering: Image synthesis using neural textures," *ACM Transactions on Graphics (TOG)*, 38(4), 2019, 1–12.

[197] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Information Fusion*, 64, 2020, 131–48.

[198] S. Tripathy, J. Kannala, and E. Rahtu, "Icface: Interpretable and controllable face reenactment using gans," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, 3385–94.

[199] A. Van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, *et al.*, "Conditional image generation with pixelcnn decoders," *Advances in Neural Information Processing Systems*, 29, 2016.

[200] A. Van Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *International Conference on Machine Learning*, PMLR, 2016, 1747–56.

[201] L. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 2020, 910–32.

[202] C.-Y. Wang, Y.-D. Lu, S.-T. Yang, and S.-H. Lai, "PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 20281–90.

[203] G. Wang, H. Han, S. Shan, and X. Chen, "Cross-domain face presentation attack detection via multi-domain disentangled representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 6678–87.

[204] G. Wang, H. Han, S. Shan, and X. Chen, "Improving cross-database face presentation attack detection via adversarial domain adaptation," in *2019 International Conference on Biometrics (ICB)*, IEEE, 2019, 1–8.

[205] G. Wang, H. Han, S. Shan, and X. Chen, "Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection," *IEEE Transactions on Information Forensics and Security*, 16, 2020, 56–69.

[206] J. Wang, Y. Sun, and J. Tang, "LiSiam: Localization Invariance Siamese Network for Deepfake Detection," *IEEE Transactions on Information Forensics and Security*, 17, 2022, 2425–36.

[207] J. Wang, J. Zhang, Y. Bian, Y. Cai, C. Wang, and S. Pu, "Self-domain adaptation for face anti-spoofing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 4, 2021, 2746–54.

[208] R. Wang, F. Juefei-Xu, L. Ma, X. Xie, Y. Huang, J. Wang, and Y. Liu, "Fakespotter: A simple yet robust baseline for spotting ai-synthesized fake faces," *arXiv preprint arXiv:1909.06122*, 2019.

[209] T.-C. Wang, A. Mallya, and M.-Y. Liu, "One-shot free-view neural talking-head synthesis for video conferencing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 10039–49.

[210] X. Wang, T. Yao, S. Ding, and L. Ma, "Face manipulation detection via auxiliary supervision," in *International Conference on Neural Information Processing*, Springer, 2020, 313–24.

[211] Z. Wang, Z. Yu, C. Zhao, X. Zhu, Y. Qin, Q. Zhou, F. Zhou, and Z. Lei, "Deep spatial gradient and temporal depth learning for face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 5042–51.

[212] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, 10(4), 2015, 746–61.

[213] X. Wu, J. Zhou, J. Liu, F. Ni, and H. Fan, "Single-shot face anti-spoofing for dual pixel camera," *IEEE Transactions on Information Forensics and Security*, 16, 2020, 1440–51.

[214] J. Xiao, Y. Tang, J. Guo, Y. Yang, X. Zhu, Z. Lei, and S. Z. Li, "3DMA: A multi-modality 3D mask face anti-spoofing database," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2019, 1–8.

[215] F. Xiong and W. AbdAlmageed, "Unknown presentation attack detection with face rgb images," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, 2018, 1–9.

[216] C. Xu, J. Zhang, M. Hua, Q. He, Z. Yi, and Y. Liu, "Region-Aware Face Swapping," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 7632–41.

[217] W. Xu, J. Liu, S. Zhang, Y. Zheng, F. Lin, J. Han, F. Xiao, and K. Ren, "RFace: Anti-spoofing facial authentication using cots," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, IEEE, 2021, 1–10.

[218] Z. Xu, S. Li, and W. Deng, "Learning temporal features using LSTM-CNN architecture for face anti-spoofing," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, IEEE, 2015, 141–5.

[219] C.-Z. Yang, J. Ma, S. Wang, and A. W.-C. Liew, "Preventing deepfake attacks on speaker authentication by dynamic lip movement analysis," *IEEE Transactions on Information Forensics and Security*, 16, 2020, 1841–54.

[220] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," *arXiv preprint arXiv:1408.5601*, 2014.

[221] X. Yang, W. Luo, L. Bao, Y. Gao, D. Gong, S. Zheng, Z. Li, and W. Liu, "Face anti-spoofing: Model matters, so does data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 3507–16.

[222] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, 8261–5.

[223] B. Yin, W. Wang, T. Yao, J. Guo, Z. Kong, S. Ding, J. Li, and C. Liu, "Adv-makeup: A new imperceptible and transferable attack on face recognition," *arXiv preprint arXiv:2105.03162*, 2021.

[224] P. Yu, J. Fei, Z. Xia, Z. Zhou, and J. Weng, "Improving Generalization by Commonality Learning in Face Forgery Detection," *IEEE Transactions on Information Forensics and Security*, 17, 2022, 547–58.

[225] Z. Yu, R. Cai, Z. Li, W. Yang, J. Shi, and A. C. Kot, "Benchmarking Joint Face Spoofing and Forgery Detection with Visual and Physiological Cues," *arXiv preprint arXiv:2208.05401*, 2022.

[226] Z. Yu, X. Li, X. Niu, J. Shi, and G. Zhao, "Face anti-spoofing with human material perception," in *European Conference on Computer Vision*, Springer, 2020, 557–75.

[227] Z. Yu, X. Li, J. Shi, Z. Xia, and G. Zhao, "Revisiting pixel-wise supervision for face anti-spoofing," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

[228] Z. Yu, X. Li, P. Wang, and G. Zhao, "Transrppg: Remote photoplethysmography transformer for 3d mask face presentation attack detection," *IEEE Signal Processing Letters*, 28, 2021, 1290–4.

[229] Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, "Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, 151–60.

[230] Z. Yu, Y. Qin, X. Li, Z. Wang, C. Zhao, Z. Lei, and G. Zhao, "Multi-modal face anti-spoofing based on central difference networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, 650–1.

[231] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep learning for face anti-spoofing: A survey," *arXiv preprint arXiv:2106.14948*, 2021.

[232] Z. Yu, Y. Qin, X. Xu, C. Zhao, Z. Wang, Z. Lei, and G. Zhao, "Auto-fas: Searching lightweight networks for face anti-spoofing," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, 996–1000.

[233] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, and G. Zhao, "NAS-FAS: Static-dynamic central difference network search for face anti-spoofing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(9), 2020, 3005–23.

[234] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, "Searching central difference convolutional networks for face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 5295–305.

[235] B. Zhang, B. Tondi, and M. Barni, "Adversarial examples for replay attacks against CNN-based face recognition with anti-spoofing capability," *Computer Vision and Image Understanding*, 197, 2020, 102988.

[236] C. Zhang, Y. Zhao, Y. Huang, M. Zeng, S. Ni, M. Budagavi, and X. Guo, "Facial: Synthesizing dynamic talking face with implicit attribute learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 3867–76.

[237] D. Zhang, C. Li, F. Lin, D. Zeng, and S. Ge, "Detecting Deepfake Videos with Temporal Dropout 3DCNN.," in *IJCAI*, 2021, 1288–94.

[238] P. Zhang, F. Zou, Z. Wu, N. Dai, S. Mark, M. Fu, J. Zhao, and K. Li, "FeatherNets: Convolutional neural networks as light as feather for face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, 0–0.

[239] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, "Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations," in *European Conference on Computer Vision*, Springer, 2020, 70–85.

[240] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, "CelebASpoof: Large-scale face anti-spoofing dataset with rich annotations," in *European Conference on Computer Vision (ECCV)*, 2020.

[241] K.-Y. Zhang, T. Yao, J. Zhang, Y. Tai, S. Ding, J. Li, F. Huang, H. Song, and L. Ma, "Face anti-spoofing via disentangled representation learning," in *European Conference on Computer Vision*, Springer, 2020, 641–57.

[242] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *2012 5th IAPR International Conference on Biometrics (ICB)*, IEEE, 2012, 26–31.

[243] H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu, "Multi-attentional deepfake detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 2185–94.

[244] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning self-consistency for deepfake detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 15023–33.

[245] Y. Zheng, J. Bao, D. Chen, M. Zeng, and F. Wen, "Exploring temporal coherence for more general video face forgery detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 15044–54.

[246] B. Zhou, Z. Xie, Y. Zhang, J. Lohokare, R. Gao, and F. Ye, "Robust Human Face Authentication Leveraging Acoustic Sensing on Smartphones," *IEEE Transactions on Mobile Computing*, 2021, 1–16.

[247] F. Zhou, C. Gao, F. Chen, C. Li, X. Li, F. Yang, and Y. Zhao, "Face anti-spoofing based on multi-layer domain adaptation," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, IEEE, 2019, 192–7.

[248] T. Zhou, W. Wang, Z. Liang, and J. Shen, "Face forensics in the wild," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 5778–88.

[249] Y. Zhou and S.-N. Lim, "Joint audio-visual deepfake detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 14800–9.

[250] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, 2223–32.

[251] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," *Advances in Neural Information Processing Systems*, 30, 2017.

[252] X. Zhu, H. Wang, H. Fei, Z. Lei, and S. Z. Li, "Face forgery detection by 3d decomposition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 2929–39.

[253] Y. Zhu, Q. Li, J. Wang, C.-Z. Xu, and Z. Sun, "One shot face swapping on megapixels," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 4834–44.