APSIPA Transactions on Signal and Information Processing, 2024, 13, e34

Overview Paper

A Survey on Deep Learning-based Face Anti-Spoofing

Pei-Kai Huang, Jun-Xiong Chong, Ming-Tsung Hsu, Fang-Yu Hsu, Cheng-Hsuan Chiang, Tzu-Hsien Chen and Chiou-Ting Hsu*

National Tsing Hua University

ABSTRACT

Face anti-spoofing (FAS) aims to distinguish live images and facial spoof attacks to defend facial recognition systems. Thanks to advancements in deep learning, recent deep learning-based FAS methods have shown promising potential, especially in effectively addressing newly developed attacks. In this survey, we first provide an overview of common challenges in FAS and then recap recent advances in deep learning-based FAS. In particular, these anti-spoofing methods generally fall into two main categories, i.e., two-class FAS and one-class FAS. Recent two-class FAS methods have employed a wide range of techniques in developing FAS models, including auxiliary supervision, local descriptor-enhanced feature learning, disentangled feature learning, meta learning, adversarial learning, data augmentation, long-range dependency learning, and multimodal learning. Meanwhile, recent one-class FAS methods have utilized reconstruction supervision, statistical learning, and generative feature learning to learn liveness features solely from live images. In this survey, we also provide an overview of publicly available FAS datasets. Finally, we summarize recent FAS development and highlight some potential future research directions.

Keywords: Face anti-spoofing, presentation attack, 3D mask attack, liveness feature, one-class detection

*Corresponding author: Chiou-Ting Hsu, cthsu@gapp.nthu.edu.tw

Received 03 August 2024; revised 02 October 2024; accepted 12 October 2024
ISSN 2048-7703; DOI 10.1561/116.20240053

1 Introduction

Ensuring authenticity of detected facial data is key to applications that heavily rely on facial recognition. Face anti-spoofing (FAS), namely 'face presentation attack detection', '3D mask attack detection' or 'face liveness detection', plays a critical role in defending facial recognition systems against fraudulent attempts from facial spoof attacks. As illustrated in Figure 1, face anti-spoofing is a binary classification task that aims to distinguish between live images and facial spoof attacks. Common facial spoof attacks often include presentation attacks, such as Print Attack (printing a face on paper) and Replay Attack (replaying a face video on digital devices), as well as 3D mask attacks (wearing a mask on a face), as shown in Figure 2. These attacks pose a serious threat to the security of facial recognition systems. They may diminish the reliability and effectiveness of facial recognition systems, potentially leading to unauthorized access or misuse of sensitive information. Therefore, many recent deep learning-based FAS methods have been developed to counter these attacks for maintaining the security of facial recognition systems.

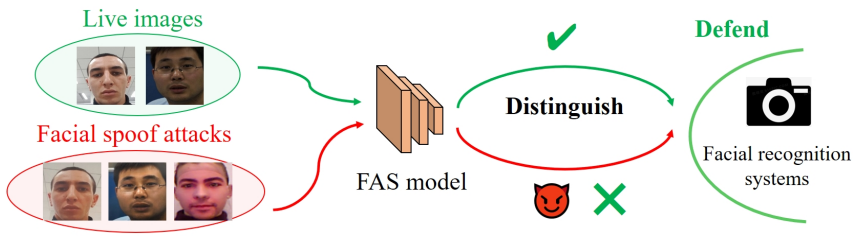


Figure 1: Illustration of face anti-spoofing (FAS).

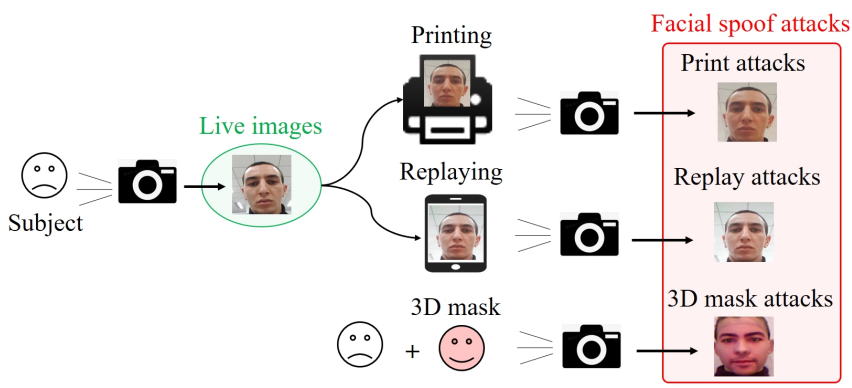


Figure 2: Illustration of live images and facial spoof attacks.

Recent learning-based FAS methods primarily focused on two settings: two-class setting, where the training data include both live images and facial spoof attacks, and one-class setting, where the training data consist only of live images. To narrow the domain gap between training and testing data, previous methods have considered different scenarios (as shown in Figure 3), including domain generalization (DG), domain adaptation (DA), source-free domain adaptation (SFDA), and test-time adaptation (TTA), to counter newly developed facial spoof attacks from unseen domains. In particular, most FAS methods adopted DG scenario to develop a generalized model capable of learning from multiple source domains. As shown in Figure 3 (a), by adopting DG scenario, these FAS methods aim to enhance the generalization capabilities of FAS models across diverse datasets and environmental conditions. In Figure 3 (b), when the target data are accessible during the training stage, DA scenario is often employed to facilitate the adaptation of source knowledge into the target domain. By leveraging DA scenario, FAS models are able to effectively bridge the domain gap between the source and target datasets, thereby enhancing the performance and generalization capabilities of the target domain. In Figure 3 (c), when the source data are inaccessible due to data privacy concerns, the concept of SFDA has emerged as a solution by directly fine-tuning a pre-existing off-the-shelf model on the target domain. By circumventing the need for access to source data during training, SFDA offers a practical and privacy-preserving alternative for adapting FAS models to specific target domains. Through this process, the fine-tuned model can better align with the characteristics and nuances of the target dataset, thereby enhancing its performance and adaptability in real-world deployment scenarios. In Figure 3 (b)-(c), both DA and SFDA need to access target data for adaptation during offline training. However, because collecting all possible attacks during this stage is impractical and impossible, the pre-adapted model may still encounter difficulties in detecting unseen attack types. Consequently, FAS models may require further adaptation during the online inference stage to effectively detect unseen novel attacks. In contrast to DA and SFDA, TTA addresses a more realistic scenario where the source domain data are either inaccessible or no longer available and only an off-the-shelf model is accessible, as shown in Figure 3 (d). Hence, TTA aims to enable online adaptation of an off-the-shelf model directly to unlabeled target data during the inference stage. By leveraging the information of unlabeled target data, TTA adaptively fine-tune the parameters of FAS models to better align with the specific characteristics of target data encountered during inference. This dynamic adaptation process allows FAS models to effectively adapt to unseen attack types and environmental variations, thereby enhancing its performance and generalization capabilities in real-world scenarios.

Early FAS datasets primarily consisted of single modality data, i.e., RGB images. However, as facial spoof attacks have evolved and become more

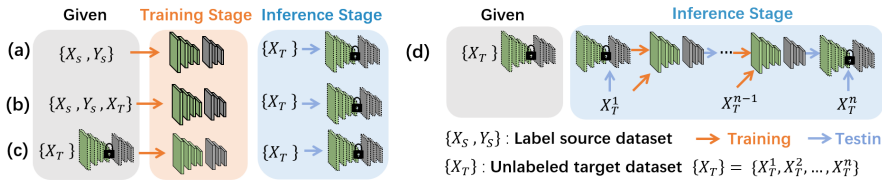


Figure 3: Illustration of different cross-domain scenarios in [41]: (a) domain generalization (DG), (b) domain adaptation (DA), (c) source-free domain adaptation (SFDA), and (d) test-time adaptation (TTA).

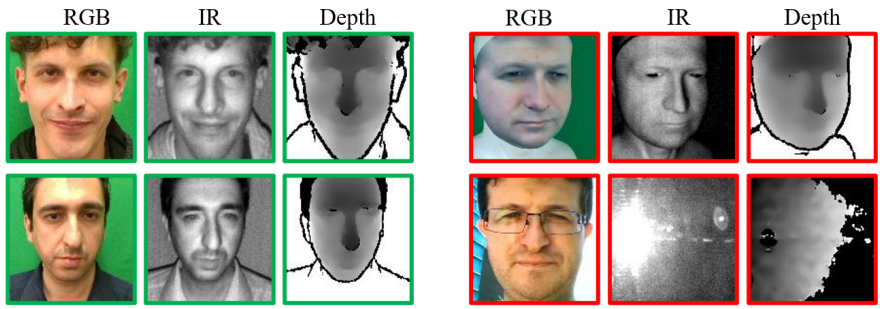


Figure 4: Different modalities of live class (marked by green box) and spoof class (marked by red box).

sophisticated, there is an increasing demand for datasets that incorporate multi-modalities to enhance the robustness and effectiveness of FAS systems. This shift towards multi-modal datasets specifies the limitations of single modality data in capturing diverse spoof cues and characteristics necessary for accurate spoofing detection. Some examples of different modalities in FAS are given in Figure 4 and show distinct characteristics in different modalities. For example, RGB images offer rich color and detailed textural information, yet they are highly sensitive to variations in illumination. By contrast, infrared (IR) images are less affected by illumination changes and are more consistent over diverse lighting conditions. On the other hand, depth images complementarily provide structural insights through depth information but usually possess little detailed textural information present in RGB images. By integrating multiple modalities, these multimodal FAS datasets offer a more comprehensive and nuanced representation of facial features, thereby enabling FAS systems to achieve higher accuracy and reliability.

In this survey, we introduce recent deep learning-based face anti-spoofing methods following the structure in Figure 5. In Section 1, we first outline two-class and one-class settings as well as different cross-domain scenarios. In

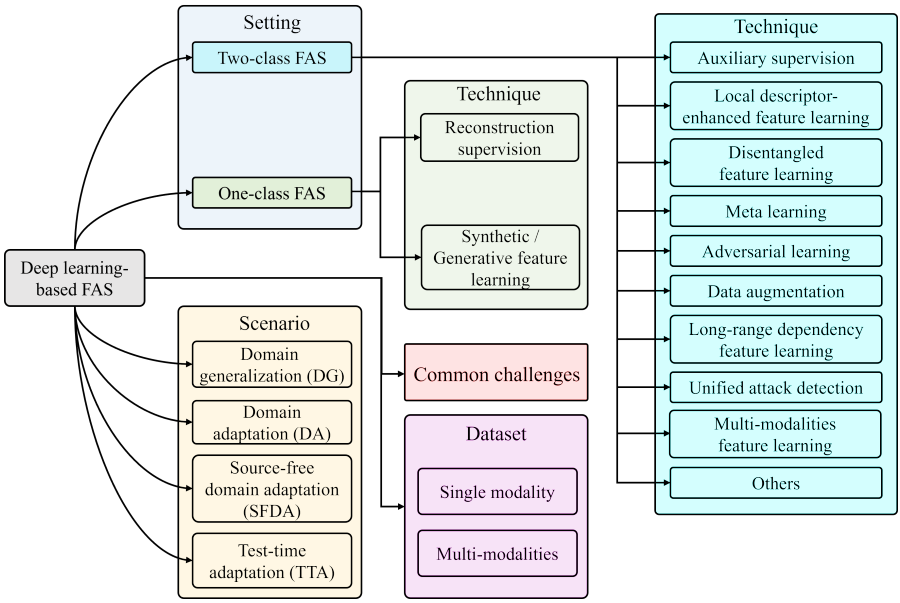


Figure 5: Structure of this survey for deep learning-based face anti-spoofing.

Section 2, we introduce common challenges in FAS. Next, we recap recent two-class FAS methods in Section 3 and one-class FAS methods in Section 4. In Section 5, we then review several popular datasets and common evaluation metrics in FAS. Finally, we conclude this survey in Section 6.

Note that, compared to previous FAS surveys [124, 3], which primarily adopt two perspectives: 1) single versus multi-modalities [124], and 2) passive versus active approaches based on user interaction requirements [3], this survey specifically emphasizes the discussion of one-class and two-class settings in previous FAS works employing similar techniques within these contexts. In Table 1, we also introduce recent scenarios in FAS, such as source-free domain adaptation (SFDA) and test-time adaptation (TTA), which have not been covered in previous surveys. Furthermore, this survey explores common challenges in FAS and highlights recent advancements, including the application of vision-language models and acoustic-based features.

To sum up, in this survey:

- We comprehensively cover recent deep leaning-based face anti-spoofing methods on common two-class and one-class settings as well as cross-domain scenarios, including domain generalization (DG), domain adaptation (DA), source-free domain adaptation (SFDA), and test-time adaptation (TTA).

Table 1: Summary of different scenarios and common benchmarks in face anti-spoofing.

Scenarios		Problem Statement	Common benchmarks and exemplary methods
Domain generalization (DG)		Given the labeled source training data $\{X_S, Y_S\}$, the objective is to train FAS models that can generalize effectively to the unlabeled target data $\{X_T\}$	1. Intra-testing[39, 38, 37]: Oulu[9], Siw[77] 2. Leave-one-dataset-out cross-testing[40, 42]: Oulu (O)[9], CASIA-MFSD (C)[135], Idiap Replay-Attack (I)[15], MSU-MFSD (M)[113] Protocols: [O, C, M] → I, [O, C, I] → M, [I, C, M] → O, and [O, I, M] → C 3. Limited source cross-domain testing[40, 42]: Oulu (O)[9], CASIA-MFSD (C)[135], Idiap Replay-Attack (I)[15], MSU-MFSD (M)[113] Protocols: [M, I] → C and [M, I] → O
Domain adaptation (DA)	Semi-domain adaptation (SDA)	Given the labeled source training data $\{X_S, Y_S\}$, the partially labeled target training data $\{X_{T_a}, Y_{T_a}\}$, and the unlabeled target data $\{X_{T_b}\}$, the objective is to train FAS models that can effectively adapt to the unlabeled target data $\{X_{Tb}\}$.	1. Cross-dataset testing[47]: CASIA-MFSD (C) [135], Idiap Replay-Attack (I)[15], MSU-MFSD (M) [113] Protocols: C → I, C → M, I → C, I → M, M → C, M → I
	Unsupervised domain adaptation (UDA)	Given the labeled source training data $\{X_S, Y_S\}$ and the unlabeled target data $\{X_T\}$, the objective is to train FAS models that can effectively adapt to the unlabeled target data$\{X_T\}$.	1. Leave-one-dataset-out cross-testing[106, 14]: Oulu (O)[9], CASIA-MFSD (C) [135], Idiap Replay-Attack (I)[15], MSU-MFSD (M) [113] Protocols: [O, C, M] → I, [O, C, I] → M, [I, C, M] → O,[O, I, M] → C 2. Limited source cross-domain testing[14]: CASIA-MFSD (C) [135], Idiap Replay-Attack (I)[15], MSU-MFSD (M) [113] Protocols: [M, I] → C, [M, I] → O 3. Cross-dataset testing[58]: Oulu (O)[9], CASIA-MFSD (C) [135], Idiap Replay-Attack (I)[15], MSU-MFSD (M), Protocols: C → I, C → M, C → Y, I → C, I → M, I → Y, M → C, M → I, M → Y, Y → C, Y → I, Y → M
Source-free domain adaptation (SFDA)		Given the off-the-shelf FAS models and unlabeled target data $\{X_T\}$, the objective is to train FAS models that can effectively adapt to the unlabeled target data $\{X_T\}$.	1. Leave-one-dataset-out cross-testing: (A). Oulu (denotes as O)[9], CASIA-MFSD (C) [135], Idiap Replay-Attack (I)[15], MSU-MFSD (M) [113] Protocols: [O, C, M] → I, [O, C, I] → M, [I, C, M] → O,[O, I, M] → C (B). Oulu (O)[9], Siw (S) [77], HKBU-MARs (H)[76], Siw-M (M) [79] Protocols: [O, M, H] → S, [O, S, H] → M, [M, S, H] → O,[M, S, O] → H
Test-time adaptation (TTA)		Given the off-the-shelf FAS models and the mini-batch of unlabeled target data $\{X_{t_1}\}$,the objective is to train FAS models that can effectively adapt to the unlabeled target data$\{X_{t_1}\}$ and directly make reliable prediction.	1. Unseen attack testing[41]: OULU-NPU(O)[9], CASIA-MFSD (C)[135], MSU-MFSD (M)[113], Idiap Replay-Attack (I)[15], 3DMAD (D)[20], HKBU-MARs (H) [76] Protocols: [O, C, I] → [M, D, H], [O, M, I] → [C, D, H], [O, C, M] → [I, D, H],[I, C, M] → [O, D, H] 2. Leave-one-attack-out testing[41]: Oulu(O)[9], CASIA-MFSD (C)[135], MSU-MFSD (M)[113], Idiap Replay-Attack (I)[15], 3DMAD (D)[20], HKBU-MARs (H) [76] Protocols :[O, M, I] → [C, D, H],[C, D, H] → [O, M, I]
Few-shot		Given the labeled source training data $\{X_S, Y_S\}$ and a limited set of k labeled target training data $\{X_{T_a}, Y_{T_a}\}$, the objective is to train FAS models that can effectively adapt to the unlabeled target data $\{X_{T_b}\}$.	1. Leave-one-dataset-out cross-testing[40, 42]: Oulu (O)[9], CASIA-MFSD (C)[135], Idiap Replay-Attack (I)[15], MSU-MFSD (M)[113] Protocols: [O, C, M] → I, [O, C, I] → M, [I, C, M] → O, [O, I, M] → C 2. Limited source cross-domain testing[40, 42]: CASIA-SURF (S)[130], CASIA-CeFA (C)[68], WMCA (W)[29] Protocols: [C, S] → W, [S, W] → C, [C, W] → S

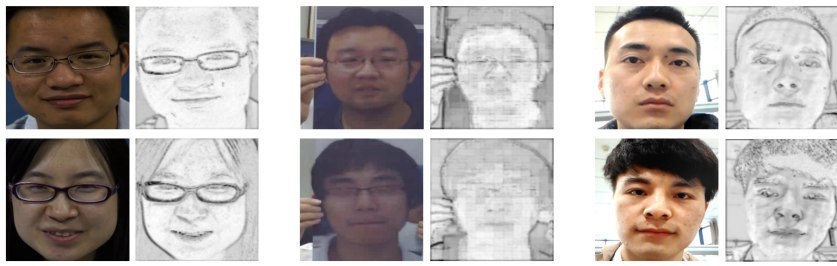
- We introduce common challenges in FAS and summarize the techniques adopted in both two-class and one-class FAS methods.
- We review popular datasets and evaluation metrics in FAS, and point out future research directions towards countering ever-evolving facial spoof attacks.

2 Common Challenges in FAS

There are three common challenges in face anti-spoofing. The first one is the issue of similar visual appearance. As shown in Figure 6, we see that genuine live faces, i.e., Figure 6 (a) and facial print attacks, i.e., Figure 6 (b)-(c), may visually resemble each other, thereby posing challenges for accurate classification. Because live and spoof faces are visually similar, face anti-spoofing often requires a more delicate representation to characterize intrinsic features associated with facial spoof attacks compared to other image classification tasks. Next, the second challenge concerns the problem of subtle spoof cues. Figure 6 shows that facial spoof attacks involve subtle spoof cues, such as the grid artifacts of facial print attacks in Figure 6 (b) and the moiré patterns of facial replay attacks in Figure 6 (c), which are imperceptible to human eyes. Therefore, face anti-spoofing is a nontrivial task that needs to capture subtle spoof cues for distinguishing live images from facial spoof attacks. Finally, the third challenge arises from the absence of prior knowledge about unseen attack types. Similar to most deep learning-based tasks, the effectiveness of deep learning-based FAS heavily depends on the quantity and quality of the labeled training dataset, which serves as the foundation for model training and generalization. However, as facial spoof attacks continue to evolve and become increasingly sophisticated, FAS detectors struggle to keep pace with the rapid emergence of new and unseen attack variations.

3 Two-class Face Anti-spoofing Methods

Following the structure in Figure 5, in this section, we review recent two-class deep learning-based FAS methods and their adopted techniques, including auxiliary supervision, local descriptor-enhanced feature learning, disentangled feature learning, meta learning, adversarial learning, data augmentation, long-range dependency learning, unified attack detection and multi-modalities feature learning.



(a) (b) (c)

Figure 6: Examples of facial images and their low-level feature maps from [42], (a) genuine live images without spoof cues, (b) facial print attacks with grid artifacts, and (c) facial replay attacks with moiré patterns.

3.1 Auxiliary Supervision

Integration of auxiliary information, such as facial depth maps [39, 77, 94, 122, 111, 52, 106, 49], Remote Photoplethysmography (rPPG) signals/maps [39, 121, 75, 60, 74], reflection maps [51, 132, 122], attention maps [42], moire maps [8], frequency maps [24], and paired captions [97], has emerged as a successful strategy in enhancing the supervision of live/spoof classification in FAS. By incorporating these additional sources of information, FAS models are able to leverage a more comprehensive understanding of facial cues and characteristics to improve their accuracy and robustness in distinguishing between genuine live faces and facial spoof attacks.

3.1.1 Depth supervision

Intuitively, live faces exhibit natural depth variations that can be perceived visually or through depth maps to reflect three-dimensional structure of faces. In contrast, certain spoof attacks, such as replay attacks, often lack these depth variations and may appear planar or artificially uniform in depth. This difference in depth serves as a distinguishing characteristic between genuine live faces and spoof attacks in FAS. By leveraging depth characteristics, the authors in [39, 77, 94, 122, 111, 52, 106, 49, 138] proposed to adopt depth supervision to enable the identification of facial spoof attacks through detecting unnatural or uniform depth information.

3.1.2 rPPG signal/map supervision

Live faces typically exhibit natural Remote Photoplethysmography (rPPG) signals, reflecting subtle changes in skin color caused by blood flow. These

physiological responses indicate characteristics of genuine faces. In contrast, facial spoof attacks, such as those involving 3D masks and printed papers/photos, often lack rPPG signal responses, because non-biological materials used in facial spoof attacks do not possess the physiological properties of genuine rPPG signals. Therefore, the authors in [39, 121, 75, 60, 74] proposed to detect the absence or abnormality of rPPG signal responses to distinguish live faces and facial spoof attacks.

3.1.3 Other auxiliary supervision

In addition to depth information and rPPG responses, recent FAS methods focused on exploring various auxiliary information from different perspectives to enhance FAS. For example, the authors in [51, 132, 122] proposed to estimate highlight surface reflections on faces to help detecting spoofing materials or artifacts that may distort the reflection patterns observed in genuine facial images. Next, in [42], the authors proposed adopting attention maps to provide regional indications of where the attacked regions are located, offering FAS models additional guidance with fine-grained information. Furthermore, the authors in [8] proposed to generate moiré pattern maps to guide FAS model in effectively countering replay attacks. In [57], the authors proposed cross-stage relation enhancement and spoof material perception to improve feature extraction across model stages. Moreover, the authors in [24] proposed adopting Discrete Cosine Transform (DCT) to decompose the frequency map from input images into auxiliary inputs, aiming to mitigate the sensitivity of FAS models to variations in capture environments (e.g., sensors or light conditions). In addition, thanks to the rapid development of language models, some recent methods [71, 97, 21] have incorporated vision-language models, such as CLIP [90], to facilitate the learning process of FAS models. CLIP (Contrastive Language-Image Pretraining) is a multimodal model that learns joint representations of images and text and is able to understand semantic relationships between visual and textual inputs. By leveraging CLIP or similar language models, the authors in [71, 97, 21] proposed to enhance the ability of FAS to discern between genuine live faces and spoof attacks by facilitating the understanding of contextual information associated with facial images.

3.2 Local Descriptor-enhanced Feature Learning

To capture subtle spoof cues, several authors in recent FAS methods [127, 126, 125, 109, 122, 13, 42, 38, 138] have proposed integrating predefined or learnable descriptors into vanilla convolutional architectures. This integration aimed to capture gradient-level information within facial images for enhancing the representation capability of FAS models.

3.2.1 Pre-defined local descriptors

In [127, 126, 125, 109, 122, 13], the authors proposed including predefined descriptors into vanilla convolution, e.g., Central Difference Convolution [127, 126], Dual-cross Central Difference Convolution [125], Sobel Convolution [109], Bilateral Convolution [122], and Eight-Direction Differential Convolution [13], to capture gradient information of subtle spoof cues to learn discriminative liveness features for FAS.

3.2.2 Learnable local descriptors

While previous methods [127, 126, 125, 109, 122, 13, 42, 38, 138] have integrated various local descriptors to enhance conventional convolutional approach, it is noteworthy that these local descriptors remain fixed and are not updated during model training. As noted in [42], the predefined and unlearnable descriptors lack flexibility in capturing diverse textural features, thereby limiting their applicability in FAS. Therefore, instead of predetermining the descriptor, in [42, 38, 138], the authors proposed including learnable descriptors into vanilla convolution, such as Learnable Descriptive Convolution [42], Decoupled-Learnable Descriptive Convolution [38], and Dynamic Kernel Generator [138].

3.3 Disentangled Feature Learning

In FAS, disentangled feature learning aims to separate or disentangle the underlying representation of liveness and domain information in facial images. By disentangling these factors, FAS models in [80, 78, 50, 134, 105, 36, 112, 40, 114, 34, 104, 129] are able to better understand the essential liveness features associated with genuine live faces and facial spoof attacks. This feature disentanglement facilitates more effective and accurate classification by reducing the interference caused by domain-specific variations, thus enhancing the robustness and generalization capability of FAS models. Through disentangled feature learning, FAS models are able to focus on capturing intrinsic characteristics of live faces while disregarding irrelevant domain-specific variations to increase the performance in detecting facial spoof attacks. In addition, in [36, 112, 40], the authors proposed adopting disentangled feature learning to augment spoof training data toward unseen domains. Similarly, in [117], the authors proposed using finer domain partition to disentangle liveness-irrelevant factors.

3.4 Meta Learning

Meta learning aims to train models to quickly adapt and learn generalized features from limited data samples. In [95, 88, 14, 45, 52, 106], the authors

proposed adopting meta learning to enable FAS models to effectively generalize its knowledge and adapt to unseen spoofing attacks during inference. When addressing the domain generalization issue via meta learning, it is necessary to ensure that the meta-optimization successively leads to stable directions towards the well-generalized model. However, as observed in [95], because of simple binary live/spoof ground-truth labels and the domain shift, the vanilla meta-learning directions are prone to be arbitrary during the meta-train and meta-testing steps. Therefore, the authors in [95, 14, 45, 52, 106, 11] proposed adopting external depth supervision or learned pattern, such as Meta Patterns to provide fine-grained information to regularize the meta learning for finding generalized FAS models.

3.5 Adversarial Learning

Adversarial learning aims to introduce adversarial examples or perturbations during the training process to enhance the robustness of models. By exposing the FAS models to these intentionally crafted challenging examples, which are designed to fool the FAS models, adversarial learning in [137, 94, 30, 34, 48] encouraged the FAS models to learn robust liveness features to improve the generalization ability of FAS models across diverse conditions and environments. For example, in [94], the authors proposed to adopt adversarial learning to learn generalized feature space from multiple source domains. Next, the authors in [137] proposed adopting adversarial learning to align the liveness features across various facial regions to effectively handle domain discrepancies. Furthermore, in [30], the authors proposed adopting adversarial learning to transfer the knowledge from teacher models to enhance the FAS models. Moreover, the authors in [34] proposed adopting adversarial learning to disentangle the spoof-specific and domain-specific features, and then mixing different spoof-specific and domain-specific features to generate new samples. Finally, the authors in [48] proposed adopting adversarial learning to align the conditional distributions across domains to learn domain-invariant conditional features.

3.6 Data Augmentation

Data augmentation in FAS involves applying various transformations and techniques to increase the diversity and volume of the training data at both the image-level [107, 32, 81, 118, 36, 103] and feature-level [40, 112, 105]. In [118], the authors proposed combining two images to simulate reflection artifacts. Also, the authors proposed using color distortion [107, 32] and weak augmentation [81, 103] to enlarge the training data. Furthermore, the authors in [36] proposed to learn disentangled features and then to generate an augmented spoof dataset via reconstructing images by remixing disentangled features. In addition, the authors in [112, 105] proposed swapping disentangled

features to augment the training data in the feature space; and the authors in [40] proposed to enrich the diversity of liveness features and to enlarge the generalization ability of domain features before generating the augmented features.

3.7 Long-range Dependency Feature Learning

With the great success of convolutional neural networks (CNNs) in many computer vision tasks, CNN-based methods have become a favorite in face anti-spoofing. However, the limited receptive field of convolutional operations in CNNs restricts their ability to capture global context and long-range dependencies from images. To address this limitation, many recent face anti-spoofing methods [62, 28, 67, 38, 35, 110, 70, 64] have adopted Vision Transformers (ViTs) as their backbone to model long-range pixel dependencies. This enables them to learn more comprehensive and distinguishing features between live and spoof faces. By leveraging the self-attention mechanism inherent in transformers, ViTs are able to effectively capture intricate patterns and relationships across different patches within the whole image. Although this capability allows modeling long-range data dependency, ViTs may not adequately capture the local intrinsic features crucial for FAS, e.g. fine-grained textures. This limitation comes from that ViTs emphasize global context but potentially overlook detailed local features.

Therefore, the authors in [38] proposed to combine CNNs and ViTs to balance the tradeoff between receptive field size and local feature capabilities from both architectures on modeling long-range and distinguishing characteristics of FAS. In addition, in [12], the authors proposed to integrate the histogram information of transformer tokens into ViT to learn more domain-invariant feature representations.

3.8 Unified Attack Detection

In [139], the authors proposed integrating SoftMoE into CLIP's image encoder to enhance its capacity for handling the sparse feature distribution in unified attack detection (UAD) tasks, thereby reducing the gap between physical attack detection (PAD) and digital attack detection (DAD). In particular, they proposed replacing SoftMoE's traditional weighting mechanism with linear attention to improve the model's ability to handle both physical and digital attacks within a unified framework. Next, in [96], the authors proposed a new benchmark to evaluate the effectiveness of multi-modal face anti-spoofing models in detecting both physical and digital attacks. In [23], the authors also proposed adopting vision-language models (VLMs) to learn joint and category-specific knowledge for Unified Attack Detection. Furthermore, the authors in [128] provided a large-scale dataset (UniAttackData) to facilitate the

research of Unified Attack Detection. In [32], the authors proposed simulating the color distortions of print attacks, the moire patterns of replay attacks, the facial artifacts of digital forgery, and the gradient noises of adversarial attacks to augment training samples for detecting both physical and digital attacks. Moreover, the authors in [121] proposed using both visual appearance and physiological rPPG cues to develop a joint detection framework for face spoofing and forgery. Finally, in [53], the authors proposed incorporating various auxiliary information, such as remote physiological signals and pseudo depth maps, to improve detection performance of unified attack detection.

3.9 Multi-modalities Feature Learning

3.9.1 Modalities between images

With the advance of spoofing attacks, many methods [123, 69, 27, 108, 67, 70, 19, 101, 59, 64, 69, 120, 31, 65] are developed to improve the performance of FAS models by incorporating modalities other than RGB, mainly infrared and depth images. In [123], the authors extended their previous single modality CDCN to a multi-modal CDCN, with each modality having its own branch to learn modality-aware features independently. The authors in [69] proposed an adversarial framework to translate features across modalities to enhance the performance through cross-modality translation. In [120], the authors proposed integrating Vision Transformers and Masked Autoencoders to enhance multi-modal feature extraction to develop more generalized FAS models that perform effectively across diverse environments. In [31], the authors proposed exploring hyperbolic space to enhance feature separability and cross-modal robustness. In [65], the authors proposed a dual cross-attention mechanism combined with a semi-fixed mixture-of-expert strategy to improve generalizability across multiple modalities in face anti-spoofing tasks.

The authors in [27] used a different approach by proposing the use of cross-modal focal loss to modulate the contribution of different modalities and to learn their complementary information. Using the design of Conv-MLP [108], the authors proposed to extract local and long-range depended features and also designed a novel moat loss to improve the extraction of discriminative features by preventing the blindly clustering of spoof features. In [59], the authors implemented a novel design of cross-modality fusion module to encourage the extraction of complementary features between modalities. The authors in [19] used dual-stream fusion method to fuse both infrared and surface normal map created using depth images, in which one stream focuses on extracting complementary global features, while the other extracts complementary fine-grained features.

On the other hand, because most previous multi-modal methods require the existence of training modalities during inference, the authors in [70, 67, 64]

proposed a practical setting of flexible modal, in which the modal is capable of inferencing even when one or more modalities are absent. In [70] the authors proposed a novel cross-modal attention module to extract modality-agnostic feature learning and to mine informative patches. In addition, the authors [70] also implemented separate classification heads for each modality. Similarly, the same authors in [67] proposed to use another cross-modal attention module to extract both modality-agnostic and complementary liveness features across modalities. The authors in [64] proposed to reduce the effect of unreliable features within each modality using cross-modality adapter to ensure the contribution of each modality during inference by using the gradient modulation strategy.

Recent advancements have further focused on addressing the challenge of missing modalities by utilizing visual prompts [119, 66]. In [119], the authors proposed leveraging visual prompts and residual contextual prompts to adapt models to varying modality availability without requiring extensive re-training, significantly improving robustness under missing-modality scenarios. Similarly, In [66], the authors proposed adapting cross-modal learning and language-guided prompts to dynamically adjusts visual features for handling missing modalities.

3.9.2 Acoustic-based features

Although modalities like IR and depth have proven effective in countering many types of attacks, they come with significant limitations. Each of these modalities requires specialized sensors for data capture, which are often not available on most face recognition devices. In addition, the variation in sensor design across different devices leads to inconsistencies between datasets, limiting the scalability and broader adoption of these modalities. To address these limitations, the authors in [136, 55, 54, 116] proposed using acoustic-based features, where echoed audio signals are utilized to capture facial characteristics and detect subtle variations in facial geometry. In [136], the authors proposed combining acoustic and visual features for user authentication. They used audio signals to capture 3D facial geometry and fuse these with visual features from CNNs to improve robustness against spoofing attempts. The authors in [55] presented Echo-FAS, an acoustic-based Face Anti-Spoofing system that uses a smartphone's speaker and microphone to emit and capture signals reflecting off the user's face, providing a cost-effective and secure alternative to traditional RGB-based systems. Their experiments show that Echo-FAS achieves nearly 99% AUC performance and can enhance liveness detection when combined with RGB models, mitigating domain gaps effectively. Continuing their previous work [54], the authors introduced M3FAS, a multimodal face anti-spoofing system that combines RGB data with acoustic signals to enhance

the accuracy and robustness of Presentation Attack Detection (PAD). They employ a hierarchical cross-attention module and a multi-head learning strategy, demonstrating through extensive experiments that this system can effectively mitigate overfitting, improve detection performance, and operate flexibly even under challenging conditions like missing modalities or poor-quality inputs. Lastly, the authors in [116] introduced AFace, an authentication system that effectively addresses the shortcomings of previous methods, which often focus on preventing 2D attacks but are vulnerable to 3D spoofing using printed models. AFace utilizes acoustic sensing with an iso-depth model that links acoustic echoes to facial structures, allowing it to distinguish between genuine users and 3D attacks. Its range-adaptive algorithm enhances flexibility by compensating for distance variations.

3.10 Other Two-class FAS Methods

Some other FAS methods have explored various techniques, including live-spoof transition alignment [99] , triplet mining [94, 42, 46], continual learning [93, 10], patch learning [133, 81, 103], and transfer learning [131, 89] and different settings, including test-time adaptation (TTA) [41], source-free domain adaptation (SFDA), and few-shot learning [35] to enhance the effectiveness and robustness of spoof detection.

In [99], the authors proposed to learn the live-to-spoof transition for generalized FAS. Next, in [94], the authors utilized triplet mining to ensure the closeness of liveness features within and across domains in comparison to the distace to spoof features. The authors in [46] focused on extracting domain-invariant features by clustering live features from various domains and by pushing them away from spoof features. Similarly, in [42], the authors focused on clustering live features, and additionally used attack type labels to cluster spoof features to learn a more refined feature representation.

To tackle the challenge of unseen data from new sources, the methods [93, 10] proposed to incorporate continual learning into face authentication systems (FAS) to store domain information while adapting to new domains. In [93], the authors designed a neural network with mechanisms to identify emerging types of attacks by managing confidence for novel inputs and updating the model with new data without losing information on past threats. In [10], the authors introduced novel convolutional adapters to enhance domain adaptation and proposed a contrastive regularization to prevent catastrophic forgetting by leveraging previous domain knowledge through proxy prototypes.

Since facial structures are generally irrelevant to the performance of face authentication systems (FAS) and may even cause difficulties, the methods [133, 81, 103] proposed to focus on image patches so as to alter the facial structure. Specifically, the authors in [133] proposed to permute and remix patches from separate classes and domains so as to ensure reliable extraction

of discriminative features. In [81], the authors introduced to merge patches of transformed images to generate identity-agnostic features for better mining spoof features. Also, in [103], the authors redefined FAS as a fine-grained patch-type recognition system and proposed to capture spoof-related features.

To address FAS in test-time adaptation scenarios, the authors in [41] focused on online adaptation using an off-the-shelf model without accessing to labeled or source data. The 3A-TTA method in [41] proposed to select reliable features and employ these features within an anti-forgetting framework. Additionally, they introduced a novel contrastive learning constraint to enhance the learning of distinct feature representation.

In [35], the authors simulated another real-world scenario and addressed a few-shot domain adaptation. That is, only a few samples of the target domain are available to adapt the model. The authors in [35] proposed using a frozen ViT and ensemble adapters during adaptation to maintain the stability of ViT while promoting the learning of diverse features within the ensemble adapters. To further enhance the training process, they also introduced feature-level augmentation using Feature-Wise Transformation to increase the diversity of training samples.

4 One-class Face Anti-spoofing Methods

In this section, we introduce recent one-class FAS methods that utilize various techniques, such as reconstruction-based feature learning [43, 63], statistical learning [84], and synthetic/generative feature learning [83, 4, 37].

In [43, 63], the authors proposed to reconstruct facial images to learn liveness information. However, due to the absence of facial spoof training images, one-class FAS models may simply learn some general facial features rather than the genuine liveness features. The authors in [84] then proposed adopting Gaussian Mixture Models (GMMs) to learn the distribution of live images by using the features of Image Quality Measures introduced in [113] for learning GMMs. Moreover, the authors in [83, 4] proposed to mix the sampled Gaussian noises with the liveness features of live images to synthesize pseudo spoof features. Nevertheless, since mixing Gaussian noises into the liveness features is far from enough to mimic the spoof latent features, the performance of this one-class FAS method seemed limited. Therefore, the authors in [37] proposed to adopt generative feature learning to generate latent spoof features. In particular, because live images exhibit no spoof cues and spoof images should exhibit visible spoof cues, the authors in [37] proposed to adopt non-zero pseudo spoof cue maps to generate pseudo latent spoof features to facilitate the learning of one-class FAS. Furthermore, the authors in [61] proposed a knowledge distillation approach, where a teacher network trained on large datasets transfers knowledge to a student network which enable effective

one-class learning in FAS by distilling spoof information without requiring explicit spoof examples during training.

5 Face Anti-spoofing Datasets and Evaluation Metrics

In this section, we introduce recent FAS datasets, including both single modality and multi-modality datasets.

5.1 Single Modality FAS Datasets

Table 2 summarizes single modality FAS datasets, which typically consist of RGB images, including both low-resolution RGB images [100, 86, 135, 15, 56, 113] and high-resolution RGB images [135, 87, 18, 76, 85, 82, 9, 58, 77, 44, 79, 102, 132, 2, 126, 72]. Grayscale images [86, 56] are more common in early single modality datasets, while color images [100, 135, 15, 113, 87, 18, 76, 85, 82, 9, 58, 77, 44, 79, 102, 132, 2, 126, 72] are more prevalent in recent single modality datasets. In addition, with the development of facial spoofing attacks, more realistic 3D mask attacks [56, 76, 82, 79, 132, 126] are becoming more popular.

5.2 Multi-modalities FAS Datasets

While FAS methods using RGB images have achieved promising results, single-modality FAS methods still find difficulties in challenging environments, because RGB images are inherently sensitive to varying lighting conditions. Hence, recent multimodal FAS methods adopted multi-modality cameras to capture different modalities, including RGB images, infrared (IR) images, depth images, thermal images, and spectroscopic images, to enhanc security in face recognition systems. As shown in Table 3, the most common modalities in multimodal FAS datasets include RGB images, IR images, and depth images. RGB images, in comparison to depth images, provide more detailed textural information but offer less structural depth information; and in comparison to IR images, RGB images provide richer color information but are more sensitive to illumination changes under various lighting conditions.

Similar to single-modality FAS datasets, the majority of multi-modal FAS datasets include RGB modality [1, 6, 7, 130, 29, 68, 33, 93, 20, 26, 98, 17, 115], and only few exceptions [91, 73] consist of only the light field modality. Also, some of the multi-modality datasets are captured in low-resolution [6, 20, 91, 26, 98, 115] and high-resolution [1, 7, 130, 29, 68, 33, 93, 17, 73] formats.

The second most prevalent modality is IR, including both near infrared (NIR) [1, 6, 7, 130, 29, 68, 33, 93, 17, 115] and shortwave infrared (SWIR) [33, 93, 98]. Depth modality is also widely represented within many datasets

Table 2: Single modality FAS datasets.

Dataset	Year	Image(I)/ Video(V)	High(H)-/ Low(L)-resolution	Color(C)/ Grayscale(G)	Live/ Spoof	Device	Print(P)/Replay(R)/ Waxworks(W) / 3D mask(M) attacks
NUAA[100]	2010	I	L (640 × 480)	C	5105/7509	Webcam	P (flat, wrapped)
YALE_Recaptured[86]	2011	I	L (64×64)	G	640/1920	Kodak C813 Samsung Omnia i900	P (flat)
CASIA-MFSD[135]	2012	V	L (640×480) H (1280×720)	C	150/450	USB camera Sony NEX-5	P (flat, wrapped, cut) R (tablet)
REPLAY-ATTACK[15]	2012	V	L (320×240)	C	200/1000	Apple Macbook Air webcam	P (flat) R (tablet, phone)
Kose and Dugelay[56]	2013	I	L	G	200/198	/	M (hard resin)
MSU-MFSD[113]	2014	V	L (640×480, 720×480)	C	70/210	Apple Macbook Air webcam Google Nexus 5 (frontal)	P (flat) R (tablet, phone)
UVAD[87]	2015	V	H (1366×768)	C	808/16268	Sony CyberShot DSC-HX Canon PowerShot SX1 IS Nikon Coolpix P100 Kodak Z981 Olympus SP 800UZ Panasonic FZ35	R (monitor)
REPLAY-Mobile[18]	2016	V	H (720×1280)	C	390/640	Apple iPad Mini 2 LG G4	P (flat) R (monitor)
HKBU-MARs V2[76]	2016	V	H (1280×720, 800×600, 1920×1080)	C	504/504	Logitech C920 Industrial Camera Canon EOS M3 Google Nexus 5 iPhone 6 Samsung S7 Sony Tablet S	M (hard resin)
MSU USSA[85]	2016	I	H (1280×960, 3264×2448)	C	1140/9120	Google Nexus 5 (frontal, rear) Others (collected online)	P (flat) R (laptop, tablet, phone)
SMAD[82]	2017	V	H	C	65/65	/	M (silicone)
OULU-NPU[9]	2017	V	H (1920×1080)	C	720/2880	Samsung Galaxy S6 edge HTC Desire EYE MEIZU X5 ASUS Zenfone Selfie Sony XPERIA C5 Ultra Dual OPPO N3	P (flat) R (phone)
Rose-Youtu[58]	2018	V	H (640×480, 1280×720)	C	500/2850	Hasee smartphone Huawei smartphone iPad 4 iPhone 5s ZTE smartphone	P (flat) R (monitor, laptop) M (paper)
SiW[77]	2018	V	H (1920×1080, 1280×720)	C	1320/3300	Logitech C920 webcam Canon EOS T6	P (flat, wrapped) R (phone, tablet, monitor)
WFFD[44]	2019	I,V	H	C	2300/2300 (I) 140/145 (V)	Others (collected online)	W (wax)
SiW-M[79]	2019	V	H (1920×1080, 1280×720)	C	660/968	Logitech C920 webcam Canon EOS T6	P (flat) R (phone, tablet, monitor) M (hard resin, plastic, silicone, paper, mannequin) Makeup (cosmetics, impersonation, obfuscation) Partial (glasses, cut paper)
Swax[102]	2020	I,V	H	C	110 (I) 1812 (V)	Others (collected online)	W (wax)
CelebA-Spoof[132]	2020	I	H	C	156384/ 469153	10 sensors	P (flat, wrapped) R (monitor, tablet, phone) M (paper)
RECOD-Mtablet[2]	2020	V	H (1920×1080)	C	450/1800	Moto G5 Moto X Style XT1572	P (flat) R (monitor)
CASIA-SURF 3DMask[126]	2020	V	H (1280×720)	C	288/864	Apple Huawei Samsung	M (plaster)
HiFiMask[72]	2021	V	H	C	13650/40950	iPhone11 iPhoneX MI10 P40 S20 Vivo HJIM	M (transparent, plaster, resin)
SuHiFiMask[22]	2022	V	H	C	10195/101	Surveillance cameras	M(Resin, plaster, silicone, paper)
CelebA-Spoof-Enroll[5]	2022	I	H	C	156384/ 469153	10 sensors	P (flat, wrapped) R (monitor, tablet, phone) M (paper)

Table 3: Multi modality FAS datasets.

Dataset	Year	Image(I)/ Video(V)	High(H)-/Low(L)-resolution	Modalities	Live/ Spoof	Device	Print(P)/Replay(R)/ 3D mask(M) attacks
3DMAD [20]	2014	V	L (640×480)	VIS Depth	170/85	Kinect	M (paper, hard resin)
GUC-LiFFAD [91]	2015	V	L (1080×1080)	Light field	1798/3028	Lytro Light Field Camera	P (flat) R (tablet)
3DFS-DB [26]	2016	V	L (D:640×480, VIS:1280x960)	VIS Depth	260/260	Kinect, Carmine 1.09	M (plastic)
BRSU [98]	2016	I	L (636×508)	VIS SWIR	102/404	/	M (silicon, plastic, resin, latex)
Msspoof [17]	2016	I	H (1280x1024)	VIS NIR	1470/3024	uEye camera	P (flat)
MLFP [1]	2017	V	H (VIS:1280×720) L (NIR:424×512, Thermal:640×480)	VIS NIR Thermal	150/1200	Android smartphones, FLIR ONE, Kinect	M (latex, paper)
ERPA [6]	2017	V	L (640×480)	VIS Depth NIR Thermal	Total 86	Xenics Gobi, Thermal camera, Intel Realsense SR300	P (flat) R (monitor) M (resin, silicone)
CSMAD [7]	2018	V+I	H (VIS:1920×1080) L (Depth, NIR:640×480, Thermal:320×240)	VIS Depth NIR Thermal	104/159	Intel RealSense SR300, Seek Thermal Compact PRO	M (silicone)
LF-SAD [73]	2019	I	H (2450x1634)	Light field	328/596	Lytro ILLUM camera	P (flat, wrapped) R (monitor)
3DMA [115]	2019	V	L (640×480)	VIS NIR	536/384	AuthenMetric binocular camera	M (plastics)
CASIA-SURF [130]	2019	V	H (VIS:1280×720), L (Depth, NIR:640×480)	VIS Depth NIR	3000/18000	Intel RealSense SR300	P (flat, wrapped, cut)
WMCA [29]	2019	V	H (1920×1080, 1260×720) L (Thermal:320×240)	VIS Depth NIR Thermal	347/1332	Intel RealSense SR300, Seek Thermal Compact PRO	P (flat) R (tablet) M (plastic, silicone, paper, mannequin) Partial (glasses)
HQ-WMCA [33]	2020	V	H (1920×1200)	VIS Depth NIR SWIR Thermal	555/2349	/	P (flat) R (tablet, phone) M (plastic, silicon, paper, mannequin) Makeup Partial (glasses, wigs, tatoo)
CeFA [68]	2021	V	H (1280×720)	VIS Depth NIR	6300/27900	Intel RealSense	P (flat, wrapped) R M (print, silica gel)
PADISI [93]	2021	V	H (1984×1264)	VIS Depth NIR SWIR Thermal	1105/924	/	P (flat) R (tablet, phone) M (plastic, silicon, transparent, mannequin) Makeup Partial (glasses,funny eye, tatoo)

because of its effectiveness in tackling replay and print attacks [6, 7, 130, 29, 68, 33, 93, 20, 26]. Another popular modality is thermal imaging, which is difficult to emulate on most surfaces due to skin temperature variations [1, 6, 7, 29, 33, 93].

Compared to single-modality datasets, multi-modality datasets provide more comprehensive information and enable more reliable FAS detection. Consequently, most multi-modality datasets are prepared to handle a wider variety of attack types, enhancing their robustness and effectiveness in real-world applications. Notably, datasets such as [1, 6, 29, 68, 33, 93, 20, 26, 98, 115] include more challenging 3D mask attacks. Additionally, the datasets [29, 33, 93] feature partial attacks, and the dataset [29] even includes an additional makeup attack.

5.3 Evaluation Metrics

In the field of face anti-spoofing (FAS), several key metrics are used to evaluate how effectively systems can differentiate between real users and spoofing attempts. Two foundational measures are the False Rejection Rate (FRR)[16], which tracks the percentage of legitimate users wrongly rejected, and the False Acceptance Rate (FAR) [25], which shows how often spoof attacks are incorrectly accepted. Common metrics for both intra- and cross-dataset evaluations include the Half Total Error Rate (HTER)[16], Equal Error Rate (EER)[92], and Area Under the Curve (AUC). HTER is calculated as the average of FRR and FAR, while EER identifies the point where these two rates are equal. AUC is used to assess the capability of the model to separate genuine users from spoof attacks across various thresholds. In addition, more specific metrics like Attack Presentation Classification Error Rate (APCER) and Bonafide Presentation Classification Error Rate (BPCER) are increasingly employed. APCER measures the error rate for incorrectly classifying spoof attempts as genuine, while BPCER focuses on errors in rejecting legitimate users. The Average Classification Error Rate (ACER), which combines APCER and BPCER, offers a useful overall performance indicator for intra-dataset testing scenarios [9, 77].

6 Conclusion and Research Directions

Before concluding this survey, we first point out some potential future research directions.

6.1 Language Guidance for FAS

As discussed in Section 3.1.3, recent advancements in CLIP [90] have significantly enhanced understanding of semantic relationships between visual and textual inputs. The works [71], [21] and [97] demonstrate promising results in FAS. Nevertheless, although their implementations maintain the main benefits of CLIP on object identification, liveness-related textual information was yet not explicitly explored on the pretrained CLIP. Future research could focus on extracting live/spoof-discriminative information from the pretrained CLIP model to enhance its capability in distinguishing between live and spoof semantic content. Future research could also involve developing methods to integrate additional textual features or fine-tuning strategies to better capture those subtle features to improve the overall performance in FAS.

6.2 Test-time Adaptation for FAS

When labeled source data are unavailable in the adaptation stage, test-time adaptation provides a more practical real-world scenario. This setting is particularly useful for existing pre-trained face recognition systems. When a system is able to adjust dynamically to new data without model retraining or updating, its flexibility and robustness to attack can be significantly improved. Such adaptability is also key to preventing unseen attacks and ensuring its continuous effectiveness on detecting new threats. Additionally, test-time adaptation may also reduce personal privacy risks, as each face image is only processed once and will not be stored, thereby minimizing data exposure. This adaption does not only mitigate the risks associated with static models but also reduce the operational overhead associated with frequent updates. There exits only few work on fully test-time FAS [41] and many related issues remain unexplored in this setting.

6.3 Conclusion

This article conducts a comprehensive survey on recent deep learning-based FAS approaches and introduces public FAS benchmark datasets, and different settings. We have provided a detailed taxonomy of these methods and have given in-depth discussion. Additionally, we have outlined some potential research direction in this field.

References

[1] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore, "Face presentation attack with latex masks in multispectral videos", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, 81–9.

[2] W. R. Almeida, F. A. Andaló, R. Padilha, G. Bertocco, W. Dias, R. d. S. Torres, J. Wainer, and A. Rocha, "Detecting face presentation attacks in mobile devices with a patch-based CNN and a sensor-aware loss function", *PloS one*, 15(9), 2020, e0238058.

[3] P. Anthony, B. Ay, and G. Aydin, "A review of face anti-spoofing methods for face recognition systems", in *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, IEEE, 2021, 1–9.

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

[5] D. Belli, D. Das, B. Major, and F. Porikli, "A personalized benchmark for face anti-spoofing", in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, 338–48.

[6] S. Bhattacharjee and S. Marcel, "What you can't see can help you-extended-range imaging for 3d-mask presentation attack detection", in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, IEEE, 2017, 1–7.

[7] S. Bhattacharjee, A. Mohammadi, and S. Marcel, "Spoofing deep face recognition with custom silicone masks", in *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, IEEE, 2018, 1–7.

[8] Y. Bian, P. Zhang, J. Wang, C. Wang, and S. Pu, "Learning multiple explainable and generalizable cues for face anti-spoofing", in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, 2310–4.

[9] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "Oulu-npu: A mobile face presentation attack database with real-world variations", in *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, IEEE, 2017, 612–8.

[10] R. Cai, Y. Cui, Z. Li, Z. Yu, H. Li, Y. Hu, and A. Kot, "Rehearsal-free domain continual face anti-spoofing: Generalize more and forget less", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, 8037–48.

[11] R. Cai, Z. Li, R. Wan, H. Li, Y. Hu, and A. C. Kot, "Learning meta pattern for face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 17, 2022, 1201–13.

[12] R. Cai, Z. Yu, C. Kong, H. Li, C. Chen, Y. Hu, and A. C. Kot, "S-adapter: Generalizing vision transformer for face anti-spoofing with statistical tokens", *IEEE Transactions on Information Forensics and Security*, 2024.

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

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

[15] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing", in *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, IEEE, 2012, 1–7.

[16] I. Chingovska, A. R. Dos Anjos, and S. Marcel, "Biometrics evaluation under spoofing attacks", *IEEE transactions on Information Forensics and Security*, 9(12), 2014, 2264–76.

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

[18] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database", in *2016 international conference of the Biometrics Special Interest Group (BIOSIG)*, IEEE, 2016, 1–7.

[19] P. Deng, C. Ge, H. Wei, Y. Sun, and X. Qiao, "Attention-aware dual-stream network for multimodal face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 2023.

[20] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3D masks", *IEEE transactions on information forensics and security*, 9(7), 2014, 1084–97.

[21] H. Fang, A. Liu, N. Jiang, Q. Lu, G. Zhao, and J. Wan, "VL-FAS: Domain Generalization via Vision-Language Model For Face Anti-Spoofing", in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2024, 4770–4.

[22] H. Fang, A. Liu, J. Wan, S. Escalera, C. Zhao, X. Zhang, S. Z. Li, and Z. Lei, "Surveillance face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 2023.

[23] H. Fang, A. Liu, H. Yuan, J. Zheng, D. Zeng, Y. Liu, J. Deng, S. Escalera, X. Liu, J. Wan, *et al.*, "Unified physical-digital face attack detection", *arXiv preprint arXiv:2401.17699*, 2024.

[24] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper, "Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection", in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, 3722–31.

[25] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features", *Future Generation Computer Systems*, 28(1), 2012, 311–21.

[26] J. Galbally and R. Satta, "Three-dimensional and two-and-a-half-dimensional face recognition spoofing using three-dimensional printed models", *IET Biometrics*, 5(2), 2016, 83–91.

[27] A. George and S. Marcel, "Cross modal focal loss for rgbd face anti-spoofing", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, 7882–91.

[28] A. George and S. Marcel, "On the effectiveness of vision transformers for zero-shot face anti-spoofing", 2021.

[29] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network", *IEEE transactions on information forensics and security*, 15, 2019, 42–55.

[30] X. Guo, Y. Liu, A. Jain, and X. Liu, "Multi-domain learning for updating face anti-spoofing models", in *European Conference on Computer Vision*, Springer, 2022, 230–49.

[31] S. Han, R. Cai, Y. Cui, Z. Yu, Y. Hu, and A. Kot, "Hyperbolic face anti-spoofing", *arXiv preprint arXiv:2308.09107*, 2023.

[32] X. He, D. Liang, S. Yang, Z. Hao, H. Ma, B. Mao, X. Li, Y. Wang, P. Yan, and A. Liu, "Joint Physical-Digital Facial Attack Detection Via Simulating Spoofing Clues", *arXiv preprint arXiv:2404.08450*, 2024.

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

[34] H. Huang, Y. Xiang, G. Yang, L. Lv, X. Li, Z. Weng, and Y. Fu, "Generalized face anti-spoofing via cross-adversarial disentanglement with mixing augmentation", in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, 2939–43.

[35] H.-P. Huang, D. Sun, Y. Liu, W.-S. Chu, T. Xiao, J. Yuan, H. Adam, and M.-H. Yang, "Adaptive transformers for robust few-shot cross-domain face anti-spoofing", in *European Conference on Computer Vision*, Springer, 2022, 37–54.

[36] P.-K. Huang, C.-L. Chang, H.-Y. Ni, and C.-T. Hsu, "Learning to augment face presentation attack dataset via disentangled feature learning from limited spoof data", in *2022 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2022, 1–6.

[37] P.-K. Huang, C.-H. Chiang, T.-H. Chen, J.-X. Chong, T.-L. Liu, and C.-T. Hsu, "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 (CVPR)*, 2024.

[38] P.-K. Huang, C.-H. Chiang, J.-X. Chong, T.-H. Chen, H.-Y. Ni, and C.-T. Hsu, "LDCformer: Incorporating Learnable Descriptive Convolution to Vision Transformer for Face Anti-Spoofing", in *2023 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2023, 121–5.

[39] P.-K. Huang, M.-C. Chin, and C.-T. Hsu, "Face anti-spoofing via robust auxiliary estimation and discriminative feature learning", in *Asian Conference on Pattern Recognition*, Springer, 2021, 443–58.

[40] P.-K. Huang, J.-X. Chong, H.-Y. Ni, T.-H. Chen, and C.-T. Hsu, "Towards diverse liveness feature representation and domain expansion for cross-domain face anti-spoofing", in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2023, 1199–204.

[41] P.-K. Huang, C.-Y. Lu, S.-J. Chang, J.-X. Chong, and C.-T. Hsu, "Test-Time Adaptation for Robust Face Anti-Spoofing", in *BMVC*, 2023.

[42] P.-K. Huang, H.-Y. Ni, Y. Ni, and C.-T. Hsu, "Learnable Descriptive Convolutional Network for Face Anti-Spoofing.", in *BMVC*, 2022, 239.

[43] X. Huang, J. Xia, and L. Shen, "One-class face anti-spoofing based on attention auto-encoder", in *Biometric Recognition: 15th Chinese Conference, CCBR 2021, Shanghai, China, September 10–12, 2021, Proceedings 15*, Springer, 2021, 365–73.

[44] S. Jia, X. Li, C. Hu, G. Guo, and Z. Xu, "3D face anti-spoofing with factorized bilinear coding", *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10), 2020, 4031–45.

[45] Y. Jia, J. Zhang, and S. Shan, "Dual-branch meta-learning network with distribution alignment for face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 17, 2021, 138–51.

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

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

[48] F. Jiang, Q. Li, P. Liu, X.-D. Zhou, and Z. Sun, "Adversarial learning domain-invariant conditional features for robust face anti-spoofing", *International Journal of Computer Vision*, 131(7), 2023, 1680–703.

[49] J. Jiang and Y. Sun, "Depth-based ensemble learning network for face anti-spoofing", in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, 2954–8.

[50] A. Jourabloo, Y. Liu, and X. Liu, "Face de-spoofing: Anti-spoofing via noise modeling", in *Proceedings of the European conference on computer vision (ECCV)*, 2018, 290–306.

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

[52] Y.-E. Kim, W.-J. Nam, K. Min, and S.-W. Lee, "Style selective normalization with meta learning for test-time adaptive face anti-spoofing", *Expert Systems with Applications*, 214, 2023, 119106.

[53] C. Kong, S. Wang, and H. Li, "Digital and Physical Face Attacks: Reviewing and One Step Further", *APSIPA Transactions on Signal and Information Processing*, 12(1), 2023, e25.

[54] C. Kong, K. Zheng, Y. Liu, S. Wang, A. Rocha, and H. Li, "M^3FAS: An Accurate and Robust MultiModal Mobile Face Anti-Spoofing System", *IEEE Transactions on Dependable and Secure Computing*, 2024.

[55] C. Kong, K. Zheng, S. Wang, A. Rocha, and H. Li, "Beyond the pixel world: A novel acoustic-based face anti-spoofing system for smartphones", *IEEE Transactions on Information Forensics and Security*, 17, 2022, 3238–53.

[56] N. Kose and J.-L. Dugelay, "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*, 2013, 111–6.

[57] D. Li, G. Chen, X. Wu, Z. Yu, and M. Tan, "Face anti-spoofing with cross-stage relation enhancement and spoof material perception", *Neural Networks*, 175, 2024, 106275.

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

[59] K. Li, H. Yang, B. Chen, P. Li, B. Wang, and D. Huang, "Learning polysemantic spoof trace: A multi-modal disentanglement network for face anti-spoofing", in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, No. 1, 2023, 1351–9.

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

[61] Z. Li, R. Cai, H. Li, K.-Y. Lam, Y. Hu, and A. C. Kot, "One-class knowledge distillation for face presentation attack detection", *IEEE Transactions on Information Forensics and Security*, 17, 2022, 2137–50.

[62] C.-H. Liao, W.-C. Chen, H.-T. Liu, Y.-R. Yeh, M.-C. Hu, and C.-S. Chen, "Domain invariant vision transformer learning for face anti-spoofing", in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, 6098–107.

[63] S. Lim, Y. Gwak, W. Kim, J.-H. Roh, and S. Cho, "One-class learning method based on live correlation loss for face anti-spoofing", in, Vol. 8, IEEE, 2020, 201635–48.

[64] X. Lin, S. Wang, R. Cai, Y. Liu, Y. Fu, Z. Yu, W. Tang, and A. Kot, "Suppress and Rebalance: Towards Generalized Multi-Modal Face Anti-Spoofing", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[65] A. Liu, "CA-MoEiT: Generalizable Face Anti-spoofing via Dual Cross-Attention and Semi-fixed Mixture-of-Expert", *International Journal of Computer Vision*, 2024, 1–14.

[66] A. Liu, M. Hui, J. Zheng, H. Yuan, X. Yu, Y. Liang, S. Escalera, J. Wan, and Z. Lei, "FM-CLIP: Flexible Modal CLIP for Face Anti-Spoofing", in *ACM Multimedia 2024*, 2024.

[67] A. Liu and Y. Liang, "Ma-vit: Modality-agnostic vision transformers for face anti-spoofing", *arXiv preprint arXiv:2304.07549*, 2023.

[68] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, and S. Z. Li, "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*, 2021, 1179–87.

[69] A. Liu, Z. Tan, J. Wan, Y. Liang, Z. Lei, G. Guo, and S. Z. Li, "Face anti-spoofing via adversarial cross-modality translation", *IEEE Transactions on Information Forensics and Security*, 16, 2021, 2759–72.

[70] A. Liu, Z. Tan, Z. Yu, C. Zhao, J. Wan, Y. L. Z. Lei, D. Zhang, S. Z. Li, and G. Guo, "Fm-vit: Flexible modal vision transformers for face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 2023.

[71] A. Liu, S. Xue, J. Gan, J. Wan, Y. Liang, J. Deng, S. Escalera, and Z. Lei, "CFPL-FAS: Class Free Prompt Learning for Generalizable Face Anti-spoofing", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

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

[73] M. Liu, H. Fu, Y. Wei, Y. A. U. Rehman, L.-m. Po, and W. L. Lo, "Light field-based face liveness detection with convolutional neural networks", *Journal of Electronic Imaging*, 28(1), 2019, 13003–3.

[74] S.-Q. Liu, X. Lan, and P. C. Yuen, "Learning temporal similarity of remote photoplethysmography for fast 3d mask face presentation attack detection", *IEEE Transactions on Information Forensics and Security*, 17, 2022, 3195–210.

[75] S.-Q. Liu, X. Lan, and P. C. Yuen, "Remote Photoplethysmography Correspondence Feature for 3D Mask Face Presentation Attack Detection", in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[76] S. Liu, P. C. Yuen, S. Zhang, and G. Zhao, "3D mask face anti-spoofing with remote photoplethysmography", in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, Springer, 2016, 85–100.

[77] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, 389–98.

[78] Y. Liu and X. Liu, "Spoof trace disentanglement for generic face anti-spoofing", in, Vol. 45, No. 3, IEEE, 2022, 3813–30.

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

[80] Y. Liu, J. Stehouwer, and X. Liu, "On disentangling spoof trace for generic face anti-spoofing", in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, Springer, 2020, 406–22.

[81] Y. Liu, Y. Chen, M. Gou, C.-T. Huang, Y. Wang, W. Dai, and H. Xiong, "Towards unsupervised domain generalization for face anti-spoofing", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, 20654–64.

[82] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar, "Detecting silicone mask-based presentation attack via deep dictionary learning", *IEEE Transactions on Information Forensics and Security*, 12(7), 2017, 1713–23.

[83] K. Narayan and V. M. Patel, "Hyp-OC: Hyperbolic One Class Classification for Face Anti-Spoofing", *arXiv preprint arXiv:2404.14406*, 2024.

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

[85] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones", *IEEE transactions on information forensics and security*, 11(10), 2016, 2268–83.

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

[87] A. Pinto, W. R. Schwartz, H. Pedrini, and A. de Rezende Rocha, "Using visual rhythms for detecting video-based facial spoof attacks", *IEEE Transactions on Information Forensics and Security*, 10(5), 2015, 1025–38.

[88] Y. Qin, Z. Yu, L. Yan, Z. Wang, C. Zhao, and Z. Lei, "Meta-teacher for face anti-spoofing", *IEEE transactions on pattern analysis and machine intelligence*, 44(10), 2021, 6311–26.

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

[90] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision", in *International conference on machine learning*, PMLR, 2021, 8748–63.

[91] R. Raghavendra, K. B. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera", *IEEE Transactions on Image Processing*, 24(3), 2015, 1060–75.

[92] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey", *ACM Computing Surveys (CSUR)*, 50(1), 2017, 1–37.

[93] M. Rostami, L. Spinoulas, M. Hussein, J. Mathai, and W. Abd-Almageed, "Detection and continual learning of novel face presentation attacks", in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, 14851–60.

[94] R. Shao, X. Lan, J. Li, and P. C. Yuen, "Multi-adversarial discriminative deep domain generalization for face presentation attack detection", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, 10023–31.

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

[96] Y. Shi, Y. Gao, Y. Lai, H. Wang, J. Feng, L. He, J. Wan, C. Chen, Z. Yu, and X. Cao, "Shield: An evaluation benchmark for face spoofing and forgery detection with multimodal large language models", *arXiv preprint arXiv:2402.04178*, 2024.

[97] K. Srivatsan, M. Naseer, and K. Nandakumar, "Flip: Cross-domain face anti-spoofing with language guidance", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, 19685–96.

[98] H. Steiner, A. Kolb, and N. Jung, "Reliable face anti-spoofing using multispectral swir imaging", in *2016 international conference on biometrics (ICB)*, IEEE, 2016, 1–8.

[99] Y. Sun, Y. Liu, X. Liu, Y. Li, and W.-S. Chu, "Rethinking domain generalization for face anti-spoofing: Separability and alignment", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, 24563–74.

[100] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model", in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI 11*, Springer, 2010, 504–17.

[101] Y. Tian, Y. Huang, K. Zhang, Y. Liu, and Z. Sun, "Polarized Image Translation from Nonpolarized Cameras for Multimodal Face Anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 2023.

[102] R. H. Vareto, A. M. Saldanha, and W. R. Schwartz, "The swax benchmark: attacking biometric systems with wax figures", in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, 986–90.

[103] C.-Y. Wang, Y.-D. Lu, S.-T. Yang, and S.-H. Lai, "Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 20281–90.

[104] Y.-C. Wang, C.-Y. Wang, and S.-H. Lai, "Disentangled representation with dual-stage feature learning for face anti-spoofing", in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, 1955–64.

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

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

[107] W. Wang, P. Liu, H. Zheng, R. Ying, and F. Wen, "Domain generalization for face anti-spoofing via negative data augmentation", *IEEE Transactions on Information Forensics and Security*, 2023.

[108] W. Wang, F. Wen, H. Zheng, R. Ying, and P. Liu, "Conv-MLP: A convolution and MLP mixed model for multimodal face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 17, 2022, 2284–97.

[109] Z. Wang, Z. Yu, C. Zhao, X. Zhu, Y. Qin, Q. Zhou, F. Zhou, and Z. Lei, "Deep spatial gradient and temporal depth learning for face anti-spoofing", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, 5042–51.

[110] Z. Wang, Q. Wang, W. Deng, and G. Guo, "Face anti-spoofing using transformers with relation-aware mechanism", *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3), 2022, 439–50.

[111] Z. Wang, Q. Wang, W. Deng, and G. Guo, "Learning multi-granularity temporal characteristics for face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 17, 2022, 1254–69.

[112] Z. Wang, Z. Wang, Z. Yu, W. Deng, J. Li, T. Gao, and Z. Wang, "Domain generalization via shuffled style assembly for face anti-spoofing", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 4123–33.

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

[114] H. Wu, D. Zeng, Y. Hu, H. Shi, and T. Mei, "Dual spoof disentanglement generation for face anti-spoofing with depth uncertainty learning", *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7), 2021, 4626–38.

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

[116] Z. Xu, T. Liu, R. Jiang, P. Hu, Z. Guo, and C. Liu, "AFace: Range-flexible Anti-spoofing Face Authentication via Smartphone Acoustic Sensing", *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(1), 2024, 1–33.

[117] J. Yang, Z. Yu, X. Ni, J. He, and H. Li, "Generalized Face Anti-spoofing via Finer Domain Partition and Disentangling Liveness-irrelevant Factors", *arXiv preprint arXiv:2407.08243*, 2024.

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

[119] Z. Yu, R. Cai, Y. Cui, A. Liu, and C. Chen, "Visual prompt flexible-modal face anti-spoofing", *arXiv preprint arXiv:2307.13958*, 2023.

[120] Z. Yu, R. Cai, Y. Cui, X. Liu, Y. Hu, and A. C. Kot, "Rethinking vision transformer and masked autoencoder in multimodal face anti-spoofing", *International Journal of Computer Vision*, 2024, 1–22.

[121] Z. Yu, R. Cai, Z. Li, W. Yang, J. Shi, and A. C. Kot, "Benchmarking joint face spoofing and forgery detection with visual and physiological cues", *IEEE Transactions on Dependable and Secure Computing*, 2024.

[122] Z. Yu, X. Li, X. Niu, J. Shi, and G. Zhao, "Face anti-spoofing with human material perception", in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, Springer, 2020, 557–75.

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

[124] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep learning for face anti-spoofing: A survey", *IEEE transactions on pattern analysis and machine intelligence*, 45(5), 2022, 5609–31.

[125] Z. Yu, Y. Qin, H. Zhao, X. Li, and G. Zhao, "Dual-Cross Central Difference Network for Face Anti-Spoofing", in *IJCAI International Joint Conference on Artificial Intelligence*, 2021.

[126] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, and G. Zhao, "Nas-fas: Static-dynamic central difference network search for face anti-spoofing", *IEEE transactions on pattern analysis and machine intelligence*, 43(9), 2020, 3005–23.

[127] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, "Searching central difference convolutional networks for face anti-spoofing", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, 5295–305.

[128] H. Yuan, A. Liu, J. Zheng, J. Wan, J. Deng, S. Escalera, H. J. Escalante, I. Guyon, and Z. Lei, "Unified physical-digital attack detection challenge", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, 919–29.

[129] H. Yue, K. Wang, G. Zhang, H. Feng, J. Han, E. Ding, and J. Wang, "Cyclically disentangled feature translation for face anti-spoofing", in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, No. 3, 2023, 3358–66.

[130] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, and S. Z. Li, "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*, 2019, 919–28.

[131] W. Zhang, H. Liu, F. Liu, R. Ramachandra, and C. Busch, "Effective presentation attack detection driven by face related task", in *European Conference on Computer Vision*, Springer, 2022, 408–23.

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

[133] K.-Y. Zhang, T. Yao, J. Zhang, S. Liu, B. Yin, S. Ding, and J. Li, "Structure destruction and content combination for face anti-spoofing", in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2021, 1–6.

[134] K.-Y. Zhang, T. Yao, J. Zhang, Y. Tai, S. Ding, J. Li, F. Huang, H. Song, and L. Ma, "Face anti-spoofing via disentangled representation learning", in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*, Springer, 2020, 641–57.

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

[136] B. Zhou, Z. Xie, and F. Ye, "Multi-modal face authentication using deep visual and acoustic features", in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, IEEE, 2019, 1–6.

[137] L. Zhou, J. Luo, X. Gao, W. Li, B. Lei, and J. Leng, "Selective domain-invariant feature alignment network for face anti-spoofing", *IEEE Transactions on Information Forensics and Security*, 16, 2021, 5352–65.

[138] Q. Zhou, K.-Y. Zhang, T. Yao, X. Lu, R. Yi, S. Ding, and L. Ma, "Instance-aware domain generalization for face anti-spoofing", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, 20453–63.

[139] H. Zou, C. Du, H. Zhang, Y. Zhang, A. Liu, J. Wan, and Z. Lei, "La-SoftMoE CLIP for Unified Physical-Digital Face Attack Detection", *arXiv preprint arXiv:2408.12793*, 2024.