IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 May 2024, accepted 27 June 2024, date of publication 1 July 2024, date of current version 18 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3421345

RESEARCH ARTICLE

Quality Matters: Boosting Face Presentation Attack Detection With Image Quality Metrics

XINWEI LIU[1], RENFANG WANG[1], WEN LIU[2], LIANGBIN ZHANG[1], AND XIAOXIA WANG[3]
[1]Department of Computer Science, College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315104, China
[2]DMJ Lab, College of Information and Intelligent Engineering, Zhejiang Wanli University, Ningbo 315104, China
[3]Library and Learning Resource Center, Zhejiang Wanli University, Ningbo 315104, China

Corresponding author: Renfang Wang (renfang_wangac@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62106228, in part by Zhejiang Natural Science Foundation under Grant LQ22F020003, in part by the Public Welfare Technology Research Program of Zhejiang Provincial under Grant LGF20F020005 and Grant LGG22E050042, in part by Ningbo Natural Science Foundation under Grant 2021J175, in part by Ningbo Key Research and Development Plan under Grant 2023Z035 and Grant 2023Z003, and in part by Ningbo Yongjiang Talent Introduction Program 2021.

ABSTRACT Face Presentation Attack Detection (PAD) is critical for enhancing the security of facial recognition systems against sophisticated attacks. This study explores the use of general Image Quality Assessment (IQA) methods in face PAD, offering an alternative strategy that deviates from traditional, face-specific PAD techniques. Our evaluation of eight widely-used IQA methods across four PAD databases is structured around three distinct experimental protocols. Preliminary findings indicate that the general IQA methods are not fully effective in differentiating between genuine and attack samples, highlighting the need for modification. Nonetheless, a notable enhancement in performance is observed following the re-training of these methods using PAD datasets, bringing their effectiveness in line with that of advanced traditional PAD methods. This study provides evidence for the potential of general IQA in bolstering the resilience of face recognition systems against presentation attacks.

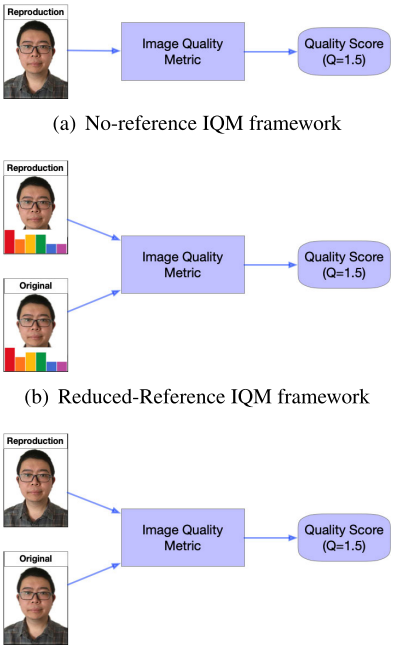
INDEX TERMS Biometrics, face recognition, image quality assessment, presentation attack detection.

I. INTRODUCTION

Within the scope of image processing, the evaluation of image quality is a critical aspect with a variety of practical applications. Image Quality Metrics (IQMs) serve as essential tools for this purpose, providing a means to quantitatively assess the fidelity of digital images. A multitude of IQMs exists, specifically tailored for gauging the quality of natural images [1]. These metrics are typically classified into three distinct categories based on the need for a reference image: full-reference, reduced-reference, and no-reference IQMs [2]. Full-reference IQMs rely on comparing the input images to pristine reference images to determine quality, whereas reduced-reference IQMs require supplementary information about the reference images for their calculations. On the other hand, no-reference IQMs perform quality estimation solely based on the input images without the need for any reference data. A visual representation of these three categories of IQMs is depicted in Figure 1.

Beyond the evaluation of natural images, IQMs have been extensively utilized across various fields, encompassing medical imaging, satellite imagery, and biometric technologies, among others. Ongoing studies in biometrics consistently underscore the significance of the quality of samples within such systems, as it directly influences the effectiveness of the biometric system [3]. In this context, the development of methodologies for the assessment of biometric sample quality has gained momentum, alongside the establishment of ISO/IEC standards that define quality parameters for biometric data like fingerprints, irises, and facial images [43], [44], [45], [46]. The principal objective of these quality assessment methods is to determine the reliability and effectiveness of biometric samples [4]. Given that biometric systems typically operate without a reference image, the application of no-reference IQMs becomes essential for the evaluation of the quality of biometric sample imagery.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Jin.



(a) No-reference IQM framework

(b) Reduced-Reference IQM framework

(c) Full-reference IQM framework

FIGURE 1. General frameworks of no-reference, reduced-reference, and full-reference IQM.

The ability to recognize faces is a daily activity that we perform almost effortlessly, thanks to its non-intrusive nature. The clarity and completeness of face images are crucial for the effectiveness of face recognition technologies, as various factors can affect the quality of the captured facial images. While Face Image Quality Assessment (FIQA) has similarities to the broader Image Quality Assessment (IQA) domain, significant differences set them apart. FIQA is a specialized subset within the larger realm of IQA, which is a widely explored area in the field of image processing. Unlike IQA, which is concerned with evaluating the overall quality of natural images, FIQA is tailored to assess the specific attributes and characteristics of faces.

The rapid progress in FIQA and the widespread use of high-quality capture devices like smartphones have increased the accessibility of clear personal face images in public domains. This widespread accessibility may significantly increase potential security threats to face recognition systems [5]. In light of this, presentation attacks—where an individual presents a fraudulent representation to a face recognition system with the intent to gain unauthorized access—pose a significant vulnerability [6]. In this context, the role of face PAD is essential; it aims to distinguish between live subject images (bona fide) and counterfeit images created through various presentation attack techniques, such as printed photographs, recorded videos, or carefully crafted replicas of authorized faces. Despite being a prevalent form of threat in the face recognition process, detecting presentation attacks remains a formidable challenge. This is due to the difficulty in developing a versatile method that can effectively adapt to diverse acquisition devices, environmental settings, and categories of attacks. Face recognition systems are increasingly being targeted by sophisticated presentation attacks, emphasizing the critical need for robust PAD techniques. While specialized PAD methods have been developed, they often struggle with generalizability and adaptability to new and emerging attack vectors.

Many face recognition systems adopt a dual-phase strategy, as shown in Figure 2, which is designed to control the quality of input face images and detect any potential presentation attacks. This mechanism allows for the rejection of face images at an early stage if they are of substandard quality or if an attack is suspected. The implementation of both FIQA and PAD is instrumental in enhancing the accuracy and security of face recognition systems. Extensive research has demonstrated discernible differences in image quality between genuine and counterfeit face images [7], [8], [10]. Much like FIQA, traditional PAD techniques are typically specific to a single biometric modality, such as face recognition. Moreover, existing PAD methods, while effective in certain contexts, have several limitations. For instance, many rely on feature sets that are tailored to specific types of attacks, reducing their effectiveness when faced with novel or unknown threats. Additionally, specialized PAD techniques often require significant computational resources and may not be feasible for real-time applications. In contrast, recognizing that variations in image quality can serve as a vital clue for detecting face presentation attacks opens up the possibility of employing generalized methods, like general IQMs initially developed for natural images, to identify such attacks. General IQMs have been designed to work across a wide range of images and conditions, making them well-suited for the diverse and evolving nature of presentation attacks. Furthermore, IQMs can be more computationally efficient, allowing for faster processing times that are crucial for real-time systems. This approach has the potential to be expanded beyond facial recognition to encompass multiple biometric modalities, such as fingerprints and iris scans, among others.

In this study, we commence by examining the efficacy of general IQMs in their unaltered state for the task of face PAD. Subsequently, we select a subset of these general IQMs and re-train them using PAD-specific face images to evaluate their modified impact when integrated with advanced PAD methodologies. Our analysis is twofold: firstly, we assess the capability of general IQMs to differentiate between authentic and attack face samples through the allocation of distinct quality scores; secondly, we scrutinize the impact of eliminating both genuine and attack face images from the recognition process based on the outcomes of the quality assessment. Furthermore, we investigate whether the efficacy of PAD algorithms is augmented by the exclusion of low-quality face samples as determined by general IQMs. Lastly, we explore the potential for score-level fusion to further refine the performance of IQMs in the context of face PAD.

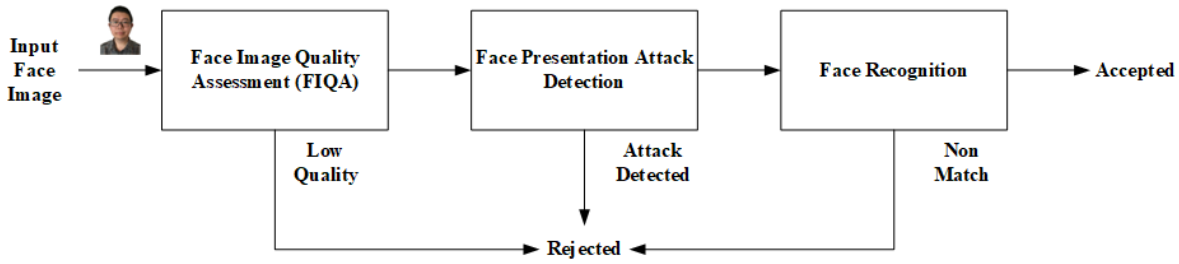


FIGURE 2. Face image quality control and presentation attack detection before recognition.

Our study leverages the broad applicability and computational efficiency of general IQMs to address the shortcomings of existing PAD methods. By retraining these metrics with PAD-specific datasets, we aim to enhance their ability to distinguish between bona fide and attack samples, thereby improving the resilience of face recognition systems against a wide array of attacks. The novelty of this manuscript lies in harnessing the power of general IQMs for PAD, offering a unique solution that combines the strengths of both worlds. This innovative research topic has the potential to significantly advance the field of biometric security by providing a more robust and adaptable PAD solution. The main contributions of this manuscript are:

- Our experiments reveal that the initial versions of general IQMs struggle to effectively differentiate between genuine and presentation attack face images;
- Upon re-training with face images specifically collected for PAD, the performance of these general IQMs aligns with that of the leading face PAD techniques;
- The preliminary application of general purpose IQMs to filter out low-quality face images prior to PAD enhances both the accuracy and the security of face recognition systems;
- The fusion of scores from multiple IQMs not only leverages the advantages of each metric but also mitigates their individual shortcomings, thereby bolstering the effectiveness of face PAD.

The layout of this manuscript proceeds as follows: Section II provides an overview of existing literature and research related to conventional methods of face PAD, the utilization of IQMs in PAD, and the pertinent databases used in these studies. Section III details the experimental configuration, encompassing the chosen IQMs, traditional PAD techniques, and the databases employed for our analysis. Section IV presents the findings from our experiments, along with a thorough discussion of these results. Section V concludes the manuscript with a summary of our conclusions and an outlook on future research directions.

II. STATE-OF-THE-ART OF FACE PAD

The primary objective of face PAD is to ascertain whether the characteristics of a given face image originate from a living individual or an artificial representation. Presentation attacks have the potential to flood the face recognition system with a multitude of unauthorized face images, thereby adversely impacting the system's efficacy. By effectively filtering out these unauthorized images, PAD techniques help to maintain a stringent security standard for face recognition systems. The precision of PAD is a crucial indicator of the reliability of a face recognition system's security in the face of presentation attacks, making it a critical and formidable challenge. Over recent years, a plethora of research has conducted systematic and exhaustive examinations of the dynamics between face presentation attacks and PAD [19], [20], [21], [22], [23], [24], [25]. These investigations underscore the necessity of PAD as a fundamental component before the widespread adoption of face recognition technology in everyday applications. While numerous PAD methods have been proposed, many are tailored to specific attack types and perform poorly when generalized to unseen attacks [19], [21]. This lack of generalizability limits their applicability in diverse real-world scenarios. Several state-of-the-art PAD techniques, such as those relying on deep learning models, require significant computational resources [25]. These high demands can be prohibitive for deployment in low-resource environments, such as mobile devices or remote sensing systems. Some PAD methods are highly sensitive to environmental conditions, such as lighting variations or background noise, which can significantly impact their performance [22], [23]. This sensitivity limits their reliability in uncontrolled settings.

IQMs can be used for face PAD is based on the assumption that:" *It is expected that a presentation attack face image acquired in an attack attempt can have image quality differences compared to a bona fide face sample image captured in the normal acquisition protocol for which the device was designed.*" Such an assumption has been proved to be true by many studies in recent years [7], [8], [9], [10]. The above mentioned quality differences can be the following: sharpness level, intensity of brightness, naturalness of color, global and local distortion artifacts, absence of information or the appearance of irregular information, and so on. For instance, a face sample image taken from the screen of a smartphone is more likely to have out of focus or reflection artifacts during the reproduction processes. In addition, presentation attack can sometimes tamper the face image just before the feature extraction by replacing the bona fide face image to a

synthetically manipulated image, which lacks of many image quality attributes usually appears in natural face image.

A. FACE PAD USING HARDWARE APPROACHES

Face PAD using hardware approaches have advanced significantly, offering robust solutions to combat presentation attacks. Among these, Erdogmus and Marcel [47] as well as Zhou et al. [48] focused on the identification of facial presentation attacks by assessing the presence of depth information in captured face images. In alternative studies, the use of reflectance maps obtained through near-infrared (NIR) imaging was explored for fake face detection. Recently, Jiang et al. [49] introduced a method involving the concatenation of visible and NIR images, which were then input into a convolutional neural network (CNN) designed for face PAD. An alternative methodology presented by Mohamed et al. [50] centered on the extraction of Local Binary Pattern (LBP) features from both visible light and near-infrared image pairs. Meanwhile, Singh et al. [51], Zhang et al. [52], and Weitzner et al. [53] capitalized on light-field cameras, which capture both disparity and depth data in a single frame, to advance face PAD. Although these hardware-assisted PAD techniques have demonstrated superior performance over methods that rely only on visible light images, they still encounter operational constraints in certain scenarios.

B. FACE PAD ANALYZING TEXTURE

Within the scope of 2D face PAD, attackers often present counterfeit faces using printed materials or electronic screens. The intrinsic constraints of such presentation mediums can result in the degradation of the counterfeit faces, characterized by blurred or distorted textures. This phenomenon enables the identification of fake face attacks through the analysis of textural inconsistencies between live and fraudulent face images. In their work, Määttä et al. [54] and Chingovska et al.. [55] utilized the LBP and its variations to capture these textural cues. In a different approach, Boulkenafet et al. [56], [57] proposed a PAD method that integrates LBP features across multiple color spaces—RGB, HSV, and YCbCr—to delineate color variations in textures. Beyond the LBP framework, Agarwal et al. [58] introduced Haralick features as an alternative for textural representation.

With the rise of deep learning, there has been a shift towards self-learned texture-based PAD methods. Li et al. [59] combined traditional LBP features with CNN filtered features, while Sun et al. [60] incorporated local ternary label supervision into their CNN training process for PAD. Liu et al. [61] introduced a tree-structured network aimed at categorizing spoof samples into semantic subgroups. Shu et al. [62] proposed a novel texture descriptor, the LDP, which enhances the traditional LBP. To capitalize on color information in face analysis, Li et al. [63] developed a compact network architecture that improves the detection of fake faces. Alshaikhli et al. [64] integrated the Resnet-50 model with spatial and channel attention mechanisms to enhance face PAD performance.

Despite the promising results of these texture-based analysis methods against low-quality fake faces, challenges emerge when confronting high-definition printing or display technologies, which can lead to a notable decrease in system performance.

C. FACE PAD ANALYZING MOTION CUE

When juxtaposed with the static nature of manipulated facial representations in printed or digital formats, live facial expressions naturally exhibit unique motion characteristics, which can be harnessed for the identification of counterfeit faces that are static in nature. For instance, Pan et al. [65], within the context of an undirected conditional random field model, developed a detection technique that focuses on the motion signals associated with eye blinking in video inputs. To capture the textural changes brought about by facial movements, Pereira et al. [66] and Phan et al. [67] respectively adopted LBP-TOP and LDP-TOP features to represent these subtle textural details. In a distinct methodology, Tirunagari et al. [68] employed the dynamic mode decomposition technique to analyze the patterns of facial muscle movements.

The advent of deep learning has facilitated the extraction of motion features from video sequences of faces. Xu et al. [69] leveraged a CNN to extract features from individual video frames, which were then combined and fed into a Long Short-Term Memory (LSTM) network to assess feature dynamics. Extending beyond the LSTM model, the 3D CNN has been integrated into face PAD systems, as demonstrated by Li et al. [70]. Rehman et al. [71] introduced a supervised CNN classifier for PAD, equipped with a layer designed to capture dynamic disparities. Focusing on the liveness signal derived from pulse detection, Li et al. [72] proposed a technique for identifying face masks by estimating facial pulsations from video sequences. In a related vein, Heusch and Marcel [73] presented a long-term spectral statistics-based approach for precisely characterizing pulsations in live facial videos.

Despite the efficacy of these techniques in mitigating static photo or mask-based presentation attacks, they are vulnerable to attacks involving the replay of face videos, where attackers capture and replay pre-recorded video clips in front of the camera.

D. FACE PAD BASED ON DEEP LEARNING AND TRANSFORMER

Rehman and Komulainen [26] proposed to improve the accuracy of face PAD using a learnable pre-processing layer for low-level deep features. These deep features are from a candidate layer in CNN and hand-crafted features of a input face sample image and can be used to generate adaptive convolutional weights for the deep features of the previous candidate layer. Qiao et al. [27] developed a fine-grained detection network for face PAD. This method

has three main steps: 1) a transformer style network using convolution mapping operation is used for local feature extraction; 2) the global features are extracted by a self-attention module; 3) a linear mapping algorithm is used for final classification. In addition, a VGG based transfer network is adopted to overcome the issue of limited training samples. Liu et al. [28] designed a new adversarial learning framework guided by physical properties to classify whether an input face sample is bona fide or imposter. Two processes are defined in this method: additive process and inpainting process. Three additive components and one inpainting component are used to describe bona fide and imposter face samples. Liao et al. [29] introduced a domain-invariant vision transformer for face PAD. Two different losses are used to enhance the generalizability of the vision transformer: a concentration loss and a separation loss. The first loss is adopted to learn a domain-invariant representation that aggregates the features of bona fide face samples. The latter loss is employed to joint different presentation attacks from various domains. Dong et al. [30] developed an end-to-end domain generalization face PAD approach for unknown presentation attack type. This approach reform face PAD in an open set framework in order to overcome the domain discrepancy issue. Therefore, a novel statistical extreme value theory and an identity-aware contrastive learning strategy are proposed to further achieve face PAD for unknown multiple attack types.

E. FACE PAD USING IQMS

Leveraging the 'quality difference' theory previously discussed, several studies have explored the applicability of general IQMs in the realm of Face Presentation Attack Detection (PAD). Galbally and Marcel [7] were among the pioneers in merging face PAD with IQA. Their approach utilized 14 image quality features, such as mean squared error, peak signal-to-noise ratio, and Laplacian mean squared error, among others, extracted from a single face image. The empirical findings demonstrated that this method was competitive with contemporary state-of-the-art face PAD techniques. Further advancing the field, Yeh et al. [10] introduced a PAD method that employs a perceptual IQA model. This technique applies the Effective Pixel Similarity Deviation model to a no-reference image quality evaluator, using the standard deviation of the gradient magnitude similarity map of selected effective pixels to construct a multi-scale descriptor for classification. Feng et al. [11] proposed a PAD method that integrates image quality features with motion cues. They developed an extendable multi-cues integration framework to enhance the generalization capability of IQA-based PAD. Image quality features were extracted through Shearlet transform, while motion-based features were derived from dense optical flow, with both sets of features being merged effectively using a bottleneck feature merging technique. In a related study, Chang et al. [12] presented a PAD method that utilizes multi-scale perceptual image quality features. They grouped handcrafted texture features from face images into three categories and employed a support vector machine to extract 21 multi-scale image quality features for effective PAD. Li et al. [13] proposed a two-stage PAD approach, where initial face samples are subjected to image quality assessment for manual categorization. A regression function then utilizes these quality features to estimate coefficients for PAD on input samples, refining the detection process.

Beyond the IQA-based face PAD algorithms previously outlined, Galbally et al. [8] have put forward a biometric PAD framework that leverages IQA across multiple modalities, including iris, fingerprint, and facial recognition. This system necessitates the extraction of a comprehensive set of 25 image quality features from the input biometric samples. Subsequently, the PAD process is executed through a pair of pre-trained classifiers, namely linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). An extensive performance evaluation across the three biometric modalities underscores the system's efficacy in detecting a variety of presentation attacks, showcasing its robustness against diverse attack vectors. In a related study, Aravena et al. [42] explored the utility of face image quality assessment techniques in the context of PAD. Their empirical findings suggest that the deployment of face quality assessment for pre-filtering low-quality images can significantly enhance the PAD process.

The integration of IQA principles into PAD methods, as reviewed in this section, underscores the potential for enhancing the precision and reliability of biometric systems. By fusing IQA insights with PAD strategies, the resultant biometric system stands to benefit from an augmented capability to discern and against sophisticated presentation attacks. This study addresses the gaps in the current body of work by leveraging general IQMs that have been demonstrated to generalize well across different conditions and attack types. By, for instance, retraining these IQMs with PAD-specific datasets, it can improve generalizability of face PAD methods without sacrificing computational efficiency. Moreover, the proposed approach in this manuscript has been designed to be robust against environmental variations, ensuring reliable performance in diverse settings. In contrast to existing methods that struggle with generalizability and computational efficiency, our approach offers a balanced solution that maintains high detection rates while remaining feasible for low-resource applications.

F. DATABASES FOR FACE PAD

The Replay-Mobile [14] dataset is a specialized collection designed for Face Presentation Attack Detection (PAD) research, comprising 1190 video sequences of faces and corresponding still images, capturing presentations of attacks from 40 distinct subjects across varying lighting scenarios. These visual records were produced using an iPad Mini2 and an LG-G4 smartphone, employing the devices' front cameras. The videos are characterized by a resolution of 720 $\times$ 1280 pixels, encapsulated in ".mov" format, and recorded at a frame rate of 25 frames per second (Hz).

Authentic facial recordings were conducted under a quintet of lighting states: controlled, adverse, direct, lateral, and diffuse. To emulate presentation attacks, high-fidelity images (captured using a Nikon Coolpix P520 camera with an 18-megapixel resolution) and videos (recorded with the LG-G4 smartphone's rear camera at 16-megapixel 1080p full high definition) were obtained from each participant under diverse conditions. Three distinct methods were employed for the attack simulations: 1) a matte-screen displayed the target face image or video, with both the presentation and the recording apparatus secured on stands; 2) the printed facial image was presented on a stand, while the recording devices were also fixed; 3) the printed image was held by a stand, with the recording apparatus being manually operated by an individual. This process resulted in the registration of 16 attack videos per subject, enriching the dataset for PAD analysis.

CASIA-MFSD [15] face PAD database contains 600 face videos and images presentation attacks from 50 subjects under different acquisition environment. Three acquisition devices were used to reflect different face quality: 1) an old USB camera to represent low face quality (video resolution is 480 × 640 pixels); 2) a high standard new USB camera to represent normal face quality (video resolution is 480 × 640 pixels); 3) a high resolution Sony NEX-5 camera to represent high face quality (video resolution is 720 × 1280 pixels). The Sony NEX-5 camera was also used to capture face images (resolution 1080×1900 pixels) for photo presentation attacks. All bona fide face videos were acquired under natural scenes and subjects were asked to blink their eyes during the recording. Three presentation attacks were conducted in the database: warped photo attack, cut photo attack, and video attack. For warped photo attack, Sony NEX-5 camera was used to take a face image and generate a video for each subject. The acquired high resolution face images were printed on the copper paper and the attacker warped this paper to simulate facial motion in the video attack. For cut photo attack, the eyes regions in the acquired high resolution face images have been cut off so that attacker can conduct eye blink behavior in the video attack. For video attack, high resolution (720 × 1280 pixels) face videos were shown on an iPad and then recorded for such an attack.

Oulu-NPU [16] face PAD database consists of 5904 face videos presentation attacks from 55 subjects under three different lighting conditions by using six smartphones: Samsung Galaxy S6 edge, HTC Desire EYE, MEIZU X5, ASUS Zenfone Selfie, Sony XPERIA C5 Ultra Dual, and OPPO N3. All videos were recorded at Full HD resolution(1920 × 1080) using the frontal cameras of the six mobile devices with the same recording application installed on each device. Three sessions under different illumination condition and background scene were conducted to acquire bona fide face videos: 1) in a big open office environment with the normal office light and natural light; 2) in a meeting room with only normal office light; 3) in a small office with the normal office light and natural light. Both print and video-replay attacks were included in the database. For print attacks, high quality face images were printed on A3 glossy paper using a Canon imagePRESS C6011 and a Canon PIXMA iX6550 printer. For video-replay attacks, high quality videos were shown on two displays: a 19 inches Dell UltraSharp 1905FP display with 1280×1024 resolution, and a 13 inches MacBook Retina display with 2560×1600 resolution. All six devices were used to generate print and video-replay attacks. Printed face image attacks were held by the attacker with stationary acquisition devices for print presentation attacks, while both videos and acquisition devices were stationary for video-replay presentationg attacks.

NUAA [17] face PAD database has 12614 printed face presentation attacks from 15 subjects by using a low price webcam camera under unconstrained lighting conditions. A series of 500 face images were acquired in two separate acquisition sessions by the camera for bona fide face samples. Each subject was required to face the camera with neutral expression and no motion (e.g. eye blink, head rotation) was allowed. A Canon camera was used to first capture face images from all subjects, and then print attacks were generated in three paper sizes by using a color HP printer: $6.8cm \times 10.2cm$, $8.9cm \times 12.7cm$, and A4 paper size. Finally the same webcam camera was used again to obtain print presentation attack face images.

SiW [18] face PAD database provides bona fide and imposter face videos from 165 subjects. Eight bona fide and no more than 20 attack videos are recorded for each subject, and a total of 4478 videos are in the SiW database. All videos are in 30 fps, about 15 second length, and 1080P resolution. The bona fide videos were collected in four sessions. In the first session, the subject moved his/her head with difference distances to the camera. In the second session, the subject changed the face angle of the head between $[-90°, 90°]$ with different face expressions. In the third and fourth sessions, the subject repeated the first two sessions under moving light source around the face from different orientations. Two print presentation attacks and four replay attacks were generated for each subject. A high quality face image (5184×3456 resolution) was acquired for each subject and printed with an HP color LaserJet M652 printer for print attacks. The print attack videos were recorded by holding printed face images still or warping in front of the cameras. Replay attack videos were illustrated on four attack devices: 1) a Samsung Galaxy S8 smartphone, iPhone 7, iPad Pro, and PC (Asus MB168B) screens. Two of the four high-quality bona fide face videos were randomly selected to display on the attack devices.

An overview of the above mentioned face PAD databases as well as other existing databases is give in Table 1.

III. EXPERIMENTAL SETUPS

A. SELECTED IQMS

Since only no-reference IQMs can be used for face PAD tasks, we selected eight commonly used and recent no-reference

TABLE 1. An overview of commonly used face PAD databases.

Dataset	Year	Subjects	Samples	Modality	Attacks
NUAA [17]	2010	15	12,641	RGB	Print, Replay
CASIA-MFSD [15]	2012	50	600	RGB	Print, Replay
ERPA [74]	2013	151	86	RGB, Depth, IR, Thermal	3D Silicon/resin Mask
MIW [75]	2013	10	154	RGB	Print, Cut
MLEP [76]	2015	88	1,330	RGB, Thermal	3D Latex, Paper Mask
MSU-MESFID [77]	2015	26	440	RGB, IR	Print, Replay
HKBU-MARS [78]	2016	17	1,067	RGB, Depth, IR	3D Rigid Mask
Replay-Mobile [14]	2016	40	141	RGB, IR	Print, Replay, Facial disguise
CASIA-M2V1 [79]	2017	55	558	7-band multi-spectral data	Print, Makeup
CIGF-PAD [80]	2017	107	416	RGB	Print, Mask
Oulu-NPU [16]	2017	15	5,940	RGB	Print, Replay
DFWV [82]	2018	1000	11,155	RGB	2D, 3D
Rose-Youtu [81]	2018	20	350	RGB	Print, Replay
SiW [18]	2018	745	2355	RGB, Depth, IR	2D/3D Mask, Makeup
WMC [83]	2019	72	456	RGB	Print, Cut
CASIA-SURF [84]	2019	30	127,440	RGB	Photo, Replay
LCC-FD3D [85]	2019	75	18,987	RGB	Print, Facial Disguise
CASIA-CeFA2 [86]	2020	679	1,630	RGB	Print, Replay, 3D Mask
CelebA-Spoof [87]	2020	1,107	625,537	RGB	Print, Replay, 3D Paper Cut
SynthASpoof [88]	2023	25,000	103,800	RGB	Print, Replay

IQMs to see their ability as well as evaluate their performance on the face PAD:

- NIMA [31]: an automatically learned image quality assessment to predict the distribution of human opinion scores using a re-trained deep object recognition neural network.
- BIQI [32]: a modular framework for constructing image quality assessment using SVM for classification.
- DBCNN [33]: a no-reference image quality assessment method using a deep bilinear CNN.
- dipIQ [34]: a no-reference quality assessment approach by learning-to-rank discriminable image pairs.
- BIECON [35]: a fully deep no-reference image quality assessment algorithm using the local quality maps as intermediate targets for CNNs.
- OGIQA [36]: a no-reference image quality assessment method using relative gradient statistics and adaboosting neural network.
- MUSIQ [37]: a multi-scale transformer to deal with native resolution images in different sizes and aspect ratios for image quality assessment.
- SSEQ [38]: a no-reference image quality assessment approach based on spatial and spectral entropies using SVM for classification.

In selecting the eight IQMs for our study, we considered several criteria, including their theoretical robustness, demonstrated efficacy in image quality assessment, computational efficiency, and their ability to capture nuances relevant to face PAD. Each of the chosen IQMs is grounded in a unique theoretical framework. For instance, the NIMA model utilizes deep learning to predict human opinion scores, while BIQI employs a modular framework with SVM for classification. These diverse approaches provide a comprehensive assessment of image quality attributes. The relevance of each IQM to PAD lies in their sensitivity to image qualities that are often compromised in presentation attacks. For example, the DBCNN model is adept at detecting local distortions and artifacts, which are common in printed or digitally manipulated attack images. We compared our chosen IQMs with other existing metrics, considering their applicability to PAD. The selected IQMs demonstrated a superior balance between accuracy and computational efficiency, which was critical for our real-time PAD system requirements. Preliminary results from our experiments provided evidence supporting the selection of these IQMs. For instance, initial tests showed that the dipIQ model effectively discriminated between high-quality and low-quality face images, which is crucial for distinguishing between bona fide and attack samples. We anticipate that the integration of these IQMs into PAD will contribute to the field by enhancing the robustness and accuracy of attack detection. The chosen metrics, with their diverse theoretical bases and proven track records, are expected to provide a multifaceted approach to PAD.

B. SELECTED FACE PAD METHODS

Eight recently proposed and commonly used face PAD methods are selected to compared their performance to selected IQMs:

- Single Side Domain Generalisation (SSDG) [39]: it improves the generalization ability of the face PAD system in the following two ways: 1) learning a compact

distribution of bona fide face features, and 2) learning a dispersed distribution of fake face features in the domain.

- ViTranZFAS [40]: the backbone of this method is a vision transformer. The last layer of the transformer is replaced by a fully connected layer with one output node and fine-tunes using binary cross-entropy loss.
- MobileNetv3 [41]: it is a smartphone CPUs based CNN. Squeeze-and-excitation and hard swish activation modules are introduced in this CNN. By adding these two modules it has better accuracy and faster computation speed for image classification compared to the earlier version. In this paper, we used ImageNet to train the MobileNetv3 first, and then changed the last layer of the network as a two-class output.
- LMFD-PAD [89]: it introduces a dual-stream CNNs framework for face PAD, addressing challenges in unseen scenarios. One stream employs learnable frequency filters to capture features resilient to sensor/illumination variations, while the other utilizes RGB images for complementary information. The proposed hierarchical attention module integrates information from both streams, enhancing generalizability across intra-dataset and cross-dataset setups.
- MADDoG [90]: it enhances face PAD generalization using a multi-adversarial discriminative deep domain generalization framework. By enforcing dual-force triplet-mining constraints, it ensures a discriminative and shared feature space across source domains, improving effectiveness against new presentation attacks. Incorporating auxiliary face depth supervision further boosts generalization.
- FRT-PAD [91]: it enhances the precision of face PAD by integrating insights from ancillary face-related tasks. The approach is characterized by the utilization of features that are tailored to the specific requirements of the task at hand. Furthermore, it deploys a cross-modal adapter, which is augmented with a graph attention network to facilitate the adaptation process. PAD is accomplished by harnessing hierarchical features extracted from a CNN-based presentation attack detector, alongside features that have been re-mapped to enhance their utility in the PAD process.
- FedSIS [93]: it introduces a privacy-preserving framework using a hybrid vision transformer architecture through Federated Learning and split learning. To enhance generalization to new domains, it employs an intermediate representation sampling strategy and distills discriminative information using a shared adapter network.
- OCKD-FacePAD [92]: it introduces a pedagogical framework that facilitates cross-domain face PAD through one-class domain adaptation. Within this framework, the role of the teacher network is to develop robust feature representations, which are cultivated through training on the data from the source domain. Concurrently, the student networks are tasked with acquiring analogous representations by leveraging a limited set of authentic samples from the target domain. In the evaluation phase, the system evaluates the similarity score derived from the juxtaposition of the teacher and student representations to effectively differentiate between attempted attacks and bona fide facial presentations.

C. FACE PAD DATABASES

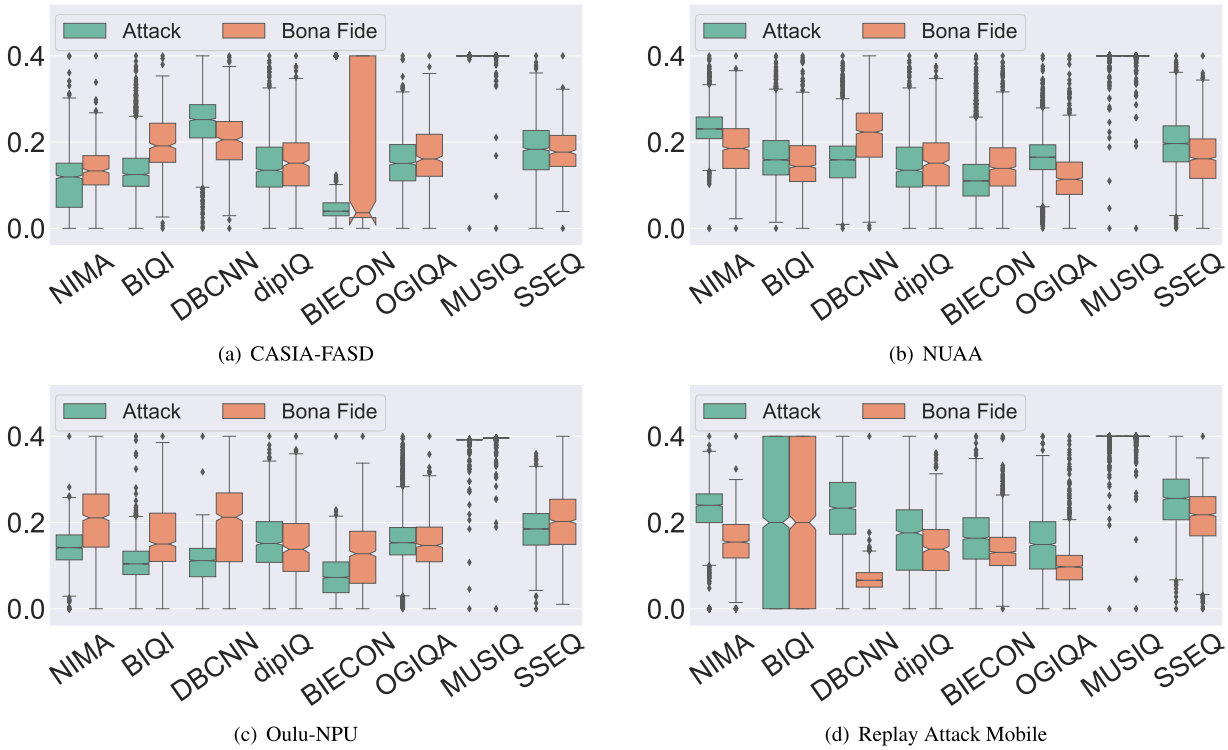
Five databases are used for the performance evaluation of face PAD methods in the experiments: CASIA-FASD, NUAA, Oulu-NPU, Replay Attack Mobile, and SiW. While the first four databases are used to evaluate the IQMs' performance and re-training, the SiW database is used for the cross-dataset validation of re-trained IQMs.

IV. EXPERIMENTS AND RESULTS

To systematically assess the efficacy of IQMs in the context of face PAD and to benchmark their performance against conventional PAD techniques, we propose a set of three distinct experimental protocols. The initial protocol entails an examination of the innate capacity of IQMs to differentiate between authentic and attack face samples through the assignment of distinct quality scores, without any re-training. The second protocol involves the re-training of select IQMs on a composite dataset of PAD images, followed by an analysis of the impact on the detection process when excluding face samples deemed of low quality based on the quality assessment outcomes. The third protocol seeks to establish whether the integration of re-trained IQMs into the PAD process can lead to an enhancement in the overall performance of the face recognition system. Expanding upon these protocols, a key aspect of our methodology is the application of score level fusion, which involves combining the quality scores assigned by multiple IQMs to improve the detection of presentation attacks. The score level fusion is predicated on the notion that no single metric can capture all the intricacies of image quality, especially in the context of PAD where the subtleties of presentation attacks can vary widely. Consequently, our strategy involves assigning weights to each IQM's output based on their demonstrated efficacy in identifying nuances pertinent to PAD. These weights are meticulously tuned to ensure that the fused score not only accentuates the individual strengths of the IQMs but also compensates for their individual limitations.

A. PROTOCOL I - IQMS EFFECT ON PAD

Figure 3 shows a box plot of the quality score statistics for both bona fide and attack face samples from four face PAD databases by using eight general non-reference IQMs. According to the quality assumption we discussed previously, quality scores from bona fide face samples should be higher than presentation attack samples. In addition, quality scores from bona fide and attack face samples should well separated from each other if the IQMs can recognize the quality



(a) CASIA-FASD (b) NUAA

(c) Oulu-NPU (d) Replay Attack Mobile

FIGURE 3. Score statistics for both bona fide and attack face samples from four face PAD database by using eight general non-reference IQMs.

differences between real and fake face samples. However, from Figure 3 we cannot observe quality scores from bona fide face samples (orange boxes) are higher than attack face samples (green boxes). On the contrary, quality scores from attack face samples by using DBCNN in Figure 3 (a), NIMA and SSEQ in Figure 3 (b), NIMA, DBCNN, BIECON, OGIQA, and SSEQ in Figure 3 (d) are higher than bona fide face samples. Amount the other IQMs, only BIQI on CASIA-FASD and Oulu-NPU databases (Figure 3 (a), (c)), DBCNN on NUAA and Oulu-NPU databases (Figure 3 (b), (c)) can obtain higher quality scores from bona fide face samples.

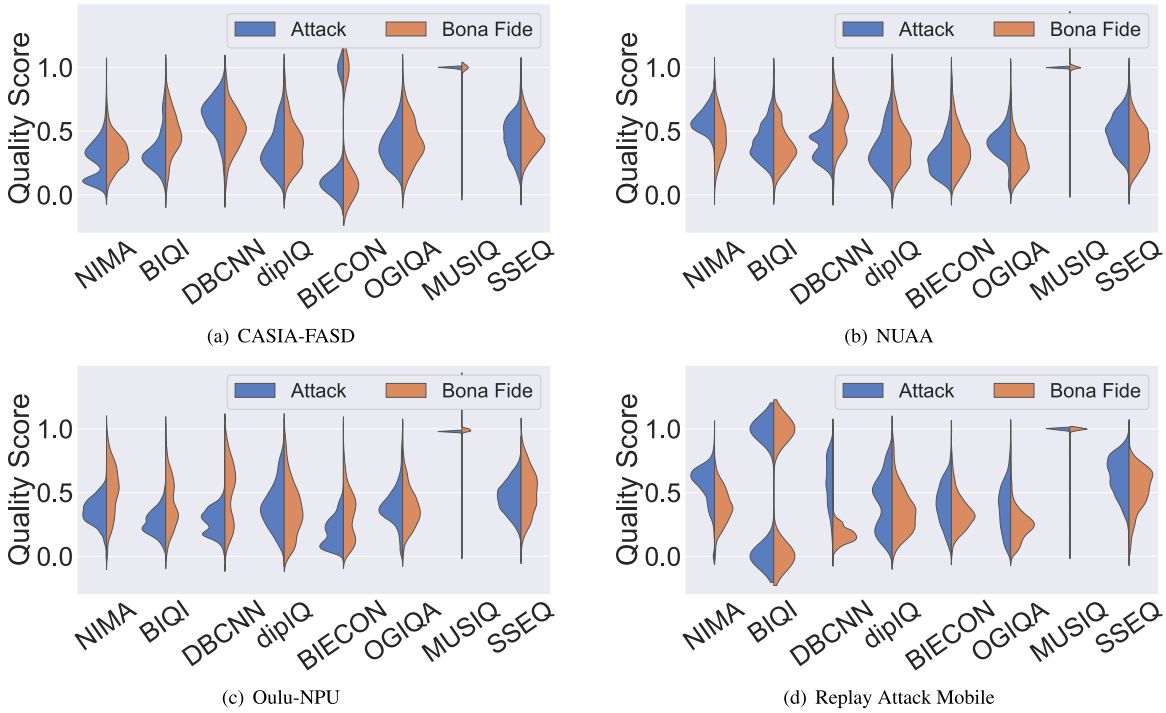
The distribution of quality scores for both bona fide and attack face samples from four face PAD database by using eight general non-reference IQMs are given in Figure 4. We can observe that the quality scores from bona fide and attack samples are highly overlapped, which indicates that tested IQMs cannot distinguish the bona fide and presentation attack face samples by assigning different quality scores. From Figure 4 we can discover that both bona fide and attack quality scores from BIECON on CASIA-FASD database (Figure 4 (a)), BIQI on Replay Attack Mobile database (Figure 4 (d)) are separated into two parts. Such a finding could because BIECON and BIQI are sensitive to the different attack types and acquisition conditions in CASIA-FASD and Replay Attack Mobile databases. Thus they assign different quality scores to both bona fide and presentation attack face samples.

From Figure 3 and 4 we can see that the distribution of quality scores is very similar for both bona fide and attack face samples. Therefore, none of the tested IQMs can be used for face PAD in their original form.

B. PROTOCOL II - RE-TRAINED IQMS EFFECT ON PAD

1) TRAINING OF IQMS AND PAD METHODS

Following the findings from the initial experimental protocol, we embarked on the re-training of six Image Quality Metrics (IQMs): NIMA, DBCNN, dipIQ, BIECON, OGIQA, and MUSIQ. This re-training process utilized a consolidated dataset, which was compiled from the CASIA-FASD, NUAA, Oulu-NPU, and Replay Attack Mobile databases. A straightforward train-test split was applied, reserving 80% of the aggregated data for the training phase and allocating the remaining 20% for the evaluation of the models. Concurrently, to benchmark the PAD capabilities, the aforementioned eight traditional face PAD techniques were also trained on the combined databases. The implementation of all six IQMs and eight face PAD methodologies was conducted in Python, utilizing their publicly accessible open-source versions. A uniform input image dimension of 300 × 300 pixels was employed across all methods. The training procedures were executed on a personal computer equipped with an Intel i7-12700 CPU, 16 GB of RAM, and an NVIDIA RTX 3060 GPU.



(a) CASIA-FASD (b) NUAA

(c) Oulu-NPU (d) Replay Attack Mobile

FIGURE 4. Distribution of quality scores for both bona fide and attack face samples from four face PAD database by using eight general non-reference IQMs.

2) IMAGE QUALITY EFFECT ON FILTERING PRESENTATION ATTACKS

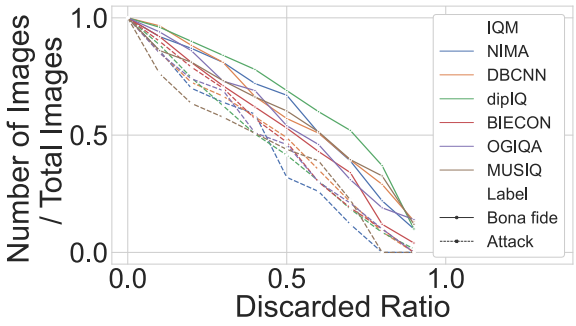
For the segment of the experimental process involving the 20% test dataset extracted from the integrated database, quality assessments were performed on each image using all six re-trained Image Quality Metrics (IQMs). Subsequently, images constituting the lowest $X\%$, with X taking on values from 0% to 90%, were systematically excluded from further consideration. This exclusion was implemented irrespective of the image label. Figure 5 (a) delineates the proportion of discarded images, normalized in relation to the total count of images for each label present in the test dataset of the merged database. Similarly, Figure 5 (b) presents this ratio for the SiW database. The data indicates that genuine face samples, depicted by solid lines, are discarded at a lower frequency compared to the face attack samples, which are represented by dashed lines. A consistent trend was observed across all six re-trained IQMs when applied to both the merged database and the SiW database. These findings suggest that, despite not being explicitly designed for face PAD, the application of these IQMs as a pre-filtering step in image processing can be advantageous. This is evidenced by the increased selection of high-quality bona fide face samples and the rejection of low-quality images that are indicative of presentation attacks following retraining.

We illustrate again the score statistics and distribution of quality scores for both bona fide and attack face samples from testing dataset in merged database and SiW database by using six re-trained IQMs in Figure 6. By comparing Figure 3 and 4 with Figure 6 we can see that, the score difference between bona fide and attack face samples is much bigger after re-training. As expected, attack samples show a lower score on average than bona fide samples. By comparing Figure 6 (b) and (d) we can discover re-trained IQMs are not only able to separate attack and bona fide face samples, but also can have expected good performance on new dataset (SiW database). It could be noted that attack samples from DBCNN have higher scores than most other IQMs in Figure 6 (d), and bona fide samples from OGIQA have lower scores than the others. In addition, it can be observed that scores from NIMA, dipIQ, and BECON have less overlapping than the other three IQMs Figure 6 (d).

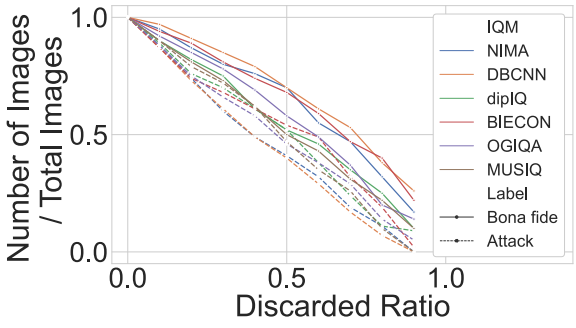
3) PERFORMANCE OF TRADITIONAL PAD METHODS VERSUS RE-TRAINED IQMS

In order to compared the face PAD performance between re-trained IQMs and traditional methods, we discarded 20% of the lowest quality face samples from both the 20% merged testing dataset and SiW dataset by using each of the IQM. The performance of traditional PAD methods was evaluated based on the ISO 30107-3 standard [94] by calculating the Bona fide Presentation Classification Error Rate (BPCER), and Attack Classification Error Rate (APCER). The definitions of these two measures are as the following:

$$BPCER = \frac{\sum_{i=1}^{N_{BF}} RES_i}{N_{BF}} \tag{1}$$



(a) Merged database



(b) SiW database

FIGURE 5. Bona fide and attack face sample images filtering by quality scores for six IQMs from merged PAD database and SiW database.

$$APCER = \frac{1}{N_{PAIS}} \sum_{i=1}^{N_{PAIS}} (1 - RES_i) \tag{2}$$

where N_{BF} is the number of bona fide face samples, N_{PAIS} is the number of attack samples, and RES_i equals to 1 is the system's response to the $i-th$ attack is classified as an attack and equals to 0 if classified as bona fide. For both IQMs and traditional face PAD methods, their performance can be a single value of BPCER for a fixed APCER. For instance, $BPCER_{10}$ represents the BPCER value is calculated where the APCER is 10%.

For a multi-factorial comparative study, we also calculated the performance indicators for traditional PAD methods and re-trained IQMs under two types of presentation attacks: print and replay. The comparative analysis of various PAD methods unveils significant insights into their performance across different metrics and datasets. Through evaluating Equal Error Rate (EER), $BPCER_{10}$, and $BPCER_{20}$ for both print and replay attacks, discernible patterns emerge in Table 2 and Figure 7. From Table 2 and Figure 7 we can observe that, in scrutinizing EER, MADDoG and NIMA consistently emerges as a robust performer, exhibiting the lowest error rates across both print and replay attacks in the 20% merged testing dataset and the SiW dataset. Conversely, OGIQA consistently lags behind, displaying comparatively higher error rates across all evaluations. Notably, re-trained general IQMs like NIMA and BIECON exhibit more reliability compared to traditional PAD approaches, suggesting a potential preference for established approaches in practical implementations.

$BPCER_{10}$ and $BPCER_{20}$ analysis further reinforces the dominance of NIMA, showcasing its ability to minimize false non-match, particularly evident in its performance across print and replay attacks in both datasets. Conversely, OGIQA consistently records higher BPCER10 and BPCER20 values, underscoring its inadequacies in effectively detecting presentation attacks. Notably, consistent performers like NIMA and BIECON demonstrate less performance degradation when transitioning from the 20% merged testing dataset to the SiW dataset, indicating superior generalization capabilities. Most of traditional PAD and IQMs are more sensitive to print attack, conversely, OGIQA exhibit higher PAD ability to replay presentation attacks, emphasizing potential PAD to specific datasets with replay presentation attacks.

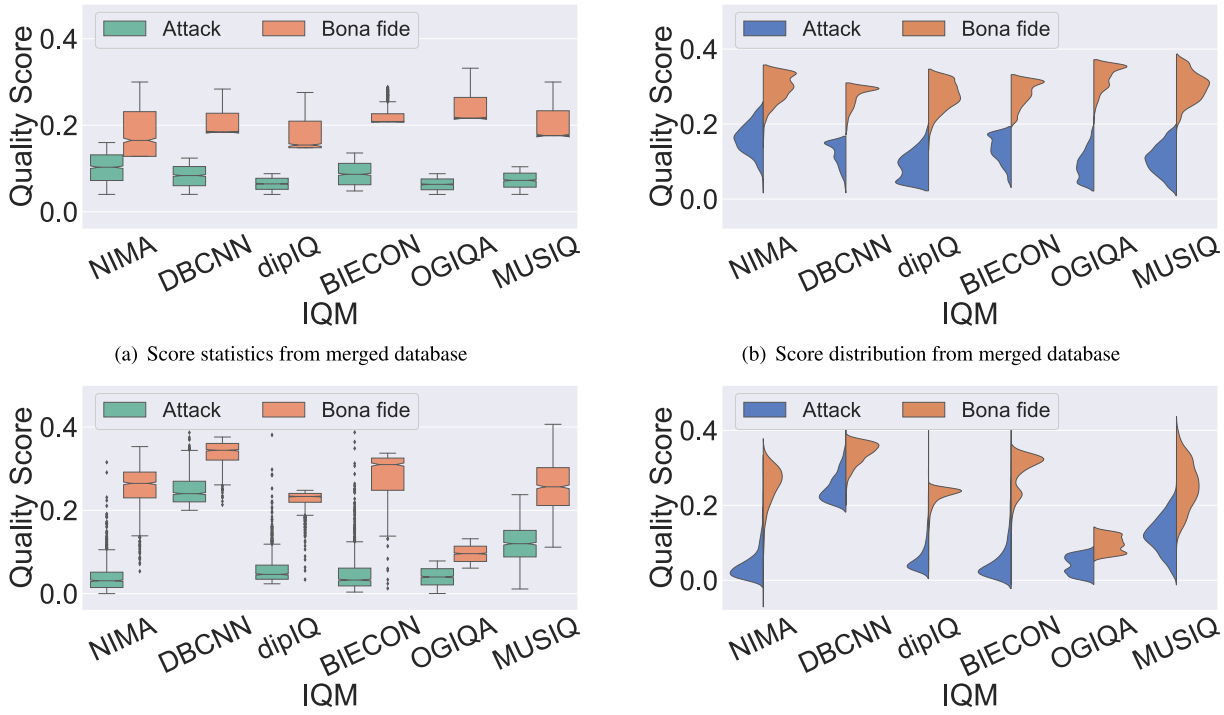
Metric-specific performance highlights a strong correlation between EER and $BPCER_{10/20}$, suggesting that methods excelling in one metric tend to perform well in the other, further underscoring the reliability of top-performing methods. In light of these findings, prioritizing robust methods such as NIMA and BIECON for practical implementations is advisable, while allocating resources towards improving underperforming methods like OGIQA is essential. Additionally, further testing on diverse datasets is recommended to validate the generalization ability of PAD methods and explore potential synergies through a balanced approach that leverages the strengths of various methods. Visual plots in Figure 7 complement the analysis, providing clear comparisons across attack types and offering a comprehensive understanding of each method's effectiveness.

Finally, by comparing the lowest error rates between traditional PAD methods and IQMs we can see that, the difference is smaller than 1% and $BPCER_{20}$ from NIMA is even lower than OCKD-FacePAD. Such an results indicates that by using both the merged dataset to re-train general purpose IQMs, their performance on both the 20% merged testing dataset and a new cross-dataset validation SiW database is comparable to the state-of-the-art traditional PAD methods.

C. PROTOCOL III - IQMS FILTERED FACE PAD

This protocol is to verify whether the system performance can be further improved by filtering low quality face samples by using re-trained IQMs for face PAD. Therefore, we first discarded 20% of the lowest quality face samples by using NIMA and BIECOM (due to their better performance in Table 2), respectively. Then, SSDG, ViTranZFAS, MobileNetv3, LMFD-PAD, MADDoG, FRT-PAD, FedSIS, and OCKD-FacePAD are used to conduct face PAD task on the filtered datasets. The experimental results are illustrated in Table 3.

The error rates without quality filtering are marked as blue (the same values as in Table 2) in Table 3. It can be discovered that the overall error rates after quality filtering are lower. After filtering 20% of the lowest quality face



(a) Score statistics from merged database

(b) Score distribution from merged database

(c) Score statistics from SiW database

(d) Score distribution from SiW database

FIGURE 6. Score statistics and distribution of quality scores for both bona fide and attack face samples by using six re-trained IQMs from merged and SiW databases.

TABLE 2. Performance of traditional PAD methods versus re-trained IQMs under different presentation attack factors. DR represents the Discarded Rate. Data in the performance indicators section: left side represents results from 20% merged testing dataset, right side represents results from SiW dataset.

Methods	DR (in %)	Performance indicators (in %)								
		EER			$BPCER_{10}$			$BPCER_{20}$		
		Print	Replay	Overall	Print	Replay	Overall	Print	Replay	Overall
SSDG	N/A	2.85/3.77	3.12/4.03	3.07/3.94	1.62/2.03	1.68/2.10	**1.64/2.06**	1.51/2.77	1.59/2.83	1.54/2.79
ViTranZFAS	N/A	4.54/5.71	4.76/5.80	4.75/5.78	2.29/3.22	2.35/3.28	2.31/3.25	2.99/4.13	3.11/4.21	3.06/4.17
MobileNetv3	N/A	5.55/7.83	5.68/7.94	5.61/7.91	4.66/5.84	4.76/5.93	4.69/5.88	6.66/8.20	6.77/8.30	6.71/8.24
LMFD-PAD	N/A	2.99/4.29	3.05/4.36	3.01/4.32	3.00/3.86	3.06/3.93	3.04/3.89	2.88/4.02	3.02/4.15	2.92/4.07
MADDoG	N/A	1.94/3.03	2.05/3.09	**1.99/3.05**	2.14/2.75	2.18/2.80	2.15/2.77	2.84/3.38	2.90/3.42	2.87/3.41
FRT-PAD	N/A	3.22/3.80	3.27/3.83	3.24/3.82	1.72/2.08	1.78/2.16	1.76/2.13	1.43/2.90	1.43/2.90	1.43/2.90
FedSIS	N/A	4.21/5.35	4.26/5.40	4.22/5.37	3.84/.03	3.88/5.07	3.86/5.04	4.51/6.30	4.58/6.36	4.55/6.33
OCKD-FPAD	N/A	2.44/3.10	2.47/3.15	2.46/3.13	1.30/2.12	1.30/2.12	1.30/2.12	1.29/2.36	1.31/2.40	**1.30/2.38**
NIMA	20	3.42/4.33	4.11/4.75	3.89/4.51	1.04/1.99	1.41/2.39	**1.22/2.23**	1.01/2.09	1.28/2.40	**1.14/2.28**
DBCNN	20	5.77/7.42	8.68/10.31	7.04/9.67	4.29/5.95	5.02/6.97	4.83/6.48	5.36/6.88	7.77/9.16	6.47/8.34
dipIQ	20	3.22/5.00	4.41/5.89	3.97/5.47	2.74/3.93	3.40/4.55	3.02/4.20	3.04/4.53	3.41/5.04	3.28/4.89
BIECON	20	3.31/4.34	3.35/4.37	**3.32/4.36**	2.44/3.87	2.50/3.96	2.47/3.91	2.49/4.70	2.49/4.70	2.49/4.70
OGIQA	20	7.59/11.43	7.14/110.28	7.41/10.66	7.00/8.35	6.51/7.87	6.88/8.04	6.94/9.79	6.69/9.50	6.88/9.63
MUSIQ	20	4.22/5.64	4.70/5.95	4.40/5.80	4.52/5.98	4.83/6.42	4.79/6.21	4.19/6.61	4.57/7.33	4.38/6.99

samples by NIMA and BIECON, the EER decreases by about 1%-2%, $BPCER_{10}$ decrease by about 1%, and $BPCER_{20}$ decreases by about 1%-3% for all eight PAD methods. Similar to the results reported in Table 2, the lowest EER is from MADDoG by using NIMA for quality filtering. NIMA remains providing lowest $BPCER_{10}$ and $BPCER_{20}$ for SSDG and OCKD-FacePAD, respectively. In addition, we show PAD performance using BPCER versus discarded ratio of face samples filtered by IQMs in Figure 8. From Figure 8 we can observe that the best performance for both $BPCER_{10}$ and $BPCER_{20}$ from SSDG is when 50% lowest quality face samples are discarded by using NIMA (BIECON is 20% for $BPCER_{10}$ and 40% for $BPCER_{20}$). The best performance for both $BPCER_{10}$ and $BPCER_{20}$ from ViTranZFAS is when 20% lowest quality face samples are discarded by using BIECON (NIMA is 20% for $BPCER_{10}$ and 60% for $BPCER_{20}$). The best performance for both $BPCER_{10}$ and $BPCER_{20}$ from MobileNetv3 is when 60% lowest quality face samples are discarded by using NIMA (BIECON is 20% for both $BPCER_{10}$ and $BPCER_{20}$). For the rest five PAD methods,

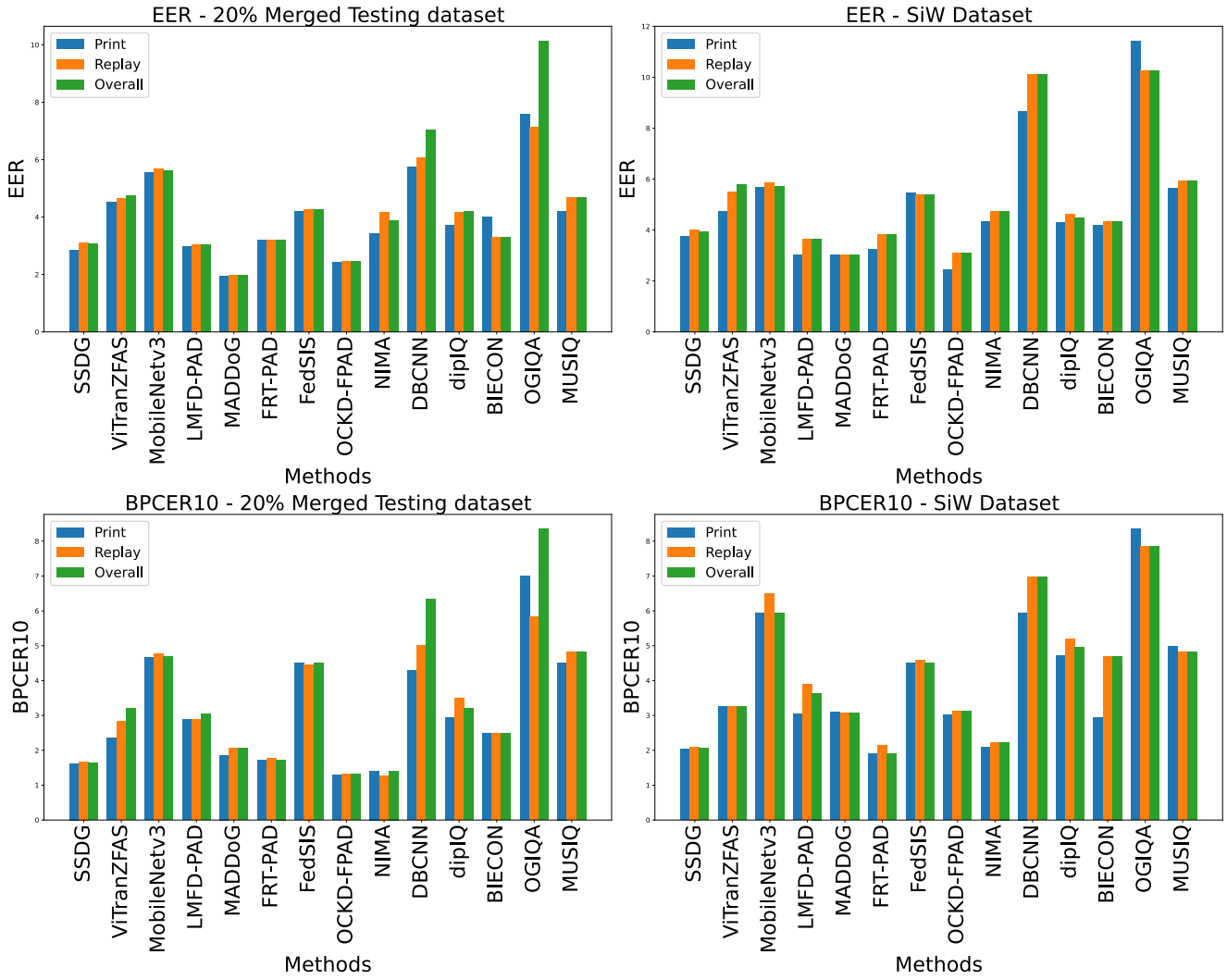


FIGURE 7. Performance comparison between traditional PAD methods and re-trained IQMs under different presentation attack factors.

the overall performance for both $BPCER_{10}$ and $BPCER_{20}$ is better when more low quality samples are discarded by using NIMA and BIECON.

The experimental results from protocol II and III indicate that re-training existing general purpose IQMs by using face PAD databases can not only improve their performance on face PAD but also make them highly competitive compared to the state-of-the-art traditional face PAD methods. By using re-trained IQMs to first discard low quality face samples, the performance of traditional face PAD methods can be further improved.

D. SCORE LEVEL FUSION FOR ENHANCED FACE PAD PERFORMANCE

In an effort to further enhance the performance of IQMs for face PAD, we have explored the application of score level fusion techniques. Recognizing the potential of combining the outputs from multiple IQM models, we hypothesized that a fused score could provide a more robust and accurate representation of the image quality and, consequently, a more reliable detection of presentation attacks.

1) WEIGHT ASSIGNMENT

The weight assignment process in our score level fusion technique is designed to reflect the relative performance of each IQM. The performance is quantified using the area under the Receiver Operating Characteristic (ROC) curve (AUC), which is derived from the FMR and FNMR. The ROC curve is a graphical representation of the FMR against the FNMR at various threshold settings. The AUC provides a single measure of performance, with higher values indicating better discrimination between bona fide and attack samples. The AUC was calculated for each IQM using the trapezoidal rule for numerical integration of the ROC curve. Given a set of predicted probabilities and their corresponding true labels,

TABLE 3. PAD performance after quality filtering.

PAD methods	IQMs	Discarded Rate (in %)	Performance indicators (in %)		
			EER	$BPCER_{10}$	$BPCER_{20}$
SSDG	N/A		3.94	2.06	2.79
	NIMA	20	3.05	**1.87**	2.22
	BIECON	20	2.54	1.93	2.61
ViTranZFAS	N/A		5.78	3.25	4.17
	NIMA	20	4.29	2.77	3.08
	BIECON	20	3.64	2.92	3.58
MobileNetv3	N/A		7.91	5.88	8.24
	NIMA	20	5.44	4.10	5.97
	BIECON	20	5.00	4.37	5.61
LMFD-PAD	N/A		4.32	3.89	4.07
	NIMA	20	4.01	3.57	3.82
	BIECON	20	3.94	3.60	3.56
MADDoG	N/A		3.05	2.77	3.41
	NIMA	20	**2.14**	2.05	3.01
	BIECON	20	2.19	2.13	2.22
FRT-PAD	N/A		3.82	2.13	2.90
	NIMA	20	3.45	1.98	2.44
	BIECON	20	3.51	2.07	2.47
FedSIS	N/A		5.37	5.04	6.33
	NIMA	20	4.99	4.61	5.24
	BIECON	20	5.13	4.54	5.20
OCKD-FacePAD	N/A		3.13	2.12	2.38
	NIMA	20	2.48	1.88	**2.09**
	BIECON	20	2.33	1.94	2.17

the AUC is computed as follows:

$$\text{AUC} = \sum_{j=1}^{n-1} \left(\frac{\text{FMR}_j + \text{FMR}_{j+1}}{2} \right) \times (\text{FNMR}_{j+1} - \text{FNMR}_j)$$

where n is the number of discrete probability points, FMR_j and FNMR_j are the False Match Rate and False Non-Match Rate, respectively, at the j-th probability threshold. To ensure that the weights sum up to one, the AUC values are normalized:

$$\omega_{\text{AUC,IQM}_i} = \frac{\text{AUC}_{\text{IQM}_i}}{\sum_{i=1}^{m} \text{AUC}_{\text{IQM}_i}}$$

Here, m denotes the total number of IQMs, and $\omega_{\text{AUC,IQM}_i}$ is the weight assigned to the i-th IQM based on its AUC value. The final weights for the score level fusion can also incorporate other factors such as the confidence in the IQM's performance or the desired balance between sensitivity and specificity. A comprehensive weight ω_{IQM_i} for the i-th IQM can be calculated as:

$$\omega_{\text{IQM}_i} = w_{\text{AUC,IQM}_i} \times w_{\text{conf,IQM}_i} \times w_{\text{bal,IQM}_i}$$

where $w_{\text{conf,IQM}_i}$ represents the confidence weight and $w_{\text{bal,IQM}_i}$ represents the balance weight. These additional weights can be determined based on further analysis or experimental outcomes. The calculated weights are then applied to the quality scores output by each IQM, resulting in a weighted score for each sample. These weighted scores are aggregated to produce the final decision score.

2) FUSION TECHNIQUE

The fusion technique employed in our study utilizes a weighted scoring system to combine the individual quality assessments from multiple IQMs. This approach aims to enhance the overall performance of face PAD by leveraging the collective predictions of the IQMs. Let $\mathcal{I} = \{i_1, i_2, \ldots, i_m\}$ denote the set of indices corresponding to the m IQMs used in the study. For each IQM indexed by i, a quality score s_i is assigned to a given face image. These scores are then combined using weights ω_i, which are assigned based on the performance of each IQM as detailed in the Weight Assignment section. The aggregated score F for a given face image, resulting from the fusion of scores from all m IQMs, is calculated using the following equation:

$$F = \sum_{i \in \mathcal{I}} \omega_i \cdot s_i$$

Here, ω_i represents the weight assigned to the i-th IQM, and s_i is the quality score given by the i-th IQM for the face image. The aggregated score F is then compared against a predetermined threshold τ to make a decision on the

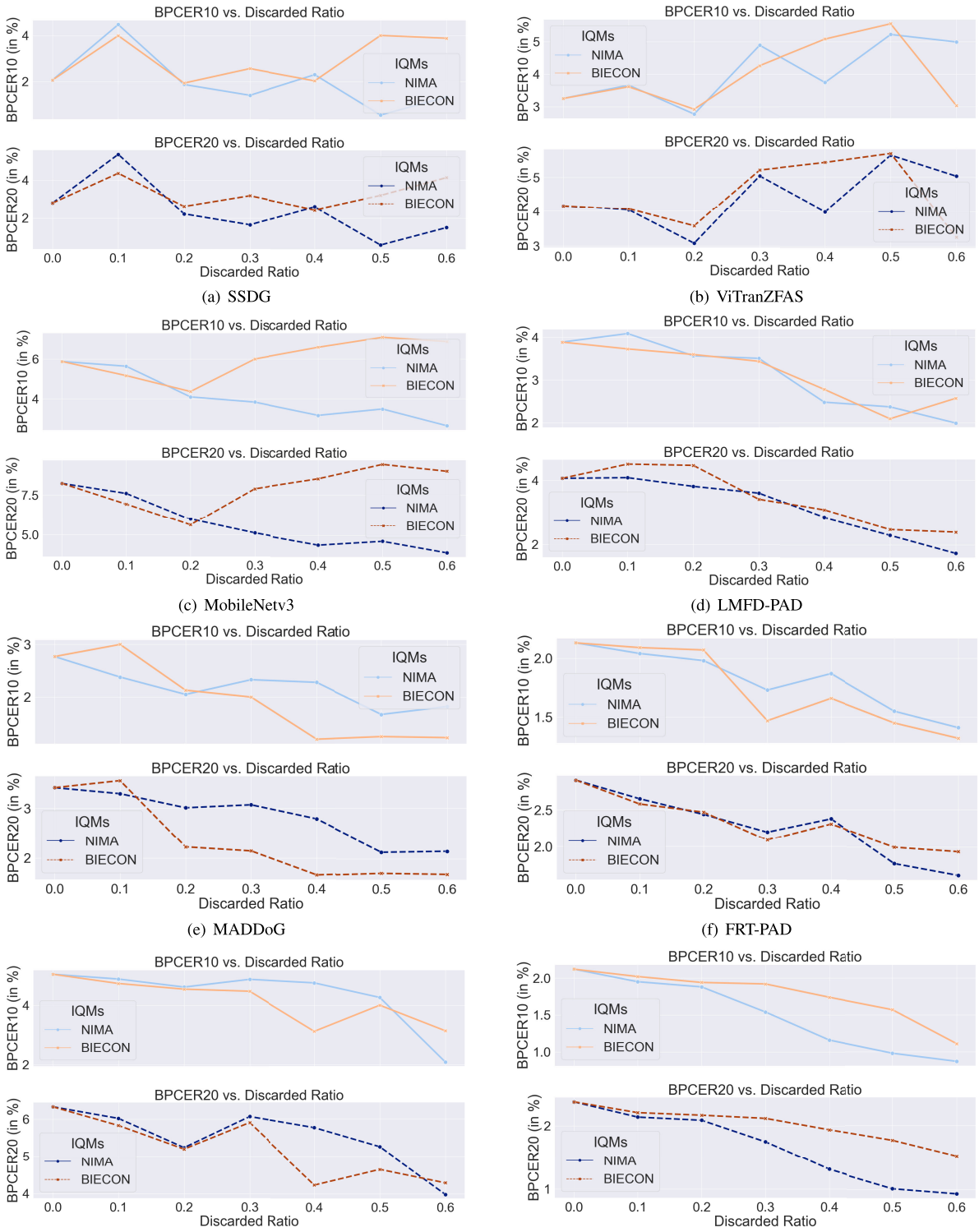


FIGURE 8. PAD performance using BPCER versus discarded ratio of face samples filtered by IQMs.

authenticity of the face image:

$$\text{Decision} = \begin{cases} \text{Bona Fide} & \text{if } F > \tau, \\ \text{Attack} & \text{if } F \leq \tau. \end{cases}$$

The threshold τ is chosen to minimize the overall error rate of the PAD system, which is defined as the sum of the FMR and FNMR:

$$\text{Error Rate} = \text{FMR} + \text{FNMR}$$

TABLE 4. Performance of traditional PAD methods versus score level fusion approach.

Methods	Performance indicators (in %)		
	EER	$BPCER_{10}$	$BPCER_{20}$
Best performance in Table 3	2.14	1.87	2.09
Fused approach	1.97	1.64	2.01

The optimization of the threshold τ can be performed using a grid search over a range of possible threshold values. For each candidate threshold, the corresponding error rate is calculated, and the threshold that results in the minimum error rate is selected as the final threshold.

3) RESULTS AND DISCUSSIONS

To ensure the effectiveness of the fusion approach, we conducted experiments using the SiW database.

The results of the score level fusion experiment are presented in Table 4. As shown in the table, the fused approach achieved a EER of 1.97%, representing a 0.17% reduction in comparison to the best-performing individual method in Table 3. Moreover, both the $BPCER_{10}$ and $BPCER_{20}$ are reduced by using the new fused approach.

The score level fusion technique leverages the collective strengths of multiple IQMs, compensating for individual weaknesses and enhancing the PAD system's robustness and accuracy. The weighted sum method, combined with an optimized threshold, allowed for a more nuanced classification of face samples. However, the success of the score level fusion relies on accurate weight assignment and threshold optimization. The validation set used to determine these parameters must be representative of the true population to avoid over-fitting.

V. CONCLUSION AND FUTURE WORK

In this study, we assessed the efficacy of general-purpose IQMs for face PAD using three distinct protocols. Firstly, we examined whether IQMs could differentiate between authentic and attack face samples by assigning distinct quality scores in their original state. Secondly, we explored whether a re-training strategy could enhance the performance of IQMs, bringing them on par with advanced traditional PAD methods. Thirdly, we investigated the impact of filtering out low-quality samples using re-trained IQMs to potentially boost the performance of face recognition systems. Additionally, we examined the potential of score-level fusion to further refine IQM performance in PAD.

The findings from our experiments, along with the fused approach, are as follows: initially, we found that the original IQMs struggled to effectively distinguish between genuine and attack face samples. However, re-training IQMs using face PAD databases significantly improved their performance, making them comparable to state-of-the-art traditional PAD methods, both on the testing dataset and on a new cross-dataset validation database. Furthermore, we discovered that filtering out low-quality face samples prior to PAD could substantially enhance the accuracy and security of face recognition systems. Specifically, NIMA and BIECON demonstrated superior performance in terms of the BPCER and the EER, respectively. Among traditional PAD methods, OCKD-FacePAD outperformed others. Lastly, the integration of score-level fusion techniques showed a noticeable improvement in PAD performance, highlighting the potential of IQMs when combined effectively.

Our approach, while demonstrating promising results, has certain limitations that should be considered. Firstly, the reliance on general IQMs may not fully capture the nuances of face presentation attacks, particularly those that are highly sophisticated or use novel techniques. Secondly, our method assumes access to a diverse set of training data, which may not be readily available in all applications. The performance of our approach is also contingent upon the quality of the re-training process, which requires careful selection of parameters and validation techniques. Additionally, introducing other types of presentation attacks images may also affect the evaluation results.

Despite these limitations, our findings have practical implications for enhancing the security of biometric systems. By integrating general IQMs into PAD, we offer a flexible and potentially cost-effective strategy for improving detection rates. This approach could be particularly beneficial in environments where specialized PAD techniques are not feasible due to budget constraints or technical complexity. However, it is important to note that the implementation of our method would require careful consideration of these limitations to ensure its effectiveness in a given context.

There is a need to develop an IQM capable of handling both face quality assessment and PAD, thereby strengthening the robustness of face recognition systems. The score level fusion experiment underscores the potential of using ensemble techniques to enhance the robustness and accuracy of biometric security systems in defending against presentation attacks. Further research is warranted for the advancement of multi-modal biometric PAD approaches.

REFERENCES

[1] G. Zhai and X. Min, "Perceptual image quality assessment: A survey," *Sci. China Inf. Sci.*, vol. 63, no. 11, pp. 1–52, Nov. 2020.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] X. Liu, "Multi-modality quality assessment for unconstrained biometric samples," Ph.D. dissertation, Normandie Univ., Gjøvik Univ. College, Norvège, Norway, 2018. [Online]. Available: https://theses.hal.science/tel-02299278/document

[4] F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "Quality measures in biometric systems," *IEEE Secur. Privacy*, vol. 10, no. 6, pp. 52–62, Nov. 2012.

[5] J. E. Tapia, A. Valenzuela, R. Lara, M. Gomez-Barrero, and C. Busch, "Selfie periocular verification using an efficient super-resolution approach," *IEEE Access*, vol. 10, pp. 67573–67589, 2022.

[6] J. Hernandez-Ortega, J. Fierrez, A. Morales, and J. Galbally, "Introduction to face presentation attack detection," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*. Cham, Switzerland: Springer, 2019, pp. 187–206.

[7] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1173–1178.

[8] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.
[9] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *ACM Comput. Surveys*, vol. 54, no. 10s, pp. 1–49, Jan. 2022.
[10] C.-H. Yeh and H.-H. Chang, "Face liveness detection based on perceptual image quality assessment features with multi-scale analysis," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 49–56.
[11] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung, "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 451–460, Jul. 2016.
[12] H.-H. Chang and C.-H. Yeh, "Face anti-spoofing detection based on multi-scale image quality assessment," *Image Vis. Comput.*, vol. 121, May 2022, Art. no. 104428.
[13] H. Li, S. Wang, and A. C. Kot, "Face spoofing detection with image quality regression," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Dec. 2016, pp. 1–6.
[14] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–7.
[15] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, 2012, pp. 26–31.
[16] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A mobile face presentation attack database with real-world variations," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG 2017)*, Sep. 2017, pp. 612–618.
[17] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Proc. ECCV*, 2010, pp. 504–517.
[18] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 389–398.
[19] M. Sebastien, M. Nixon, and S. Li, *Handbook of Biometric Anti-spoofing: Trusted Biometrics Under Spoofing Attacks*. Cham, Switzerland: Springer, 2014.
[20] S. Kumar, S. Singh, and J. Kumar, "A comparative study on face spoofing attacks," in *Proc. Int. Conf. Comput., Commun. Autom. (ICCCA)*, 2017, pp. 1104–1108.
[21] P. Anthony, B. Ay, and G. Aydin, "A review of face anti-spoofing methods for face recognition systems," in *Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, 2021, pp. 1–9.
[22] Z. Ming, M. Visani, M. M. Luqman, and J.-C. Burie, "A survey on anti-spoofing methods for facial recognition with RGB cameras of generic consumer devices," *J. Imag.*, vol. 6, no. 12, p. 139, Dec. 2020.
[23] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.
[24] L. J. González-Soler, M. Gomez-Barrero, and C. Busch, "On the generalisation capabilities of Fisher vector-based face presentation attack detection," *IET Biometrics*, vol. 10, no. 5, pp. 480–496, Sep. 2021.
[25] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep learning for face anti-spoofing: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5609–5631, 2022.
[26] Y. A. U. Rehman, L.-M. Po, and J. Komulainen, "Enhancing deep discriminative feature maps via perturbation for face presentation attack detection," *Image Vis. Comput.*, vol. 94, Feb. 2020, Art. no. 103858.
[27] T. Qiao, J. Wu, N. Zheng, M. Xu, and X. Luo, "FGDNet: Fine-grained detection network towards face anti-spoofing," *IEEE Trans. Multimedia*, vol. 25, pp. 7350–7363, 2022.
[28] Y. Liu and X. Liu, "Physics-guided spoof trace disentanglement for generic face anti-spoofing," 2020, *arXiv:2012.05185*.
[29] 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 *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 6087–6096.
[30] X. Dong, H. Liu, W. Cai, P. Lv, and Z. Yu, "Open set face anti-spoofing in unseen attacks," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 4082–4090.
[31] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.
[32] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
[33] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020.
[34] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "DipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.
[35] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
[36] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Process., Image Commun.*, vol. 40, pp. 1–15, Jan. 2016.
[37] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Jun. 2021, pp. 5148–5157.
[38] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, Sep. 2014.
[39] Y. Jia, J. Zhang, S. Shan, and X. Chen, "Single-side domain generalization for face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8481–8490.
[40] A. George and S. Marcel, "On the effectiveness of vision transformers for zero-shot face anti-spoofing," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Aug. 2021, pp. 1–8.
[41] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Nov. 2019, pp. 1314–1324.
[42] C. Aravena, D. Pasmino, J. E. Tapia, and C. Busch, "Impact of face image quality estimation on presentation attack detection," 2022, *arXiv:2209.15489*.
[43] D. Benini, *Biometric Quality Framework*, Standard ISO/IEC 29794-1, 2006.
[44] *Information Technology—Biometric Sample Quality—Part 4: Finger Image Data*, Standard IEC T. 29794-4: 2010, 2010.
[45] *Information Technology—Biometric Sample Quality—Part 5: Face Image Data*, Standard IEC, T. 29794-5, 2004.
[46] *Information Technology—Biometric Sample Quality—Part 5: Iris Image Data*, Standard IEC T. 29794-6, 2015.
[47] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–6.
[48] J. Zhou, C. Ge, J. Yang, H. Yao, X. Qiao, and P. Deng, "Research and application of face anti-spoofing based on depth camera," in *Proc. 2nd China Symp. Cognit. Comput. Hybrid Intell.*, Sep. 2019, pp. 225–229.
[49] F. Jiang, P. Liu, and X. Zhou, "Multilevel fusing paired visible light and near-infrared spectral images for face anti-spoofing," *Pattern Recognit. Lett.*, vol. 128, pp. 30–37, Dec. 2019.
[50] S. Mohamed, A. Ghoneim, and A. Youssif, "Visible/infrared face spoofing detection using texture descriptors," in *Proc. MATEC Web Conf.*, 2019, p. 400.
[51] A. Singh, G. Jaswal, and A. Nigam, "FDSNet: Finger dorsal image spoof detection network using light field camera," in *Proc. IEEE 5th Int. Conf. Identity, Secur., Behav. Anal. (ISBA)*, Jan. 2019, pp. 1–9.
[52] P. Zhang, F. Zou, Z. Wu, N. Dai, S. Mark, M. Fu, J. Zhao, and K. Li, "FeatherNets: Convolutional neural networks as light as feather for face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1574–1583.
[53] D. Weitzner, D. Mendlovic, and R. Giryes, "Face authentication from grayscale coded light field," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 2611–2615.
[54] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–7.
[55] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2012, pp. 1–7.
[56] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 2636–2640.

[57] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016.

[58] A. Agarwal, R. Singh, and M. Vatsa, "Face anti-spoofing using Haralick features," in *Proc. IEEE 8th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2016, pp. 1–6.

[59] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot, "Learning generalized deep feature representation for face anti-spoofing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 10, pp. 2639–2652, Oct. 2018.

[60] W. Sun, Y. Song, C. Chen, J. Huang, and A. C. Kot, "Face spoofing detection based on local ternary label supervision in fully convolutional networks," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3181–3196, 2020.

[61] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4675–4684.

[62] X. Shu, H. Tang, and S. Huang, "Face spoofing detection based on chromatic ED-LBP texture feature," *Multimedia Syst.*, vol. 27, pp. 161–176, Jul. 2021.

[63] L. Li, Z. Xia, X. Jiang, F. Roli, and X. Feng, "CompactNet: Learning a compact space for face presentation attack detection," *Neurocomputing*, vol. 409, pp. 191–207, Oct. 2020.

[64] M. Alshaikhli, O. Elharrouss, S. Al-Maadeed, and A. Bouridane, "Face-Fake-Net: The deep learning method for image face anti-spoofing detection : Paper ID 45," in *Proc. 9th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Jun. 2021, pp. 1–6.

[65] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Aug. 2007, pp. 1–8.

[66] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "*LBP − TOP* based countermeasure against face spoofing attacks," in *Computer Vision—ACCV*. Cham, Switzerland: Springer, 2012, pp. 121–132.

[67] Q. T. Phan, D. T. Dang-Nguyen, G. Boato, and F. G. B De Natale, "FACE spoofing detection using LDP-TOP," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Aug. 2016, pp. 404–408.

[68] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 762–777, Apr. 2015.

[69] Z. Xu, S. Li, and W. Deng, "Learning temporal features using LSTM-CNN architecture for face anti-spoofing," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2015, pp. 141–145.

[70] L. Li and X. Feng, "Face anti-spoofing via deep local binary pattern," *Deep Learn. Object Detection Recognit.*, vol. 1, pp. 91–111, Jul. 2019.

[71] Y. A. U. Rehman, L.-M. Po, and M. Liu, "SLNet: Stereo face liveness detection via dynamic disparity-maps and convolutional neural network," *Expert Syst. Appl.*, vol. 142, Mar. 2020, Art. no. 113002.

[72] X. Li, J. Komulainen, G. Zhao, P. C. Yuen, and M. Pietikainen, "Generalized face anti-spoofing by detecting pulse from face videos," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, 2016, pp. 4244–4249.

[73] G. Heusch and S. Marcel, "Pulse-based features for face presentation attack detection," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–8.

[74] S. Bhattacharjee and S. Marcel, "What you can't see can help you–extended-range imaging for 3D-mask presentation attack detection," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2017, pp. 1–7.

[75] C. Chen, A. Dantcheva, and A. Ross, "Automatic facial makeup detection with application in face recognition," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–8.

[76] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore, "Face presentation attack with latex masks in multispectral videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 275–283.

[77] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3D masks," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1084–1097, Jul. 2014.

[78] S. Liu, B. Yang, P. C. Yuen, and G. Zhao, "A 3D mask face anti-spoofing database with real world variations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2016, pp. 100–106.

[79] R. Raghavendra, K. B. Raja, S. Venkatesh, F. A. Cheikh, and C. Busch, "On the vulnerability of extended multispectral face recognition systems towards presentation attacks," in *Proc. IEEE Int. Conf. Identity, Secur. Behav. Anal. (ISBA)*, Feb. 2017, pp. 1–8.

[80] C. Chen, A. Dantcheva, T. Swearingen, and A. Ross, "Spoofing faces using makeup: An investigative study," in *Proc. IEEE Int. Conf. Identity, Secur. Behav. Anal. (ISBA)*, Feb. 2017, pp. 1–8.

[81] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1794–1809, Jul. 2018.

[82] M. Singh, R. Singh, M. Vatsa, N. K. Ratha, and R. Chellappa, "Recognizing disguised faces in the wild," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 1, no. 2, pp. 97–108, Apr. 2019.

[83] K. Kotwal, Z. Mostaani, and S. Marcel, "Detection of age-induced makeup attacks on face recognition systems using multi-layer deep features," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 1, pp. 15–25, Jan. 2020.

[84] S. Zhang, A. Liu, J. Wan, Y. Liang, G. Guo, S. Escalera, H. J. Escalante, and S. Z. Li, "CASIA-SURF: A large-scale multi-modal benchmark for face anti-spoofing," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 2, pp. 182–193, Apr. 2020.

[85] J. Y. Bok, K. H. Suh, and E. C. Lee, "Verifying the effectiveness of new face spoofing DB with capture angle and distance," *Electronics*, vol. 9, no. 4, p. 661, Apr. 2020.

[86] 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 *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1178–1186.

[87] Y. Zhang, Z. F. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, "CelebA-Spoof: Large-scale face anti-spoofing dataset with rich annotations," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 70–85.

[88] M. Fang, M. Huber, and N. Damer, "SynthASpoof: Developing face presentation attack detection based on privacy-friendly synthetic data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 1061–1070.

[89] 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 *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1131–1140.

[90] R. Shao, X. Lan, J. Li, and P. C. Yuen, "Multi-adversarial discriminative deep domain generalization for face presentation attack detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10015–10023.

[91] W. Zhang, H. Liu, F. Liu, R. Ramachandra, and C. Busch, "Effective presentation attack detection driven by face related task," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 408–423.

[92] 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 Trans. Inf. Forensics Security*, vol. 17, pp. 2137–2150, 2022.

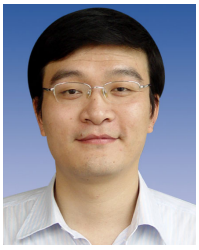
[93] N. Alkhunaizi, K. Srivatsan, F. Almalik, I. Almakky, and K. Nandakumar, "FedSIS: Federated split learning with intermediate representation sampling for privacy-preserving generalized face presentation attack detection," 2023, *arXiv:2308.10236*.

[94] *Information Technology Biometric Presentation Attack Detection Part 3: Testing and Reporting*, Standard ISO/IEC 30107-3, 2017.

XINWEI LIU received the B.Sc. degree in electronic and information engineering from Hubei University of Technology, Wuhan, China, in 2011, the M.S. degree in media technology (MiT) from Gjøvik University College, Norway, in 2014, and the double Ph.D. degrees in computer science from the Norwegian University of Science and Technology (NTNU), Norway, in 2018, and the University of Caen (UNICAEN), France, in 2019. He is currently a Lecturer with the College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo, China. His research interests include image processing and biometrics.

RENFANG WANG received the B.S. degree in computer application from the Huazhong University of Science and Technology, Wuhan, China, in 2000, the M.S. degree in computer application technology from Henan University of Science and Technology, Luoyang, China, in 2004, and the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2008. He joined the College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo, China, in 2008, as a Professor of computer science and technology. His current research interests include big data analysis and computer vision.

WEN LIU is the Director with the Digital Manufacturing Research Institute, a Visiting Scholar with Hong Kong Polytechnic University and Singapore Nanyang Polytechnic University, a Leading Talent of Ningbo City, a Leader of Entrepreneurship Team of ''South the Taihu Lake Elite'' and the Innovation Team of ''Green Valley Elite,'' a Postdoctoral Cooperative Supervisor of Ningbo University, and the Master's Supervisor of Zhejiang University and Ocean University of China. He mainly engaged in research related to precision plastic forming, fluid control, and artificial intelligence. His representative research achievements include flexible internal high-pressure technology and precision pneumatic valve technology. He has published over 20 papers, granted 48 invention patents, and authored eight textbooks. He has hosted one national key research and development project, one provincial/municipal fund each and participated in 36 major special projects in Ningbo City in 2025.



LIANGBIN ZHANG is an Associate Professor and a Faculty Member with the School of Big Data and Software Engineering. He has presided over one basic public welfare research project in Zhejiang Province and one provincial experimental teaching project. He has presided over five municipal and departmental level scientific research and teaching projects and three horizontal projects with enterprises and has participated in two provincial and ministerial level projects. He has published over 20 articles in Chinese core journals, international journals, and conferences; and received six software copyrights. His main research directions include cloud computing, design and optimization of network architecture, and fractal image coding.

XIAOXIA WANG is mainly engaged in research on the theory and application of modern educational technology, the development and application effects of new technologies and new equipment, and the development and training of teachers' teaching skills.