applied sciences

IMPACT FACTOR 2.5

CITESCORE 5.5

Review

Face Anti-Spoofing Based on Deep Learning: A Comprehensive Survey

Huifen Xing, Siok Yee Tan, Faizan Qamar and Yuqing Jiao

Special Issue

Deep Learning in Object Detection

Edited by

Dr. Yang Lu

MDPI

https://doi.org/10.3390/app15126891

applied sciences

MDPI

Review

Face Anti-Spoofing Based on Deep Learning: A Comprehensive Survey

Huifen Xing [1,2], Siok Yee Tan [1,*], Faizan Qamar [3] and Yuqing Jiao [1,2]

[1] Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia; p131528@siswa.ukm.edu.my (H.X.); p129517@siswa.ukm.edu.my (Y.J.)
[2] School of Computer Science and Artifcial Intelligence, Chaohu University, Hefei 238024, China
[3] Center for Cyber Security, Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Malaysia; faizanqamar@ukm.edu.my
* Correspondence: esther@ukm.edu.my

Abstract: Face recognition has achieved tremendous success in both its theory and technology. However, with increasingly realistic attacks, such as print photos, replay videos, and 3D masks, as well as new attack methods like AI-generated faces or videos, face recognition systems are confronted with significant challenges and risks. Distinguishing between real and fake faces, i.e., face anti-spoofing (FAS), is crucial to the security of face recognition systems. With the advent of large-scale academic datasets in recent years, FAS based on deep learning has achieved a remarkable level of performance and now dominates the field. This paper systematically reviews the latest advancements in FAS based on deep learning. First, it provides an overview of the background, basic concepts, and types of FAS attacks. Then, it categorizes existing FAS methods from the perspectives of RGB (red, green and blue) modality and other modalities, discussing the main concepts, the types of attacks that can be detected, their advantages and disadvantages, and so on. Next, it introduces popular datasets used in FAS research and highlights their characteristics. Finally, it summarizes the current research challenges and future directions for FAS, such as its limited generalization for unknown attacks, the insufficient multi-modal research, the spatiotemporal efficiency of algorithms, and unified detection for presentation attacks and deepfakes. We aim to provide a comprehensive reference in this field and to inspire progress within the FAS community, guiding researchers toward promising directions for future work.

Keywords: face anti-spoofing; presentation attacks; deep learning; multi-modal; domain generalization

Academic Editor: Yang Lu

Received: 15 April 2025
Revised: 21 May 2025
Accepted: 28 May 2025
Published: 18 June 2025

Citation: Xing, H.; Tan, S.Y.; Qamar, F.; Jiao, Y. Face Anti-Spoofing Based on Deep Learning: A Comprehensive Survey. *Appl. Sci.* **2025**, *15*, 6891. https://doi.org/10.3390/app15126891

1. Introduction

In the current era of digitalization and intelligence, authentication methods have undergone significant changes. Traditional identity verification methods that rely on physical devices like keys and access cards, which are at risk of loss, so are now being replaced by biometric authentication [1]. Common biometric recognition options include the face [2,3], fingerprint, iris, palm print, ear, voice [4], finger vein, and gait [5]. Fingerprint data are easy to collect but can be affected by the condition of a user's skin, reducing accuracy. Iris recognition requires specific equipment, making collection difficult. Finger vein data offers high levels of security, but the equipment is not widely available and contact is required, which limits efficiency. Due to the rapid proliferation of image capture devices, facial data are no longer difficult to acquire. Face recognition (FR) [6–8], due to

its simplicity, non-contact nature, and high accuracy, is now widely used in areas such as airport and station access, telecom service processing, financial authentication, and employee attendance management [9].

However, the rapid growth of FR has also introduced new security risks. Using forged faces to bypass a face recognition system (FRS) can lead to face spoofing attacks. According to a report by China Central Television (CCTV) Finance, in the first half of 2024, multiple cases of criminals wearing silicone face masks to impersonate others and commit theft occurred in Shanghai and Xuzhou, Jiangsu. Using silicone face masks as a disguise during criminal activities is becoming a new tactic for some criminals. Particularly in the era of generative artificial intelligence (GAI), almost nothing seems impervious to being overturned. Even long-held beliefs like "seeing is believing" and "pictures as proof" are losing credibility. In February 2024, the Hong Kong police disclosed an AI-powered "multi-face swapping" fraud case involving a loss of 200 million Hong Kong dollars. These cases illustrate that FRSs without anti-spoofing verification cannot ensure security, posing a significant threat to users' privacy and property. Accurately determining the authenticity of faces is crucial for user security.

Today, significant breakthroughs in face recognition have been made. However, face recognition still has potential risks and faces various attacks, including those in both the digital and physical domains [10]. These attacks have raised urgent public security concerns such as user privacy leaks, identity theft, and financial fraud. As shown in Figure 1, in steps ① and ②, respectively, FRSs are susceptible to both physical and digital face attacks [10]. Malicious attackers can easily launch various physical presentation attacks during the face capture stage (i.e., step ①) and can use various digital forgery means (i.e., step ②) to attack a FRS.

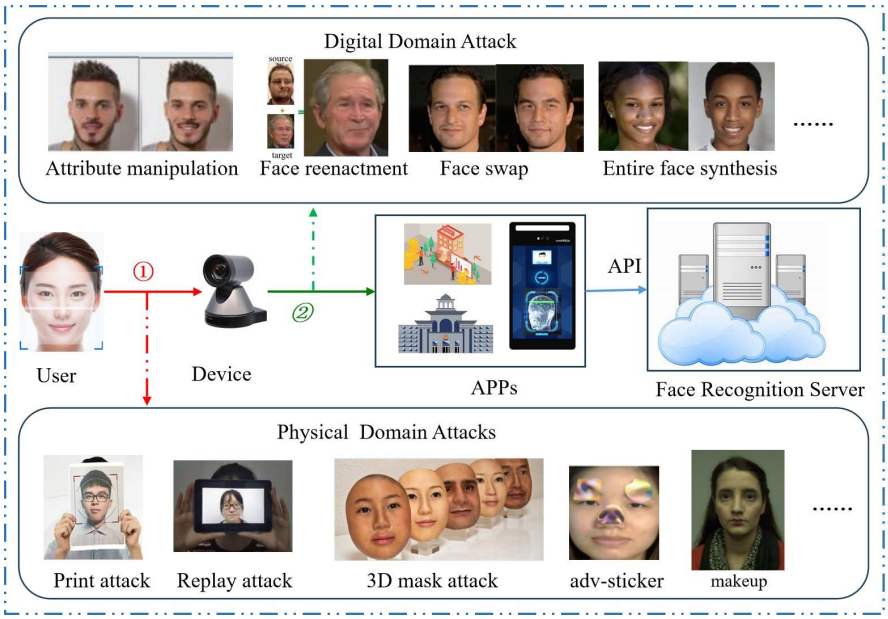

Figure 1. Overview of the FR process. Step①and step②are vulnerable to physical and digital face attacks.

With the rapid development of generative adversarial networks (GANs) [11], face deepfakes have been dramatically improved and forged faces and videos [12] are becoming more and more sophisticated. As shown in the upper part of Figure 1, face deepfakes refer to a face manipulation technology based on deep learning (DL). It can purposefully manipulate faces in images and videos, including through face attribute manipulation [13], face reenactment, face swapping, entire face synthesis, etc. [14]. However, in real environments, FRSs typically capture images through devices and input them directly into the system.

Conducting various attacks on digital images during transmission is challenging, whereas physical attacks are easier to implement. In the physical realm, FRSs can be compromised by presentation attacks (PAs), including print attacks, replay attacks, 3D mask attacks, physical adversarial attacks, as well as makeup and tattoos [15]. Therefore, we focus on discussing presentation attacks and detection methods in the physical domain.

1.1. Related Works

The existing literature reviews or surveys on FAS tend to focus on a single topic or lack comprehensiveness. Raheem et al. [16] discussed various indicators of face liveness detection (FLD) in detail, providing a reference for devising suitable solutions to face spoofing problems. However, it did not cover DL for FAS. Jia et al. [17] discussed 3D mask face anti-spoofing methods in detail from both software and hardware perspectives. Although DL algorithms were explored, the number of referenced studies was limited. Safaa et al. [18] centred on the application of a convolutional neural networks (CNNs) without addressing the generalization of face anti-spoofing. Abdullakutty et al. [19] concentrated on DL-based methods and multi-modal fusion face anti-spoofing algorithms, but lacked sufficient a discussion on generalization aspects. Khairnaret et al. [20] mainly paid attention to the bibliometric analysis of PAs based on domain adaptation.

Recent surveys have provided new insights into FAS. For instance, a review [15] outlined DL techniques but lacked a focus on traditional methods. Similarly, a review on face adversarial attacks [21] analyzed adversarial techniques but overlooked areas like multimodal fusion and domain generalization. The comprehensive review [22] offered an overview of both DL-based and traditional hand-crafted approaches, but did not discuss cross-domain learning in-depth. Furthermore, a review [23] concentrated on liveness detection in smartphone scenarios yet did not address the broader FAS field. Additionally, the review [24] mentioned domain generalization issues in FAS but did not thoroughly analyze the specific implementation challenges or limitations involved.

Table 1 summarizes the characteristics and limitations of the existing literature reviews on FAS. Overall, current reviews in the FAS field tend to have a narrow focus and lack a holistic and integrative perspective. Key areas, such as the systematic exploration of domain generalization, multimodal fusion, the spatiotemporal efficiency of algorithms, and the unified detection of presentation attacks and deepfakes remain underexplored. In addition, discussions on generalization capability, handling of complex attack scenarios, and emerging approaches such as self-supervised and few-shot learning could be further expanded. Our review aims to provide a more comprehensive and in-depth analysis of the FAS field, offering valuable insights to guide future research directions.

1.2. Motivation and Contribution

In recent years, DL has significantly advanced the field of FAS, especially with the support of large-scale datasets. However, existing reviews often focus on single topics and need a more comprehensive analysis of different modalities, the generalization capabilities of algorithms, and spatiotemporal efficiency. Therefore, a systematic review of FAS based on DL holds academic and practical significance in exploring current research bottlenecks and future directions.

Firstly, DL techniques, such as CNNs, GANs, and contrastive learning have achieved notable success in FAS. However, challenges remain in addressing their generalization to unseen attacks and cross-domain applications. A comprehensive review of these methods will facilitate an understanding of their strengths and weaknesses, promoting further development in the field. Secondly, integrating multi-modal information (e.g., RGB, depth images, infrared) shows potential for enhancing FAS detection capabilities, but it also

increases computational complexity. Summarizing the pros and cons of multi-modal methods can guide future research toward balancing accuracy and efficiency. Finally, the real-time performance and efficiency of FAS algorithms in smart devices require attention. A review can analyze the performance of algorithms across different devices, helping researchers understand how to maintain detection accuracy while reducing computational costs and ensuring real-time operation.

Table 1. A summary of the previous literature reviews.

Ref.	Year	Ref. Count	Topics and Observation	Disadvantages	Our Article
[16]	2019	208	Various indicators of FLD	It does not cover DL-based FAS algorithms.	We provide a systematic introduction to DL-based methods.
[17]	2020	57	FAS methods on 3D masks	The number of DL-based algorithms was very small.	
[18]	2020	58	The application of CNN in face and iris attack detection	Did not address the generalization of FAS methods.	We provide a detailed introduction to the cross-domain FAS, including DA, domain generalization (DG), zero-/few-shot learning, and anomaly detection.
[19]	2021	163	DL-based methods and FAS on multi-modal fusion	Sufficient discussion on generalization.	
[20]	2021	23	Face presentation attacks and detection based on domain adaptation (DA)	Lacked other classification	
[21]	2021	115	A literature review on face adversarial attacks and defenses	Lack of a consistent evaluation framework, and omission of experimental datasets.	Our review summarizes detection methods for various types of presentation attacks, including adversarial attacks.
[10]	2022	253	A literature review on face attacks and detection methods from the perspectives of digital and physical domains.	The introduction of face adversarial attacks and defenses methods in general.	
[15]	2022	252	A comprehensive review on FAS based on DL	The discussion of cross-domain FAS issues requires a more in-depth examination, and the analysis of FAS on smartphones requires improvement.	We include not only DL-based methods but also FAS methods combining with handcrafted and DL approaches, as well as smartphone-based detection methods.
[22]	2023	188	A comprehensive review of FAS based on DL and traditional handcraft features	The discussion on cross-domain FAS needs to be more in-depth. Moreover, the analysis of FAS on smartphones needs to be improved.	
[23]	2023	15	A survey on FAS techniques for smartphones	Lack of discussion on standardized evaluation and insufficient exploration of future directions.	
[24]	2023	275	A survey on DG methods across various application domains	Although the review includes a summary of methods for FAS, it needs a systematic approach.	We provide a detailed introduction to the cross-domain generalization of FAS

Based on previous research, this article comprehensively introduces the types of face presentation attacks and detection methods based on DL. Compared to previous reviews, this survey is unique in the following three aspects:

(1) This article comprehensively surveys face presentation attacks and detection methods based on DL, including face anti-spoofing on smartphones.
(2) This survey covers 229 multiple articles, categorizing them into smaller sub-topics, and provides a comprehensive discussion.
(3) This survey reveals the current challenges and future research trends about face anti-spoofing and provides a reference basis for research in face authentication.

The rest of this survey is organized as follows. Section 2 reviews face presentation attacks designed to deceive a FRS. Next, we introduce detection methods against face

presentation attacks in Section 3. We discuss the experimental datasets in Section 4. Furthermore, we discuss current challenges and potential future research directions in Section 5. At the end, Section 6 is the conclusion.

2. Attack Types

Presentation attacks [15] refer to an attempt by an attacker to deceive an FRS by presenting spoof face images, videos, or masks to trick the system into incorrectly verifying or recognizing the identity. Presentation attacks [15] mainly include impersonation attacks and obfuscation attacks. Impersonation attacks mainly include print, replay, and 3D mask attacks. Obfuscation attacks primarily consist of makeup attacks, tattoo attacks, funny glasses attacks, wig attacks, and physical adversarial attacks [21], which the attacker can directly initiate. The topology of face presentation attack types is shown in Figure 2. The details are introduced below.

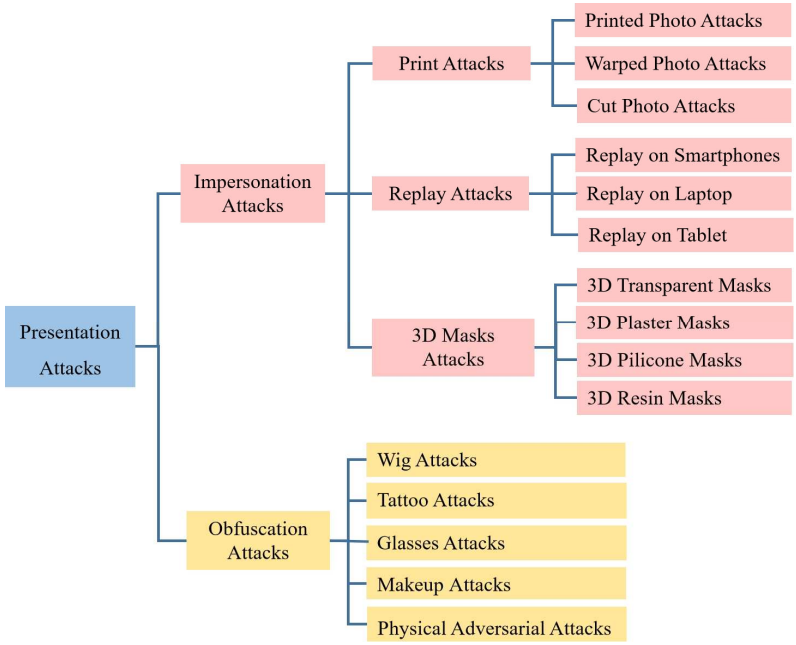

Figure 2. Topology of face presentation attack types.

2.1. Impersonation Attacks

Print attacks refer to an attack method in which the attacker prints the target user's face on paper and presents it in front of the camera to deceive the FRS. This attack method is straightforward and typical. Some attackers will cut out the eyes, nose, mouth, and other parts of the photo to expose the attacker's organs and simulate the blinking and opening of a real person's mouth, significantly increasing the difficulty of liveness detection. Examples of print attacks are shown in Figure 3a. For example, in the CASIA-MFSD [25] database, print attacks are collected by photographing printed photos. Specifically, a face image is first printed, and the photo is then distorted or cropped to "simulate" an honest face. Finally, the distorted or cropped photo is recorded to generate a spoofing face video. CASIA-MFSD used three cameras to capture face images at three different quality levels. Low-quality videos were captured using a long-time-used USB camera with a resolution of 640×480 pixels. Normal-quality videos were recorded with a newly purchased USB camera at 480×640 pixels. Furthermore, high-quality videos were recorded using a high-resolution Sony NEX-5 camera with a maximum resolution of 1280×720.

Reply attacks refer to an attack method in which an attacker uses the replay function of an electronic device to play a video of the target user and presents it in front of the camera

of the capture device. This attack method is more threatening than a photo printing attack. Video replay attacks can easily invalidate algorithms that detect facial movements, posing a considerable challenge to the versatility of liveness detection algorithms. Examples of replay attacks are shown in Figure 3b.

In the CASIA-MFSD [25] database, a reply attack is performed by displaying high-resolution real videos on an iPad. Due to the iPad screen's resolution limitation, the device inevitably downscales the original high-resolution (1280× 720) videos. The MSU-MFSD dataset [26] divides replay attacks into two cases. One involves recording with a Canon 550D camera and replaying the video on an iPad Air, and the other involves recording with an iPhone 5S and replaying it on the same device.

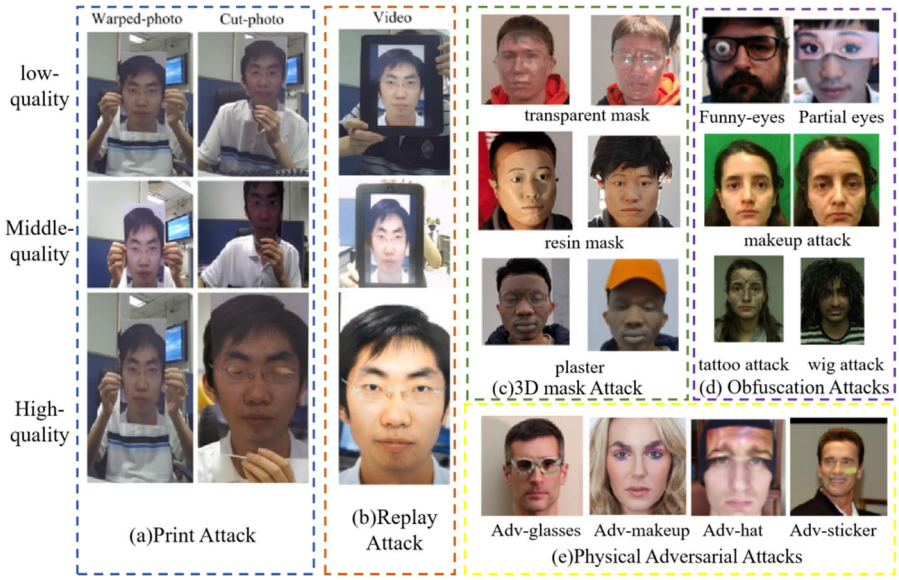

Figure 3. Various attack examples: (**a**) print attacks from CASIA-MFSD [25]; (**b**) replay attacks from CASIA-MFSD [25]; (**c**) 3D masks from HiFiMask [27]; (**d**) obfuscation attacks [28]; (**e**) physical adversarial attacks: Adv-glasses [29], Adv-makeup [30], Adv-hat [31], Adv-sticker [32].

A 3D masks attacks involves creating masks from materials like resin, plaster, or silicon to mimic a target user's face and deceive a FRS. These attacks are rare and complex due to the challenges in obtaining accurate 3D facial data, which requires specialized equipment, as well as the high cost and time involved in producing realistic masks. Inferior masks can be easily detected using simple texture features, but advanced masks with realistic textures can often bypass most liveness detection algorithms. However, due to the difficulties and costs associated with creating 3D masks, detection methods still primarily focus on photo and video replay attacks. Examples of 3D mask attacks are shown in Figure 3c. For example, the HiFiMask [27] dataset includes 25 subjects with yellow, white, and black skin tones to promote a fair AI and reduce bias caused by skin colour (a total of 75 subjects).

2.2. Obfuscation Attacks

Obfuscation attacks involve disguising facial features using various methods to interfere with or deceive a FRS. For example, the SiW-M [28] dataset comprises 13 spoof types, including print attacks, replay attacks, 3D masks, makeup attacks, and partial attacks. Examples of obfuscation attacks are shown in Figure 3d. These attacks pose significant threats as they can effectively disrupt or mislead such systems, impairing their ability to accurately identify individuals. Typically, obfuscation attacks use visual disguises that force the FRS to retrain on standard features, thereby reducing system security and reliability.

Tattoo attacks involve applying tattoos to the face or other visible areas to create disruptive patterns that affect the performance of a FRS.

Funny glasses attacks involve wearing specially designed glasses with exaggerated patterns or colors to interfere with facial recognition algorithms.

Wig attacks involve using wigs to change the appearance of hair, which affecting the system's assessment of hair and overall facial features.

Makeup attacks refer to attackers using makeup to deceive to alter facial features in an attempt to deceive a FRS. For example, heavy makeup may cover key features like the eyes or nose.

Physical adversarial attacks [21] refer to introducing subtle, carefully designed physical perturbations to face images in the real world, aiming to mislead the FRS. Unlike digital adversarial attacks, physical adversarial attacks involve disrupting the face in actual environments to cause the FRS to produce incorrect identity recognition results. Examples of physical adversarial attacks are shown in Figure 3e.

3. Face Anti-Spoofing

Face attacks and detection methods are an arms race, with attackers continuously developing new technologies to breach the FRS, including face presentation attacks, face deepfakes, and adversarial attack examples. FAS has evolved from handcrafted feature extraction to DL-based models, gradually addressing more complex attack techniques. Researchers have introduced domain adaptation and adversarial learning methods to enhance cross-domain generalization. Multi-modal fusion and 3D information are also employed to tackle more sophisticated 3D mask attacks. Recent advancements include meta-learning and self-supervised learning, which aim to improve the model adaptability to meet practical application requirements. As shown in Figure 4, we have summarized the topology of FAS.

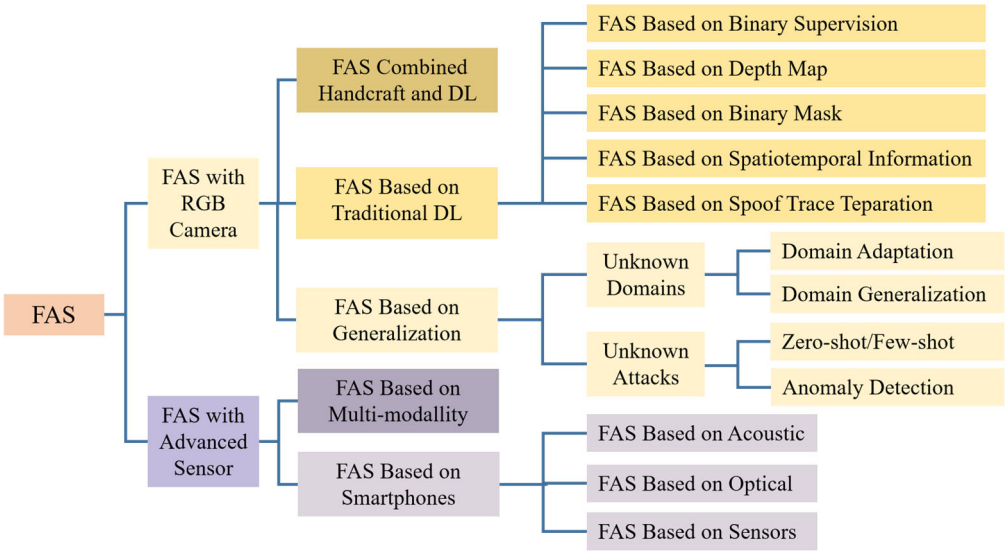

Figure 4. Topology of FAS based on DL.

3.1. FAS with RGB Camera

3.1.1. FAS Combining Handcraft and DL

It can be seen from the reviews [15,22] that the research on FAS prior to 2018 primarily focused on traditional manual feature extraction, with its primary strategy being to identify deception clues from the prostheses, which required extensive prior knowledge. Some methods were based on random motion, such as gaze tracking [33,34], eye blinking [35,36],

nodding and smiling [37], lip reading [38], etc. Some methods focused on life information, for example, remote Photoplethysmography (rPPG) [39–41], optical flow [42,43], and face micro-movements [44,45]. Moreover, texture features were extracted for FAS, including local binary pattern (LBP) [46], local binary pattern on three orthogonal planes (LBP-TOP) [47], color distortion texture [48], scale-invariant feature transform (SIFT) [49], speeded-up robust features (SURF) [50], histogram of oriented gradients (HOG) [51], difference of Gaussian (DoG) [52],and so on. Some researchers have paid attention to image quality features, including specular reflection, color distribution and sharpness [26,53]. In addition, some sensors were used for FAS, such as infrared images [54,55], light fields [56,57], and depth maps [58]. Typical handcrafted methods are shown in Table 2. As shown in Table 2, these algorithms are categorized based on the types of attacks they are generally effective at detecting, along with an analysis of their advantages and disadvantages. Despite their effectiveness in specific scenarios, these traditional methods often require improvement, such as necessitating user cooperation, being susceptible to external conditions, and needing additional hardware. With the development of deep neural networks (DNNs), many researchers have begun to use DNNs to extract the features of face images for FAS.

Table 2. Methods based on traditional handcrafted features. For simplicity, 2D-P represents print attacks, 2D-R represents replay attacks, 3D-M represents 3D masks attacks, OA represents obfuscation attacks, and Adv represents physical adversarial attacks. All subsequent tables are similar.

Methods	Main Idea	Ref.	Attacks	Advantages	Disadvantages
Methods based on motion	Gaze tracking Eye blinking Nodding and Smiling Lip reading	[33,34] [35,36] [37] [38]	2D-P	Taking into account motion characteristics, they can effectively resist print attacks.	It is difficult to deal with replay videos attacks, longer detection time.
Methods based on life information	rPPG Optical flow Micro-movements	[39–41] [42,43] [44,45]	3D-M 2D-P, 2D-R 2D-P, 3D-M	High accuracy under specific constraints. High accuracy and strong generalization for print attacks.	Video input is required but lacks robustness when highly affected by external lighting and individual movements.
Methods based on texture	LBP, LBP-TOP Color texture SIFT SURF HOG DoG	[46,47] [48] [49] [50] [51] [52]	2D-P	With a small amount of calculation and a relatively stable environment, it can effectively resist print and replay attacks.	It is easily affected by recording equipment, lighting conditions, and image quality, and its generalization capability across datasets is not strong.
Methods based on image quality	Image specular reflection, color distribution, and sharpness	[26,53]	2D-P	Relatively strong cross-dataset generalization ability for single-type spoofed faces while offering a fast processing speed.	It is more sensitive to changes in the external environment and difficult to resist high-resolution matte photo and video attacks.
Methods based on hardware	Infrared image Light field Depth map	[54,55] [56,57] [58]	2D-P, 2D-R	High accuracy	It requires the addition of expensive hardware devices, and the time of processing images increases.

Yang et al. [59] introduced DL to FAS for the first time in 2014, demonstrating the positive impact of CNN on face liveness detection. Since then, researchers have integrated traditional features with DNN to enhance the performance of FAS. By combining handcrafted features like LBP [46] and HOG [51] with DNN, these algorithms use methods like cascade fusion, feature-level fusion, and to improve spoofing detection. Traditional features capture detailed textures, while DL-based models identify complex patterns, improving the algorithm's accuracy and robustness. However, this may increase computational complexity and processing time.

Cascade model fusion: Feng et al. [60] proposed a layered neural network to fuse shearlet-based image quality and optical flow-based motion cues for face liveness detection. Li et al. [61] fully utilized the differences in reflective characteristics between real faces and 3D masks. Li et al. [62] combined optical flow with video frames as CNN input, employing spatial and channel attention for adaptive classification. Li et al. [63] leveraged VGG

for extracting features, applied principal component analysis (PCA) for dimensionality reduction, and employed Support Vector Machine (SVM) for classification. Asim et al. [64] cascaded LBP-TOP with CNN to capture spatio-temporal features via histograms from CNN layers, and classified by SVM with nonlinear RBF kernel. Shao et al. [65] extracted features with CNN, used optical flow for facial motion detection, and jointly learned spatial and channel features. Agarwal et al. [66] utilized the power of CNN filters and texture encoding for silicone mask attacks. Ya et al. [67] integrated face image quality assessment (FIQA) with neural network decomposition for improving generalization.

Feature-level fusion: Khammari et al. [68] proposed a dual-stream approach combining LBP and simplified weber local descriptor (SWLD) encoded CNN. The fused features ensured the preservation of the local intensity information and the orientations of the edges. Yu et al. [69] developed a method using rPPG signals and vision transformer (ViT) to detect 3D mask attacks by analyzing the physiological differences between real faces and spoofed masks. Li et al. [70] addressed replay video attacks via motion blur analysis, extracting temporal blur descriptors with a 1D CNN and blur width using local similarity pattern (LSP) features, which were fused for detection. Chen et al. [71] introduced a method combining deep features extracted by CNN and color texture features by using rotation-invariant local binary pattern (RI-LBP), applying the fused data to an SVM for classification.

Decision-level fusion: CHEN et al. [72] proposed a FAS algorithm that utilized an improved Retinex-LBP to extract key texture and illumination features from the image, combined with R-CNN for deep feature extraction. Sharifi et al. [73] used CNN and overlapped histograms of local binary patterns (OVLBP) methods to extract facial features and employed score-level and decision-level fusion strategies to achieve more accurate FAS detection. Solomon et al. [74] separately extracted image quality features and deep features using ResNet50 for classification and combined their confidence scores to make the final decision.

Table 3 lists hybrid FAS using handcrafted features and DL. As shown in Table 3, combining different techniques with deep learning enables the detection of more complex types of attacks. For example, in [62], by integrating facial motion and texture cues, it was possible to detect 2D print, 2D replay, and 3D mask attacks.

3.1.2. FAS Based on Traditional DL

Early training of DNN in FAS struggled to outperform traditional handcrafted methods due to the need for large-scale data. However, advancements in architectures like U-Net [75] and DenseNet [76], along with large-scale datasets such as CelebA-spoof [77] and HiFiMaskt [27], have propelled end-to-end DL methods to dominate the FAS field. Unlike hybrid methods that integrate partially handcrafted features and have no learnable parameters, traditional end-to-end DL methods directly learn the mapping function from facial input to spoof detection. During model training, deep models can use binary loss supervision, pixel-level supervision, pseudo-depth labels, binary mask labels, and other labels to distinguish between real and spoof faces. In addition to using spatial information for facial spoof detection, temporal sequences between video frames can also be used to extract spoofing features. Based on the type or source of the spoofing cues, we categorize traditional DL-based FAS methods into five types: binary supervision methods, depth map methods, binary mask methods, spatiotemporal information fusion methods, and spoof trace separation methods. Below, we provide a detailed explanation of these methods.

Table 3. Overview of hybrid FAS: The symbol "√" indicates the presence of the corresponding attack type within the dataset. All subsequent tables are similar.

Strategies	Main Idea and Ref.	Datasets	Attacks			Advantages	Disadvantages
			2D-P	2D-R	3D-M		
Cascade model fusion: Traditionally, handcrafted feature extraction is used as a preprocessing step, followed by further processing of the extracted features using a DL to achieve finer features.	Dense optical flow-based motion features, shearlet-based image quality features and DNN [60]	CASIA MFSD	√	√		Cascaded models can progressively refine and extract more complex features, thereby enhancing their classification capability.	(1) If the feature extraction in the preceding model has issues, subsequent steps will be affected, leading to a decline in the overall performance. (2) There may be feature gaps or incompatibilities between handcrafted features and deep features, resulting in performance saturation.
		Replay-Attack	√	√			
		3D-MAD			√		
	Image reflectance and 1D CNN [61]	3D-MAD			√		
		HKBU-M V2			√		
	The optical flow-based motion features and texture cues, attention module and CNN [62]	Replay-Attack	√	√			
		OULU-NPU	√	√			
		HKBU-MARs V1			√		
Cascade model fusion: The DL and handcrafted feature extraction processes are sequential. The DL model is used to extract deep features, which are then subjected to handcrafted feature processing for further operations or analysis.	VGG features, PCA and SVM [63]	CASIA MFSD	√	√			
		Replay-Attack	√	√			
	CNN, LBP-TOP and SVM [64]	CASIA MFSD	√	√			
		Replay-Attack	√				
	CNN features and optical flow-based motion features [65]	3DMAD			√		
	CNN filters, texure encoding and SVM [66]	SMAD			√		
	FIQA and Resnet [67]	SiW	√	√			
		SiW-M	√	√	√		
Feature-level fusion: During the feature extraction stage, traditional handcrafted features are fused with deep features extracted by DL models.	LBP, SWLD, and CNN [68]	CASIA MFSD	√	√		The synergy between features enhances the model's capabilities.	Aligning and processing features from different sources is necessary, which increases the complexity and reduces the model's interpretability.
		Replay-Attack	√	√			
	Combination rPPG and ViT [69]	3D-MAD			√		
		HKBU-M V2			√		
	1D CNN and motion blur features with LSP [70]	Replay-Attack	√	√			
		OULU-NPU	√	√			
	Color texture features with RI-LBP and CNN [71]	NUAA [52]	√				
		CASIA MFSD	√	√			
Decision-level fusion: Handcrafted features and deep features are processed separately by independent classifiers (or detectors). The final prediction result is obtained through decision fusion, such as voting or weighted voting.	Retinex-LBP and R-CNN [72]	CASIA MFSD	√	√		Fusion occurs only at the final decision stage, without the need to align or process the features themselves, offering high flexibility.	Independent processing may lead to information redundancy or conflict, and the algorithm highly depends on the fusion strategy.
		Replay-Attack	√	√			
		OULU-NPU	√	√			
	CNN features and OVLBP [73]	Replay-Attack	√	√			
		OULU-NPU	√	√			
	Image quality features and deep features using ResNet50 [74]	CASIA-MASD	√	√			
		Replay-Attack	√	√			
		MSU MFSD	√	√			
		SiW	√	√			

A. FAS Based on Binary Supervision

In the early stages of applying DL to FAS, researchers treat the task of distinguishing between real and fake faces as a simple binary classification problem. When training neural networks, the training data is labelled with 0 and 1 to represent fake faces and real faces, respectively, and classification is achieved by calculating the difference between the network's predictions and the 0/1 labels using Binary Cross Entropy (BCE) loss. The flowchart of this type of method is shown in Figure 5.

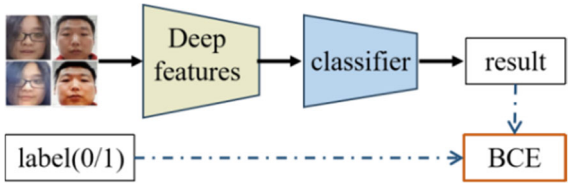

Figure 5. The flowchart of FAS based on binary supervision.

Lucena et al. [78] proposed a transfer learning method that used the sigmoid function for the binary classification of real and fake faces, achieving state-of-the-art results on the 3DMAD and Replay-Attack datasets. Nagpal et al. [79] adjusted Inception-v3 [80], ResNet-50 [81], and ResNet-152 [81] for binary classification, evaluated their performance on the MSU-MFSD dataset, and provided evaluation results and application suggestions for the three models. To address the difficulty of obtaining spoofed face data for DL, Guo et al. [82] first converted 2D printed photos into 3D objects, then simulated operations

such as bending and rotating the printed photos in 3D space to synthesize a large number of 3D virtual spoofed photos. Finally, they used the synthesized samples to train a modified ResNet-50 and present the binary classification results. On the other hand, considering the weak intra-class and inter-class constraints of BCE loss, some studies modify the BCE loss to provide CNN with more discriminative supervisory signals. Xu et al. [83] reformulated FAS as a fine-grained classification problem instead of a binary classification problem and used type labels (e.g., bonafide, print, and replay) for multi-class supervision. However, FAS models supervised with multi-level CE loss may fail when faced with high-fidelity PAs. To learn a compact space with small intra-class distances and large inter-class distances, Almeida et al. [84] introduced contrastive loss and triplet loss. However, unlike visual retrieval tasks, bonafide and PAs in FAS tasks usually have asymmetric internal distributions (more compact and diverse, respectively). Based on this evidence, Wang et al. [85] proposed using asymmetric angular margin softmax loss to supervise FAS patch models, thereby relaxing the intra-class constraints among PAs.

Table 4 shows an overview of the methods for FAS based on binary supervision. As shown in Table 4, FAS methods based on binary supervision are suitable for detecting common 2D print and 2D replay attacks in environments where the attack types are known, labels are clear, and both spoof and genuine samples are available. They are well-suited for deployment in fixed-data scenarios that require only a binary real/spoof judgment rather than a detailed attack-type classification, typically in low- to medium-risk applications.

Table 4. Overview of methods for FAS based on binary supervision.

Ref.	Main Idea	Supervision	Datasets	Attacks				Advantage	Disadvantage
				2D-P	2D-R	3D-M	OA		
[78]	Transfer learning and CNN	BCE	Replay-Attack	√	√				
			3DMAD			√			
[79]	Analysis of effectiveness on CNN	BCE	MSU-MFSD	√	√				
[82]	3D virtual synthesis as input	BCE	Replay-Attack	√	√				
			CASIA-MASD	√	√				
[83]	Fine-grained multi-class supervision	Multi-level BCE	SiW	√	√				
			OULU-NPU	√	√				
			SiW-M	√	√	√	√	The method is simple and intuitive, providing clear training signals and stable training based on mature technology.	It is prone to failure when faced with high-fidelity attacks, such as high-definition spoofing videos.
[84]	Learning compact spaces with small intra-class distances and large inter-class distances	Contrast loss, triple loss	Replay-Attack	√	√				
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			RECORD-MPAD	√	√				
[85]	Supervising FAS patch model via asymmetric corner softmax loss to relax intra-class constraints	softmax loss	Replay-Attack	√	√				
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			MSU-MFSD	√	√				
			SiW	√	√				

B. FAS Based on Depth Maps

In the field of computer vision, a depth map is an image or image channel that contains information about the distance from the surface of the target object to the viewpoint. When applied to FAS, the depth map highlights the differences in distance between the surface of real and spoofed faces and the camera. For real faces, which have a three-dimensional structure, there are significant distance variations between different regions (such as the nose and ears) and the camera. In contrast, spoofed faces are typically planar, resulting in little to no variation in the distances between different regions of the spoofed face and the camera. The specific training process is illustrated in Figure 6. Initially, the depth maps corresponding to real or spoofed face images are used as supervisory signals to guide the model in predicting the depth map of the input image. Subsequently, a depth supervision loss is employed to make the final classification decision.

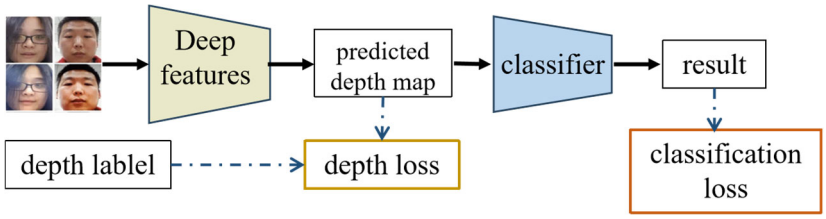

Figure 6. The flowchart of FAS based on depth maps.

Atoum et al. [86] proposed a dual CNN method combining global and local facial features. First, one CNN was used to obtain a depth map of the image globally (i.e., spatial cues) and provide a "live" score based on this information. Then, another CNN learned distinguishable texture features from randomly selected small local patches of the image and provided another score. Finally, these two scores were fused to obtain the final decision, enhancing the accuracy of the judgment. However, FAS methods face two significant challenges: extracting dynamic features from multi-frame image sequences increases the model's time complexity, and expert-designed network structures struggle to capture fine-grained information in images and quickly adapt to different environments.

To enhance the automatic design and fine-grained depth feature acquisition capabilities, Yu et al. [87] first introduced neural architecture search (NAS) into FAS to automatically find the optimal backbone architecture for capturing central difference convolution (CDC). They also designed multiscale attention fusion (MAFM) to refine and fuse low, mid, and high-level CDC features through spatial attention. On this basis, they extend CDCN to a more automated and better-performing CDCN++. To ensure NAS for FAS generalize well to new environments and new types of attacks, Yu et al. [88] further developed CDC and pooling operators, exploited an efficient static-dynamic representation for mining the FAS-aware spatio-temporal discrepancy, and proposed domain/type-aware Meta-NAS for robust searching. Zheng et al. [89] designed a dual-stream spatial-temporal network to explore potential depth and multiscale information. In this network, a temporal shift module was introduced to extract temporal information, and symmetric loss was used for more accurate auxiliary supervision in depth information estimation network. Finally, a scale-level attention module was introduced to fuse information from the dual-stream network to determine the authenticity of the image.

All in all, FAS based on depth map supervision can better capture the three-dimensional structure and details of the face, reduce artifacts caused by illumination changes, angle changes, etc., and improve the recognition rate. However, synthesizing 3D shape labels for each training sample is expensive and inaccurate, and it also lacks the rationality for real-depth PA (such as 3D masks and mannequins). Table 5 shows an overview of methods for FAS based on a depth map.

C. FAS Based on Binary Masks

Compared to depth map supervision, binary mask labels are easier to generate. With binary mask labels, we can determine whether a PA occurs in the corresponding patch, which is attack-type agnostic, spatially interpretable, and easier to generalize to all PAs. The flowchart of this type of method is shown in Figure 7.

George et al. [90] trained the model using both binary and deep pixel-wise supervision, relying solely on frame-level information. As a result, the approach attained low computational and time overhead. However, due to the limited amount of training data available, the method suffered from a weak generalization capability. Hossain et al. [91] introduced a learnable attention module to refine features, mitigating the limitations of noisy binary mask labels, especially in partial attacks. Sun et al. [92] demonstrated that local label supervision outperformed global labels for small datasets, introducing ternary

labels (real, spoofed, and uncertain background) to improve pixel-level predictions. Other researchers, based on the difference in facial material-related reflectance between live skin and spoof media, proposed using both depth and reflection labels to supervise depth models [93,94]. Yu et al. [93] leveraged material differences between real skin and spoof media, using bilateral convolutional networks (BCNs) with depth, reflection [95], and patch maps for supervision. Despite significant progress in facial depth auxiliary supervision, there remain considerable challenges in modelling lightweight temporal networks. Yu et al. [96] proposed a novel pyramid supervision, guiding depth models to learn local details and global semantics from multiscale spatial contexts. Further advancements incorporated auxiliary supervision from Fourier maps [97], LBP texture maps [98], and sparse 3D point clouds [99], enhancing textural and geometric discrimination. Table 6 summarizes these FAS methods based on binary mask and extended supervision.

Table 5. Overview of methods for FAS based on depth map.

Ref.	Main Idea	Supervision	Datasets	Attacks				Advantages	Disadvantages
				2D-P	2D-R	3D-M	OA		
[86]	Dual CNN method combining global and local features of face.	Depth map	Replay-Attack	√	√				
			CASIA-MFSD	√	√				
			MSU-USSA	√	√				
			Replay-Attack	√	√				
[87]	CDCN can capture detailed patterns via aggregating intensity and gradient information, and the MAFM module obtain afiner-grained features.	Depth map	CASIA-MFSD	√	√			It can better capture the 3D structure and details of the human face; reduce artifacts caused by changes in lighting and angles, etc.; and has strong applicability.	The process of synthesizing 3D shape labels is costly and time-intensive. For 3D mask attacks or mannequin model attacks, since these attacks also have real depth information, they are more difficult to detect.
			OULU-NPU	√	√				
			MSU-MFSD	√	√				
			SiW	√	√				
			SiW-M	√	√	√	√		
[88]	An efficient static-dynamic representation for mining the FAS-aware spatio-temporal discrepancy and Meta-NAS for robust searching.	Depth map	Replay-Attack	√	√				
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			MSU-MFSD	√	√				
			SiW	√	√				
			SiW-M	√	√	√	√		
			CASIA-SURF			√			
			3DMAD			√			
			HKBU-M V2			√			
[89]	A dual-stream spatial-temporal network was designed to explore depth information and multiscale information.	Time, depth map	Replay-Attack	√	√				
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			SiW	√	√				
			NUAA	√					

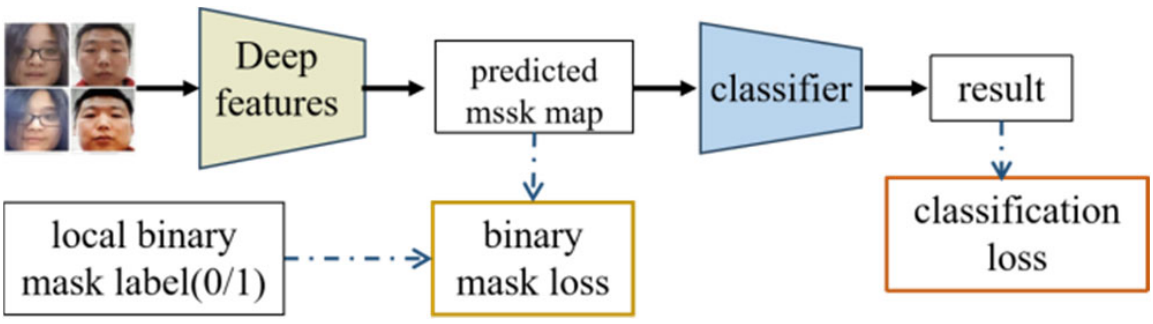

Figure 7. The flowchart of FAS based on binary mask.

Table 6. Overview of methods for FAS based on binary mask and extended supervision.

Ref.	Main Idea	Supervision	Datasets	Attacks				Advantages	Disadvantages
				2D-P	2D-R	3D-M	OA		
[90]	Binary mask and deep pixel-wise supervision	Binary mask label	Replay-Attack	√	√			Binary mask labels are easier to generate, able to identify facial regions and backgrounds, capture subtle features, and are attack-type agnostic.	Complex attacks (such as cutting out eyes, nose, and mouth and simulating real human blinking and mouth movements) may not be accurately detected.
[91]	Angular margin-based BCE, a learnable attention module, and binary masks supervision	Binary mask label	OULU-NPU	√	√				
			OULU-NPU	√	√				
			Replay-Mobile	√	√				
[92]	Pixel-level supervised learning of facial images using local ternary labels (real face, fake, background)	Local ternary label	Replay-Attack	√	√			When there is not enough training data, more refined classification can be achieved.	Local ternary labels need to be accurately annotated.
[93]	Intrinsicmaterial-based patterns via aggregating multi-level bilateral micro-information	Depth map, reflection and patch map	CASIA-MFSD	√	√			New perspective: material perception	The limited scope of the datasets cannot fully demonstrate the effectiveness of material perception.
			OULU-NPU	√	√				
			SiW	√	√				
			OULU-NPU	√	√				
			SiW	√	√				
			CASIA-MFSD	√	√				
			Replay-Attack	√	√				
			MSU-MFSD	√	√				
			SiW-M	√	√	√	√		
[94]	Auxiliary supervisions	Depth map, reflection map	SiW	√	√			Focusing on the complementary information of the face enriches the feature representation.	Dependent on the quality of auxiliary supervision.
			OuluNPU	√	√				
			CASIA-FASD	√	√				
			Replay-Attack	√	√				
			CASIA-MFSD	√	√				
[96]	Plugged-and-played pyramid supervision	Depth map, Predicted Depth Maps	Replay-Attack	√	√			Both performance improvement and interpretability enhancement	Needs a large amount of data with fine-grained pixel/patch-level annotated labels
			MSU-MFSD	√	√				
			OULU-NPU	√	√				
			SiW-M	√	√	√	√		
[97]	Bi-directional feature pyramid network and fourier spectra supervision	Fourier spectra BCE	OULU-NPU	√	√			Introducing Fourier spectra enhances the accuracy of spoof detection.	The model's generalization ability is limited.
			Replay-Mobile	√	√				
[98]	Liveness and content features via disentangled representation learning	Depth map Texture map	CASIA-MFSD	√	√			Separating spoof and real features enhances feature discrimination ability and detection accuracy.	Disentanglement relies on precise data annotation.
			Replay-Attack	√	√				
			SiW	√	√				
			OULU-NPU	√	√				
[99]	Discriminative features via fine-grained 3D point cloud supervision	3D point cloud	CASIA-MFSD	√	√			Capturing facial depth and geometric features and handling various types of spoofing attacks.	Obtaining 3D point cloud data is challenging.
			Replay-Attack	√	√				
			SiW	√	√				
			OULU-NPU	√	√				

D. FAS Based on Spatiotemporal Information

The information generated by facial movements is also captured within consecutive multiple frames of a video stream. This temporal data embedded within successive frames is conducive to face detection. The flowchart for this method is depicted in Figure 8. It should be noted that the spatiotemporal information derived from facial movements still belongs to passive, one-way detection methods. In this process, the system passively receives video or image streams and then employs models such as CNNs or LSTMs to learn temporal patterns for detecting spoofs.

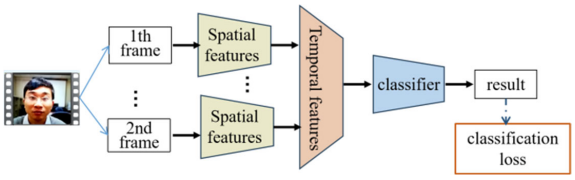

Figure 8. The flowchart of FAS based on spatiotemporal information.

Lin et al. [100] utilized planar homography to detect the correlated motion patterns in replayed videos that were absent in real faces. By dividing adjacent frames into regions and calculating homography relationships, they identified spoofed faces. Yang et al. [101] proposed a FAS method considering global temporal and local spatial cues combined CNN-LSTM for global spatiotemporal feature fusion and a region attention module for identifying critical local regions, thereby enhancing robustness and interpretability. Wang et al. [102] proposed a method that integrated spatial gradient and temporal depth learning, employing a residual spatial gradient block (RSGB) for extracting single-frame spatial features and a cascade of short-term spatiotemporal blocks followed by ConvGRU for multi-frame depth map generation. Cai et al. [103] presented a DRL-FAS framework. This method used ResNet18 for global spatial features and utilized RNN for local temporal learning, followed by feature fusion for classification. Wang et al. [104] introduced a

temporal transformation network (TTN) for learning temporal cues, leveraged temporal difference attention (TDA) for utilize their complementarity on multiple frames and used temporal depth difference loss (TDL) for extracting depth information. Liu et al. [105] fused motion trajectories features, history cues and texture difference features from facial key points, using attention and skip connections to enhance robustness under varying lighting and attack conditions. In summary, spatiotemporal-based FAS methods effectively combine global and local features to distinguish genuine faces from spoofed ones. Nevertheless, their reliance on multi-frame processing increases computational complexity, thereby affecting the real-time performance. Table 7 provides an overview of these methods. As indicated in Table 7, this type of algorithm is capable of detecting common 2D print, 2D replay, and low-quality 3D mask attacks. However, due to the high computational overhead of multi-frame processing, it is not suitable for low-power real-time systems.

Table 7. Overview of methods for FAS based on spatiotemporal information.

Ref.	Main Idea	Backbone	Datasets	Attacks 2D-P	Attacks 2D-R	Attacks 3D-M	Advantages	Disadvantages
[100]	Utilizing local planar homography for capturing fine-grained facial motion cues.	Resnet	Replay-Attack	√	√		Contributing to the verification accuracy, the MPEM enhances the recall rate of the attack videos.	Local planar homography involving complex calculations
[101]	Fusing global temporal and local spatial information from the video stream	CNN LSTM	OULU-NPU	√	√		Automatically focuses on important regions, enabling network behavior analysis and improving model interpretability.	Lots of experiments and dataset requirements complicate the research process and make reproduction more difficult.
			Replay-Attack	√	√			
[102]	Combining spatial gradient and temporal depth information	CNN	CASIA-MFSD	√	√		The framework may process spatial gradient and temporal depth simultaneously to improve detection accuracy.	Processing spatial and temporal features is computationally intensive.
			OULU-NPU	√	√			
			SiW	√	√			
			Replay-Attack	√	√			
			CASIA-MFSD	√	√			
			OULU-NPU	√	√			
			SiW	√	√			
			3DMAD			√		
[103]	Deep reinforcement learning(DRL), global spatial features and local temporal learning	ResNet18 RNN	Replay-Attack	√	√		Reinforcement learning improves recognition accuracy.	Weak generalization ability.
			CASIA-MFSD	√	√			
			OULU-NPU	√	√			
			MSU-MFSD	√	√			
			SiW	√	√			
			ROSE-YOUTU	√	√	√		
[104]	Multi-granularity temporal characteristics learning using ViT	ViT	Replay-Attack	√	√		Multi-granularity temporal features may capture more dynamic information and effectively improve the accuracy of FAS.	Increasing the computational complexity and requiring higher hardware resources.
			CASIA-MFSD	√	√			
			OULU-NPU	√	√			
			MSU-MFSD	√	√			
			SiW	√	√			
			CASIA-SURF	√				
			CFA	√	√	√		
			MLFP			√		
[105]	Merging motion trajectories, history cues, and texture differences around the facial local key points with a attention module and skip fusion strategy.	Swin-Transformer	FAAD [105]	√	√	√	The model performs excellently under varying lighting conditions and motion scales.	Performance is influenced by the datasets, and user interaction may be needed.
			MMI [105]	√	√			

E. FAS Based on Spoof Trace Separation

We refer to accurately isolating the essential spoofing features from a fake face as spoof trace separation. This type of FAS algorithm involves three basic steps, as seen in Figure 9. First, we input the fake face I' into the generator G to produce the spoof trace image $G(I')$. Next, we subtract the spoof trace image $G(I')$ from the fake face I'm to generate a synthetic real image $I' - G(I')$. Finally, the final classification result will be computed based on the separated spoof traces.

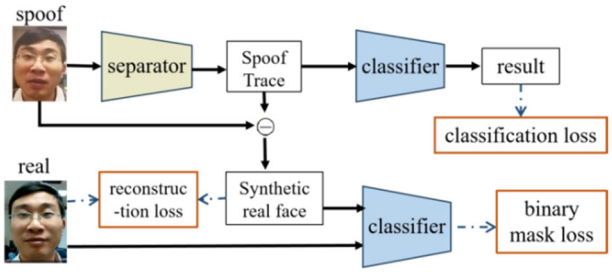

Figure 9. The flowchart of FAS based on spoof trace separation.

Jourabloo et al. [106] introduced a three-part CNN structure, DS-Net for separating spoof traces and reconstruct real faces, DQ-Net for estimating depth information, and VQ-Net for differentiating reconstructed real faces from original ones. These networks operated iteratively, improving spoof detection. Similarly, Liu et al. [107] extended on this concept with the spoof trace disentanglement network (STDN), which not only reconstructed real faces but also generated synthetic spoofed faces, offering enhanced interpretability and robustness against diverse attacks. Compared to [106], ref. [107] had stronger interpretability and resistance to diverse attacks. Feng et al. [108] proposed a network with a spoof trace generator and auxiliary classifier, which addressed overfitting and enhanced generalization. Wu et al. [109] utilized a dual spoof disentanglement generation (DSDG) framework with a variational autoencoder (VAE) to disentangle latent features and enhance robustness. Similarly to [98], Wang et al. [110] combined dual-stage learning with reconstruction techniques to separate spoof-related features, improving the recognition of various attacks. Ge et al. [111] addressed domain shift with the DiffFAS framework, which employed generative diffusion models to create high-fidelity spoof faces, thereby enhancing data diversity and generalization. Yu et al. [112] used Fourier frequency disentanglement and data augmentation to improve cross-domain generalization by extracting frequency features and enhancing training samples.

FAS algorithms, which utilize spoof trace separation, enhance accuracy and interpretability by disentangling spoof traces from facial features, increasing robustness against complex attacks. However, these algorithms require high computational resources, complex structures, and cumbersome training processes. Therefore, they are best suited for scenarios that require high security levels, such as financial systems, facial payment, and government applications. Table 8 provides a summary of these methods.

FAS based on DL have achieved better performance than the traditional manual feature methods, greatly improving accuracy, especially when applied to a single domain. It more effectively detects 3D mask attacks, significantly advancing the development of FAS technology. Due to the inexplicability and limited generalization ability of DL-based methods, determining how to further improve the interpretability, controllability, and generalization ability of DL-based FAS remains a topic worthy of ongoing research.

3.1.3. FAS Based on Generalization

The current DL-based FAS methods have achieved satisfactory results in the evaluation of various datasets. However, these models are based on the assumption that the source and target datasets are independent and identically distributed (i.i.d.) [24], which often results in the trained model using a large number of domain-specific features for classification, such as background, lighting. The inability to learn features that can truly distinguish between real faces and spoof faces easily leads to the model overfitting on various datasets. In the real world, there are a lot of differences between the samples used for inference testing and the data used for training, such as environments, attack methods. Therefore, algorithms that overfit on the training datasets cannot be generalized to real usage sce-

narios. In other words, the model faces the problem of a weak generalization ability in out-of-domain scenarios.

Table 8. Summary of FAS based on spoof trace separation.

Ref.	Main Idea	Backbone	Datasets	Attacks 2D-P	2D-R	3D-M	OA	Advantages	Disadvantages
[106]	Utilizing noise modeling to obtain a spoof noise, and using it for classification	CNN	Replay-Attack	√	√			Utilizing multi-layer feature decoupling to improve model robustness on specific datasets.	The only defense against printed photos and video replays
[107]	Using adversarial learning to disentangle the spoof traces	GAN	CASIA-MFSD	√	√			Resisting diverse attacks and improving the detection accuracy.	Adversarial training needs high computing resources.
			OULU-NPU	√	√				
			OULU-NPU	√	√				
			SiW	√	√				
			SiW-M	√	√	√	√		
[108]	Utilizing U-Net to generate spoof cue maps with anomaly detection, and a auxiliary classifier made spoof cues more discriminative.	ResNet18 U-Net	Replay-Attack	√	√			General spoof cues can be learned to solve the problem of over-fitting and improve generalization.	Models require large amounts of diverse data.
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			SiW	√	√				
[109]	Learning a joint distribution of the identity representation and the spoofing pattern representation in the latent space	CDCN	Replay-Attack	√	√			Improves the model's generalization ability and detection accuracy.	Generating paired live and spoofing images increase the complexity and time cost of training.
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			SiW	√	√				
			SiW-M	√	√	√	√		
[110]	Effectively separating spoof-related features from irrelevant ones by incorporating reconstruction techniques.	ResNet18 U-Net [76]	Replay-Attack	√	√			Recognizes various spoofing attacks and improves their generalization across different attack types and conditions	Reliance on reconstruction may introduce additional complexity and computational overhead
			CASIA-MFSD	√	√				
			MSU-MFSD	√	√				
			SiW	√	√				
			SiW-M	√	√	√	√		
[111]	Using image style, image quality, and diffusion models for domain shift	UNet	Replay-Attack	√	√			To solve the problem of domain shift, and also alleviate the scarcity of labeled data with novel type attacks	Generating high-fidelity images may require significant computational resources.
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			MSU-MFSD	√	√				
			WMCA	√	√	√	√		
			PADISI	√	√	√	√		
[112]	Fourier frequency space disentanglement and data augmentation	DepthNet	Replay-Attack	√	√			The frequency domain analysis can capture subtle spoofing patterns.	Relying on low-level texture features may impact detection performance.
			CASIA-MFSD	√	√				
			OULU-NPU	√	√				
			MSU-MFSD	√	√				

This section will systematically summarize and analyze recent research progress on generalization models in the field of DL-based FAS. Starting from the difficulties and challenges encountered by FAS in real-world scenarios, the problems are divided into two categories: encountering unknown domains and encountering unknown attacks. Different methods for each problem are summarized, and then the methods for solving unknown domains are divided into two categories, domain adaptation and domain generalization. The methods for solving unknown attacks are divided into zero-shot/few-shot learning and anomaly detection. In this review, the term unknown domains refer to external variations unrelated to spoofing (e.g., background, lighting, and sensor noise) that nevertheless affect the visual quality of images. Generally speaking, the most significant domain discrepancy is covariate shift [113]. In contrast, unknown attacks typically refer to attack types that do not appear during the training phase and possess novel physical properties, such as material and geometry. The principles, advantages, and disadvantages of representative methods are detailed.

A. Domain Adaptation

The purpose of domain adaptation (DA) is to train the model on the source domain to perform as well as possible on the target domain by using the given target domain's knowledge [114]. Given a labeled source domain $D_s = \{x_i, y_i\}$ and a target domain $D_t = \{x_i, y_i\}$, we assume that their feature space and class space are the same, but their joint distribution is different, that is, $X_s = X_t$, $Y_s = Y_t$, $P_s(x, y) \neq P_t(x, y)$. The task of domain adaptation is to use source domain data to learn the prediction function $f : x_t \rightarrow y_t$ so that f has the smallest prediction error in the target domain. In unsupervised domain adaptation, where the labels are not available in the target domain, that is, Y is unknown. The schematic diagram of the model is shown in Figure 10a.

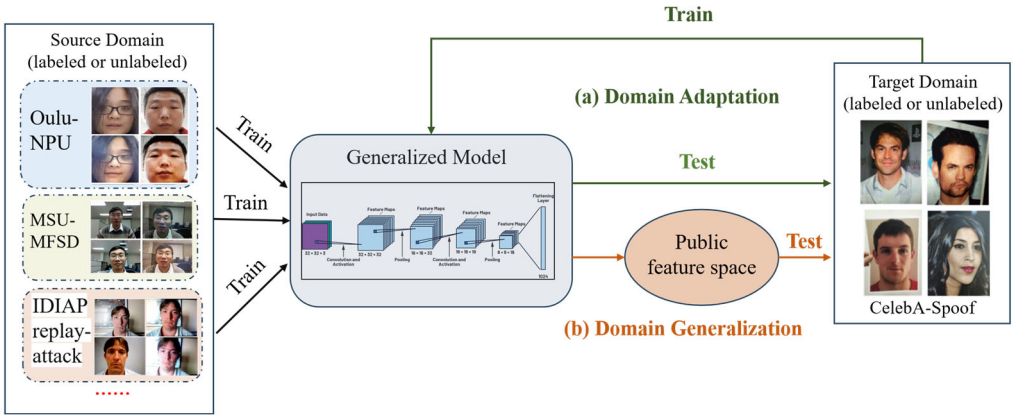

Figure 10. Framework comparison among domain adaptation (DA) and domain generalization (DG) [15]. (**a**) The DA methods need the (unlabeled) target domain data to learn the model, (**b**) DG methods learn a generalized model without relying on knowledge from the target domain.

Domain distribution discrepancy: Li et al. [115] proposed minimizing the maximum mean discrepancy (MMD) between the latent features in source and target domains was to learn a more generalized classifier. Li et al. [116] further proposed a novel optimization function for network distillation by combining cross-entropy loss, the MMD, and paired sample similarity embedding to effectively capture spoofing information. However, merely reducing the MMD distance between domains may not sufficiently exploit useful information from the source domain. This is why adversarial transfer learning has recently become a research hotspot.

Adversarial training: Wang et al. [117] proposed a disentangled representation learning net (DR-Net), which enabled the transfer of domain-independent knowledge to distinguish live and spoof faces in unlabeled target domains, with extensive evaluations on public datasets. Jia et al. [118] introduced a unified unsupervised and semi-supervised domain adaptation network (USDAN). It employed a marginal and conditional distribution alignment module for adversarial learning to minimize domain discrepancies and improve generalization. A Conditional Distribution Alignment (CDA) module could be flexibly used for semi-supervised or unsupervised settings. El-Din et al. [119] combined adversarial training with deep clustering to generate pseudo-labels for auxiliary training, addressing domain and device differences. Zhou et al. [120] utilized GANs to stylize target data to match source domain style, aligning features and reducing domain discrepancies to improve generalization and accuracy.

Other methods: Tu et al. [121] proposed a total pairwise confusion loss (TPC) and a fast DA module to improve generalization in live face detection and recognition, mitigating the negative impact of domain changes on prosthetic attack representation. Wang et al. [122] introduced a self-domain adaptation framework that utilizesd meta-learning to learn discriminative features across multiple source domains and adapt to target domain data during inference. Quan et al. [123] developed an adaptive transfer mechanism to reduce domain bias by gradually increasing the contribution of unlabeled target domain data during training. DA typically assume access to source and target data, but this is often unrealistic. To reduce reliance on source data source-free domain adaptation (SFDA) [124] was proposed. Liu et al. [125] introduced a SFDA framework for FAS (SDA-FAS), which utilized source-oriented regularization and contrastive domain alignment to reduce domain discrepancies. They extended on this with SDA-FASv2 [126], incorporating PatchMix data augmentation and supervised contrastive loss to mitigate domain and identity bias.

DA-based Methods can significantly improve the generalization ability of the model, but it is still necessary to employ the data of the target domain and uncover the relationship

between the source and the target domains. Unfortunately, this requirement is not always met. Table 9 compares and summarizes DA-based FAS in terms of the main idea, backbone, datasets, attack types, advantages, and disadvantages.

Table 9. Summary of FAS methods based on domain adaptation.

Ref.	Main Idea	Backbone	Datasets	Attacks				Advantages	Disadvantages
				2D-P	2D-R	3D-M	OA		
[115]	MMD, Kernel Hilbert Space, and kernel mean matching (KMM)	AlexNet	Replay-Attack CASIA-MFSD MSU-MFSD ROSE-Youtu	√ √ √ √	√ √ √ √	√		Reduces the statistical distance between domains and has strong interpretability	MMD distance can not fully reflect the differences among domains.
[116]	Knowledge distillation and MMD	AlexNet	CASIA-MFSD Replay-Attack MSU-MFSD MSU-USSA ROSE-Youtu	√ √ √ √ √	√ √ √ √ √	√			
[117]	Adversarial training, disentangled representation learning	ResNet18	Replay-Attack CASIA-MFSD MSU-MFSD ROSE-Youtu Oulu-NPU CASIA-SUR	√ √ √ √ √ √	√ √ √ √ √	√		Uses domain-independent features for classification	The decoupling strategy is simple and does not combine with prior knowledge.
[118]	Marginal and conditional distribution alignment	ResNet18	CASIA-MFSD Replay-Attack MSU-MFSD CASIA-SURF	√ √ √ √	√ √ √			Flexibly switches between unsupervised and semi-supervised.	Conditional distribution is relatively simple.
[119]	Deep clustering and DA	MobileNetV3	Replay-Attack MSU-MFSD Replay-Mobile OULU-NPU	√ √ √ √	√ √ √ √			Combined with deep depth, the distinguishing ability of feature representation is improved.	The model requires a large amount of calculation.
[120]	Directly stylizes the target data in the source-domain style via GAN translation.	GAN CNN	CASIA-MFSD Replay-Attack MSU-MFSD MSU-MFSD	√ √ √ √	√ √ √ √			Improving the generalization by feature alignment from source and target domains using GAN	Self-generating adversarial samples takes a lot of time.
[121]	Total pairwise confusion loss and fast domain adaption	CNN VGG-16	CASIA-MFSD Replay-Attack OULU-NPU SiW WMCA	√ √ √ √ √	√ √ √ √ √	√	√	Jointly addresses face recognition and face anti-spoofing in a mutually boosting way.	The multi-task training slightly decreases the interest performance of face anti-spoofing.
[122]	Meta-learning to achieve self-domain adaptation without a large amount of labeled data	CNN	MSU-MFSD CASIA-MFSD Replay-Attack OULU-NPU	√ √ √ √	√ √ √ √			Unsupervised meta-learning reduces reliance on annotated data, lowering annotation costs and time.	Depending on differences between source and target domains
[123]	Semi-supervised transfer, the unlabeled data with reliable pseudo labels, temporal consistency	CNN ResNet18	MSU-MFSD CASIA-MFSD Replay-Attack OULU-NPU MSU-MFSD	√ √ √ √ √	√ √ √ √ √			Only a few pieces of labeled training data.	Noise or inconsistency may affect the quality of pseudo labels.
[126]	Source-free domain adaptation, source-oriented regularization, and contrastive domain alignment	DeiT-S [127]	CASIA-MFSD Replay-Attack OULU-NPU Celebi-Spoof 3DMAD HKBU-M V2 CASIA-SURF 3D	√ √ √ √	√ √ √ √	√ √ √ √		It enables effective domain adaptation without access to source domain data through generalized pretraining.	When the number of real or fake faces is extremely imbalanced, this work might not be robust enough.

B. Domain generalization

Domain generalization (DG) [24,128] aims to train a model on source domains that can generalize to unseen target domains, without access to target domain data during training. Unlike DA, where target domain data is available, DG focuses on learning a model that minimizes prediction error in the unknown target domain. DG is more challenging because target domain data is unavailable and the model must be robust against out-of-distribution scenarios. Most DG research focuses on domain-invariant representation learning, using techniques like feature alignment, style alignment, gradient alignment, and meta-learning to identify representations that are effective across different domains. The schematic diagram of the model is illustrated in Figure 10b.

Feature alignment: Shao et al. [129] developed generalized feature extractors for multiple source domains through adversarial training, supplemented by auxiliary deep supervision to enhance generalization. They applied dual-force triple mining constraints to achieve compact within-class and dispersed between-class results. Jia et al. [130] argued that narrowing the distribution of attacks across domains might result in the neglect of domain-specific information. Consequently, they used adversarial learning and inequality three-tuple loss to treat attacks and live faces as separate categories. Wang et al. [131] utilized GANs to transfer images from RGB to the depth domain, addressing challenges like lighting and device variations. Liu et al. [132] introduced adaptive normalized representation learning (ANRL) to select domain-agnostic features adaptively, improving generalization with dual calibrated constraints (DCC) to aggregate the multiple source samples of the

same class. Chen et al. [133] focused on camera-invariant spoofing features through high-frequency feature decomposition and enhanced discriminative capability by combining reconstructed images with high and low-frequency information. Wang et al. [134] applied consistency regularization to improve generalization across different environments, thereby reducing overfitting. Liao et al. [135] introduced a domain-invariant ViT (DiVT) for FAS that used a concentration loss to learn domain-invariant representations and a separation loss to distinguish attack types across domains. However, ViT tends to capture noisy features from background objects and other irrelevant details. Liu et al. [136] developed CFPL-FAS, which employed class-free prompt learning (CFPL) to learn different semantic prompts based on content and style features separately, with the aim of enhancing the model's performance on unseen domains. Fang et al. [137] proposed VL-FAS, using fine-grained natural language descriptions and a sample-level visual text optimization (SLVT) module for improving domain generalization, though it did not fully leverage language supervision. Zhang et al. [138] introduced TF-FAS, which combined material and depth features with a semantic guidance mechanism for better cross-domain generalization.

Style alignment: Wang et al. [139] introduced the shuffled style assembly network (SSAN) to separate content and style in feature representations, creating a stylized feature space. They also built a large-scale FAS benchmark to further evaluate algorithm performance. Zhou et al. [140] proposed an instance-aware domain generalization framework that reduced sensitivity to instance-specific styles, aligning features without domain labels. Their method also included a dynamic kernel generator and categorical style assembly to improve style insensitivity. Zhou et al. [141] presented the TTDG framework, which enhanced model generalization by using test data. The test-time style projection (TTSP) module mapped unseen data styles into the training domain, and the diverse style shifts simulation (DSSS) module simulated various style changes, all without requiring the model to update at test time. This method worked with both CNN and ViT backbones.

Gradient alignment: Sun et al. [142] proposed SA-FAS, which encourages domain separability while aligning the live-to-spoof transition across all domains. This method learned domain-variant features but used a domain-invariant classifier, proving its effectiveness. Le and Woo [143] introduced GAC-FAS, using gradient alignment to encourage the model to converge to an optimal flat minimum for improving the model generalization without additional learning modules. GAC-FAS aligned gradient updates with empirical risk minimization (ERM) at key points in each domain.

In addition, Cai et al. [144] developed a hierarchical fusion module that combined RGB images and meta-pattern (MP) features across multiple levels to enhance performance and generalization for unknown domains. Zheng et al. [145] proposed FAMIM, a framework using frequency domain features and self-supervised learning to boost cross-domain generalization in FAS. Liu et al. [146] introduced UDGFAS, an unsupervised domain generalization framework that utilized identity-agnostic representations and nearest neighbor search to learn generalized domain-invariant features.

Overall, these studies improved the generalization performance of FAS by employing innovative techniques such as GANs, style-guided DA, and meta-learning. Table 10 compares and summarizes FAS based on domain generalization, outlining the main idea, backbone, advantages, and disadvantages. At present, the four widely used datasets are OULU-NPU (O), CASIA-MFSD (C), Idiap Replay Attack (I), and MSU-MFSD (M). Generally, each dataset is regarded as one domain, and the leave-one-out test protocol is applied to evaluate their cross-domain generalization. Specifically, O&C&I→M refers to the protocol that trains on OUlU-NPU, CASIA-MFSD, and Idiap Replay Attack, and tests on MSU-MFSD. O&M&I→C,O&C&M→I, and I&C&M→O are defined in a similar fashion.

Table 11 compares state-of-the-art DG methods on four testing domains. In Table 11, We adopt the Half Total Error Rate (HTER), Area Under Curve (AUC) as evaluation matrix.

Table 10. Summary of FAS methods based on domain generalization.

Ref.	Main Idea	Backbone	Advantages	Disadvantages
[129]	Using adversarial training, deep supervision and dual-force triple-mining constraints	DepthNet	Extract domain-independent discriminative features	Difficulty in fitting multiple discriminators
[130]	Leveraging single-side adversarial learning and asymmetric triplet loss for enhancing the generalization ability	ResNet18	Asymmetric processing improves the generalization performance	Asymmetric design needs to be improved
[131]	Transferring an input face image from the RGB domain to the depth domain	DenseNet	Alleviates the difficulties caused by appearance variations in lighting, image quality, and image capturing device.	Ignoring the temporal information may not work well for 3D spoofing attaches such as silicone masks.
[132]	To utilize AFNM to adaptively fuse normalized features from BN and IN and use DCC to aggregate the multiple source samples of the same class	CNN	Obtains domain-agnostic and discriminative representations for FAS.	To evaluate the method with extremely limited source domains.
[133]	Camera-invariant features and reconstructing images by low-frequency and high-frequency.	CNN	The camera-invariant feature learning enables the model to maintain stable performance under different camera devices.	Training and deployment costs are high.
[134]	Consistency regularization	CDCN	Consistency regularization ensures the consistency of feature representations.	It heavily relies on consistency under specific conditions, such as viewpoints and lighting.
[135]	A domain-invariant ViT, a concentration loss, and a separation loss	ViT	ViT reduces differences between domains and improves detection accuracy and robustness.	ViT has large data requirements and high computational complexity in visible light.
[136]	Using CFPL to learn semantic prompts based on content and style features	ViT	Learns more generalizable feature representations without relying on class labels.	It relies on a large amount of data.
[137]	Vision-language model (VLM) and a sample-level visual text optimization model	ViT-B	The method improves the precision and discriminative capability of feature extraction.	VLM may require substantial computational resources
[138]	Combining twofold element fine-grained semantic guidance: material and depth features	ViT-B/16	Strong generalization capability.	The dual feature and semantic mechanisms increase the model's complexity.
[139]	Domain generalization through shuffled style assembly and recombination	DepthNet ResNet-18	Through domain style alignment, the negative impact of style differences across various domains is effectively mitigated. The model can maintain high robustness and generalization when facing different environmental conditions, such as camera devices and lighting variations.	Due to environmental changes such as variations in lighting, camera quality, and angles, style alignment may be inaccurate. This can cause the model to struggle in learning domain-invariant features effectively, impacting its generalization capability.
[140]	Learns the generalizable feature by weakening the features' sensitivity to instance-specific styles	CNN		
[141]	Test-time style projection (TTSL) and diverse style shifts simulation (DSSS)	ResNet18 ViT		
[142]	Encouraging domain separability while aligning the live-to-spoof transition across all domains.	ResNet-18	This is achieved by aligning gradient updates or loss functions to enhance the model's robustness across different domains.	The effectiveness is limited when facing highly diverse data domains, and significant tuning may be required to achieve optimal performance.
[143]	By aligning gradient updates for the model converging to an optimal flat minimum	ResNet-18		
[144]	Designs a hierarchical fusion network to extract meta-features using meta-learning.	ResNet50	Integrating RGB images and MP information has advanced the research on hybrid methods.	The meta-model extractor needs optimization.
[145]	Introduces frequency-domain augmentation and a self-supervised learning	ViT-Tiny	--	Insufficient experimental validation.
[146]	Identity-agnostic representations, in-domain nearest neighbors	ResNet50	Fully unsupervised	Limited sample size may affect the model's adaptability to complex scenarios.

Table 11. Comparison with the state-of-the-art FAS methods on four testing domains:'↓' and '↑' indicate that superior performance is associated with a lower HTER and a higher AUC.

Ref.	O&C&I→M		O&M&I→C		O&C&M→I		I&C&M→O	
	HTER↓	AUC↑	HTER↓	AUC↑	HTER↓	AUC↑	HTER↓	AUC↑
[129]	27.98	80.02	22.19	84.99	17.69	88.06	24.50	84.51
SSDG-M [130]	25.17	81.83	18.21	94.61	16.67	90.47	23.11	85.45
SSDG-R [132]	15.61	91.54	11.71	96.59	7.38	97.17	10.44	95.94
[131]	18.26	89.40	21.43	88.81	19.40	86.87	22.03	87.71
[132]	15.67	91.90	16.03	91.04	10.83	96.75	17.85	89.26
[134]	14.4	93.3	13.5	96.0	10.0	95.5	17.8	89.8
[135]	13.06	94.04	3.71	99.29	2.86	99.14	8.67	96.92
[136]	2.50	99.42	5.43	98.41	1.43	99.28	2.56	99.10
[137]	7.92	97.05	5.00	98.90	3.13	99.31	4.00	98.64
[138]	3.44	99.42	0.81	99.92	2.24	99.67	2.26	99.48
[139]	13.72	93.63	8.88	96.79	6.67	98.75	10.00	96.67
[140]	8.86	97.14	10.62	94.50	5.41	98.19	8.70	96.40
[141]	10.00	96.15	6.50	97.78	7.91	96.83	8.14	96.49
[142]	5.95	96.55	8.78	95.37	6.58	97.54	10.00	96.23
[143]	8.60	97.16	4.29	98.87	5.00	97.56	8.20	95.16
[144]	5.24	97.28	9.11	96.09	15.35	90.67	12.40	94.26
[145]	5.71	97.71	9.89	95.60	8.42	96.05	12.66	94.40
[146]	7.14	97.31	11.44	95.59	6.28	98.61	12.18	94.36

C. Generalization Methods on Unknown attacks

Most previous DL-based methods treat FAS as a closed-set problem, concentrating on predefined spoofing attacks. However, it is unrealistic to cover all potential attacks within the dataset, leading to models that overfit on known attacks and remain vulnerable to unknown ones. Recently, researchers have shifted their focus towards making FAS models more robust to unknown attacks, with zero-shot/few-shot learning and anomaly detection being two popular approaches.

Methods Based on Zero-shot/Few-shot:

Zero-shot/few-shot learning, in theory, originate from DA, and methodologically fall under representation learning or meta-learning. Many existing approaches use spoofing samples of the same type during both training and testing, such as print attacks or replay attacks. However, these methods may perform poorly when encountering novel attacks that were not seen during training. Therefore, it is necessary to develop network models that can leverage previously learned knowledge to handle unknown attacks. To address this challenge, zero-shot/few-shot learning have emerged as important solutions. The key difference between the two lies in data availability. Zero-shot learning focuses solely on learning from existing spoofing types without access to any samples of novel attacks, whereas few-shot learning involves a small number of samples from the novel attack types.

Liu Y et al. [28] introduced the FAS database, which included 13 types of spoof attacks, and studied the zero-shot FAS problem. They proposed an unsupervised deep tree network (DTN) to classify unknown attacks. Pérez-Cabo et al. [147] proposed a meta-learning framework following the continual few-shot learning paradigm to address the problem of catastrophic forgetting, while enabling continuous model learning when faced with sequentially arriving novel attack data. Qin et al. [148] defined FAS as a zero-shot/few-shot problem and introduced an adaptive method that detected unknown attacks by learning from real and fake faces and new attack samples. Yang et al. [149] proposed a few-shot domain expansion strategy that aligned the source and target domains in the semantic space in an unsupervised manner, enabling the model to perform well in their joint expansion domain. George et al. [150] demonstrated that fine-tuned ViT models performed well on unknown attacks and cross-dataset evaluation. Nguyen et al. [151] presented a self-attention adversarial learning framework (ADGN) to create a generic yet distinct latent space (GDL-space) for improving generalization against rare, and unseen attacks. Huang et al. [152] introduced ensemble adapters and feature-wise transformation layers

in ViT for robust performance with few samples. Table 12 summarizes the FAS methods based on zero-shot/few-shot learning.

Table 12. Overview of generalization methods based on zero-shot/few-shot.

Ref.	Main Idea	Backbone	Datasets	Attacks				Advantages	Disadvantages
				2D-P	2D-R	3D-M	OA		
[28]	Deep tree learning, zero-Shot learning, and multiple classification	CNN	Replay-Attack	√	√				
			CASIA-MFSD	√	√				
			MSU-MFSD	√	√				
			SiW-M	√	√	√	√		
[147]	Leverage meta-learning models to adaptively update in the face of new attacks.	ResNet	CASIA-MFSD	√	√				
			Replay-Attack	√	√				
			3DMAD			√			
			MSU-MFSD	√	√				
			Replay-Mobile	√	√				
			HKBU-M V2			√			
			Oulu-NPU	√	√				
			Rose-Youtu	√	√	√			
			SiW	√	√				
[148]	By training a meta-model, it can quickly adapt to new tasks and new environments and maintain good performance even with zero or few samples.	PR-Net	CSMAD			√		Zero-shot/ few-shot learning can use existing spoofing types without access to any samples of novel attacks or a small number of samples. Through technologies such as meta-learning or transfer learning, these methods can learn and infer, allowing for better adaptation to new or unknown attacks.	When attack samples are scarce or of poor quality, they may cause performance fluctuations. A small number of samples may limit feature expression capabilities. Zero-shot/few-shot learning (such as meta-learning or GAN) may increase the demand for computational resource and time requirements.
			CASIA-MFSD	√	√				
			MSU-MFSD	√	√				
			MSU-USSA	√	√				
			SiW	√	√				
			3DMAD			√			
			Oulu-NPU	√	√				
			Replay-Attack	√	√				
			CASIA-SURF	√					
[149]	Models can quickly scale and adapt to small amounts of new domain data	ResNet-18	Celeba-Spoof	√	√	√			
			MSU-MFSD	√	√				
			Oulu-NPU	√	√				
			SiW	√	√				
[150]	Using unsupervised pretraining and finetuning with a small amount of annotated data	ViT	WMCA	√	√	√	√		
			SiW-M	√	√	√	√		
[151]	Self-attention mechanism	Resnet-50	3DMAD			√			
			CSMAD			√			
			SiW-M	√	√	√	√		
[152]	Adaptive Transformer performs feature extraction and attention mechanism optimization through hierarchical structures.	ViT	CASIA-MFSD	√	√				
			Replay attack	√	√				
			MSU-MFSD	√	√				
			Oulu-NPU	√	√				
			CASIA-SURF	√					
			CASIA-CeFA	√	√	√			
			WMCA	√	√	√	√		

Methods based on anomaly detection:

Anomaly detection is designed to to identify abnormal samples that deviate from normality during testing. In face anti-spoofing, live faces are considered normal, while attacks are abnormal. Unlike the traditional two-class classification, anomaly detection usually employs a one-class classification, using only real faces for training, as attack types are often unknown in practical scenarios.

Pérez-Cabo et al. [153] introduced a deep metric learning model with triplet loss and a few-shot a posteriori probability estimation for anomaly detection, improving feature representations. Fatemifar et al. [154] extended anomaly detection by incorporating client-specific thresholds and utilizing a one-class classifier based on Gaussian distribution based statistical hypothesis to enhance performance. Baweja et al. [155] proposed an end-to-end anomaly detection model that introduced a pseudo-negative class using a Gaussian distribution. However, it faced issues with poor robustness. Similarly, George et al. [156] used a one-class classifier with multi-channel CNNs for anomaly detection, relying on normal face data. Fatemifar et al. [157] further developed a client-specific anomaly detection method, adapting classifiers to each user's data for better accuracy and generalization. Huang et al. [158] proposed OC-SCMNet, a one-class method that uses spoof cue maps (SCMs) to learn spoof features from live faces, aiming for zero responses for real faces while dynamically preserving spoof features in a memory bank.Besides, Jiang et al. [159] employed evidence theory to handle the confidence of model outputs, quantifying uncertainty and enabling the model to be more robust against unknown attacks. They also established a semantic consistency loss to ensure that feature representations remain consistent across different scenarios. Table 13 compares and summarizes generalization methods based on anomaly detection, detailing the main idea, backbone, datasets, attack types, advantages, and disadvantages.

Table 13. Overview of FAS based on anomaly detection.

Ref.	Main idea	Backbone	Datasets	Attacks				Advantages	Disadvantages
				2D-P	2D-R	3D-M	OA		
[153]	Anomaly detection, metric learning, and introducing a few-shot a posteriori probability estimation	ResNet-50	CASIA-MFSD	√	√			By learning normal facial features and detecting significant abnormal samples, the anomaly detection method maintains good performance in different datasets and scenarios. It can identify attack types that do not appear in the training set, and have strong adaptability and generalization capabilities.	It relies on high-quality normal samples and may be sensitive to environmental changes and the diversity of normal faces, resulting in increased false positive rates.
			Replay-Attack	√	√				
			MSU-MFSD	√	√				
			GRAD-GPAD [160]	√	√	√			
			UVAD		√				
[154]	A one-class classifier with Gaussian distribution based statistical hypothesis and client-specific extension of anomaly detection	VGG16	Replay-Attack	√	√				
			Replay-Mobile	√	√				
			Rose-Youtu	√	√	√			
			Replay-Attack	√	√				
[155]	Pseudo-negative class using Gaussian distribution and pairwise confusion loss	VGGFace	Rose-Youtu	√	√	√			
			Oulu-NPU	√	√				
			SiW	√	√				
			WMCA	√	√	√	√		
[156]	One-class representation and a Gaussian Mixture Model	MCCNN	SiW-M	√	√	√	√		
			MLFP			√			
			Replay-Mobile	√	√				
[157]	Using client-specific information in a one-class anomaly detection and representations derived from CNNs	VGG16	Replay-Attack	√	√				
			Rose-Youtu	√	√	√			
			CASIA-MFSD	√	√				
			Replay-Attack	√	√				
			MSU-MFSD	√	√			By not relying on spoof attacks, estimating SCMs and guiding feature learning enhance the detection capability for unseen attacks.	The design of SCMs guidance and the memory bank increases the complexity of the model.
[158]	One-class representation using SCMs for zero responses for real faces	CNN	Oulu-NPU	√	√				
			SiW	√	√				
			3DMAD			√			
			HKBU-MARs			√			
			CASIA-SURF	√					
			CASIA-MFSD	√	√				
			Replay-Attack	√	√				
[159]	A regularized evidential DL strategy and an entropy optimization-based semantic consistency learning strategy	ResNet-18	MSU-MFSD	√	√			It could generalize well to cross-scenario target domains with both known and unknown PAs with different types and quantities.	--
			Oulu-NPU	√	√				
			Rose-Youtu	√	√	√			
			HQ-WMCA	√	√	√	√		

3.2. FAS with Advanced Sensors

3.2.1. FAS Based on Multi-Modality

Multi-modality refers to information derived from different sensors and presented in different forms, including RGB images, depth images, and infrared images, with each being a modality. For the same object, multi-modality can provide heterogeneous information that is semantically related and complementary in content, allowing for features that would remain undetected through a single modality. Therefore, combining multiple modalities can effectively counteract forgery attacks in a single modality and enhance the detection capability of FAS.

Currently, FAS datasets are composed of RGB datasets, though Zhang et al. [161] released the first multi-modal large-scale dataset, CASIA-SURF, which included RGB images, depth images, and infrared images. In the Chalearn LAP multi-modal FAS attack detection challenge at CVPR2019 [162], Parkin et al. [163] used an existing face recognition network to detect multi-modal spoofing attacks, achieving first place on the CASIA-SURF dataset, while the method had a heavy parameter load. Shen et al. [164] extracted discriminative features using patch-based feature learning and applied multi-stream fusion with a modal feature erasing (MFE) layer to enhance performance, achieving second place with a score of 99.8052% on the above dataset. In addition to various modalities, racial differences will affect the ability to generalize FAS. To address this, CeFA [165] expanded CASIA-SURF with racial images from East Asia, Central Asia, and Africa, becoming the largest cross-racial multi-modal dataset currently. Yu et al. [166] extracted high-frequency features from face images in the CeFA dataset using CDCN and integrated information from different modalities, enhancing the model's capability in cross-racial scenarios. To address the challenge of high-quality masks in multi-modal methods, Yang et al. [167] proposed PipeNet, which selected the most suitable branch network structure for different modality images provided by CeFA, maximizing the utilization of multi-modal information.

Next, researchers have explored multi-modal algorithms for FAS using datasets like CASIA-SURF and CeFA. A. Liu et al. [168] proposed the cross-modal auxiliary (CMA) framework, which combined a modality translation network (MT-Net) and a modality assistance network (MA-Net), eliminating the need for additional modalities during testing.

W. Liu et al. [169] introduced a two-stage cascade framework that combined feature-level and decision-level fusion for improved accuracy and robustness. A. Liu et al. developed MA-ViT [170] and FM-ViT [171], integrating ViT with multi-modality to enhance generalization and adaptability, despite the fact that they had high computational costs. He et al. [172] proposed a lightweight multi-modal FAS using a patch-level feature extraction network, which reduced parameters while maintaining accuracy. Lin et al. [173] introduced a multi-modal domain generalization framework, which utilized an uncertainty-guided cross-adapter (U-Adapter) and the rebalanced modality gradient modulation (ReGrad) to address modality unreliability and imbalance. Yu et al. [174] combined local descriptors with a self-supervised pretraining model for multi-modal FAS. Antil et al. [175] used ViT with overlapping patches to fuse multi-modal features effectively. Table 14 summarizes multi-modal FAS methods, including main ideas, backbones, datasets, attack types, advantages, and disadvantages.

Table 14. Overview of FAS based on multi-modality.

Ref.	Main Idea	Backbone	Datasets	Attacks				Advantages	Disadvantages
				2D-P	2D-R	3D-M	OA		
[163]	Leveraging complementary information across modalities.	ResNet18 ResNet34 ResNet50	CASIA-SURF	√				High detection accuracy	Having a heavy parameter load
[164]	Bag-of-local-features and randomly erasing a certain modality	ResNext	CASIA-SURF	√				MFE enhances generalization	High computational overhead and training time.
[166]	High-frequency features and multi-modal features	CDCN	CFA	√	√	√		Improving the model's ability under different racial conditions	---
[167]	Using selective modality pipeline fusion network	CNN SENet154	CFA	√	√	√		Selective fusion reduces the impact of redundant information	The modality selection strategy may lead to instability in performance.
[168]	Adversarial cross-modality alignment	CycleGAN	CASIA-SURF CFA WMCA SiW OULU-NPU CASIA-MFSD Replay-Attack MSU-MFSD MIPD [169]	√ √ √ √ √ √ √ √ √	√ √ √ √ √ √	√	√	Adversarial cross-modality alignment enhances the model's generalization ability	The adversarial training process is complex and computationally expensive.
[169]	Nonlinearly fuses low-level and high-level features from different modalities.	ResNet50						Good performance against multi-material 3D mask spoofing	2D spoofing detection may be subpar in complex scenarios.
[170]	Paying attention to the modal independence	ViT	CASIA-SURF CASIA-SURF	√ √				Improving generalization for different modal data	Higher computational complexity
[171]	Flexibly adjusting the structure for different modalities	ViT	MSU-MFSD CASIA-SURF MSU-MFSD	√ √ √	√ √			Improving adaptability to different modalities	Same as above
[172]	Patch-level images from multi-modal data and uses a lightweight network	LMFFNet [172]	CASIA-SURF Replay-Attack CASIA-MASD	√ √ √	√ √			Reducing parameters while maintaining accuracy	Weaker robustness to illumination changes and complex backgrounds
[173]	U-Adapter and ReGrad strategy to address modality unreliability and imbalance	ViT	CFA PADISI-Face CASIA-SURF	√ √ √	√ √	√ √	√	Enhancing the generalization by addressing the modality unreliability and imbalance	Needs to further adopt more efficient uncertainty estimation techniques.
[174]	A modality-asymmetric masked autoencoder, self-supervised without costly annotated labels	ViT	WMCA WMCA	√ √	√ √	√ √	√ √	Being robust under various missing-modality cases	Weaker generalization and discriminative capacity of this method
[175]	Uses overlapping patches, parameter sharing in the ViT and a hybrid feature block	ViT	CASIA-SURF CFA WMCA CASIA-SURF	√ √ √ √	√ √	√ √	√	Computationally efficient	Need further enhancement of generalization.

3.2.2. FAS Based on Smartphones

Compared with general-purpose FAS methods that focus on accuracy and generalization issues [15], FAS on smartphones places more emphasis on practical issues in real-world applications, such as security, usability, and practicality. Therefore, FAS on smartphones can utilize various sensors (e.g., camera, microphone, speaker, accelerometer, etc.) to detect spoof faces. Meanwhile, FAS on smartphones often requires participants to evaluate their systems in the real-world using prototypes they implement on smartphones.

It is also important to note that in current FAS research, most detection methods remain one-way, where the system passively receives input images or video streams and then performs liveness or spoof detection. In contrast, FAS methods on smartphones are often interactive and two-way in nature. That is, the system actively issues instructions or stimuli, such as playing sounds, emitting light, or gaze responses, and the user is required to produce a genuine response. This interaction enables more reliable spoof detection.

FAS based on acoustic: The Echoprint method [176] was the first to combine RGB visual and acoustic modalities for user authentication, emitting sound signals to capture 3D facial contours, yet it overlooked face anti-spoofing. Chen et al. [177] introduced Echoface, which utilized acoustic signals to differentiate real faces from flat forgeries with over 96% accuracy across varying conditions. EchoFAS [178] enhanced performance through the use of short-time Fourier transform (STFT), CNN, fast Fourier transform (FFT), and ViT for feature extraction. Zhang et al. [179] proposed SonarGuard, which combined ultrasonic micro-Doppler features and lip motion trajectories for robust FLD on mobile devices. Xu et al. [180] presented AFace, leveraging sound wave reflectivity for adaptable and easy-to-implement face anti-spoofing.

FAS based on optical: Studies like [181,182] used flash for FLD based on reflective properties, yet they faced a poor user experience. Farrukh et al. [183] proposed FaceRevelio, which utilized photometric stereo to recover 3D facial depth features for real/fake classification. Zhang et al. [184] introduced Aurora Guard (AG), leveraging light reflection to extract depth and material features and adding a light CAPTCHA (i.e., the random light parameters sequence) for enhancing security. Kim et al. [185] developed a LiDAR-based FAS algorithm that resisted light changes and replayed attacks by providing 3D depth data and using random light patterns as a security mechanism.

FAS based on sensors: Xu et al. [186] leveraged the penetrability, material sensitivity, and fine-grained sensing capability of millimeter wave (mmWave) to propose a FAS algorithm named mmFace. By capturing facial biometric features and structural characteristics from mmWave signals reflected off the face, mask attacks can be detected. Later, Xu et al. [187] extracted both the 3D geometry and inner biomaterial features of faces using a cost radio frequency identification(RFID) tag array, which could effectively identify 2D print/replay and 3D mask attacks with high accuracy.

On top of this, Zheng et al. [188] designed GazeGuard, which captured a user's gaze responses and corresponding periocular features using random dots to ensure accurate face authentication. Yu et al. [189] introduced Auto-FAS, based on neural architecture search (NAS), to find lightweight networks suitable for mobile-level FAS and real-time applications. Kong et al. [190] proposed M3FAS, a multi-modal mobile system that integrated visual and auditory modalities with a dual-branch neural network and multi-head training strategy, improving accuracy, generalization, and robustness. Table 15 compares FAS on smartphones in terms of spoofing clues, advantages, disadvantages, attack types, and hardware.

Table 15. Overview of FAS on smartphones: 2D represents print and replay attacks, while 3D represents 3D mask attacks and Adv represents physical adversarial attacks.

Methods	Ref.	Clues	Advantages	Disadvantages	Attacks	Special Hardware
FAS based on acoustic	[176]	3D geometric structure	Acoustic-based FAS can effectively detect spoofing attacks using readily available smartphone sensors without additional hardware.	The effectiveness of acoustic signals may be affected by environmental noise or interference.	2D, adv	No
	[177]	3D geometric structure			2D, adv	No
	[178]	RGB-based models and 3D geometric structure			2D, 3D	No
	[179]	3D geometric structure			2D, 3D	No
	[180]	3D geometric structure			2D, 3D	No

Table 15. *Cont.*

Methods	Ref.	Clues	Advantages	Disadvantages	Attacks	Special Hardware
FAS based on optical	[181]	Motion, 3D geometric structure	This can be performed using existing cameras and flashes, making it easy to implement	Strong dependence on lighting and poor user experience	2D, 3D, adv	No
	[182]	3D geometric structure			2D, 3D	No
	[183]	3D geometric structure	Simple and fast	Not detecting 3D attacks	2D	No
	[184]	The depth map and the material map	Simple, fast, yet effective	Needing an extra light source to generate the reflection frames	2D, 3D	No
	[185]	3D geometric structure	LiDAR is insensitive to variations in lighting.	Limited by the availability of LiDAR devices.	2D, 3D	LiDAR
FAS based on sensors	[186]	Material differences	Offering strong resistance to mask spoofing attacks	Relying on the widespread adoption and cost of millimeter-wave radar	2D, 3D, adv	mmWave radar
	[187]	Material differences	RFID device support, lower hardware requirements	Compatibility may be limited	2D, 3D, adv	RFID
Others	[188]	Motion, biometric	Eye movement patterns are highly individual-specific and difficult to forge	Need for high-precision eye-tracking equipment	2D, 3D, adv	No
	[189]	Neural architecture search (NAS) and pixel-wise binary supervision	Lightweight and fast	Not detecting 3D attacks	2D	No
	[190]	Combining visual and auditory modalities	Enhancing generalization capability	The computation is relatively complex.	2D	No

4. Datasets for Face Anti-Spoofing

The total amount of datasets, the richness of data types, the data collection equipment, the collection environment, and other factors will all affect the performance of FAS. Most of the reviewed datasets are publicly available, while a small number are not. When introducing different datasets, we mainly focus on the characteristics of the datasets, the year, the number of living faces or spoof faces, including the number of samples, types of attacks, and the influencing factors such as lighting, posture, and prosthetic face material considered during recording. Table 16 provides an overview and comparative analysis of FAS datasets in RGB modality. Table 17 provides an overview and comparative analysis of multimodal datasets.

Table 16. Overview of FAS datasets in the RGB modality: '#Sub.' is short for subjects, typically referring to participants in data collection, individuals, or forged facial entities. The fifth column's RGB refers to RGB images. In the seventh column labeled '#Live/Spoof', 'I' and 'V' denote 'images' and 'videos', respectively.

No.	Datasets	Year	Attack Types	Model	#Sub	#Live/Spoof	Setup
1	CASIA-MFSD [25]	2012	Print (flat, wrapped, cut), Replay (tablet)	RGB	50	150/450 (V)	Seven scenarios and three levels of image quality
2	Replay-Attack [191]	2012	Print (flat), Replay (tablet, phone)	RGB	50	200/1000 (V)	Lighting and holding
3	Kose and Dugelay [192]	2013	Print (flat), Replay (tablet, phone)	RGB	35	200/198 (I)	--
4	MSU-MFSD [68]	2015	Replay (monitor)	RGB	404	70/210(V)	Indoor scenario; two types of cameras
5	UVAD [193]	2015	Replay (monitor)	RGB	404	808/16,268 (V)	Different lighting, background and places in two sections
6	REPLAY-Mobile [194]	2016	Print (flat), Replay (monitor)	RGB	--	390/640 (V)	Five lighting conditions

Table 16. *Cont.*

No.	Datasets	Year	Attack Types	Model	#Sub	#Live/ Spoof	Setup
7	MSU-USSA [49]	2016	Print (flat), Replay (laptop, tablet, phone)	RGB	1140	1140/ 9120 (I)	Uncontrolled; two types of cameras
8	HKBU-MV2 [40]	2016	Mask (hard resin) from Thatsmyface and REAL-f	RGB	12	504/ 504 (V)	Seven cameras from stationary and mobile devices and six lighting settings
9	OULU-NPU [195]	2017	Print (flat), Replay (phone)	RGB	55	720/ 2880 (V)	Lighting & background in 3 section
10	SiW [196]	2018	Print (flat, wrapped), Replay (phone, tablet, monitor	RGB	165	1320/ 3300 (V)	Four sessions with variations of distance, pose, illumination and expression
11	Rose-Youtu [115]	2018	Print (flat), Replay (monitor, laptop), Mask (paper, crop-paper)	RGB	20	500/2850 (V)	Five front-facing phone camera; Five different illumination conditions
12	CelebA-Spoof [77]	2020	Print (flat, wrapped), Replay (monitor, tablet, phone), Mask (paper).	RGB	10,177	6384/ 469,153 (I)	Four illumination conditions; indoor & outdoor; rich annotations
13	CASIA-SURF 3DMask [87]	2020	Mask (mannequin with 3D print).	RGB	48	288/ 864(V)	High-quality identity-preserved; Three decorations and six environments
14	RECOD-MPAD [84]	2020	Print (flat), Replay (monitor).	RGB	45	450/ 1800 (V)	Outdoor environment and low-light & dynamic sessions
15	HiFiMask [27]	2021	Mask (transparent, plaster, resin).	RGB	75	13650/ 40,950 (V)	three mask decorations; seven recording devices; six lighting conditions (periodic/random); six scenes
16	SiW-M [28]	2022	Print (flat), Replay, Mask (hard resin, plastic, silicone, paper, Mannequin), Makeup (cosmetics, impersonation, Obfuscation), Partial (glasses, cut paper).	RGB	493	660/ 968 (V)	Indoor environment with pose, lighting and expression variations
17	SuHiFiMask [197]	2023	2D image, Video replay, 3D Mask with materials Resin, Plaster, Silicone, Paper	RGB	101	10195/ 10,195 (V)	Long distance using surveillance cameras, recording in three scenes, three lighting types, and four types of weather
18	WFAS [198]	2023	Print (newspaper, poster, photo, album, picture book, scan photo, packging, cloth), Display (phone, tablet, TV, computer), Mask, 3D Model (garage kit, doll, adult doll, waxwork).	RGB	469,920	529,571/ 853,729 (I)	Internet, unconstrained settings.
19	SynthASpoof [199]	2023	1 Print, 3 Replay.	RGB	25,000	25,000/ 78,800 (I&V)	--

Table 17. Overview of multimodal datasets: '#Sub.' is short for subjects, typically referring to participants in data collection, individuals, or forged facial entities. The fifth column's RGB, Depth, NIR, Thermal, and SSI refer to RGB images, depth images, near-infrared images and Thermal images, and Snapshot Spectral Imaging (SSI), respectively. In the seventh column '#Live/Spoof', 'I', 'V', and 'A' denotes 'images', 'videos', and 'acoustic signal', respectively.

No.	Datasets	Year	Attack Types	Modal	#Sub	#Live/ Spoof	Setup
1	3DMAD [200]	2014	Mask (paper, hard resin)	RGB Depth	17	170/ 85 (V)	Three sessions (2 weeks interval)
2	GUC-LiFFAD [201]	2015	Print (Inkjet paper, Laserjet paper), Replay (tablet)	Light field	80	1798/ 3028 (V)	Distance of 1.5∼2 m in constrained conditions
3	3DFS-DB [202]	2016	Mask (plastic)	RGB Depth	--	260/260 (V)	Head movement with rich angles
4	ERPA [203]	2017	Print (flat), Replay (monitor), Mask (resin, silicone)	RGB Depth, NIR, Thermal	5	Total 86 (V)	Subject positioned close (0.3∼0.5 m) to the two types of cameras
5	MLFP [204]	2017	Mask (latex, paper)	VIS, NIR, Thermal	10	150/1200 (V)	Indoor and outdoor with fixed and random backgrounds
6	CSMAD [205]	2018	Mask (custom silicone)	RGB, Depth, NIR, Thermal	14	104/ 159 (V + I)	Four lighting conditions

Table 17. *Cont.*

No.	Datasets	Year	Attack Types	Modal	#Sub	#Live/ Spoof	Setup
7	WMCA [206]	2019	Print (flat), Replay (tablet), Partial (glasses), Mask (plastic, silicone, and paper, Mannequin)	RGB, Depth, NIR, Therma	72	347/1332 (V)	Six sessions with different backgrounds and illumination; pulse data for bonafide recordings
8	CASIA-SURF [161]	2019	Print (flat, wrapped, cut)	RGB, Depth, NIR	1000	3000/ 18,000 (V)	Background removed; Randomly cut eyes, nose or mouth areas
9	CASIA-SURF CeFA [165]	2021	Print (flat, wrapped), Replay, Mask (3D print, silica gel)	RGB, Depth, NIR	1607	300/ 27,900 (V)	Three ethnicities; outdoor & indoor; decoration with wig and glasses
10	PADISI-Face [207]	2021	Print (flat), Replay (tablet, phone), Partial (glasses, funny eye), Mask (plastic, silicone, transparent, Mannequin	RGB, Depth, NIR, SWIR, Thermal	360	1105/ 924 (V)	Indoor, fixed green background, 60-frame sequence of 1984 × 1264 pixel images
11	HySpeFAS [208]	2024	Print attacks, 3D mask attacks, Medical masks, Makeup attacks	SSI	6760	Train: 520/ 3380 (I) VAL: 208/ 728 (I) Test: 1924	Hyperspectral images were expertly reconstructed from SSI images utilizing the TwIST [209]. Each hyperspectral image boasts 30 spectral channels, offering a rich and diverse spectral representation.
12	UniAttackData [210]	2024	Physical Attacks: Print (flat, wrapped), Replay, Mask (3D print, silica gel). Digital Attacks: Adv (6), DeepFake (6)	RGB, Depth, NIR	1800	29706 (V) (Live: 1800, Fake: 27,906)	--
13	Echoface-Spoof [190]	2024	Print (A4 paper), Replay (Pad Pro (2388 × 1668), iPad Air 3 (2224 × 1668))	RGB, Acoustic	--	250,000 (I, A) (82,715, 166,637)	Samsung Edge Note (2560 × 1440) Samsung Galaxy S9 (3264 × 2448) Samsung Galaxy S21 (4216 × 2371) Xiaomi Redmi7 (3264 × 2448)

5. Research Challenges and Future Research Directions

5.1. FAS-Based Generalization

Challenge: From the analysis of FAS methods in Section 4, it is evident that several effective methods have been proposed within the realm of generalization research. These include adversarial domain adaptation, generative domain adaptation, meta-learning, style alignment, gradient alignment, and various deep representation learning techniques. These methods have significantly improved the performance of FAS in cross-domain scenarios. Moreover, zero-shot learning, few-shot learning, and anomaly detection show promise in handling unknown attacks. These traditional domain adaptation methods typically require simultaneous access to source and target domain data. However, source domain data may not be shared in real-world scenarios due to privacy concerns, security concerns, or other reasons. Therefore, data privacy and security issues need further exploration and resolution in the field of FAS.

Future Research Directions: (1) Source-free domain adaptation (SFDA): SFDA has become a research hotspot in recent years, aiming to perform domain adaptation when source domain data is unavailable. SFDA methods adapt to the target domain using only pre-trained source models and target domain data. Liu et al. [126] combined domain-generalized pretraining with SFDA to enhance the cross-domain generalization ability of FAS models. Some studies have introduced federated learning [211] and prompt learning [136,212] into the DA scenario. For example, Liu et al. [136] introduced class-free prompt learning (CFPL) to improve the cross-domain generalization of FAS. Therefore, future research could explore methods integrating federated learning, prompt learning, and DG techniques for domain adaptation. (2) Multi-domain adaptation: Future research could explore adapting to multiple target domains and effectively transferring knowledge

simultaneously. Multi-domain adaptation can help models perform consistently across different devices and environments, improving their generalization capability.

5.2. Multimodal Domain Generalization

Challenge: ① Modality distribution differences: Data distribution varies significantly across modalities (e.g., RGB, depth, infrared), and differences in sensor quality and resolution can lead to unreliable multi-modal live/spoof feature extraction [213]. This may mislead other modalities through cross-modal fusion. ② Imbalance and scarcity of training Samples: The imbalance between genuine and attack samples is more pronounced in multimodal settings. Particularly for new attack types that are scarce, models often excessively rely on the dominant modality with the fastest convergence (i.e., fast modality), which hampers the full utilization of other slower modalities [214]. ③ Insufficient unified multimodal datasets: There are a limited number of datasets available for cross-modal research, especially exceptionally high-quality multimodal datasets, which restricts in-depth exploration in this area.

Future Research Directions: First, create richer multimodal datasets. Second, it is essential to further develop more effective multimodal domain generalization methods for addressing modality unreliability and imbalance. Finally, exploring more lightweight and efficient multimodal FAS algorithms (such as EfficientFormer-V2 [215]) aims to meet real-time detection and low-computation resource requirements, ensuring the safe and reliable application of cross-modal FAS technologies in practical scenarios.

5.3. FAS Based Unsupervised Domain Generalization

Challenge: Currently, the majority of FAS algorithms are based on supervised learning, primarily relying on labeled datasets to distinguish between genuine faces and attack samples. However, obtaining labeled data is labor-intensive and costly, leading to the well-known problem of data scarcity in FAS, which hinders practical applications. In recent years, some studies have begun to explore semi-supervised [118] or self-supervised [125,134,216,217] learning methods aimed at reducing dependence on fully supervised data by leveraging limited labeled data or self-generated pseudo-labels. The model's generalization capability may be poor if the self-supervised pretext tasks are not highly correlated with the FAS task. A significant challenge for few-shot weakly supervised learning is whether the small number of samples can effectively cover various attack types and ensure the model's generalization ability. However, these methods are based on pretext tasks, resulting in limited performance and an inability to alleviate actual domain shifts. Furthermore, unsupervised domain generalization (UDG) benchmarks for various scenarios, such as cross-attack types, have yet to be established. Therefore, unsupervised algorithms show great potential in reducing data labeling costs and improving cross-domain generalization capabilities.

Future Research Directions: UDG [218,219] aims to learn a model from an unlabeled set of source domains such that it can semantically cluster images in an unseen target domain. Recently, Qi et al. [220] applied the batch and instance normalization techniques to UDG where there were no labels in the training domains to acquire invariant and transferable features. Liu et al. [146] proposed the first unsupervised domain generalization framework for FAS. In addition, contrastive learning (such as MoCo [221] and SimCLR [222]) can enforce invariance to various augmentations. Therefore, combining contrastive learning with other unsupervised learning strategies to investigate UDG FAS algorithms is possible.

5.4. Timeliness Issues

Challenge: Firstly, most existing algorithms emphasize detection performance and generalization but often overlook the model complexity and computational resources, which limits their application on mobile or low-power devices. For example, competitions

and databases have been released to boost research progress for FAS [27,161,165]. These challenges have focused on specific problems that affect the accuracy of the methods, such as multimodal information [162], 3D mask attacks [27], and surveillance scenarios [197]. However, constraints have been imposed on the model's compactness and efficiency. Secondly, transformer models (such as ViT [135] and Swin-Transformer [223]) suffer from large model sizes and require substantial computational resources. However, increasing accuracy often means more complex models, which can reduce real-time processing capability. Thirdly, there is a trade-off between detection accuracy and speed, thereby making it a key challenge to balance high accuracy with efficient algorithms. Lastly, real-time detection systems must adapt to various practical scenarios, such as lighting conditions, camera quality, and background changes. Existing models often need to improve in response to these dynamic environmental changes.

Future Research Directions: Firstly, to enhance robustness and generalization in real-time detection, techniques like model compression and pruning can be employed to lightweight models should be designed and develop efficient FAS algorithms suitable for low-resource devices, thereby improving real-time performance. In recent years, several lightweight neural networks have been proposed, such as MobileNet-V2 [224], MobileViT [225], and EfficientFormer-V2. These networks will contribute to designing lightweight FAS algorithms. Secondly, online learning and adaptive systems should be researched to enable them to dynamically adjust in changing environments, thereby enhancing model performance and adaptability in complex scenarios.

5.5. Unified Detection

Many existing FRSs treat FLD, deepfake detection, and FR as separate tasks while they are related tasks. So far, only a few efforts have focused on this unified detection task. Deb et al. [226] proposed a unified approach to detect digital attacks (such as deepfakes and adversarial examples) and physical attacks (such as photo, video replay, and 3D mask attacks). Al-Refai et al. [227] designed a multitask ViT-based model that achieved high accuracy for both face presentation attack detection and face representation and matching (FRM). Fang et al. [210] collected a unified physical-digital attack dataset (called UniAttackData) and proposed a unified attack detection framework based on a vision-language model (VLM). They validated the superiority of this unified face attack detection method on the dataset. Yu et al. [228] established the first joint benchmark for spoofing and deepfakes detection. They proposed a new multimodal framework that combined rPPG signals and RGB facial images, achieving optimal detection performance. Shi et al. [229] proposed a new benchmark using a multi-attribute chain of thought (MA-COT) paradigm to evaluate the capabilities of Multi-Modal large language models (MLLMs) in FAS and forgery detection, thereby advancing the development of multimodal and joint detection tasks. The literature [228,229] demonstrated that joint training can significantly enhance generalization capability as spoofing and deepfake detection are highly related tasks. Therefore, joint detection is a novel topic that requires more attention. Future work should establish standard protocols for this task to promote the development of new models. In addition to benchmarks and protocols, extracting more the generalized features and intrinsic cues between these highly related tasks is important to further improve generalization capability.

6. Conclusions

We investigated the evolutionary advancements in FAS algorithms over the past decade, focusing on the significant achievements and scientific trends of DL-based FAS algorithms in the last five years. Early successful attempts focused on constructing efficient classifiers through robust, handcrafted features. Since 2014, there has been a shift from

traditional, handcrafted features to modern, data-driven neural network models. Around 2018, research shifted towards domain generalization, including robustness and generalization to unknown domains and attacks. According to statistical results from this review, research on generalization has achieved significant breakthroughs and has expanded to cover various types of attacks and datasets. However, current research still primarily focuses on single-modal (RGB) photo and replay video attacks. To successfully deploy FAS models in more complex scenarios, such as cross-domains, cross-modalities, and mobile device environments, further and deeper investigations are still required. Finally, we analyze the current issues and future research trends in face anti-spoofing algorithms, providing references for forensic research and development within the face community.

Author Contributions: Conceptualization, H.X. and S.Y.T.; methodology, S.Y.T. and F.Q.; formal analysis, Y.J.; writing—original draft preparation, H.X.; writing—review and editing, H.X., S.Y.T. and F.Q.; visualization, S.Y.T. and F.Q.; supervision, S.Y.T. and F.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Fundamental Research Grant Scheme under grant number FRGS/1/2022/ICT10/UKM02/2; in part by the Ministry of Education Malaysia and Natural Science Research Project of Anhui Educational Committee Key Project, grant number 2024AH051339; the Young and Middle-aged Teacher Development Action Project of Anhui Province—Outstanding Young Teacher Training Project, grant number YQYB2024064; and the Natural Science Research Project of Anhui Educational Committee Key Project, grant number 2023AH052100.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: We would like to acknowledge Mixed Reality and Pervasive Computing Lab for providing access to their facilities and technical support throughout this research.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

DL	Deep learning
CNN	Conventional neural network
GAN	Generative adversarial network
FR	Face recognition
FRS	Face recognition system
ViT	Vision Transformer
PA	Presentation attack
FAS	Face anti-spoofing
FLD	Face liveness detection
BCE	Binary Cross Entropy

References

1. Ibrahim, R.M.; Elkelany, M.; Ake, A.; El-Afifi, M.I. Trends in Biometric Authentication: A review. *Nile J. Commun. Comput. Sci.* **2023**, *6*, 63–75.
2. Korchenko, O.; Tereikovskyi, I.; Ziubina, R.; Tereikovska, L.; Korystin, O.; Tereikovskyi, O.; Karpinskyi, V. Modular Neural Network Model for Biometric Authentication of Personnel in Critical Infrastructure Facilities Based on Facial Images. *Appl. Sci.* **2025**, *15*, 2553. [CrossRef]
3. Pramanik, S.; Dahlan, H.A.B. Face age estimation using shortcut identity connection of convolutional neural network. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 514–521. [CrossRef]
4. Guo, J.; Zhao, Y.; Wang, H. Generalized spoof detection and incremental algorithm recognition for voice spoofing. *Appl. Sci.* **2023**, *13*, 7773. [CrossRef]

5. Minaee, S.; Abdolrashidi, A.; Su, H.; Bennamoun, M.; Zhang, D. Biometrics recognition using deep learning: A survey. *Artif. Intell. Rev.* **2023**, *56*, 8647–8695. [CrossRef]
6. Pecolt, S.; Błażejewski, A.; Królikowski, T.; Maciejewski, I.; Gierula, K.; Glowinski, S. Personal Identification Using Embedded Raspberry Pi-Based Face Recognition Systems. *Appl. Sci.* **2025**, *15*, 887. [CrossRef]
7. Jaber, A.G.; Muniyandi, R.C.; Usman, O.L.; Singh, H.K.R. A Hybrid Method of Enhancing Accuracy of Facial Recognition System Using Gabor Filter and Stacked Sparse Autoencoders Deep Neural Network. *Appl. Sci.* **2022**, *12*, 11052. [CrossRef]
8. Abdurrahim, S.H.; Samad, S.A.; Huddin, A.B. Review on the effects of age, gender, and race demographics on automatic face recognition. *Vis. Comput.* **2018**, *34*, 1617–1630. [CrossRef]
9. Wang, M.; Deng, W. Deep face recognition: A survey. *Neurocomputing* **2021**, *429*, 215–244. [CrossRef]
10. Kong, C.; Wang, S.; Li, H. Digital and physical face attacks: Reviewing and one step further. *APSIPA Trans. Signal Inf. Process.* **2022**, *12*, e25. [CrossRef]
11. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *2*, 2672–2680.
12. Alrawahneh, A.A.-M.; Abdullah, S.N.A.S.; Abdullah, S.N.H.S.; Kamarudin, N.H.; Taylor, S.K. Video authentication detection using deep learning: A systematic literature review. *Appl. Intell.* **2025**, *55*, 239. [CrossRef]
13. Lim, J.S.; Stofa, M.M.; Koo, S.M.; Zulkifley, M.A. Micro Expression Recognition: Multi-scale Approach to Automatic Emotion Recognition by using Spatial Pyramid Pooling Module. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*.
14. Dang, M.; Nguyen, T.N. Digital face manipulation creation and detection: A systematic review. *Electronics* **2023**, *12*, 3407. [CrossRef]
15. Yu, Z.; Qin, Y.; Li, X.; Zhao, C.; Lei, Z.; Zhao, G. Deep learning for face anti-spoofing: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 5609–5631. [CrossRef]
16. Raheem, E.A.; Ahmad, S.M.S.; Adnan, W.A.W. Insight on face liveness detection: A systematic literature review. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 5165–5175. [CrossRef]
17. Jia, S.; Guo, G.; Xu, Z. A survey on 3D mask presentation attack detection and countermeasures. *Pattern Recognit.* **2020**, *98*, 107032. [CrossRef]
18. Safaa El-Din, Y.; Moustafa, M.N.; Mahdi, H. Deep convolutional neural networks for face and iris presentation attack detection: Survey and case study. *IET Biom.* **2020**, *9*, 179–193. [CrossRef]
19. Abdullakutty, F.; Elyan, E.; Johnston, P. A review of state-of-the-art in Face Presentation Attack Detection: From early development to advanced deep learning and multi-modal fusion methods. *Inf. Fusion* **2021**, *75*, 55–69. [CrossRef]
20. Khairnar, S.C.; Gite, S.S.; Thepade, S.D. A Bibliometric Analysis of Face Presentation Attacks Based on Domain Adaptation. 2021. Available online: https://digitalcommons.unl.edu/libphilprac/5454/ (accessed on 20 February 2025).
21. Vakhshiteh, F.; Nickabadi, A.; Ramachandra, R. Adversarial attacks against face recognition: A comprehensive study. *IEEE Access* **2021**, *9*, 92735–92756. [CrossRef]
22. Sharma, D.; Selwal, A. A survey on face presentation attack detection mechanisms: Hitherto and future perspectives. *Multimed. Syst.* **2023**, *29*, 1527–1577. [CrossRef] [PubMed]
23. Zheng, Z.; Wang, Q.; Wang, C. Spoofing attacks and anti-spoofing methods for face authentication over smartphones. *IEEE Commun. Mag.* **2023**, *61*, 213–219. [CrossRef]
24. Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; Loy, C.C. Domain Generalization: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 4396–4415. [CrossRef]
25. Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S.Z. A face antispoofing database with diverse attacks. In Proceedings of the 2012 5th IAPR International Conference on Biometrics (ICB), New Delhi, India, 29 March–1 April 2012; pp. 26–31.
26. Di, W.; Hu, H.; Jain, A.K. Face Spoof Detection With Image Distortion Analysis. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 746–761. [CrossRef]
27. Liu, A.; Zhao, C.; Yu, Z.; Wan, J.; Su, A.; Liu, X.; Tan, Z.; Escalera, S.; Xing, J.; Liang, Y. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 2497–2507. [CrossRef]
28. Liu, Y.; Stehouwer, J.; Jourabloo, A.; Liu, X. Deep tree learning for zero-shot face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4680–4689.
29. Sharif, M.; Bhagavatula, S.; Bauer, L.; Reiter, M.K. A general framework for adversarial examples with objectives. *ACM Trans. Priv. Secur. (TOPS)* **2019**, *22*, 1–30. [CrossRef]
30. Yin, B.; Wang, W.; Yao, T.; Guo, J.; Kong, Z.; Ding, S.; Li, J.; Liu, C. Adv-makeup: A new imperceptible and transferable attack on face recognition. *arXiv* **2021**, arXiv:2105.03162. [CrossRef]
31. Komkov, S.; Petiushko, A. Advhat: Real-world adversarial attack on arcface face id system. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 819–826.
32. Wei, X.; Guo, Y.; Yu, J. Adversarial sticker: A stealthy attack method in the physical world. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 2711–2725. [CrossRef]

33. Bigun, J.; Fronthaler, H.; Kollreider, K. Assuring liveness in biometric identity authentication by real-time face tracking. In Proceedings of the 2004 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, CIHSPS, Venice, Italy, 21–22 July 2004; pp. 104–111.
34. Ali, A.; Deravi, F.; Hoque, S. Liveness Detection Using Gaze Collinearity. In Proceedings of the 2012 Third International Conference on Emerging Security Technologies, Lisbon, Portugal, 5–7 September 2012; pp. 62–65.
35. Pan, G.; Sun, L.; Wu, Z.; Lao, S. Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio De Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
36. Jiang-Wei, L. Eye blink detection based on multiple Gabor response waves. In Proceedings of the 2008 International Conference on Machine Learning and Cybernetics, Kunming, China, 12–15 July 2008; pp. 2852–2856.
37. Wang, L.; Ding, X.; Fang, C. Face live detection method based on physiological motion analysis. *Tsinghua Sci. Technol.* **2009**, *14*, 685–690. [CrossRef]
38. Kollreider, K.; Fronthaler, H.; Faraj, M.I.; Bigun, J. Real-Time Face Detection and Motion Analysis With Application in "Liveness" Assessment. *IEEE Trans. Inf. Forensics Secur.* **2007**, *2*, 548–558. [CrossRef]
39. Nowara, E.M.; Sabharwal, A.; Veeraraghavan, A. PPGSecure: Biometric Presentation Attack Detection Using Photopletysmograms. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 56–62.
40. Liu, S.; Yuen, P.C.; Zhang, S.; Zhao, G. 3D mask face anti-spoofing with remote photoplethysmography. In *Computer Vision–ECCV 2016, Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016, Proceedings, Part VII 14*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 85–100.
41. Heusch, G.; Marcel, S. Pulse-based Features for Face Presentation Attack Detection. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22–25 October 2018; pp. 1–8.
42. Wei, B.; Hong, L.; Nan, L.; Wei, J. A liveness detection method for face recognition based on optical flow field. In Proceedings of the 2009 International Conference on Image Analysis and Signal Processing, Kuala Lumpur, Malaysia, 18–19 November 2009; pp. 233–236.
43. Kollreider, K.; Fronthaler, H.; Bigun, J. Non-intrusive liveness detection by face images. *Image Vis. Comput.* **2009**, *27*, 233–244. [CrossRef]
44. Bharadwaj, S.; Dhamecha, T.I.; Vatsa, M.; Singh, R. Computationally Efficient Face Spoofing Detection with Motion Magnification. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 105–110.
45. Tirunagari, S.; Poh, N.; Windridge, D.; Iorliam, A.; Suki, N.; Ho, A.T.S. Detection of Face Spoofing Using Visual Dynamics. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 762–777. [CrossRef]
46. Maatta, J.; Hadid, A.; Pietikainen, M. Face spoofing detection from single images using micro-texture analysis. In Proceedings of the 2011 International Joint Conference on Biometrics (IJCB), Washington, DC, USA, 11–13 October 2011; pp. 1–7.
47. de Freitas Pereira, T.; Anjos, A.; De Martino, J.M.; Marcel, S. LBP − TOP Based Countermeasure against Face Spoofing Attacks. In *Computer Vision—ACCV 2012 Workshops*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 121–132.
48. Boulkenafet, Z.; Komulainen, J.; Hadid, A. Face Spoofing Detection Using Colour Texture Analysis. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1818–1830. [CrossRef]
49. Patel, K.; Han, H.; Jain, A.K. Secure Face Unlock: Spoof Detection on Smartphones. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 2268–2283. [CrossRef]
50. Boulkenafet, Z.; Komulainen, J.; Hadid, A. Face Anti-Spoofing using Speeded-Up Robust Features and Fisher Vector Encoding. *IEEE Signal Process. Lett.* **2016**, *24*, 141–145. [CrossRef]
51. Komulainen, J.; Hadid, A.; Pietikainen, M. Context based face anti-spoofing. In Proceedings of the 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 29 September–2 October 2013; pp. 1–8.
52. Tan, X.; Li, Y.; Liu, J.; Jiang, L. Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model. In *Computer Vision—ECCV 2010, Proceedings of the 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010, Proceedings, Part VI 11*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 504–517.
53. Galbally, J.; Marcel, S. Face Anti-spoofing Based on General Image Quality Assessment. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 1173–1178.
54. Sun, X.; Huang, L.; Liu, C. Multispectral face spoofing detection using VIS–NIR imaging correlation. *Int. J. Wavelets Multiresolution Inf. Process.* **2018**, *16*, 1840003. [CrossRef]
55. Steiner, H.; Kolb, A.; Jung, N. Reliable face anti-spoofing using multispectral SWIR imaging. In Proceedings of the 2016 International Conference on Biometrics (ICB), Halmstad, Sweden, 13–16 June 2016; pp. 1–8.
56. Kim, S.; Ban, Y.; Lee, S. Face Liveness Detection Using a Light Field Camera. *Sensors* **2014**, *14*, 22471–22499. [CrossRef]

57. Sepas-Moghaddam, A.; Malhadas, L.; Correia, P.L.; Pereira, F. Face spoofing detection using a light field imaging framework. *IET Biom.* **2018**, *7*, 39–48. [CrossRef]
58. Wang, Y.; Nian, F.; Li, T.; Meng, Z.; Wang, K. Robust face anti-spoofing with depth information. *J. Vis. Commun. Image Represent.* **2017**, *49*, 332–337. [CrossRef]
59. Yang, J.; Lei, Z.; Li, S.Z. Learn convolutional neural network for face anti-spoofing. *arXiv* **2014**, arXiv:1408.5601. [CrossRef]
60. Feng, L.; Po, L.-M.; Li, Y.; Xu, X.; Yuan, F.; Cheung, T.C.-H.; Cheung, K.-W. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *J. Vis. Commun. Image Represent.* **2016**, *38*, 451–460. [CrossRef]
61. Li, L.; Xia, Z.; Jiang, X.; Ma, Y.; Roli, F.; Feng, X. 3D face mask presentation attack detection based on intrinsic image analysis. *Iet Biom.* **2020**, *9*, 100–108. [CrossRef]
62. Li, L.; Xia, Z.; Wu, J.; Yang, L.; Han, H. Face presentation attack detection based on optical flow and texture analysis. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 1455–1467. [CrossRef]
63. Li, L.; Feng, X.; Boulkenafet, Z.; Xia, Z.; Li, M.; Hadid, A. An original face anti-spoofing approach using partial convolutional neural network. In Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016; pp. 1–6.
64. Asim, M.; Ming, Z.; Javed, M.Y. CNN based spatio-temporal feature extraction for face anti-spoofing. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; pp. 234–238.
65. Shao, R.; Lan, X.; Yuen, P.C. Joint discriminative learning of deep dynamic textures for 3D mask face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2018**, *14*, 923–938. [CrossRef]
66. Agarwal, A.; Vatsa, M.; Singh, R. CHIF: Convoluted histogram image features for detecting silicone mask based face presentation attack. In Proceedings of the 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), Tampa, FL, USA, 23–26 September 2019; pp. 1–5.
67. Liang, Y.-C.; Qiu, M.-X.; Lai, S.-H. FIQA-FAS: Face Image Quality Assessment Based Face Anti-Spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 1462–1470.
68. Khammari, M. Robust face anti-spoofing using CNN with LBP and WLD. *IET Image Process.* **2019**, *13*, 1880–1884. [CrossRef]
69. Yu, Z.; Li, X.; Wang, P.; Zhao, G. Transrppg: Remote photoplethysmography transformer for 3d mask face presentation attack detection. *IEEE Signal Process. Lett.* **2021**, *28*, 1290–1294. [CrossRef]
70. Li, L.; Xia, Z.; Hadid, A.; Jiang, X.; Zhang, H.; Feng, X. Replayed video attack detection based on motion blur analysis. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2246–2261. [CrossRef]
71. Chen, F.M.; Wen, C.; Xie, K.; Wen, F.Q.; Sheng, G.Q.; Tang, X.G. Face liveness detection: Fusing colour texture feature and deep feature. *IET Biom.* **2019**, *8*, 369–377. [CrossRef]
72. Chen, H.; Chen, Y.; Tian, X.; Jiang, R. A cascade face spoofing detector based on face anti-spoofing R-CNN and improved retinex LBP. *IEEE Access* **2019**, *7*, 170116–170133. [CrossRef]
73. Sharifi, O. Score-level-based face anti-spoofing system using handcrafted and deep learned characteristics. *Int. J. Image Graph. Signal Process.* **2019**, *14*, 15. [CrossRef]
74. Solomon, E.; Cios, K.J. FASS: Face anti-spoofing system using image quality features and deep learning. *Electronics* **2023**, *12*, 2199. [CrossRef]
75. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III 18*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
76. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
77. Zhang, Y.; Yin, Z.; Li, Y.; Yin, G.; Yan, J.; Shao, J.; Liu, Z. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In *Computer Vision–ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020, Proceedings, Part XII 16*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 70–85.
78. Lucena, O.; Junior, A.; Moia, V.; Souza, R.; Valle, E.; Lotufo, R. Transfer learning using convolutional neural networks for face anti-spoofing. In *Image Analysis and Recognition, Proceedings of the 14th International Conference, ICIAR 2017, Montreal, QC, Canada, 5–7 July 2017, Proceedings 14*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 27–34.
79. Nagpal, C.; Dubey, S.R. A performance evaluation of convolutional neural networks for face anti spoofing. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
80. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
81. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
82. Guo, J.; Zhu, X.; Xiao, J.; Lei, Z.; Wan, G.; Li, S.Z. Improving face anti-spoofing by 3D virtual synthesis. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–8.

83. Xu, X.; Xiong, Y.; Xia, W. On improving temporal consistency for online face liveness detection system. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 824–833.
84. Almeida, W.R.; Andaló, F.A.; Padilha, R.; Bertocco, G.; Dias, W.; Torres, R.d.S.; Wainer, J.; Rocha, A. Detecting face presentation attacks in mobile devices with a patch-based CNN and a sensor-aware loss function. *PLoS ONE* **2020**, *15*, e0238058. [CrossRef]
85. Wang, C.-Y.; Lu, Y.-D.; Yang, S.-T.; Lai, S.-H. 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, New Orleans, LA, USA, 18–24 June 2022; pp. 20281–20290.
86. Atoum, Y.; Liu, Y.; Jourabloo, A.; Liu, X. Face anti-spoofing using patch and depth-based CNNs. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 319–328.
87. Yu, Z.; Zhao, C.; Wang, Z.; Qin, Y.; Su, Z.; Li, X.; Zhou, F.; Zhao, G. Searching central difference convolutional networks for face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5295–5305.
88. Yu, Z.; Wan, J.; Qin, Y.; Li, X.; Li, S.Z.; Zhao, G. NAS-FAS: Static-dynamic central difference network search for face anti-spoofing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3005–3023. [CrossRef] [PubMed]
89. Zheng, W.; Yue, M.; Zhao, S.; Liu, S. Attention-based spatial-temporal multi-scale network for face anti-spoofing. *IEEE Trans. Biom. Behav. Identity Sci.* **2021**, *3*, 296–307. [CrossRef]
90. George, A.; Marcel, S. Deep pixel-wise binary supervision for face presentation attack detection. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–8.
91. Hossain, M.S.; Rupty, L.; Roy, K.; Hasan, M.; Sengupta, S.; Mohammed, N. A-DeepPixBis: Attentional angular margin for face anti-spoofing. In Proceedings of the 2020 Digital Image Computing: Techniques and Applications (DICTA), Melbourne, Australia, 29 November–2 December 2020; pp. 1–8.
92. Sun, W.; Song, Y.; Chen, C.; Huang, J.; Kot, A.C. Face spoofing detection based on local ternary label supervision in fully convolutional networks. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3181–3196. [CrossRef]
93. Yu, Z.; Li, X.; Niu, X.; Shi, J.; Zhao, G. Face anti-spoofing with human material perception. In *Computer Vision–ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020, Proceedings, Part VII 16*; Springer: Cham, Switzerland, 2020; pp. 557–575.
94. Kim, T.; Kim, Y.; Kim, I.; Kim, D. Basn: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019; pp. 494–503.
95. Zhang, X.; Ng, R.; Chen, Q. Single image reflection separation with perceptual losses. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4786–4794.
96. Yu, Z.; Li, X.; Shi, J.; Xia, Z.; Zhao, G. Revisiting pixel-wise supervision for face anti-spoofing. *IEEE Trans. Biom. Behav. Identity Sci.* **2021**, *3*, 285–295. [CrossRef]
97. Roy, K.; Hasan, M.; Rupty, L.; Hossain, M.S.; Sengupta, S.; Taus, S.N.; Mohammed, N. Bi-fpnfas: Bi-directional feature pyramid network for pixel-wise face anti-spoofing by leveraging fourier spectra. *Sensors* **2021**, *21*, 2799. [CrossRef]
98. Zhang, K.-Y.; Yao, T.; Zhang, J.; Tai, Y.; Ding, S.; Li, J.; Huang, F.; Song, H.; Ma, L. Face anti-spoofing via disentangled representation learning. In *Computer Vision–ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020, Proceedings, Part XIX 16*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 641–657.
99. Li, X.; Wan, J.; Jin, Y.; Liu, A.; Guo, G.; Li, S.Z. 3DPC-Net: 3D point cloud network for face anti-spoofing. In Proceedings of the 2020 IEEE International Joint Conference on Biometrics (IJCB), Houston, TX, USA, 28 September–1 October 2020; IEEE: Piscataway, NJ, USA; pp. 1–8.
100. Lin, C.; Liao, Z.; Zhou, P.; Hu, J.; Ni, B. Live Face Verification with Multiple Instantialized Local Homographic Parameterization. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 814–820.
101. Yang, X.; Luo, W.; Bao, L.; Gao, Y.; Gong, D.; Zheng, S.; Li, Z.; Liu, W. Face anti-spoofing: Model matters, so does data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3507–3516.
102. Wang, Z.; Yu, Z.; Zhao, C.; Zhu, X.; Qin, Y.; Zhou, Q.; Zhou, F.; Lei, Z. Deep spatial gradient and temporal depth learning for face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5042–5051.
103. Cai, R.; Li, H.; Wang, S.; Chen, C.; Kot, A.C. DRL-FAS: A novel framework based on deep reinforcement learning for face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2020**, *16*, 937–951. [CrossRef]
104. Wang, Z.; Wang, Q.; Deng, W.; Guo, G. Learning multi-granularity temporal characteristics for face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 1254–1269. [CrossRef]

105. Liu, W.; Pan, Y. Spatio-Temporal Based Action Face Anti-Spoofing Detection via Fusing Dynamics and Texture Face Keypoints Cues. *IEEE Trans. Consum. Electron.* **2024**, *70*, 2401–2413. [CrossRef]
106. Jourabloo, A.; Liu, Y.; Liu, X. Face de-spoofing: Anti-spoofing via noise modeling. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 290–306.
107. Liu, Y.; Stehouwer, J.; Liu, X. On disentangling spoof trace for generic face anti-spoofing. In *Computer Vision–ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020, Proceedings, Part XVIII 16*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 406–422.
108. Feng, H.; Hong, Z.; Yue, H.; Chen, Y.; Wang, K.; Han, J.; Liu, J.; Ding, E. Learning generalized spoof cues for face anti-spoofing. *arXiv* **2020**, arXiv:2005.03922. [CrossRef]
109. Wu, H.; Zeng, D.; Hu, Y.; Shi, H.; Mei, T. Dual spoof disentanglement generation for face anti-spoofing with depth uncertainty learning. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 4626–4638. [CrossRef]
110. Wang, Y.-C.; Wang, C.-Y.; Lai, S.-H. Disentangled representation with dual-stage feature learning for face anti-spoofing. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 1–6 January 2024; pp. 1955–1964.
111. Ge, X.; Liu, X.; Yu, Z.; Shi, J.; Qi, C.; Li, J.; Kälviäinen, H. DiffFAS: Face Anti-Spoofing via Generative Diffusion Models. *arXiv* **2024**, arXiv:2409.08572. [CrossRef]
112. Yu, Y.; Du, Z.; Luo, H.; Xiao, C.; Hu, J. Fourier-Based Frequency Space Disentanglement and Augmentation for Generalizable Face Anti-Spoofing. *IEEE J. Biomed. Health Inform.* 2024; *in press*. [CrossRef]
113. Yang, Q.; Pan, S.J. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
114. HassanPour Zonoozi, M.; Seydi, V. A survey on adversarial domain adaptation. *Neural Process. Lett.* **2023**, *55*, 2429–2469. [CrossRef]
115. Li, H.; Li, W.; Cao, H.; Wang, S.; Huang, F.; Kot, A.C. Unsupervised domain adaptation for face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 1794–1809. [CrossRef]
116. Li, H.; Wang, S.; He, P.; Rocha, A. Face anti-spoofing with deep neural network distillation. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 933–946. [CrossRef]
117. Wang, G.; Han, H.; Shan, S.; Chen, X. Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. *IEEE Trans. Inf. Forensics Secur.* **2020**, *16*, 56–69. [CrossRef]
118. Jia, Y.; Zhang, J.; Shan, S.; Chen, X. Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing. *Pattern Recognit.* **2021**, *115*, 107888. [CrossRef]
119. El-Din, Y.S.; Moustafa, M.N.; Mahdi, H. Adversarial unsupervised domain adaptation guided with deep clustering for face presentation attack detection. *arXiv* **2021**, arXiv:2102.06864. [CrossRef]
120. Zhou, Q.; Zhang, K.-Y.; Yao, T.; Yi, R.; Sheng, K.; Ding, S.; Ma, L. Generative domain adaptation for face anti-spoofing. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 335–356.
121. Tu, X.; Ma, Z.; Zhao, J.; Du, G.; Xie, M.; Feng, J. Learning generalizable and identity-discriminative representations for face anti-spoofing. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 1–19. [CrossRef]
122. Wang, J.; Zhang, J.; Bian, Y.; Cai, Y.; Wang, C.; Pu, S. Self-domain adaptation for face anti-spoofing. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 19–21 May 2021; pp. 2746–2754.
123. Quan, R.; Wu, Y.; Yu, X.; Yang, Y. Progressive transfer learning for face anti-spoofing. *IEEE Trans. Image Process.* **2021**, *30*, 3946–3955. [CrossRef] [PubMed]
124. Li, J.; Yu, Z.; Du, Z.; Zhu, L.; Shen, H.T. A comprehensive survey on source-free domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 5743–5762. [CrossRef]
125. Liu, Y.; Chen, Y.; Dai, W.; Gou, M.; Huang, C.-T.; Xiong, H. Source-Free Domain Adaptation with Contrastive Domain Alignment and Self-Supervised Exploration for Face Anti-Spoofing. Springer Nature: Cham, Switzerland; pp. 511–528.
126. Liu, Y.; Chen, Y.; Dai, W.; Gou, M.; Huang, C.-T.; Xiong, H. Source-free domain adaptation with domain generalized pretraining for face anti-spoofing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 5430–5448. [CrossRef]
127. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; J'egou, H.E. Training data-efficient image transformers & distillation through attention. In Proceedings of the International Conference on Machine Learning, Virtual Event, 13–18 July 2020.
128. Wang, J.; Lan, C.; Liu, C.; Ouyang, Y.; Qin, T.; Lu, W.; Chen, Y.; Zeng, W.; Philip, S.Y. Generalizing to unseen domains: A survey on domain generalization. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 8052–8072. [CrossRef]
129. Shao, R.; Lan, X.; Li, J.; Yuen, P.C. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10023–10031.
130. Jia, Y.; Zhang, J.; Shan, S.; Chen, X. Single-side domain generalization for face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2019; pp. 8484–8493.

131. Wang, Y.; Song, X.; Xu, T.; Feng, Z.; Wu, X.-J. From RGB to depth: Domain transfer network for face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 4280–4290. [CrossRef]
132. Liu, S.; Zhang, K.-Y.; Yao, T.; Bi, M.; Ding, S.; Li, J.; Huang, F.; Ma, L. Adaptive normalized representation learning for generalizable face anti-spoofing. In Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 20–24 October 2021; pp. 1469–1477.
133. Chen, B.; Yang, W.; Li, H.; Wang, S.; Kwong, S. Camera invariant feature learning for generalized face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 2477–2492. [CrossRef]
134. Wang, Z.; Yu, Z.; Wang, X.; Qin, Y.; Li, J.; Zhao, C.; Liu, X.; Lei, Z. Consistency regularization for deep face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 1127–1140. [CrossRef]
135. Liao, C.-H.; Chen, W.-C.; Liu, H.-T.; Yeh, Y.-R.; Hu, M.-C.; Chen, C.-S. Domain invariant vision transformer learning for face anti-spoofing. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 6098–6107.
136. Liu, A.; Xue, S.; Gan, J.; Wan, J.; Liang, Y.; Deng, J.; Escalera, S.; Lei, Z. CFPL-FAS: Class Free Prompt Learning for Generalizable Face Anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–23 June 2023; pp. 222–232.
137. Fang, H.; Liu, A.; Jiang, N.; Lu, Q.; Zhao, G.; Wan, J. VL-FAS: Domain Generalization via Vision-Language Model For Face Anti-Spoofing. In Proceedings of the ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 4770–4774.
138. Wang, X.; Zhang, K.-Y.; Yao, T.; Zhou, Q.; Ding, S.; Dai, P.; Ji, R. TF-FAS: Twofold-element fine-grained semantic guidance for generalizable face anti-spoofing. In Proceedings of the European Conference on Computer Vision, Paris, France, 26–27 March 2025; pp. 148–168.
139. Wang, Z.; Wang, Z.; Yu, Z.; Deng, W.; Li, J.; Gao, T.; Wang, Z. Domain generalization via shuffled style assembly for face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4123–4133.
140. Zhou, Q.; Zhang, K.-Y.; Yao, T.; Lu, X.; Yi, R.; Ding, S.; Ma, L. Instance-aware domain generalization for face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–23 June 2023; pp. 20453–20463.
141. Zhou, Q.; Zhang, K.-Y.; Yao, T.; Lu, X.; Ding, S.; Ma, L. Test-time domain generalization for face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–23 June 2023; pp. 175–187.
142. Sun, Y.; Liu, Y.; Liu, X.; Li, Y.; Chu, W.-S. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–23 June 2023; pp. 24563–24574.
143. Le, B.M.; Woo, S.S. Gradient alignment for cross-domain face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–23 June 2023; pp. 188–199.
144. Cai, R.; Li, Z.; Wan, R.; Li, H.; Hu, Y.; Kot, A.C. Learning meta pattern for face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 1201–1213. [CrossRef]
145. Zheng, T.; Yu, Q.; Chen, Z.; Wang, J. FAMIM: A Novel Frequency-Domain Augmentation Masked Image Model Framework for Domain Generalizable Face Anti-Spoofing. In Proceedings of the ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 4470–4474.
146. Liu, Y.; Chen, Y.; Gou, M.; Huang, C.-T.; Wang, Y.; Dai, W.; Xiong, H. Towards unsupervised domain generalization for face anti-spoofing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 20654–20664.
147. Pérez-Cabo, D.; Jiménez-Cabello, D.; Costa-Pazo, A.; López-Sastre, R.J. Learning to learn face-pad: A lifelong learning approach. In Proceedings of the 2020 IEEE International Joint Conference on Biometrics (IJCB), Houston, TX, USA, 28 September–1 October 2020; pp. 1–9.
148. Qin, Y.; Zhao, C.; Zhu, X.; Wang, Z.; Yu, Z.; Fu, T.; Zhou, F.; Shi, J.; Lei, Z. Learning meta model for zero-and few-shot face anti-spoofing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 11916–11923.
149. Yang, B.; Zhang, J.; Yin, Z.; Shao, J. Few-shot domain expansion for face anti-spoofing. *arXiv* **2021**, arXiv:2106.14162. [CrossRef]
150. George, A.; Marcel, S. On the effectiveness of vision transformers for zero-shot face anti-spoofing. In Proceedings of the 2021 IEEE International Joint Conference on Biometrics (IJCB), Shenzhen, China, 4–7 August 2021; pp. 1–8.
151. Nguyen, S.M.; Tran, L.D.; Le, D.V.; Masayuki, A. Self-attention generative distribution adversarial network for few-and zero-shot face anti-spoofing. In Proceedings of the 2022 IEEE International Joint Conference on Biometrics (IJCB), Ljubljana, Slovenia, 25–28 September 2023; pp. 1–9.

152. Huang, H.-P.; Sun, D.; Liu, Y.; Chu, W.-S.; Xiao, T.; Yuan, J.; Adam, H.; Yang, M.-H. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 37–54.
153. Pérez-Cabo, D.; Jiménez-Cabello, D.; Costa-Pazo, A.; López-Sastre, R.J. Deep anomaly detection for generalized face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
154. Fatemifar, S.; Arashloo, S.R.; Awais, M.; Kittler, J. Spoofing attack detection by anomaly detection. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 14–17 May 2019; pp. 8464–8468.
155. Baweja, Y.; Oza, P.; Perera, P.; Patel, V.M. Anomaly detection-based unknown face presentation attack detection. In Proceedings of the 2020 IEEE International Joint Conference on Biometrics (IJCB), Houston, TX, USA, 28 September–1 October 2020; pp. 1–9.
156. George, A.; Marcel, S. Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks. *IEEE Trans. Inf. Forensics Secur.* **2020**, *16*, 361–375. [CrossRef]
157. Fatemifar, S.; Arashloo, S.R.; Awais, M.; Kittler, J. Client-specific anomaly detection for face presentation attack detection. *Pattern Recognit.* **2021**, *112*, 107696. [CrossRef]
158. Huang, P.-K.; Chiang, C.-H.; Chen, T.-H.; Chong, J.-X.; Liu, T.-L.; Hsu, C.-T. One-Class Face Anti-spoofing via Spoof Cue Map-Guided Feature Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–23 June 2023; pp. 277–286.
159. Jiang, F.; Liu, Y.; Si, H.; Meng, J.; Li, Q. Cross-Scenario Unknown-Aware Face Anti-Spoofing With Evidential Semantic Consistency Learning. *IEEE Trans. Inf. Forensics Secur.* **2024**, *19*, 3093–3108. [CrossRef]
160. Costa-Pazo, A.; Jiménez-Cabello, D.; Vázquez-Fernández, E.; Alba-Castro, J.L.; López-Sastre, R.J. Generalized presentation attack detection: A face anti-spoofing evaluation proposal. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–8.
161. Zhang, S.; Wang, X.; Liu, A.; Zhao, C.; Wan, J.; Escalera, S.; Shi, H.; Wang, Z.; Li, S.Z. A dataset and benchmark for large-scale multi-modal face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 919–928.
162. Liu, A.; Wan, J.; Escalera, S.; Jair Escalante, H.; Tan, Z.; Yuan, Q.; Wang, K.; Lin, C.; Guo, G.; Guyon, I. Multi-modal face anti-spoofing attack detection challenge at cvpr2019. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
163. Parkin, A.; Grinchuk, O. Recognizing multi-modal face spoofing with face recognition networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
164. Shen, T.; Huang, Y.; Tong, Z. 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, Long Beach, CA, USA, 16–17 June 2019.
165. Liu, A.; Tan, Z.; Wan, J.; Escalera, S.; Guo, G.; Li, S.Z. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In Proceedings of the IEEE/CVF winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 1179–1187.
166. Yu, Z.; Qin, Y.; Li, X.; Wang, Z.; Zhao, C.; Lei, Z.; Zhao, G. Multi-modal face anti-spoofing based on central difference networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 650–651.
167. Yang, Q.; Zhu, X.; Fwu, J.-K.; Ye, Y.; You, G.; Zhu, Y. PipeNet: Selective modal pipeline of fusion network for multi-modal face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 644–645.
168. Liu, A.; Tan, Z.; Wan, J.; Liang, Y.; Lei, Z.; Guo, G.; Li, S.Z. Face anti-spoofing via adversarial cross-modality translation. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 2759–2772. [CrossRef]
169. Liu, W.; Wei, X.; Lei, T.; Wang, X.; Meng, H.; Nandi, A.K. Data-fusion-based two-stage cascade framework for multimodality face anti-spoofing. *IEEE Trans. Cogn. Dev. Syst.* **2021**, *14*, 672–683. [CrossRef]
170. Liu, A.; Liang, Y. Ma-vit: Modality-agnostic vision transformers for face anti-spoofing. *arXiv* **2023**, arXiv:2304.07549. [CrossRef]
171. Liu, A.; Tan, Z.; Yu, Z.; Zhao, C.; Wan, J.; Liang, Y.; Lei, Z.; Zhang, D.; Li, S.Z.; Guo, G. Fm-vit: Flexible modal vision transformers for face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 4775–4786. [CrossRef]
172. He, D.; He, X.; Yuan, R.; Li, Y.; Shen, C. Lightweight network-based multi-modal feature fusion for face anti-spoofing. *Vis. Comput.* **2023**, *39*, 1423–1435. [CrossRef]
173. Lin, X.; Wang, S.; Cai, R.; Liu, Y.; Fu, Y.; Tang, W.; Yu, Z.; Kot, A. Suppress and Rebalance: Towards Generalized Multi-Modal Face Anti-Spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–23 June 2023; pp. 211–221.

174. Yu, Z.; Cai, R.; Cui, Y.; Liu, X.; Hu, Y.; Kot, A.C. Rethinking vision transformer and masked autoencoder in multimodal face anti-spoofing. *Int. J. Comput. Vis.* **2024**, *132*, 5217–5238. [CrossRef]
175. Antil, A.; Dhiman, C. MF2ShrT: Multimodal feature fusion using shared layered transformer for face anti-spoofing. *ACM Trans. Multimed. Comput. Commun. Appl.* **2024**, *20*, 1–21. [CrossRef]
176. Zhou, B.; Lohokare, J.; Gao, R.; Ye, F. EchoPrint: Two-factor authentication using acoustics and vision on smartphones. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, New Delhi, India, 29 October–2 November 2018; pp. 321–336.
177. Chen, H.; Wang, W.; Zhang, J.; Zhang, Q. Echoface: Acoustic sensor-based media attack detection for face authentication. *IEEE Internet Things J.* **2019**, *7*, 2152–2159. [CrossRef]
178. Kong, C.; Zheng, K.; Wang, S.; Rocha, A.; Li, H. Beyond the pixel world: A novel acoustic-based face anti-spoofing system for smartphones. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 3238–3253. [CrossRef]
179. Zhang, D.; Meng, J.; Zhang, J.; Deng, X.; Ding, S.; Zhou, M.; Wang, Q.; Li, Q.; Chen, Y. Sonarguard: Ultrasonic face liveness detection on mobile devices. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 4401–4414. [CrossRef]
180. Xu, Z.; Liu, T.; Jiang, R.; Hu, P.; Guo, Z.; Liu, C. AFace: Range-flexible Anti-spoofing Face Authentication via Smartphone Acoustic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2024**, *8*, 1–33. [CrossRef]
181. Tang, D.; Zhou, Z.; Zhang, Y.; Zhang, K. Face flashing: A secure liveness detection protocol based on light reflections. *arXiv* **2018**, arXiv:1801.01949. [CrossRef]
182. Ebihara, A.F.; Sakurai, K.; Imaoka, H. Efficient face spoofing detection with flash. *IEEE Trans. Biom. Behav. Identity Sci.* **2021**, *3*, 535–549. [CrossRef]
183. Farrukh, H.; Aburas, R.M.; Cao, S.; Wang, H. 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, London, UK, 21–25 September 2020; pp. 1–13.
184. Zhang, J.; Tai, Y.; Yao, T.; Meng, J.; Ding, S.; Wang, C.; Li, J.; Huang, F.; Ji, R. Aurora guard: Reliable face anti-spoofing via mobile lighting system. *arXiv* **2021**, arXiv:2102.00713. [CrossRef]
185. Kim, Y.; Gwak, H.; Oh, J.; Kang, M.; Kim, J.; Kwon, H.; Kim, S. CloudNet: A LiDAR-based face anti-spoofing model that is robust against light variation. *IEEE Access* **2023**, *11*, 16984–16993. [CrossRef]
186. Xu, W.; Song, W.; Liu, J.; Liu, Y.; Cui, X.; Zheng, Y.; Han, J.; Wang, X.; Ren, K. Mask does not matter: Anti-spoofing face authentication using mmWave without on-site registration. In Proceedings of the 28th Annual International Conference on Mobile Computing and Networking, Sydney, NSW, Australia, 17–21 October 2022; pp. 310–323.
187. Xu, W.; Liu, J.; Zhang, S.; Zheng, Y.; Lin, F.; Xiao, F.; Han, J. Anti-spoofing facial authentication based on cots rfid. *IEEE Trans. Mob. Comput.* **2023**, *23*, 4228–4245. [CrossRef]
188. Zheng, Z.; Wang, Q.; Wang, C.; Zhou, M.; Zhao, Y.; Li, Q.; Shen, C. Where are the dots: Hardening face authentication on smartphones with unforgeable eye movement patterns. *IEEE Trans. Inf. Forensics Secur.* **2022**, *18*, 1295–1308. [CrossRef]
189. Yu, Z.; Qin, Y.; Xu, X.; Zhao, C.; Wang, Z.; Lei, Z.; Zhao, G. Auto-fas: Searching lightweight networks for face anti-spoofing. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Virtual, 4–9 May 2020; pp. 996–1000.
190. Kong, C.; Zheng, K.; Liu, Y.; Wang, S.; Rocha, A.; Li, H. M3 FAS: An Accurate and Robust MultiModal Mobile Face Anti-Spoofing System. *IEEE Trans. Dependable Secur. Comput.* **2024**, *21*, 5650–5666. [CrossRef]
191. Chingovska, I.; Anjos, A.; Marcel, S. On the effectiveness of local binary patterns in face anti-spoofing. In Proceedings of the 2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 6–7 September 2012; pp. 1–7.
192. Kose, N.; Dugelay, J.-L. Shape and texture based countermeasure to protect face recognition systems against mask attacks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 111–116.
193. Pinto, A.; Schwartz, W.R.; Pedrini, H.; de Rezende Rocha, A. Using visual rhythms for detecting video-based facial spoof attacks. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 1025–1038. [CrossRef]
194. Costa-Pazo, A.; Bhattacharjee, S.; Vazquez-Fernandez, E.; Marcel, S. The replay-mobile face presentation-attack database. In Proceedings of the 2016 international conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 21–23 September 2016; pp. 1–7.
195. Boulkenafet, Z.; Komulainen, J.; Li, L.; Feng, X.; Hadid, A. OULU-NPU: A mobile face presentation attack database with real-world variations. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 612–618.
196. Liu, Y.; Jourabloo, A.; Liu, X. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 389–398.

197. Fang, H.; Liu, A.; Wan, J.; Escalera, S.; Zhao, C.; Zhang, X.; Li, S.Z.; Lei, Z. Surveillance face anti-spoofing. *IEEE Trans. Inf. Forensics Secur.* **2023**, *19*, 1535–1546. [CrossRef]
198. Wang, D.; Guo, J.; Shao, Q.; He, H.; Chen, Z.; Xiao, C.; Liu, A.; Escalera, S.; Escalante, H.J.; Lei, Z. Wild face anti-spoofing challenge 2023: Benchmark and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–23 June 2023; pp. 6380–6391.
199. Fang, M.; Huber, M.; Damer, N. Synthaspoof: Developing face presentation attack detection based on privacy-friendly synthetic data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–23 June 2023; pp. 1061–1070.
200. Erdogmus, N.; Marcel, S. Spoofing face recognition with 3D masks. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1084–1097. [CrossRef]
201. Raghavendra, R.; Raja, K.B.; Busch, C. Presentation attack detection for face recognition using light field camera. *IEEE Trans. Image Process.* **2015**, *24*, 1060–1075. [CrossRef]
202. Galbally, J.; Satta, R. Three-dimensional and two-and-a-half-dimensional face recognition spoofing using three-dimensional printed models. *IET Biom.* **2016**, *5*, 83–91. [CrossRef]
203. Bhattacharjee, S.; Marcel, S. What you can't see can help you-extended-range imaging for 3d-mask presentation attack detection. In Proceedings of the 2017 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 20–22 September 2017; pp. 1–7.
204. Agarwal, A.; Yadav, D.; Kohli, N.; Singh, R.; Vatsa, M.; Noore, A. Face presentation attack with latex masks in multispectral videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 81–89.
205. Bhattacharjee, S.; Mohammadi, A.; Marcel, S. Spoofing deep face recognition with custom silicone masks. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22–25 October 2018; pp. 1–7.
206. George, A.; Mostaani, Z.; Geissenbuhler, D.; Nikisins, O.; Anjos, A.; Marcel, S. Biometric face presentation attack detection with multi-channel convolutional neural network. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 42–55. [CrossRef]
207. Rostami, M.; Spinoulas, L.; Hussein, M.; Mathai, J.; Abd-Almageed, W. Detection and continual learning of novel face presentation attacks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 14851–14860.
208. Song, C.; Hong, Y.; Lan, J.; Zhu, H.; Wang, W.; Zhang, J. Supervised Contrastive Learning for Snapshot Spectral Imaging Face Anti-Spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–23 June 2023; pp. 980–985.
209. Bioucas-Dias, J.M.; Figueiredo, M.A. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* **2007**, *16*, 2992–3004. [CrossRef]
210. Fang, H.; Liu, A.; Yuan, H.; Zheng, J.; Zeng, D.; Liu, Y.; Deng, J.; Escalera, S.; Liu, X.; Wan, J. Unified physical-digital face attack detection. *arXiv* **2024**, arXiv:2401.17699. [CrossRef]
211. Zhang, R.; Xu, Q.; Yao, J.; Zhang, Y.; Tian, Q.; Wang, Y. Federated domain generalization with generalization adjustment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–23 June 2023; pp. 3954–3963.
212. Cho, J.; Nam, G.; Kim, S.; Yang, H.; Kwak, S. Promptstyler: Prompt-driven style generation for source-free domain generalization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 15702–15712.
213. Wu, J.; Yu, X.; Liu, B.; Wang, Z.; Chandraker, M. Uncertainty-aware physically-guided proxy tasks for unseen domain face anti-spoofing. *arXiv* **2020**, arXiv:2011.14054. [CrossRef]
214. Peng, X.; Wei, Y.; Deng, A.; Wang, D.; Hu, D. Balanced multimodal learning via on-the-fly gradient modulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8238–8247.
215. Li, Y.; Hu, J.; Wen, Y.; Evangelidis, G.; Salahi, K.; Wang, Y.; Tulyakov, S.; Ren, J. Rethinking vision transformers for mobilenet size and speed. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 16889–16900.
216. Liu, H.; Kong, Z.; Ramachandra, R.; Liu, F.; Shen, L.; Busch, C. Taming self-supervised learning for presentation attack detection: In-image de-folding and out-of-image de-mixing. *arXiv* **2021**, arXiv:2109.04100. [CrossRef]
217. Muhammad, U.; Yu, Z.; Komulainen, J. Self-supervised 2d face presentation attack detection via temporal sequence sampling. *Pattern Recognit. Lett.* **2022**, *156*, 15–22. [CrossRef]
218. Zhang, X.; Zhou, L.; Xu, R.; Cui, P.; Shen, Z.; Liu, H. Towards unsupervised domain generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4910–4920.
219. Harary, S.; Schwartz, E.; Arbelle, A.; Staar, P.; Abu-Hussein, S.; Amrani, E.; Herzig, R.; Alfassy, A.; Giryes, R.; Kuehne, H. Unsupervised domain generalization by learning a bridge across domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5280–5290.

220. Qi, L.; Liu, J.; Wang, L.; Shi, Y.; Geng, X. Unsupervised domain generalization for person re-identification: A domain-specific adaptive framework. *arXiv* **2021**, arXiv:2111.15077. [CrossRef]
221. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2019; pp. 9729–9738.
222. Chen, T.; Kornblith, S.; Swersky, K.; Norouzi, M.; Hinton, G.E. Big self-supervised models are strong semi-supervised learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 22243–22255.
223. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
224. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
225. Mehta, S.; Rastegari, M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178. [CrossRef]
226. Deb, D.; Liu, X.; Jain, A.K. Unified detection of digital and physical face attacks. In Proceedings of the 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, HI, USA, 5–8 January 2023; pp. 1–8.
227. Al-Refai, R.; Nandakumar, K. A unified model for face matching and presentation attack detection using an ensemble of vision transformer features. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 662–671.
228. Yu, Z.; Cai, R.; Li, Z.; Yang, W.; Shi, J.; Kot, A.C. Benchmarking joint face spoofing and forgery detection with visual and physiological cues. *IEEE Trans. Dependable Secur. Comput.* **2024**, *21*, 4327–4342. [CrossRef]
229. Shi, Y.; Gao, Y.; Lai, Y.; Wang, H.; Feng, J.; He, L.; Wan, J.; Chen, C.; Yu, Z.; Cao, X. Shield: An evaluation benchmark for face spoofing and forgery detection with multimodal large language models. *arXiv* **2024**, arXiv:2402.04178. [CrossRef]