Pattern Recognition Letters 145 (2021) 103–109

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Face spoofing detection under super-realistic 3D wax face attacks

Shan Jia[a], Chuanbo Hu[b,*], Xin Li[b], Zhengquan Xu[a]

[a] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[b] Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA

ARTICLE INFO

Article history:
Received 21 July 2020
Revised 28 December 2020
Accepted 24 January 2021
Available online 3 February 2021

MSC:
41A05
41A10
65D05
65D17

Keywords:
Face anti-spoofing
3D face presentation attack
Wax figure face database
Residual Attention Network

ABSTRACT

Face spoofing attacks based on 3D face images have posed a severe security risk to face recognition systems. Despite the great effort made by the technical community in recent years, existing 3D face spoofing databases, mostly based on 3D masks, still suffer from small sample size, low diversity, or poor authenticity due to the production difficulty and high cost. To fill in this gap, we introduce a new database in this paper with 4-000 single wax figure faces, named SWFFD (Single Wax Figure Face Database), as a type of super-realistic 3D face presentation attack. Collected from online resources, this database has high diversity in terms of subjects, lighting conditions, facial poses, and recording devices. We have also designed a new detection method, which combines attention-aware features from different face scales to generate discriminative representations for realistic face spoofing attack detection. Extensive experiments have been conducted on the SWFFD as well as the CelebA-HQ database (containing real faces from the online collection). Experimental results have demonstrated the effectiveness of the proposed method in both intra-database and cross-database testing scenarios.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, great progress has been made to address the vulnerability of existing face recognition systems to various face spoofing attacks (a.k.a. presentation attacks) [10]. As of today, 2D modality based attacks, which present printed photos or recorded videos to the biometric data capture subsystem, have drawn much attention due to their simplicity, efficiency, and low cost [33]. Accordingly, 2D face anti-spoofing detection has been extensively studied in the literature [26]. Existing face anti-spoofing methods mostly explore the effects of spoofing medium (e.g., the printed paper, the displaying screen) or the geometric differences between a fake 2D planar face and a real 3D structured face. However, an increasing number of studies have found that a variety of face recognition systems, even taking face spoofing detection into consideration, can still be fooled by more powerful 3D face spoofing attacks [6].

Empowered by 3D structures or materials similar to real faces, 3D face presentation attacks are more realistic and therefore more difficult to be detected by face recognition systems. Existing 3D face spoofing attacks can be realized by wearing a face mask, presenting a synthetic model, or wearing makeup, as shown in Fig. 1. When compared with 2D attacks, 3D face spoofing is much more difficult and expensive to manufacture, often requiring special devices and materials. However, the rapid advances of 3D printing technologies and services in recent years have opened up opportunities for making more affordable and higher-quality 3D face spoofing attacks. Several 3D face spoofing attack databases have been created using third-party 3D printing services, based on self-manufactured masks or from online collections. For example, as the first public 3D mask spoofing database, 3DMAD [5] used the services of ThatsMyFace[1] to generate 17 masks of users, and recorded 255 video sequences with color and depth information for both real faces and mask spoofing attacks. Similarly, HKBU-MARs database [22] obtained 8 customized masks from two mask manufacture companies and included 120 videos with lighting variations to simulate the real world scenarios. Taking another example, 3DFS-DB [7] is a self-manufactured 3D face spoofing database based on 26 printed models using two 3D printers. Likewise, Rose-Youtu face liveness detection dataset [18] and WMCA database [8] contained different types of face spoofing attacks from 25 and 72 subjects respectively. In addition to 2D face spoofing, they have both designed paper masks (WMCA also includes rigid and flexible masks) as 3D face spoofing attacks. Last, by taking advantage of the rich online resources,

* Corresponding author.
E-mail addresses: chuanbo.hu@mail.wvu.edu, cbhu@whu.edu.cn (C. Hu).

[1] http://thatsmyface.com/.

https://doi.org/10.1016/j.patrec.2021.01.021
0167-8655/© 2021 Elsevier B.V. All rights reserved.

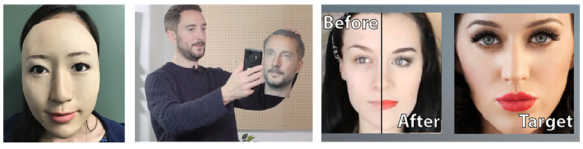

Fig. 1. Examples of 3D face spoofing attacks, (a) wearing face masks, (b) presenting a synthetic model, (c) wearing makeup.

SMAD database [24] has collected 65 videos of people wearing silicone masks and 65 genuine access videos of people auditioning, interviewing, and hosting shows.

In view of increasing attention to 3D face spoofing, a variety of studies have been devoted to 3D face spoofing detection methods. Different from 2D face anti-spoofing, existing 3D spoofing detection schemes are mainly based on the subtle differences between real face skin and mask materials. The reflectance/multispectral properties have been widely studied in the early years due to apparent visual differences of different object surfaces. Using multiple illumination wavelengths, methods [15,38] have achieved over 96% accuracies on their private dataset. Instead of requiring special devices to acquire multispectral images, texture-based methods explore the texture patterns in visible images to distinguish real faces from spoofed ones. Local Binary Pattern (LBP) was extracted in several studies [5,28,32] based on its discriminative power and computational simplicity. However, their robustness to different qualities of mask spoofing attacks remains to be further improved. Haralick features [1] show promising performance in both 2D and 3D mask spoofing databases (e.g., with 0% error rate on 3DMAD database). Besides, using intrinsic liveness signals for 3D face anti-spoofing has also attracted great research interest in recent years. Several liveness cues have been studied, including heartbeat signals [21,22], thermal signature [4], and gaze information [3]. This class of methods can achieve good performance in distinguishing real faces from masks, but their performances rely a lot on video settings. Deep learning features have also become increasingly popular for face spoofing detection. Various work [19,20,27] based on different convolutional neural network (CNN) architectures have shown high detection accuracy in telling 3D face spoofing attacks apart from real faces.

Despite some progress, several recent studies [11,21,25,30] have shown that the 3D face anti-spoofing methods will suffer from performance degradation when dealing with more diverse and realistic 3D face spoofing attacks. Mostly based on facial masks, existing 3D face spoofing databases are restricted to small data sizes (mostly less than 30 subjects), low mask quality (e.g., using 2D paper masks [2,18], or not user-customized masks [2,24]), and low diversity in subjects, facial poses, and recording environment. To address these limitations, our previous work [12] first introduced diverse and super-realistic 3D face spoofing attacks based on wax figure faces from online resources. Totally 4400 images were collected, with 2200 real faces and 2200 wax figure faces. Inspired by the powerful spoofing capability of wax figure faces, we have further collected a new database in this paper with 4000 single wax figure faces, named SWFFD (Single Wax Figure Face Database), to promote more effective 3D face spoofing detection methods. A new detection method is also designed to distinguish these realistic 3D face spoofing attacks from real faces. Combining attention-aware features from multiscale face images, the proposed method can generate subtle and discriminative representations, which has achieved outstanding detection performance under both intra-database and cross-database testing in our experiments.

2. The single wax figure face database (SWFFD)

By taking advantage of the rich and open online resources, we have collected a large number of wax figure faces from the Internet to construct a realistic 3D face spoofing database with a large size and high diversity. We first downloaded as many celebrity wax figure faces as possible, and then cleaned the dataset manually based on our own selection criterion. During the dataset cleaning, images without frontal faces, with face dimensions smaller than 50×50, or with a face with over half occlusion or embedded text (see the examples in Fig. 2(a)), have been excluded from the dataset.

Finally, a total of 4000 images with single wax figure faces from 1457 subjects were collected as the newly constructed SWFFD database. The resolutions of face images are in the range of 50×50 to 2000×2000 (with 25.20% less than 200, 64.92% between 200 and 600, 8.80% between 600 and 1000, and 1.08% over 1000).

For real faces, we have combined SWFFD with the publicly available CelebA-HQ dataset [14], which consists of 30,000 celebrity images obtained from the Internet. To reduce the quality discrepancy among different data sources, we have followed the same procedure for data cleaning (see the excluded examples in Fig. 2(b)), and finally obtained 28,000 images, which are further randomly divided into 7 sessions (4000 images in each session) to reach a balance in size with wax figure faces in SWFFD. Fig. 3 shows image examples in our combined SWFFD and CelebA-HQ databases. The statistical information about the subject's age, gender, and race (detected by Deepface [35]) of these two datasets is shown in Fig. 4. It can be seen that the faces in SWFFD are gender-balanced and have a high diversity in terms of subject age and race, which is almost consistent with the distribution of CelebA-HQ dataset.

Based on the combined database, we have designed a new data protocol based on cross-validation for performance evaluation. Specifically, we first combine the wax figure faces in SWFFD with each session of real faces in CelebA-HQ to construct seven evaluation subsets, each with 8000 face images. Then each evaluation subset is randomly divided into training, validation, and testing subsets by a ratio of 2:1:1. The average result on the seven evaluation subsets is taken as the final detection performance.

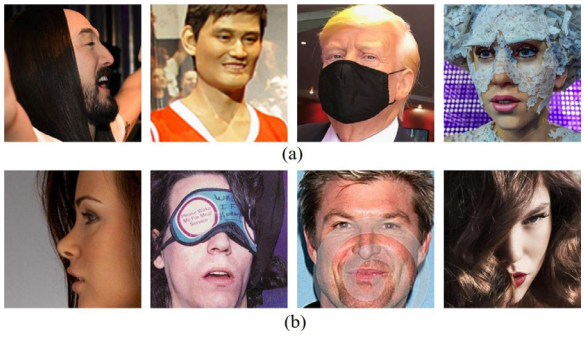

Fig. 2. Examples of excluded faces in data cleaning process. (a) SWFFD, (b) CelebA-HQ database.

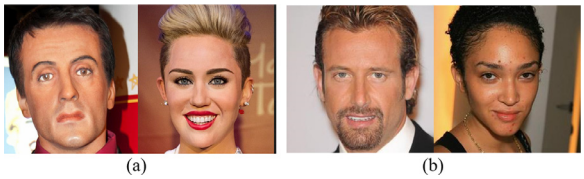

Fig. 3. Examples of face images in (a) SWFFD, (b) CelebA-HQ database.

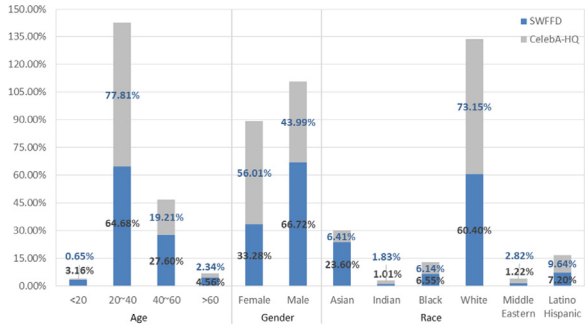

Fig. 4. Statistics of face images in SWFFD and CelebA-HQ databases.

Fig. 5. Comparison of images in SWFFD and WFFD [12] datasets. (a) SWFFD, (b) WFFD.

To sum up, the differences between the new SWFFD dataset and the previous WFFD [12] are threefold. First, SWFFD includes more wax figure faces, almost twice the size (4000) of the WFFD dataset (2200). Second, SWFFD collects images with higher qualities in terms of face size and pose (as shown in Fig. 5(a)). Containing images with both a wax figure face and a real face recorded in the same environment (see Fig. 5(b)), WFFD has more images with small faces and diverse facial poses. Third, instead of providing matched face pairs (with a wax figure face and a real face from the same subject) as WFFD did, the SWFFD dataset, combined with higher-quality real faces from CelebA-HQ, aims to provide rich data for real vs. fake face detection. We will make this newly-constructed database publicly available[2] to support the research on 3D face anti-spoofing.

3. Proposed face anti-spoofing method

In this section, we propose an effective detection method, which combines attention-aware features across different scales to enhance the discriminative power of CNN-based representations for 3D face spoofing detection. Inspired by recent advances in attention mechanism-based deep neural networks, we propose to use the Residual Attention Network (RAN) [36] in our work to generate attention-aware features from different face scales. Through the integration of the mixed attention mechanism and stacking attention modules, RAN has shown good performance in several image processing tasks such as image classification [36], visual tracking [37], and image super-resolution [41]. To the best of our knowledge, this work represents the first effort of leveraging RAN to adaptively learn more useful features for face spoofing detection. Furthermore, considering the importance of face scales to characterize different appearance features in face analysis, we propose to generate more discriminative features from multiscale face representations for realistic 3D face anti-spoofing. Overall, we hypothesize that both residual and multiscale representations will facilitate the task of revealing subtle appearance differences between real faces and fake ones.

The overall architecture of the residual attention network based on multiscale face representations has been shown in Fig. 6 (note that we have shown three different scales as an example for input, but the generalization to less or more scales is straightforward). The faces cropped by MTCNN [39] are first resampled to different sizes, such as $256 \times 256 \times 3$, $288 \times 288 \times 3$, and $320 \times 320 \times 3$, and then they are re-cropped to $224 \times 224 \times 3$ to get different face scales as the input of the RAN models. The RAN model [36] is constructed by stacking multiple residual blocks and attention modules. Designed to alleviate the vanishing gradient problem [9] in the training process, the residual block extracts and presents the basic and important features of images. It first feeds a given input feature into three convolution layers with the kernel sizes of 1×1, 3×3, 1×1, respectively, and then add this output to the original input or input after a 1×1 convolution in an element-wise manner to get the new feature map. The details of such a feature addition procedure are illustrated in Fig. 7(a).

The attention module consists of two parts - i.e., the trunk branch and the soft-mask branch, as shown in Fig. 7(b). It devotes to adaptively enhancing the useful features while suppressing the less useful ones from the trunk branch. The trunk branch performs feature processing and is constructed by residual units. The soft-mask branch, however, contains fast feed-forward sweep and top-down feedback steps with the strategies of downsampling and upsampling to softly combine the trunk branch output (i.e., it serves as a feature selector). More specifically, the soft-mask branch first downsamples the input features based on max-pooling and residual units modules, then uses linear interpolation to upsample the features to get the output with the same size as the input feature map. The output is further normalized with a sigmoid layer after two 1×1 convolution layers. Putting things together, given the trunk branch output $T(x)$ from the input x and the learned mask $M(x)$ with the same size and in range of [0, 1], the output of attention module is $H(x) = (1 + M(x)) * T(x)$. Accordingly, with two or more face scales as the input of the RAN module, we can first obtain their multiscale attention-aware feature maps (but with the same size). Then the feature maps are fused by concatenation, followed by a fully connected (FC) layer for final classification using a softmax classifier.

To further show the significance of fusing multiscale faces, we have compared the saliency maps of the RAN model on both real and wax figure faces in Fig. 8. The visualization tool FlashTorch[3] is used to better explore how the RAN network "perceives" faces with different scales in the scenario of face anti-spoofing. Five face scales have been considered, including 224×224, 256×256, 288×288, 320×320, and 352×352 (Note that for simplicity, all the scales are shortened to two dimensions in the following writing). From the gradient maps in each column of Fig. 8, we can observe that the RAN network pays attention to different regions for different scales of faces. For smaller-size faces, the network seems to focus more on global face regions due to the existence of background, while for larger-size faces, more attention is paid to specific regions, such as around the mouth and eyes. When comparing different columns, we can see the difference of the network's focus on real faces and wax figure faces. Specifically, the network tends to focus on more areas in wax figure faces than in real faces, such

[2] https://github.com/shanface33/Wax_Figure_Face_DB.

[3] https://github.com/MisaOgura/flashtorch.

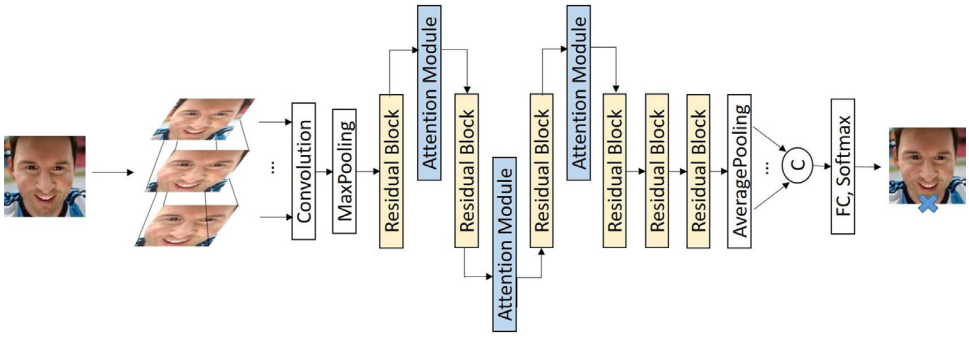

Fig. 6. Architecture of deep residual attention network based on multiscale faces for face spoofing detection. 'C' denotes Concatenation, and 'FC' denotes Fully Connected Layer. Note that three different scales are shown as an example for input, but the generalization to less or more scales is straightforward.

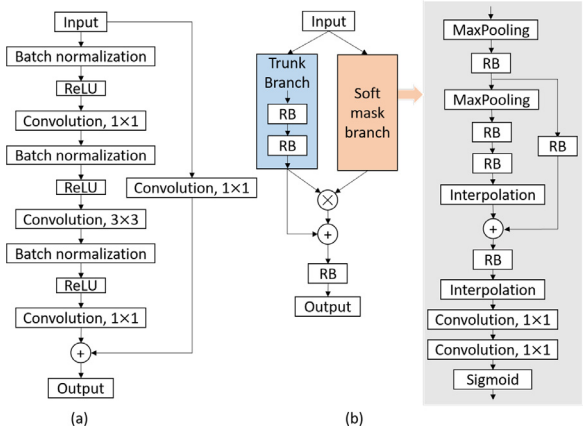

Fig. 7. Residual block and attention module in RAN [36]. (a) Residual block, (b) Attention module with a combination of trunk branch and soft mask branch. 'RB' denotes Residual Block.

as more regions around the nose for smaller face scales and more specific regions for larger face scales. Therefore, the combination of these scales will contribute to generating more discriminative features for realistic face spoofing detection.

4. Experiments

In this section, we evaluate the performance of the proposed RAN-based method on 3D face spoofing detection. Both intra-database testing and cross-database testing are conducted to justify the effectiveness of the proposed method.

4.1. Databases and metrics

Four databases are used in our experiments - namely, the new SWFFD, WFFD [12], and two existing 3D mask face spoofing databases, 3DMAD [5] (the most widely-used), and HKBU-MARs-V1 [22] (with hyper-real 3D masks). The WFFD database contains three protocols: Protocol I with 1000 pairs of heterogeneous wax figure faces and real faces, Protocol II with 1200 pairs of homologous faces, and Protocol III combining the previous two protocols to simulate real-world operational scenarios. Both 3DMAD and HKBU-MARs-V1 datasets contain videos of 300 frames (3DMAD with 255 videos and HKBU-MARs-V1 with 120). We randomly selected 10 frames and averaged their scores as the final result in spoofing detection. For performance evaluation, we report the ISO/IEC 30107-3 metrics [34] - i.e., Attack Presentation Classification Error Rate (APCER), Bona Fide Presentation Classification Error Rate (BPCER), and Average Classification Error Rate (ACER). The detection accuracy is also used in our comparison.

4.2. Implementation details

We have followed the same method as the previous work [36] for weight initialization, and trained the proposed network using Adam optimizer [16] with the batch size of 16, β_1, and β_2 equal to 0.9, and 0.999, respectively. We set the initial learning rate to be 0.01 (decreased by a factor of 10 for every 90 epochs), and opt to terminate the training at 300 epochs. All

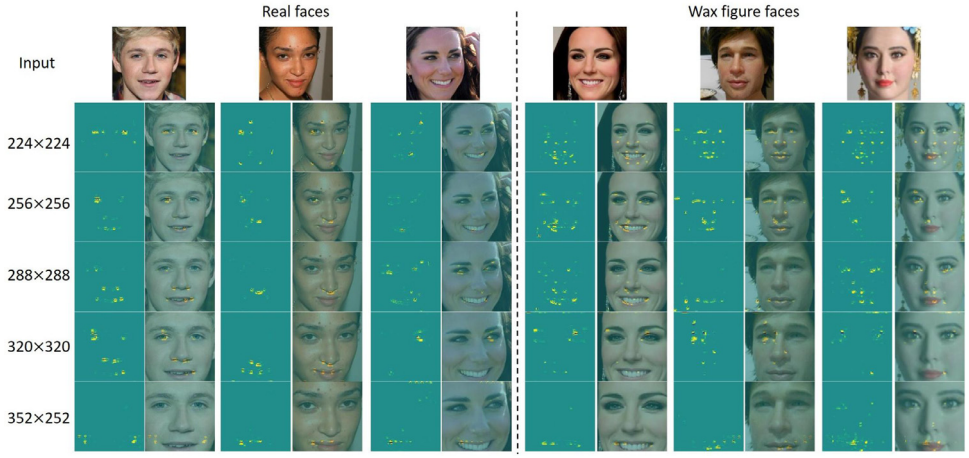

Fig. 8. Visualization of RAN network on wax figure face and real face classification based on saliency maps.

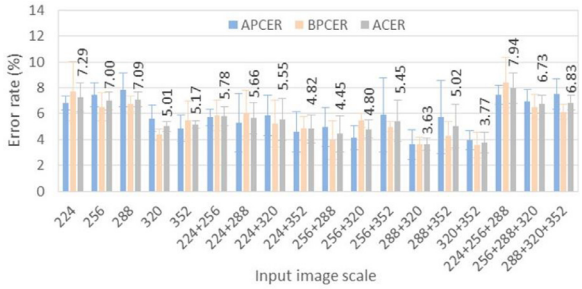

Fig. 9. Comparison of error rates with different input face scales. ACER values are labelled.

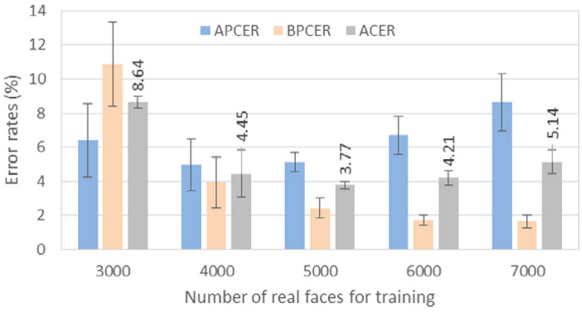

Fig. 10. Comparison of error rates with different real face numbers for training under 256 × 256 and 288 × 288 face scales fusion scheme. ACER values are labelled.

experiments are conducted using PyTorch on a workstation with four Titan XP GPUs.

4.3. Ablation study

Impact of face scale. We first show the influence of input face scale on wax figure face anti-spoofing. Fig. 9 presents the comparison results of using single face scales and combining two or three face scales as the input. Five face scales have been considered, including 224 × 224, 256 × 256, 288 × 288, 320 × 320, and 352 × 352. It can be observed that among different single face scales, larger face regions achieve lower error rates, especially for the 320 × 320 scale (with ACER of 5.01%). The fusion of two face scales results in the reduction of error rates. Especially for the fusion of 288 × 288 and 320 × 320, the final performance has achieved 3.63% for APCER, BPCER and ACER on SWFFD, sightly lower than combining 320 × 320 with 352 × 352 (with ACER of 3.77%) and using 256 × 256 and 288 × 288 (with ACER of 4.45%). However, we can see from the results in the last three groups that using three scales can not further improve the detection performance. We conjecture the reason is 'the curse of dimensionality' [17] that higher feature dimensionality tends to have the overfitting problem. Therefore, we suggest to use two scales to get promising performance. In the following experiment, we will report all results of both combining 288 × 288 with 320 × 320 scales and combining 256 × 256 with 288 × 288 scales for comparison.

Impact of database size. Since the real face dataset using CelebA-HQ is much larger than the introduced SWFFD, we have divided it into seven sessions each with the same size (4000) as the SWFFD considering the balance of fake and real faces for performance evaluation. Here we fixed the size of SWFFD and changed the real face size for training from 3000 to 7000 with a step size of 1000 to study the influence of training set size. The comparison results under 256 × 256 and 288 × 288 face scales fusion scheme are shown in Fig. 10. It is clear that unbalanced real and fake face samples lead to unbalanced APCER and BPCER rates, using 3000 with the highest BPCER (over 10%) while using a larger size of real faces resulting in smaller BPCER but larger APCER rates. With the same size as the SWFFD, the proposed method has achieved the most balanced results, although the average ACER of 4.45% is slightly higher than using 5000 real faces.

4.4. Intra-database testing

Intra-database testing is carried out on the proposed SWFFD dataset and compared with several state-of-the-art face anti-spoofing methods to show how they work for super-realistic 3D wax figure face spoofing attacks. These methods include two hand-crafted methods with promising performance on 3D mask databases: the multiscale LBP (MsLBP) [5] and Haralick features based [1]. Additionally we have included four deep learning-based methods into our benchmark: the FaceDs [13] based on noise modeling, Feathernets [40] using streaming module, and FaceBagNet [31] based on patch-based features, along with the original RAN method [36]. All these methods have public available codes so we can readily test them on the new SWFFD dataset.

As shown in Table 1, we can first observe the big differences among these methods. Two hand-crafted features and the noise modeling based method show obvious performance degradation (with accuracy less than 80% and ACER higher than 20%) in distinguishing between wax figure faces and real faces due to the high diversity and authenticity of attacks in the SWFFD dataset. The Feathernets, FaceBagNet, and RAN networks performed better with higher detection accuracy rates and lower error rates. Especially, the proposed method achieved the best performance under both fusion schemes, with accuracy over 95.5% and ACER lower than 4.5%. The FaceBagNet method ranked third thanks to the patch-based features learned from independent sub-networks, with the accuracy of 93.8%.

4.5. Cross-database testing on wax figure faces

In cross-database testing, we first show how well the proposed scheme and existing methods can perform in detecting unknown wax figure faces and real faces on SWFFD and WFFD datasets. In this experiment, training is conducted on SWFFD (with 4000 images in each subset) and testing on the Protocol II on WFFD dataset (with 2400 images). The comparison results in Table 2 illustrates the apparent degraded performance of all methods, with the ACER ranging from 25.69% to 41.55%. This can be attributed to the lower and more diverse quality of faces in WFFD than SWFFD because the collected wax figure and real faces were recorded in the same environment with the same camera in WFFD. Thanks to the attention-aware features of multiscale faces, our proposed method fusing 256 × 256 with 288 × 288 face scales has achieved the best performance with the lowest error rates, with a 6% error rate below the FaceBagNet method, and improving the original RAN-based method by as much as 10%.

We have also shown the cross-database testing results of using WFFD as the training set but SWFFD as the testing set in Table 3 (i.e., swap the role of WFFD and SWFFD). Using more diverse samples for training, all methods achieved higher classification accuracy and lower error rates under this testing scenario when compared with the results in Table 2. Similar performance differences but smaller gaps can be observed among different methods. Our proposed scheme fusing 288 × 288 with 320 × 320 face scales obtained the highest accuracy of 78.59%, and the lowest ACER of 21.41%. Such findings suggest that face spoofing detection performance degrades rapidly when the characteristics of the dataset vary. A promising solution to such cross-database cases is transfer learning [23,29], which we have left as the future research.

Table 1
Comparison results (%) of intra-database testing on SWFFD.

Method	Accuracy	APCER	BPCER	ACER
MsLBP [5]	78.81 ± 0.80	20.03 ± 1.05	22.36 ± 1.47	21.19 ± 0.80
Haralick [1]	74.52 ± 0.97	26.19 ± 2.06	24.77 ± 1.01	25.48 ± 0.97
RAN [36]	92.71 ± 1.12	6.83 ± 0.56	7.74 ± 2.27	7.29 ± 1.12
FaceDs [13]	75.36 ± 0.77	21.36 ± 0.70	41.91 ± 1.75	26.64 ± 0.77
Feathernets [40]	89.88 ± 1.40	9.01 ± 1.22	11.23 ± 1.70	10.12 ± 1.40
FaceBagNet [31]	93.80 ± 0.64	5.47 ± 0.73	6.93 ± 0.95	6.20 ± 0.64
Ours_256+288	95.55 ± 1.41	4.96 ± 1.54	3.94 ± 1.53	4.45 ± 1.41
Ours_288+320	**96.37 ± 0.48**	**3.63 ± 1.15**	**3.63 ± 0.55**	**3.63** ± 0.48

Table 2
Comparison results (%) of cross-database testing on WFFD.

Method	Accuracy	APCER	BPCER	ACER
MsLBP [5]	61.83 ± 0.28	30.25 ± 1.46	46.09 ± 1.41	38.17 ± 0.28
Haralick [1]	58.73 ± 0.50	39.68 ± 2.34	42.86 ± 1.95	41.27 ± 0.50
RAN [36]	64.32 ± 1.19	33.95 ± 8.16	37.40 ± 8.99	35.67 ± 1.19
FaceDs [13]	58.45 ± 0.29	21.55 ± 0.80	62.05 ± 1.37	41.55 ± 0.29
Feathernets [40]	63.24 ± 0.97	21.81 ± 1.28	51.71 ± 2.22	36.76 ± 0.97
FaceBagNet [31]	68.55 ± 1.28	21.48 ± 5.34	41.41 ± 3.79	31.45 ± 1.28
Ours_256+288	**74.31 ± 1.14**	**21.46 ± 3.64**	29.81 ± 3.77	**25.69 ± 1.14**
Ours_288+320	72.17 ± 2.08	26.07 ± 3.69	**29.59 ± 4.22**	27.83 ± 2.08

Table 3
Comparison results (%) of cross-database testing on SWFFD.

Method	Accuracy	APCER	BPCER	ACER
MsLBP [5]	62.86 ± 1.11	45.07 ± 1.86	29.20 ± 0.71	37.14 ± 1.11
Haralick [1]	63.01 ± 1.04	37.11 ± 1.72	36.86 ± 1.13	36.99 ± 1.04
RAN [36]	74.85 ± 1.48	30.63 ± 7.98	19.63 ± 8.98	25.15 ± 1.48
FaceDs [13]	63.13 ± 0.60	48.91 ± 1.65	24.83 ± 1.36	36.87 ± 0.60
Feathernets [40]	67.83 ± 1.11	32.60 ± 3.94	31.74 ± 3.45	32.17 ± 1.11
FaceBagNet [31]	75.94 ± 1.59	28.74 ± 3.50	**19.37 ± 2.07**	24.06 ± 1.59
Ours_256 + 288	77.33 ± 1.61	23.21 ± 3.16	22.13 ± 3.05	22.67 ± 1.61
Ours_288 + 320	**78.59 ± 2.32**	**22.53 ± 4.37**	20.28 ± 3.86	**21.41 ± 2.32**

Table 4
Comparison results (%) of cross-database testing on other 3D face spoofing datasets.

Method	SWFFD → 3DMAD			SWFFD → HKBU-MARs-V1		
	APCER	BPCER	ACER	APCER	BPCER	ACER
MsLBP [5]	41.93 ± 11.01	50.87 ± 10.14	46.40 ± 3.35	26.32 ± 20.19	59.69 ± 27.00	43.01 ± 10.18
Haralick [1]	35.11 ± 12.70	53.31 ± 11.63	44.21 ± 5.69	42.50 ± 2.07	51.14 ± 4.45	46.82 ± 1.84
RAN [36]	29.18 ± 19.56	36.34 ± 26.53	33.26 ± 3.81	53.00 ± 19.54	30.52 ± 19.39	41.76 ± 10.33
FaceDs [13]	65.16 ± 4.97	23.37 ± 3.91	44.26 ± 2.18	60.57 ± 1.34	**24.16 ± 1.05**	42.37 ± 0.56
Feathernets [40]	60.86 ± 29.13	35.41 ± 13.05	48.13 ± 9.81	26.75 ± 21.35	50.16 ± 21.04	38.46 ± 6.04
FaceBagNet [31]	**26.19 ± 15.93**	47.90 ± 24.93	37.05 ± 6.07	31.10 ± 13.00	58.20 ± 10.23	44.65 ± 4.05
Ours_256 + 288	32.81 ± 23.00	**23.34 ± 10.35**	28.07 ± 9.54	47.00 ± 12.34	26.37 ± 14.24	36.69 ± 1.95
Ours_288 + 320	29.90 ± 21.29	24.51 ± 19.06	**27.20 ± 7.24**	**23.79 ± 12.83**	35.88 ± 10.48	**29.84 ± 6.83**

4.6. Cross-database testing on other 3D face spoofing attacks

To further show the generalization ability of the proposed scheme and existing methods in detecting other 3D face spoofing attacks, we have conducted cross-database experiments with training on SWFFD (with 4000 images in each subset) and testing on 3DMAD and HKBU-MARs-V1 datasets (using all the selected frames). The comparison results in Table 4 show similar performance differences for most methods with those in Table 2 using the same data for training. The proposed method demonstrates the best generalizability. Specifically, using 288 × 288 and 320 × 320 face scales achieved the lowest ACER values with 27.20% on 3DMAD and 29.84% on HKBU-MARs-V1, quite close to the result on WFFD dataset (with ACER of 27.88%), while combining 256 × 256 with 288 × 288 ranked second in detecting different kinds of unknown 3D face spoofing attacks.

5. Conclusions

To address the limitations in existing 3D face spoofing attack databases, we have introduced a new database, SWFFD, composed of large-scale single wax figure faces with high diversity from online resources as super-realistic 3D face spoofing attacks. We also propose an effective method to learn discriminative attention-aware features from different face scales for the detection of wax figure faces and real ones. By combining the SWFFD with real faces in CelebA-HQ for performance evaluation, we have found that the proposed method demonstrates promising accuracy and robustness performance under both intra-database and cross-database testing. The benefits of multiscale fusion (especially using larger face scales) and database size balancing have also been verified through our ablation studies. It should be noted that the best performance achieved by the proposed scheme under cross-database setting still

has the error rate of over 20%. Super-realistic wax figure faces are indeed difficult to distinguish from real ones even for humans, especially with unknown attack patterns. Therefore, how to further improve the generalization property of existing methods deserves further study for not only 2D face anti-spoofing but also 3D realistic face anti-spoofing in the future. Recently developed transfer learning techniques such as domain adaptation deserve systematic study for cross-database anti-spoofing detection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is partially supported by the DoJ/NIJ under grant NIJ 2018-75-CX-0032, NSF under grant OAC-1839909, IIS-1951504 and the WV Higher Education Policy Commission Grant (HEPC.dsr.18.5).

References

[1] A. Agarwal, R. Singh, M. Vatsa, Face anti-spoofing using Haralick features, in: Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, IEEE, 2016, pp. 1–6.
[2] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, 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, pp. 81–89.
[3] A. Ali, S. Hoque, F. Deravi, Gaze stability for liveness detection, Pattern Anal. Appl. 21 (2) (2018) 437–449.
[4] S. Bhattacharjee, S. Marcel, What you can't see can help you–extended-range imaging for 3D-mask presentation attack detection, in: Proceedings of the 16th International Conference on Biometrics Special Interest Group., in: EPFL–CONF-231840, Gesellschaft fuer Informatik eV (GI), 2017.
[5] N. Erdogmus, S. Marcel, Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect, in: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, IEEE, 2013, pp. 1–6.
[6] N. Erdogmus, S. Marcel, Spoofing face recognition with 3D masks, IEEE Trans. Inf. Forensics Secur. 9 (7) (2014) 1084–1097.
[7] J. Galbally, R. Satta, Three-dimensional and two-and-a-half-dimensional face recognition spoofing using three-dimensional printed models, IET Biom. 5 (2) (2016) 83–91.
[8] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, S. Marcel, Biometric face presentation attack detection with multi-channel convolutional neural network, IEEE Trans. Inf. Forensics Secur. 15 (2019) 42–55.
[9] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[10] ISO, I.O.f.S., Information technology biometric presentation attack detection - part 1: framework, ISO/IEC JTC 1/SC 37, ISO/IEC 30107-1:2016(E)(2016).
[11] S. Jia, G. Guo, Z. Xu, A survey on 3D mask presentation attack detection and countermeasures, Pattern Recognit. 98 (2020) 107032.
[12] S. Jia, C. Hu, G. Guo, Z. Xu, A database for face presentation attack using wax figure faces, in: International Conference on Image Analysis and Processing, Springer, 2019, pp. 39–47.
[13] A. Jourabloo, Y. Liu, X. Liu, Face de-spoofing: anti-spoofing via noise modeling, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 290–306.
[14] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation, arXiv preprint arXiv:1710.10196 (2017).
[15] Y. Kim, J. Na, S. Yoon, J. Yi, Masked fake face detection using radiance measurements, JOSA A 26 (4) (2009) 760–766.
[16] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: International Conference on Learning Representations, May 7–9, 2015, San Diego, CA, ICLR, 2015.
[17] F.Y. Kuo, I.H. Sloan, Lifting the curse of dimensionality, Not. AMS 52 (11) (2005) 1320–1328.
[18] H. Li, W. Li, H. Cao, S. Wang, F. Huang, A.C. Kot, Unsupervised domain adaptation for face anti-spoofing, IEEE Trans. Inf. Forensics Secur. 13 (7) (2018) 1794–1809.
[19] L. Li, Z. Xia, X. Jiang, Y. Ma, F. Roli, X. Feng, 3D face mask presentation attack detection based on intrinsic image analysis, IET Biom. 9 (3) (2020) 100–108.
[20] J. Liu, A. Kumar, Detecting presentation attacks from 3D face masks under multispectral imaging, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 47–52.
[21] S. Liu, X. Lan, P. Yuen, Temporal similarity analysis of remote photoplethysmography for fast 3D mask face presentation attack detection, in: The IEEE Winter Conference on Applications of Computer Vision, 2020, pp. 2608–2616.
[22] S. Liu, B. Yang, P.C. Yuen, G. Zhao, A 3D mask face anti-spoofing database with real world variations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 100–106.
[23] O. Lucena, A. Junior, V. Moia, R. Souza, E. Valle, R. Lotufo, Transfer learning using convolutional neural networks for face anti-spoofing, in: International Conference Image Analysis and Recognition, Springer, 2017, pp. 27–34.
[24] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, A. Majumdar, Detecting silicone mask-based presentation attack via deep dictionary learning, IEEE Trans. Inf. Forensics Secur. 12 (7) (2017) 1713–1723.
[25] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, A. Majumdar, Detecting silicone mask-based presentation attack via deep dictionary learning, IEEE Trans. Inf. Forensics Secur. 12 (7) (2017) 1713–1723.
[26] S. Marcel, M.S. Nixon, S.Z. Li, Handbook of Biometric Anti-Spoofing, 1, Springer, 2014.
[27] D. Menotti, G. Chiachia, A. Pinto, W.R. Schwartz, H. Pedrini, A.X. Falcao, A. Rocha, Deep representations for iris, face, and fingerprint spoofing detection, IEEE Trans. Inf. Forensics Secur. 10 (4) (2015) 864–879.
[28] S. Naveen, R.S. Fathima, R. Moni, Face recognition and authentication using LBP and BSIF mask detection and elimination, in: Communication Systems and Networks, International Conference on, IEEE, 2016, pp. 99–102.
[29] K. Patel, H. Han, A.K. Jain, Cross-database face antispoofing with robust feature representation, in: Chinese Conference on Biometric Recognition, Springer, 2016, pp. 611–619.
[30] R. Shao, X. Lan, P.C. Yuen, Deep convolutional dynamic texture learning with adaptive channel-discriminability for 3D mask face anti-spoofing, in: Biometrics (IJCB), 2017 IEEE International Joint Conference on, IEEE, 2017, pp. 748–755.
[31] T. Shen, Y. Huang, Z. Tong, Facebagnet: Bag-of-local-features model for multi-modal face anti-spoofing, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.
[32] T.A. Siddiqui, S. Bharadwaj, T.I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, N. Ratha, Face anti-spoofing with multifeature videolet aggregation, in: Pattern recognition (ICPR), 2016 23rd international conference on, IEEE, 2016, pp. 1035–1040.
[33] L. Souza, L. Oliveira, M. Pamplona, J. Papa, How far did we get in face spoofing detection? Eng. Appl. Artif. Intell. 72 (2018) 368–381.
[34] I.O. for Standardization, Information technology biometric presentation attack detection - part 3: testing and reporting, ISO/IEC JTC 1/SC 37, ISO/IEC 30107-3:2017(2017).
[35] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
[36] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.
[37] Q. Wang, Z. Teng, J. Xing, J. Gao, W. Hu, S. Maybank, Learning attentions: residual attentional siamese network for high performance online visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4854–4863.
[38] Y. Wang, X. Hao, Y. Hou, C. Guo, A new multispectral method for face liveness detection, in: Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on, IEEE, 2013, pp. 922–926.
[39] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Process. Lett. 23 (10) (2016) 1499–1503.
[40] P. Zhang, F. Zou, Z. Wu, N. Dai, S. Mark, M. Fu, J. Zhao, K. Li, Feathernets: Convolutional neural networks as light as feather for face anti-spoofing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019. 0–0
[41] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 286–301.