‘Rehearsal-Free Domain Continual Face Anti-Spoofing: Generalize More and Forget Less’ (Appendix)

Rizhao Cai[1] Yawen Cui[2] Zhi Li[3] Zitong Yu[4,5*]
Haoliang Li[6] Yongjian Hu[7] Alex Kot[1]
[1]Rapid-Rich Object Search (ROSE) Lab, NTU-PKU Joint Research Institute, Nanyang Technological University
[2]University of Oulu [3]Bytedance Ltd. [4]School of Computing and Information Technology, Great Bay University
[5] Great Bay Institute for Advanced Study [6]City University of Hong Kong [7]South China University of Technology
rzcai@ntu.edu.sg, yawen.cui@oulu.fi, zhi.li.2023@bytedance.com, yuzitong@gbu.edu.cn
haoliang.li@cityu.edu.hk, eeyjhu@scut.edu.cn, eackot@ntu.edu.sg

A.1. Overview

The Appendix is organized as follows
- A.2: Results of joint training with previous rehearsal data;
- A.3: Discussion on different implementations of learnable θ;
- A.4: Exploration of few-shot continual learning;
- A.5: Details of public datasets used in our work;
- A.6: Detailed experimental results corresponding to Figure 3 and Figure 4 in the paper.

A.2. Results of joint training with previous rehearsal data

To obtain the upper bound of the proposed domain continual learning setting, we conduct joint training (JT), where all previous domain data is available and jointly used for training models whenever new domain data is coming. The results can be used to indicate the gap between the method with/without rehearsal data. The results are shown in Table a1, where ‘JT’ includes all rehearsal data, and ‘PPCR’ has no rehearsal data. It is obvious that ‘JT’ generally has less forgetting than the PPCR as it has all the seen domain data. However, with all seen domain data, the unseen performance of ‘JT’ is not necessarily better PPCR. Therefore, cross-domain performance is important and should not be ignored

Table a1. Experiments of results (%) of Joint training (JT)

Backbone	Method	P1			P2		
		mAA	mABT	mAGA	mAA	mABT	mAGA
ViT-ConvA	JT	95.53	-0.56	80.22	94.47	-0.12	72.66
ViT-ConvA	PPCR	93.04	-3.08	80.86	93.91	-1.76	79.14
ViT-CDCA	JT	96.03	-0.36	81.69	94.43	-1.88	74.52
ViT-CDCA	PPCR	92.12	-3.92	77.17	94.26	-1.25	75.73
ViT-DCDCA	JT	96.06	-0.14	82.33	94.58	1.83	78.73
ViT-DCDCA	PPCR	93.54	-3.48	82.06	94.03	**-1.72**	80.08

*Corresponding author

A.3. Different implementations of learnable θ

A naive implementation of training a learnable θ is to define $\theta = \text{Sigmoid}(\Theta)$, where Θ is a variable that can be trained and updated in backward propagation, and Sigmoid function is used to constrain the value in $[0, 1]$. We name the above implementation as 'Naive'. Though simple, the optimization of the Naive implementation could be sub-optimal that is trapped in local minima. The Proposed optimization in the paper (named 'Proposed') samples θ from a learnable distribution, and the sampling with randomness can help θ to avoid being trapped in local minima and get better generalization performance.

Table a2 compares the results of our ViT-DCDCA with the 'Naive' and 'Proposed' methods. In Protocol-2, the implementation of learnable has less forgetting ($mABT$) than the proposed algorithm, but the generalization performance $mAGA$ is poorer than our Proposed implementation. Moreover, In Protocol-1, our Proposed implementation is better than Naive in terms of the three metrics. As such, we can see the Proposed implementation is better than the Naive implementation, especially in terms of the generalization performance.

Table a2. Experimental results of ViT-DCDCA with the Naive implementation and our Proposed implementation.

ViT-DCDCA	Protoco-1			Prococol-2		
	mAA (%)	$mABT$ (%)	$mAGA$ (%)	mAA (%)	$mABT$ (%)	$mAGA$ (%)
Naive	92.72	-4.37%	74.70	**94.86**	**-1.49**	75.64
Proposed	**93.23**	**-3.59%**	**78.70**	93.79	-1.82	**78.09**

A.4. Few-shot data in continual learning sessions

In the main paper, we create a low-shot (50-shot) scenario during the continual sessions, where there are 50 real face examples and 50 spoofing face examples for training. In the Appendix, we explore the few-shot scenario and conduct 5-shot experiments, where there are only 5 real face examples and 5 spoofing face examples for training. The results are presented in Table a3. We can see from Table a3 that our ViT-DCDCA-PPCR consistently performs better than ViT with vanilla linear adapters by a clear margin in the few-shot scenario.

Table a3. 5-shot experiments in our proposed continual learning Protocol-1 and Protocol-2.

5-shot	Protoco-1			Prococol-2		
	mAA	$mABT$	$mAGA$	mAA	$mABT$	$mAGA$
ViT-Adapter	74.94%	-11.31%	56.41%	77.69%	-5.06%	57.02%
ViT-DCDCA-PPCR (ours)	**90.07%**	**-0.83%**	**82.80%**	**87.57%**	**-1.06%**	**85.56%**

A.5. Details of public datasets used in our work

In our work, we use VIS (RGB) data from 15 public datasets to construct the two continual learning protocols. As shown in Table a4, the base session contains 2D and 3D attack examples from SiW [10], CelebA-Spoof [14] , and HiFiMask [8] datasets. In the continual learning, 10 datasets are used, including IDIAP REPLAY-ATTACK [3], CASIA-FASD [15], MSU MFSD [11], HKBU MARsV2[9], OULU-NPU[2], CSMAD[1], CASIA-SURF[13], WFFD[5], WMCA[4], and CASIA-SURF 3DMASK (CASIA-3DMASK) [12]. The ROSE-YOU [6] and CeFA [7] are used as the unseen domain datasets for testing the unsee domain generalization performance. More information about the above datasets is summarized in Table a4, and Figure a1-a15 show examples of the above datasets.

In our proposed PPCR, 2D attack and 3D attack examples are clustered separately. We describe our way of defining 2D and 3D attack examples below. The 2D attack usually means the attack is launched by printed photos and digital displays, which are flat and 2D. Thus, the Print and Replay attacks in SiW, REPLAY-ATTACK, CASIA-FASD, MSU-MFSD, OULU-NPU, CASIA-SURF, WMCA, ROSE-YOUTU are defined as 2D attacks. We are also aware that in CelebA-Spoof (*e.g.*, Figure a2(h)), CASIA-FASD (*e.g.*, Figure a5(f)), CASIA-SURF (*e.g.*, Table a4. A summary of available datasets used in our work. The first column lists the datasets. † means the dataset contains 2D attacks and ‡ means the dataset contain 3D attacks

Dataset	Year	#Live/#Spoof	#Subjects	Modality	Attacks
			Base		
SiW [10]†	2018	1320/3300	20	VIS	Print(flat, wrapped)
CelebA-Spoof [14]†	2020	156384/469153	10177	VIS	Print(flat,wrapped),Replay(monito, tablet, phone), Mask(paper)
HiFi-Mask[8] ‡	2021	13650/40950	75	VIS	Mask(transparent, plaster, resin)
			Continual		
REPLAY-ATTACK [3] †	2012	200/100	50	VIS	Print(flat), Replay(tablet, phone)
CASIA-FASD [15] †	2012	150/450	50	VIS	Print(flat, wrapped, cut), Replay(tablet)
MSU-MFSD [11]†	2014	720/210	35	VIS	Print(flat), Replay(tablet, phone)
HKBU MARsV2 [9]‡	2016	502/502	12	VIS	Mask(hard resin) from Thatsmyface and REAL-f
OULU-NPU [2]†	2017	720/2880	55	VIS	Print(flat), Replay(phone)
CSMAD[1] ‡	2018	104/159	14	VIS/Depth/NIR/Thermal	Mask(custom silicone)
CASIA-SURF[13] †	2019	3000/18000	1000	VIS/Depth/NIR	Print(flat, wrapped, cut)
WFFD [5] ‡	2019	2300/2300	745	VIS	Waxworks(wax)
WMCA [4]† ‡	2019	347/1332	72	VIS/Depth/NIR/Thermal	Print(flat), Replay(tablet), Partial(glasses), Mask(plastic, silicone, and paper, Mannequin)
CASIA-SURF 3DMASK [12] ‡	2020	288/864	48	VIS	Mask(mannequin with 3D print)
			Unseen		
ROSE-YOUTU [6]†	2018	500/2850	20	VIS	Print(flat), Replay(monitor, laptop), Mask(paper, crop-paper)
CeFA [7]† ‡	2020	6300/27900	1607	VIS/Depth/NIR	Print(flat, wrapped), Replay, Mask(3D print, silica gel)

Figure a10(e)-(h)), and ROSE-YOUTU (*e.g.*, Figure a14(e)-(f)), the printed photo/paper can be wrapped or cropped as a paper mask. These examples may not be flat, but we still define such examples as 2D attacks based on three concerns. First, some of the wrapped or cropped paper still looks nearly flat, textite.g., Figure a5(f), Figure a14(f). Furthermore, such wrapped or paper masks do not have delicate 3D information about the human face structure. Second, both wrapped paper and cropped paper attacks are the same as flat print paper attacks in terms of paper materials. Third, annotating flat paper, wrapped paper, or cropped paper attacks requires extract cost. Given the above concerns, we still define these wrapped paper and cropped paper attacks as 2D attacks, which can avoid extra and expensive data annotations.

On the other hand, the 3D attack is defined by non-paper masks or face mannequins, which contain 3D information of human face structure, such as Resin masks (*e.g.* Figure a3(f), Figure a3(e)-(h)), Silicone masks (*e.g.* Figure a9 (e)-(h)) Face mannequins (*e.g.* Figure a11(e)-(h), Figure a13(e)-(h)), and so on.

A.6. Detailed experimental results

In the paper, the performance curves of Protocol-1 and Protocol-2 are shown in Figure 3 and Figure 4 of the paper respectively. In Figure 3 and Figure 4 of the paper, the exact numbers are not displayed because of the limited space. To encourage reproduction and comparison, we provide detailed experimental results in the below tables. Table a5, Table a6, Table a7, Table a8, Table a9, and Table a10 provide the results of ResNet18, ViT-Adapter, ViT-ConvA, ViT-CDCA, ViT-DCDCA, and ViT-DCDCA-PPCR respectively in Protocol-1. Similarly, Table a11, Table a12, Table a13, Table a14, Table a15, and Table a16 are the corresponding results in Protocol-2. More specifically, results in column 0 (BASE) of Table a5 correspond to the curve of ResNet18 in Figure 3(a) in the paper. Likewise, results in column 1 (REPLAY-ATTACK) of Table a5 correspond to the curve of ResNet18 in Figure 3(b) in the paper, and so on. Also, the columns 'ROSE-YOUTU' and 'CeFA' correspond to the curves of ResNet18 in Figure 3(l) and Figure 4(m) respectively. The above rule also applies to the results in Table a6- a16 and Figure 3 and Figure 4.

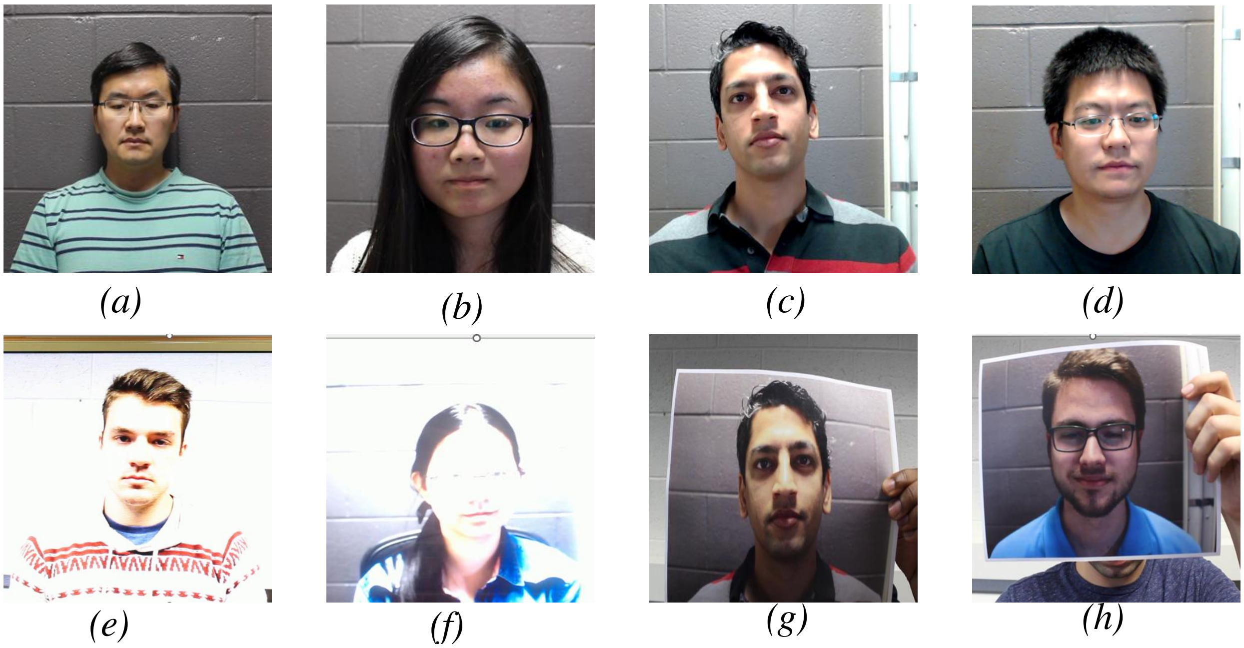

Figure a1. Examples of the SiW dataset [10]. The top row (a,b,c,d) are the real face/bona fide examples. (e) and (f) are Replay attack examples. (g) and (h) are Print attack examples.

Figure a2. Examples of the CelebA-Spoof dataset [14]. The top row (a,b,c,d) are the real face/bona fide examples. (e) is a Print photo. (f) and (g) are Replay attack examples.(h) is a paper mask attack example.

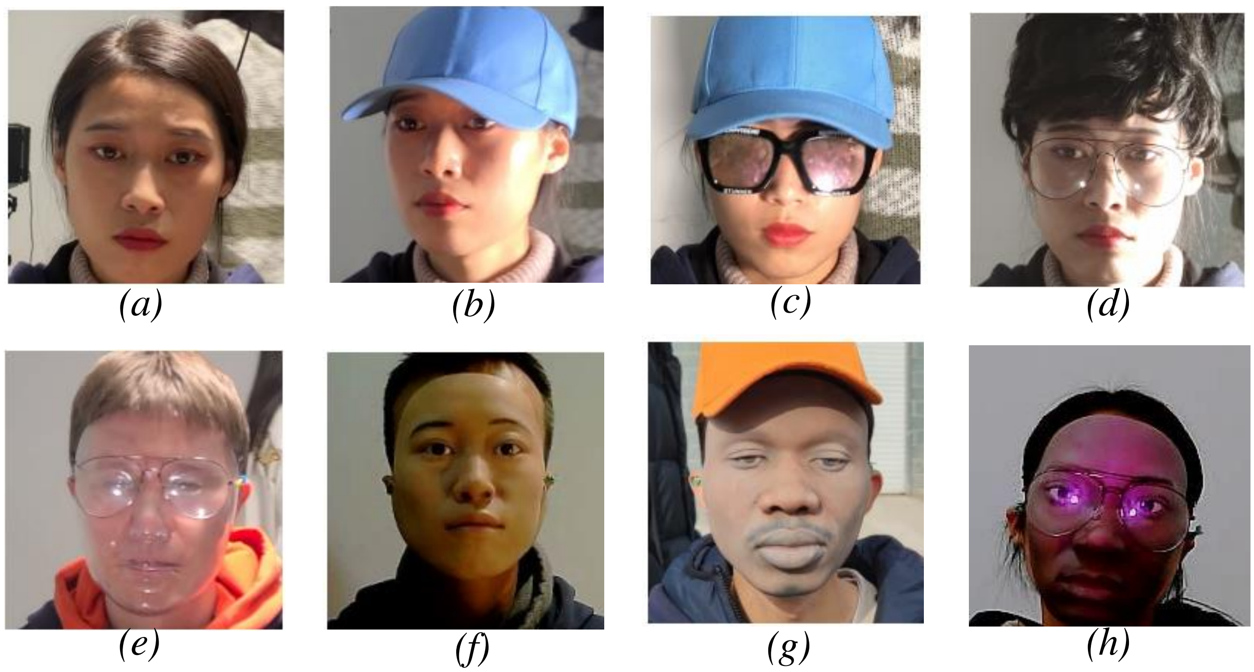

Figure a3. Examples of the HiFi-Mask dataset [8]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e) is a transparent mask attack. (f) is a Resin mask attack, (g) and (h) are plaster mask attack examples.

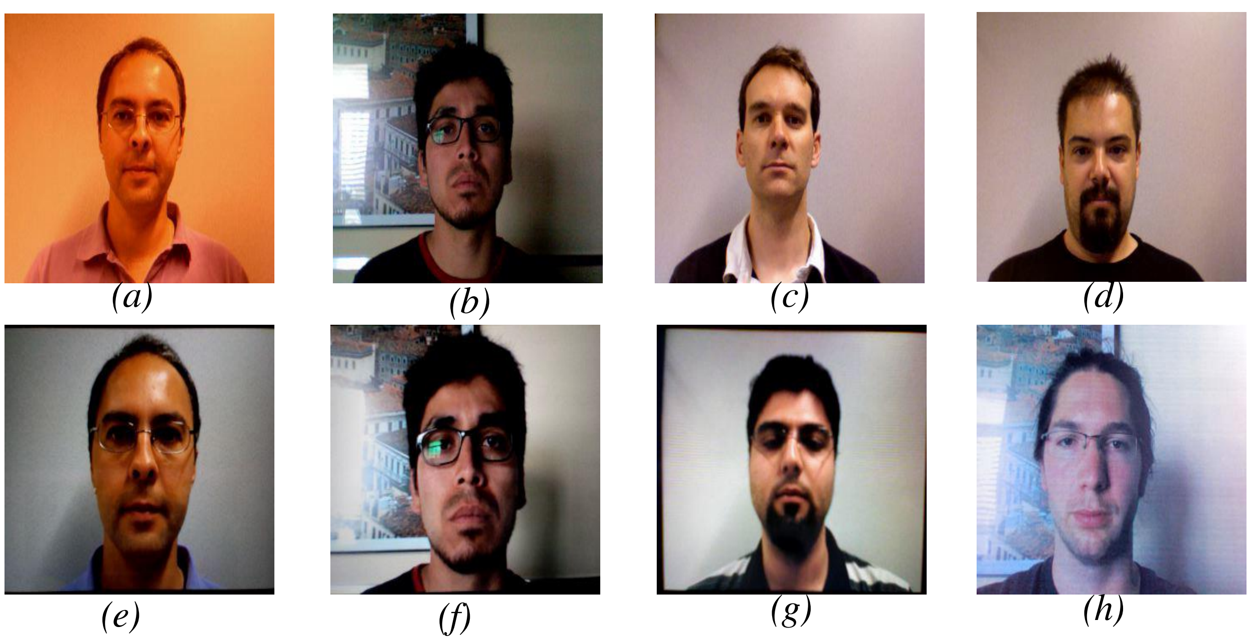

Figure a4. Examples of the IDIAP REPLAY-ATTACK dataset [3]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e), (f), (g), and (h) are Replay attack samples.

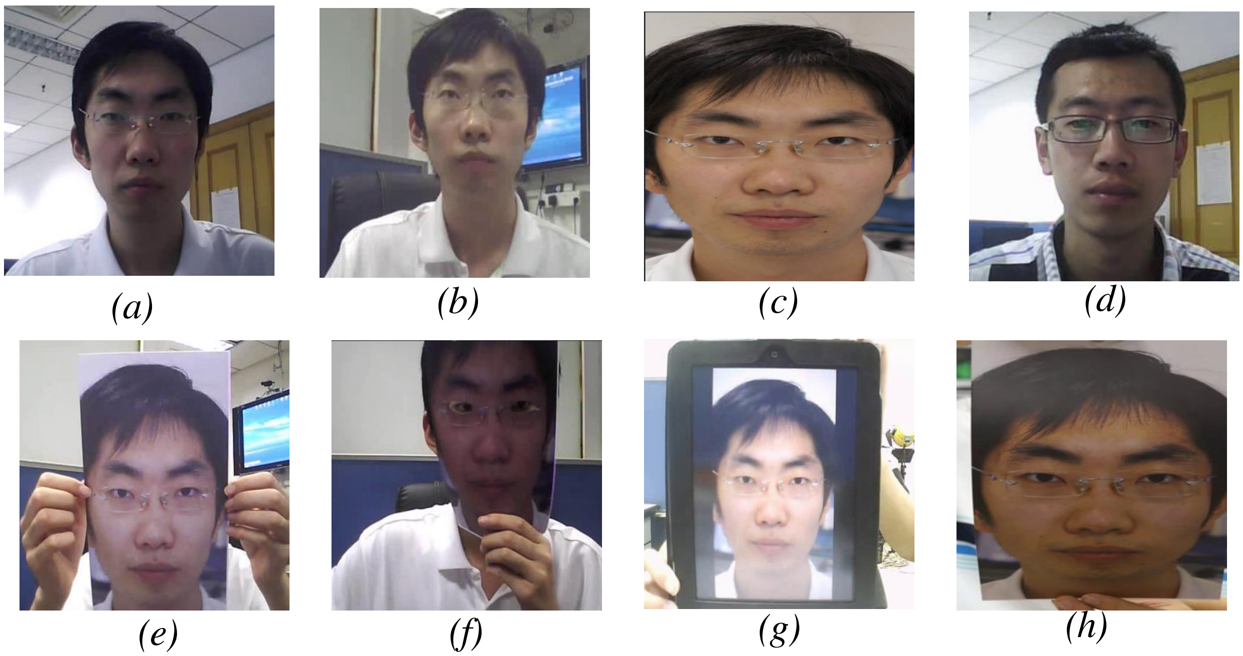

Figure a5. Examples of the CASIA-FASD dataset [15]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e) and (h) are (flat) Print photo attack examples. (f) is a wrapped paper attack example. (g) is a Replay attack example.

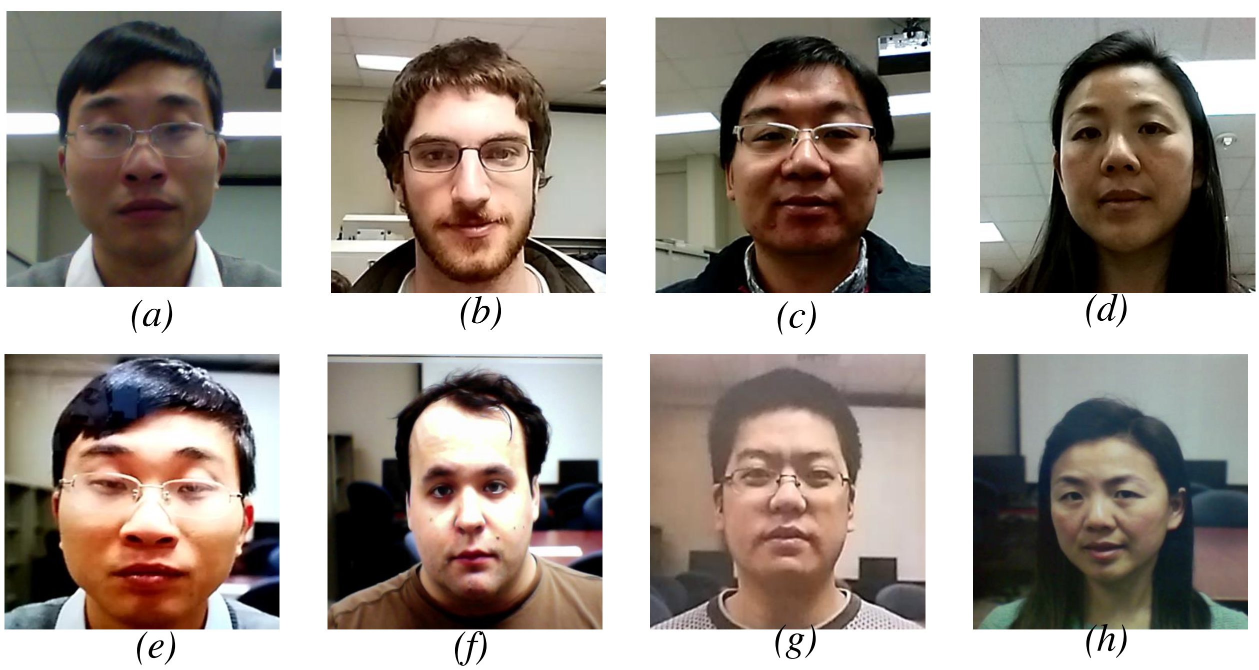

Figure a6. Examples of the MSU MFSD dataset [11]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e) and (h) are (flat) Print photo attack examples. (f) is a wrapped paper attack example. (g) is a Replay attack example.

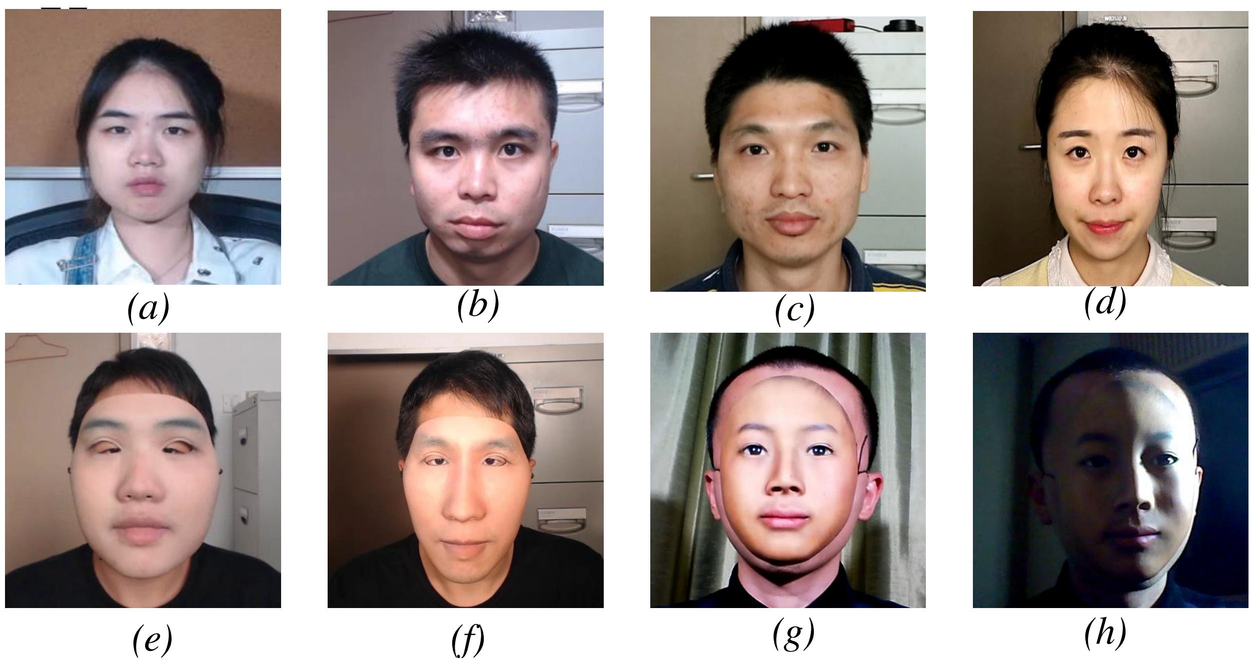

Figure a7. Examples of the HKBU MARsV2 dataset [9]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e), (f), (g) and (h) are hard resin mask attack examples.

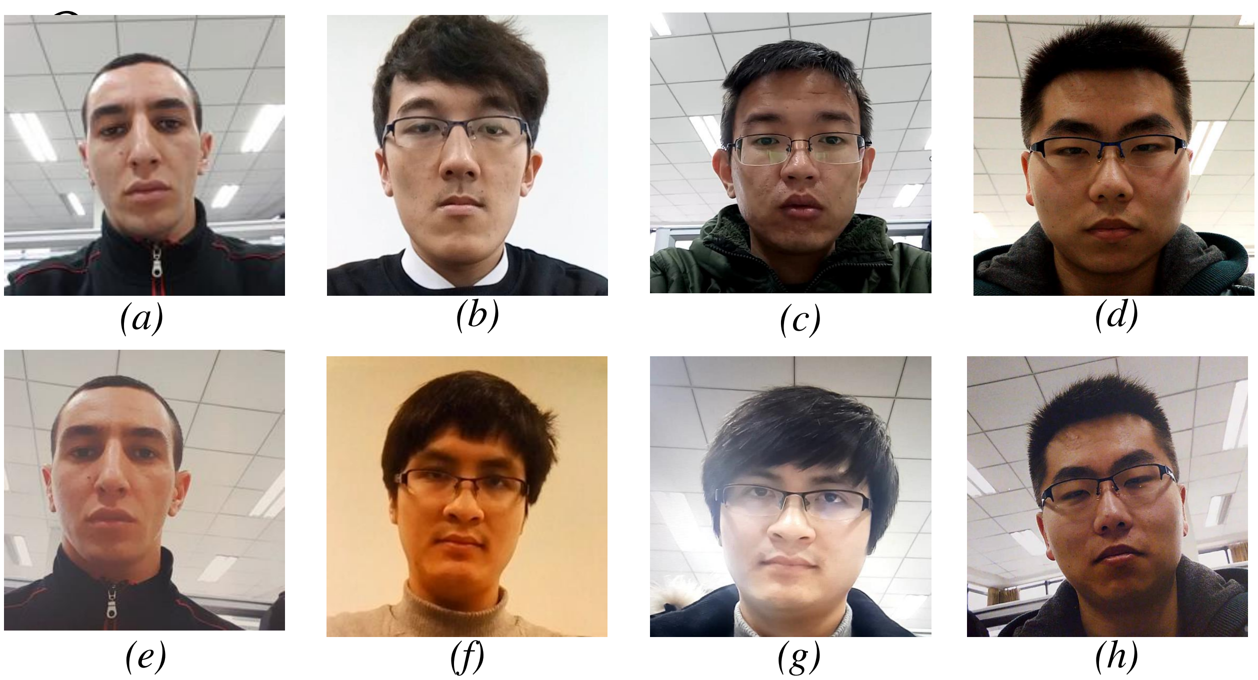

Figure a8. Examples of the OULU-NPU dataset [2]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e) and (f) are Print photo attack examples. (g) and (h) are Replay attack examples.

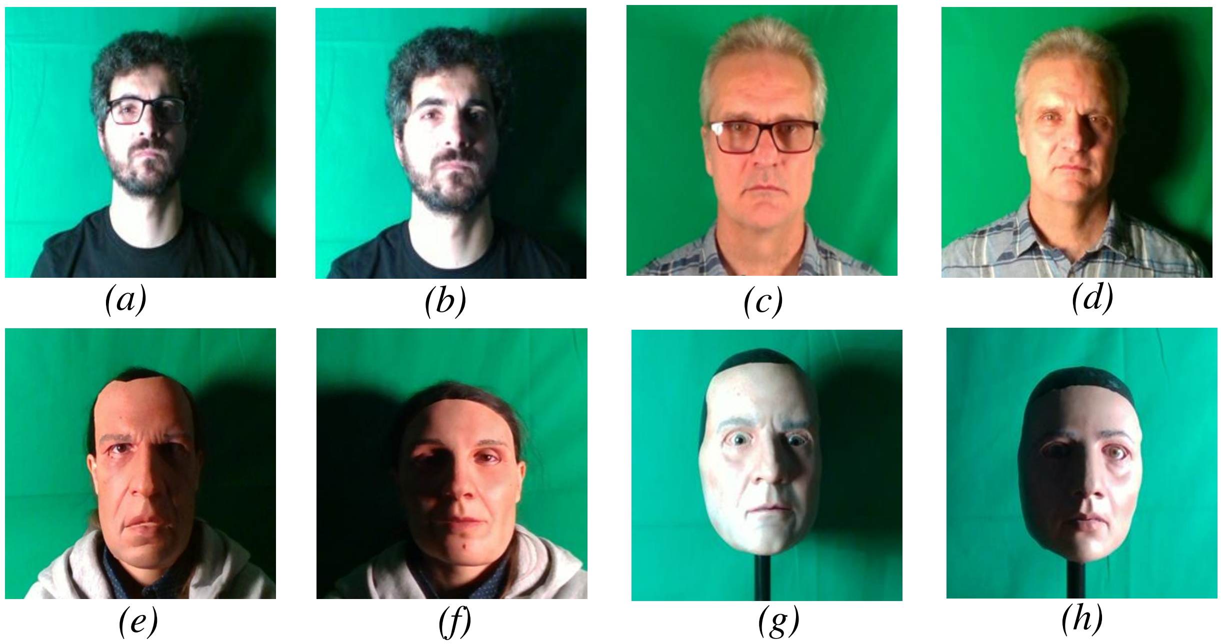

Figure a9. Examples of CSMAD dataset [1]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e) and (f) are silicon masks worn by subjects. (g) and (h) are stand silicon mask attack examples.

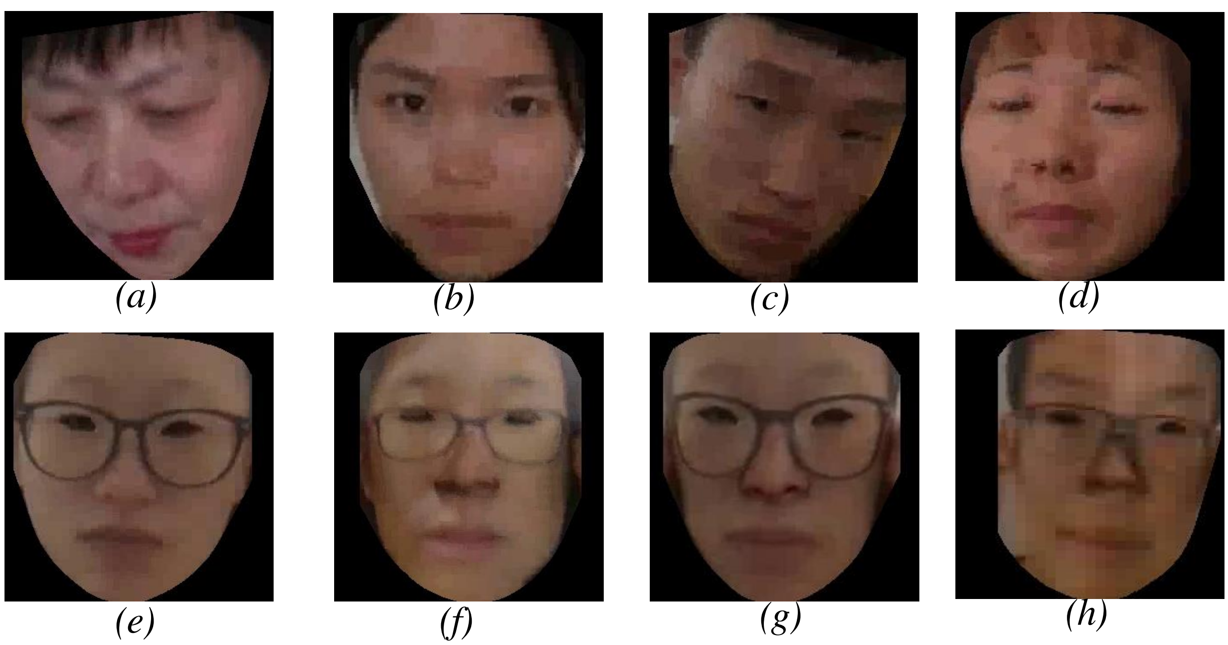

Figure a10. Examples of CASIA-SURF dataset [13]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e), (f), (g) and (h) are cropped and printed paper attack examples.

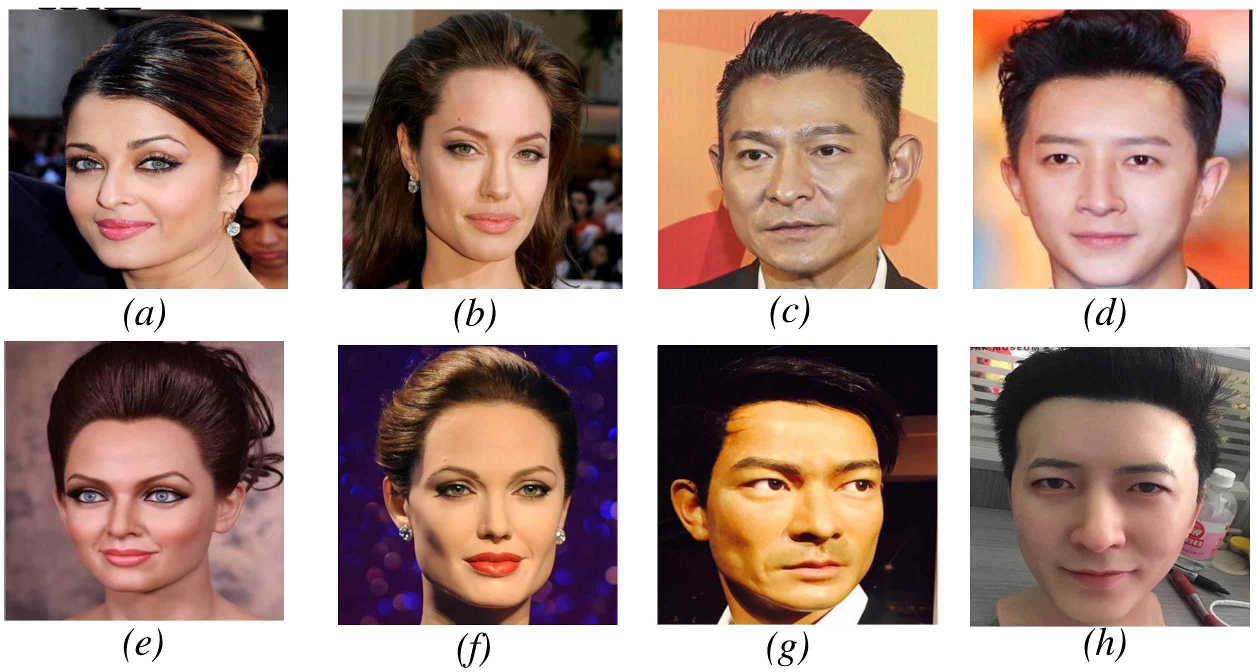

Figure a11. Examples of WFFD dataset [5]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e), (f), (g) and (h) are attack examples of waxworks.

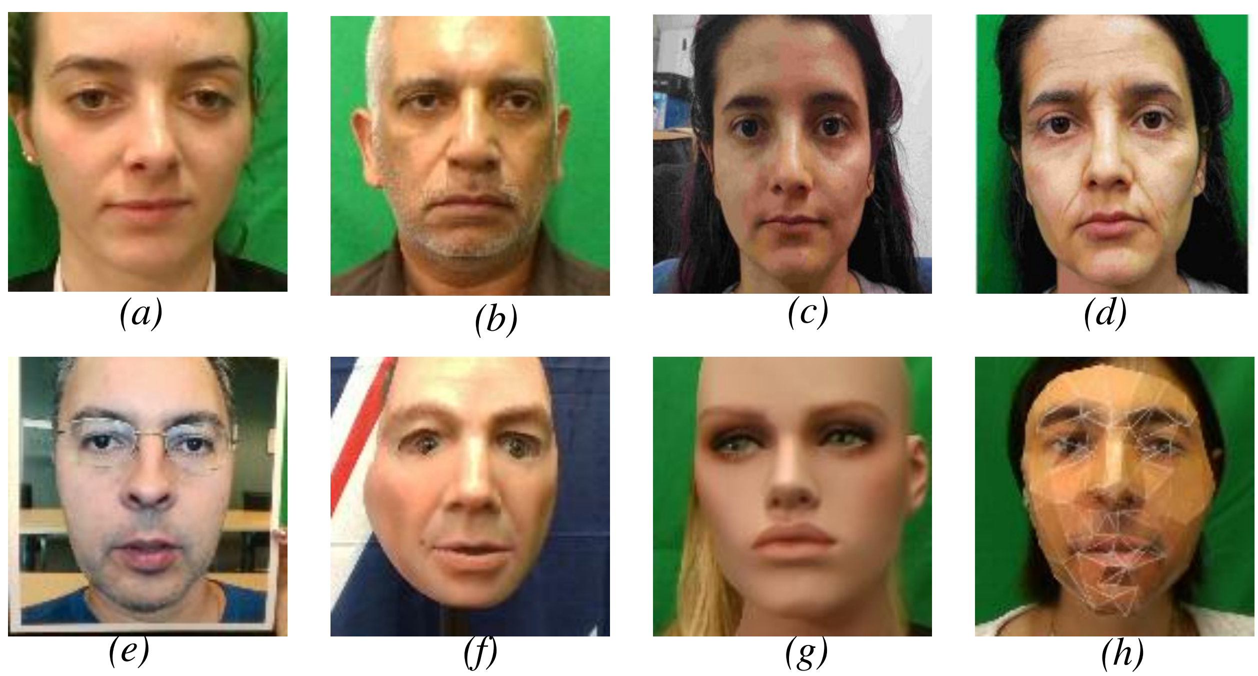

Figure a12. Examples of WMCA dataset [4]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e), (f), (g) and (h) are different 3D mask attacks.

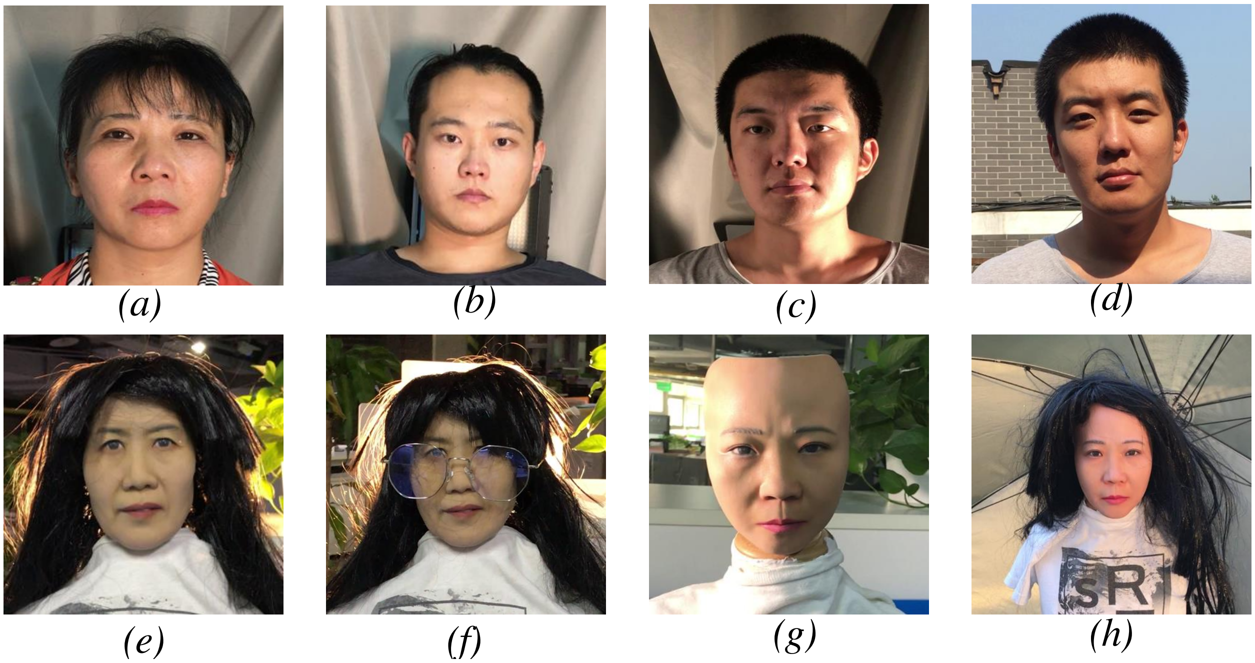

Figure a13. Examples of CASIA-SURF 3DMASK dataset [12]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e), (f), (g) and (h) are attack examples of mannequins manufactured by 3D printing.

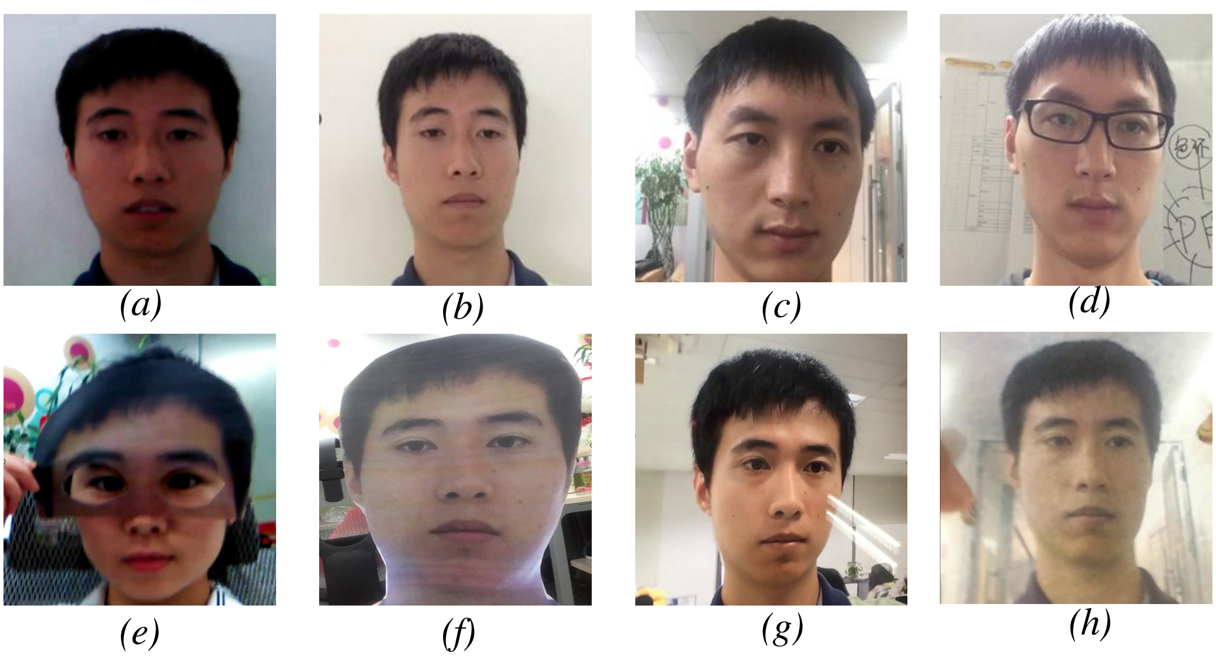

Figure a14. Examples of ROSE-YOUTU dataset [6]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e) and (f) are cropped paper mask attacks. (g) is a Replay attack example. (h) is a flat printed photo attack example.

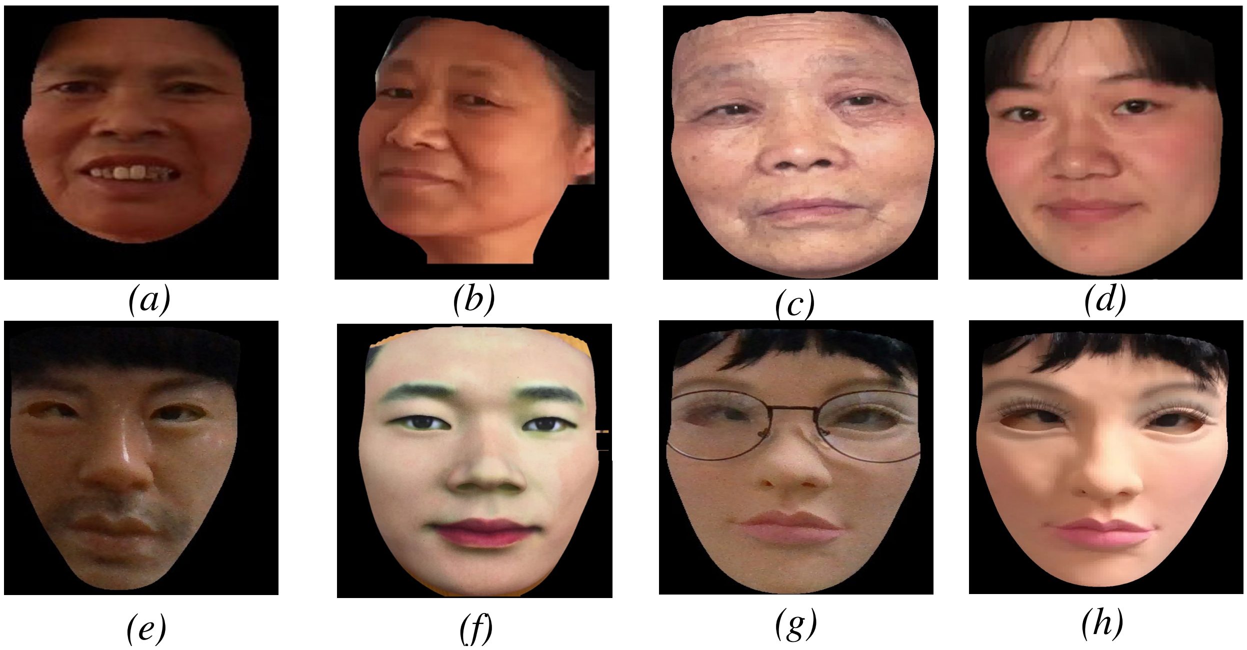

Figure a15. Examples of the CeFA dataset [7]. The top row (a,b,c,d) are the real face/bona fide examples captured in different environments. (e),(f), (g), and (h) are different 3D mask attack examples.

Table a5. Detail experimental results of each session of ResNet18 in Protocol-1. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ResNet18 in Protocol 1													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	REPLAY ATTACK	CASIA FASD	MSU MFSD	HKBU MarV2	OULU NPU	CSMAD	CASIA -SURF	WFFD	WMCA	CASIA 3DMASK	ROSE YOUTU	CeFA
0	97.39%	-	-	-	-	-	-	-	-	-	-	52.81%	81.58%
1	74.95%	89.18%	-	-	-	-	-	-	-	-	-	47.19%	73.68%
2	81.03%	67.64%	91.62%	-	-	-	-	-	-	-	-	43.54%	60.11%
3	92.57%	87.95%	92.88%	95.76%	-	-	-	-	-	-	-	44.54%	72.98%
4	92.24%	80.35%	90.81%	90.98%	96.42%	-	-	-	-	-	-	42.58%	72.41%
5	86.92%	72.04%	87.01%	86.99%	97.39%	90.42%	-	-	-	-	-	46.22%	69.38%
6	89.64%	63.26%	85.55%	88.88%	97.16%	85.48%	94.67%	-	-	-	-	52.31%	61.77%
7	80.82%	66.64%	69.69%	76.36%	81.19%	84.60%	72.41%	80.94%	-	-	-	51.60%	63.38%
8	77.65%	62.10%	79.60%	73.62%	91.31%	66.83%	92.52%	67.38%	66.14%	-	-	55.81%	55.66%
9	81.74%	73.91%	76.41%	81.53%	90.79%	69.29%	91.42%	66.37%	62.43%	87.80%	-	58.20%	57.12%
10	94.63%	68.21%	82.01%	94.00%	97.14%	89.86%	91.41%	80.40%	65.08%	88.27%	94.79%	53.82%	65.61%

Table a6. Detail experimental results of each session of ViT-Adapter in Protocol-1. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-Adapter in Protocol 1													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	REPLAY ATTACK	CASIA FASD	MSU MFSD	HKBU MarV2	OULU NPU	CSMAD	CASIA -SURF	WFFD	WMCA	CASIA 3DMASK	ROSE YOUTU	CeFA
0	97.22%	-	-	-	-	-	-	-	-	-	-	56.11%	61.39%
1	97.36%	98.05%	-	-	-	-	-	-	-	-	-	62.44%	64.98%
2	96.32%	93.96%	96.12%	-	-	-	-	-	-	-	-	61.08%	62.62%
3	91.64%	85.13%	89.50%	75.50%	-	-	-	-	-	-	-	52.30%	55.94%
4	93.72%	85.03%	92.34%	76.29%	87.95%	-	-	-	-	-	-	56.89%	49.16%
5	90.37%	76.85%	91.66%	71.21%	91.48%	97.60%	-	-	-	-	-	49.63%	54.41%
6	88.15%	75.90%	90.05%	69.76%	90.40%	93.14%	98.49%	-	-	-	-	73.21%	66.87%
7	80.17%	82.74%	88.91%	63.07%	76.66%	94.18%	70.64%	97.16%	-	-	-	51.08%	41.16%
8	85.08%	82.79%	89.17%	65.60%	81.90%	89.95%	91.57%	89.52%	97.30%	-	-	63.99%	55.34%
9	85.34%	81.51%	91.16%	64.67%	86.76%	91.40%	94.67%	93.10%	98.01%	99.42%	-	69.28%	67.69%
10	67.43%	69.77%	74.28%	49.38%	74.59%	75.49%	79.75%	67.91%	96.05%	88.58%	99.78%	57.92%	76.10%

Table a7. Detail experimental results of each session of ViT-ConvA in Protocol-1. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-ConvA in Protocol 1													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	REPLAY ATTACK	CASIA FASD	MSU MFSD	HKBU MarV2	OULU NPU	CSMAD	CASIA -SURF	WFFD	WMCA	CASIA 3DMASK	ROSE YOUTU	CeFA
0	99.64%	-	-	-	-	-	-	-	-	- -	-	92.67%	68.68%
1	97.90%	99.80% -	-	-	-	-	-	-	-	-	-	88.49%	75.99%
2	96.79%	99.29%	99.58% -	-	-	-	-	-	-	- -	-	91.15%	82.24%
3	95.77%	98.68%	99.55%	99.63% -	-	-	-	-	-	-	-	91.93%	81.26%
4	98.47%	92.46%	98.00%	95.02%	100.00% -	-	-	-	-	- -	-	87.14%	65.80%
5	96.25%	93.12%	98.03%	91.31%	100.00%	96.94% -	-	-	-	-	-	86.62%	72.48%
6	96.97%	93.50%	98.03%	89.09%	99.99%	93.51%	98.83% -	-	-	- -	-	84.03%	67.06%
7	97.16%	89.21%	96.59%	87.43%	99.98%	90.45%	96.27%	94.31% -	-	-	-	86.92%	60.12%
8	93.23%	78.75%	87.25%	72.71%	99.96%	79.00%	98.80%	92.10%	85.26%	-	-	80.63%	49.71%
9	93.40%	83.66%	91.40%	80.72%	99.80%	83.40%	98.24%	90.91%	84.61%	97.42%	-	82.89%	56.73%
10	93.33%	85.30%	91.79%	79.04%	99.96%	79.52%	96.08%	91.21%	85.71%	96.23%	97.18%	84.34%	51.21%

Table a8. Detail experimental results of each session of ViT-CDCA in Protocol-1. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-CDCA in Protocol 1													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	REPLAY ATTACK	CASIA FASD	MSU MFSD	HKBU MarV2	OULU NPU	CSMAD	CASIA -SURF	WFFD	WMCA	CASIA 3DMASK	ROSE YOUTU	CeFA
0	99.64%	-	-	-		-		-		-	-	92.67%	68.68%
1	99.49%	99.17%	-	-	-	-	-	-	-	-	-	94.35%	75.24%
2	99.53%	98.70%	99.32%	-	-	-	-	-	-	-	-	95.00%	75.95%
3	99.52%	98.35%	99.19%	99.29%	-	-	-	-	-	-	-	93.82%	74.58%
4	99.56%	95.45%	98.55%	97.85%	99.30%	-	-	-	-	-	-	92.50%	66.70%
5	99.39%	95.60%	99.09%	96.57%	97.31%	98.23%	-	-	-	-	-	95.34%	67.28%
6	98.45%	89.09%	98.41%	95.14%	99.78%	90.97%	98.13%	-	-	-	-	90.34%	56.34%
7	98.12%	83.13%	96.18%	92.99%	99.74%	90.56%	97.74%	94.75%	-	-	-	88.56%	61.21%
8	92.07%	69.05%	83.67%	79.89%	99.53%	74.65%	99.06%	93.30%	82.77%	-	-	83.55%	47.05%
9	93.14%	77.74%	91.10%	84.26%	98.68%	78.59%	98.73%	94.74%	82.38%	97.00%	-	85.47%	49.75%
10	95.08%	66.49%	89.09%	90.52%	99.09%	83.46%	95.32%	92.38%	79.66%	96.04%	96.89%	87.73%	52.12%

Table a9. Detail experimental results of each session of ViT-DCDCA in Protocol-1. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-DCDCA in Protocol 1													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	REPLAY ATTACK	CASIA FASD	MSU MFSD	HKBU MarV2	OULU NPU	CSMAD	CASIA -SURF	WFFD	WMCA	CASIA 3DMASK	ROSE YOUTU	CeFA
0	99.37%	-	-	-	-	-	-	-	-	-	-	90.02%	64.62%
1	99.30%	98.59%	-	-	-	-	-	-	-	-	-	93.26%	67.93%
2	99.30%	97.61%	98.95%	-	-	-	-	-	-	-	-	95.14%	72.42%
3	99.44%	97.65%	98.92%	98.59%	-	-	-	-	-	-	-	94.81%	75.70%
4	99.45%	92.58%	98.21%	92.47%	99.21%	-	-	-	-	-	-	92.85%	66.71%
5	98.74%	91.48%	99.54%	93.70%	97.68%	98.21%	-	-	-	-	-	96.27%	71.65%
6	98.98%	88.48%	98.93%	91.13%	99.89%	94.91%	99.45%	-	-	-	-	91.70%	62.33%
7	98.76%	83.76%	97.82%	88.44%	99.85%	94.28%	98.67%	94.72%	-	-	-	91.56%	69.31%
8	97.78%	72.52%	94.89%	76.64%	99.97%	84.99%	99.37%	91.34%	80.46%	-	-	88.82%	51.34%
9	96.91%	76.40%	96.27%	80.85%	99.69%	84.88%	99.07%	91.62%	78.62%	97.68%	-	88.47%	57.20%
10	97.94%	71.10%	95.59%	81.98%	99.93%	85.81%	98.19%	90.92%	78.11%	97.89%	96.24%	91.02%	58.38%

References

[1] Sushil Bhattacharjee, Amir Mohammadi, and Sébastien Marcel. Spoofing deep face recognition with custom silicone masks. In *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pages 1–7. IEEE, 2018. 2, 3, 8

[2] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, pages 612–618. IEEE, 2017. 2, 3, 7

[3] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, pages 1–7. IEEE, 2012. 2, 3, 5

Table a10. Detail experimental results of each session of ViT-DCDCA-PPCR in Protocol-1. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-DCDCA-PPCR in Protocol 1													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	REPLAY ATTACK	CASIA FASD	MSU MFSD	HKBU MarV2	OULU NPU	CSMAD	CASIA -SURF	WFFD	WMCA	CASIA 3DMASK	ROSE YOUTU	CeFA
0	99.63%	-		-		-		-		-	-	92.71%	70.37%
1	98.90%	99.48%	-	-	-	-	-	-	-	-	-	92.84%	78.79%
2	97.98%	98.54%	99.36%	-	-	-	-	-	-	-	-	93.43%	83.99%
3	95.76%	99.22%	98.78%	99.33%	-	-	-	-	-	-	-	93.23%	86.73%
4	99.12%	95.46%	97.20%	96.47%	99.91%	-	-	-	-	-	-	91.76%	74.21%
5	96.56%	96.09%	99.62%	96.87%	98.94%	98.43%	-	-	-	-	-	92.03%	81.02%
6	99.19%	93.37%	98.46%	95.27%	99.94%	95.18%	99.17%	-	-	-	-	89.59%	71.73%
7	98.54%	83.91%	96.88%	92.82%	99.83%	93.08%	98.76%	94.64%	-	-	-	89.65%	68.92%
8	95.70%	70.36%	88.32%	75.03%	99.85%	83.72%	98.94%	89.13%	79.92%	-	-	80.57%	59.41%
9	97.25%	82.09%	94.23%	82.64%	99.54%	89.16%	98.94%	89.40%	77.74%	98.28%	-	84.12%	69.04%
10	98.63%	85.07%	95.48%	89.83%	99.57%	91.73%	97.87%	90.77%	76.13%	98.38%	95.66%	90.99%	70.23%

Table a11. Detail experimental results of each session of ResNet18 in Protocol-2. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ResNet18 in Protocol 2													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	CASIA 3DMASK	WMCA	WFFD	CASIA -SURF	CSMAD	OULU NPU	HKBU MarV2	MSU MFSD	CASIA FASD	REPLAY ATTACK	ROSE YOUTU	CeFA
0	97.39%	-	-	-	-	-	-	-	-	-	-	52.81%	81.58%
1	95.60%	95.69%	-	-	-	-	-	-	-	-	-	48.31%	66.34%
2	85.21%	66.22%	91.86%	-	-	-	-	-	-	-	-	46.21%	56.07%
3	89.95%	67.42%	91.93%	62.75%	-	-	-	-	-	-	-	51.88%	59.19%
4	84.57%	81.33%	62.73%	58.11%	80.28%	-	-	-	-	-	-	52.63%	54.58%
5	88.21%	74.97%	88.93%	63.32%	67.98%	92.85%	-	-	-	-	-	47.05%	55.39%
6	90.20%	83.12%	81.64%	66.06%	73.24%	91.12%	89.16%	-	-	-	-	49.79%	65.89%
7	92.41%	83.54%	83.10%	60.39%	75.03%	92.68%	86.06%	96.63%	-	-	-	45.05%	60.40%
8	89.68%	73.95%	93.35%	66.85%	64.16%	87.57%	90.35%	99.03%	92.78%	-	-	47.62%	67.25%
9	84.68%	76.34%	84.49%	70.60%	63.43%	90.90%	87.65%	91.88%	88.39%	94.38%	-	53.61%	61.62%
10	84.73%	64.00%	89.11%	56.72%	61.27%	88.52%	80.29%	87.67%	89.91%	90.24%	95.90%	47.53%	77.75%

Table a12. Detail experimental results of each session of ViT-Adapter in Protocol-2. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	CASIA 3DMASK	WMCA	WFFD	CASIA -SURF	CSMAD	OULU NPU	HKBU MarV2	MSU MFSD	CASIA FASD	REPLAY ATTACK	ROSE YOUTU	CeFA
0	97.22%	-	-	-	-	-	-	-	-	-	-	56.11%	61.39%
1	97.36%	98.05%	-	-	-	-	-	-	-	-	-	62.44%	64.98%
2	96.32%	93.96%	96.12%	-	-	-	-	-	-	-	-	61.08%	62.62%
3	91.64%	85.13%	89.50%	75.50%	-	-	-	-	-	-	-	52.30%	55.94%
4	93.72%	85.03%	92.34%	76.29%	87.95%	-	-	-	-	-	-	56.89%	49.16%
5	90.37%	76.85%	91.66%	71.21%	91.48%	97.60%	-	-	-	-	-	49.63%	54.41%
6	88.15%	75.90%	90.05%	69.76%	90.40%	93.14%	98.49%	-	-	-	-	73.21%	66.87%
7	80.17%	82.74%	88.91%	63.07%	76.66%	94.18%	70.64%	97.16%	-	-	-	51.08%	41.16%
8	85.08%	82.79%	89.17%	65.60%	81.90%	89.95%	91.57%	89.52%	97.30%	-	-	63.99%	55.34%
9	85.34%	81.51%	91.16%	64.67%	86.76%	91.40%	94.67%	93.10%	98.01%	99.42%	-	69.28%	67.69%
10	67.43%	69.77%	74.28%	49.38%	74.59%	75.49%	79.75%	67.91%	96.05%	88.58%	99.78%	57.92%	76.10%

[4] Anjith George, Zohreh Mostaani, David Geissenbuhler, Olegs Nikisins, André Anjos, and Sébastien Marcel. Biometric face presentation attack detection with multi-channel convolutional neural network. *IEEE Transactions on Information Forensics and Security*, 15:42–55, 2019. 2, 3, 9

[5] Shan Jia, Chuanbo Hu, Guodong Guo, and Zhengquan Xu. A database for face presentation attack using wax figure faces. In *International Conference on Image Analysis and Processing*, pages 39–47. Springer, 2019. 2, 3, 9

[6] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot. Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(7):1794–1809, 2018. 2, 3, 10

[7] Ajian Liu, Zichang Tan, Jun Wan, Sergio Escalera, Guodong Guo, and Stan Z Li. CASIA-SUR CeFA: A Benchmark for Multi-Modal Cross-Ethnicity Face Anti-Spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1179–1187, 2021. 2, 3, 11

[8] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang Tan, Sergio Escalera, Junliang Xing, Yanyan Liang, et al. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 17:2497–2507, 2022. 2, 3, 5

[9] Siqi Liu, Baoyao Yang, Pong C Yuen, and Guoying Zhao. A 3d mask face anti-spoofing database with real world

Table a13. Detail experimental results of each session of ViT-ConvA in Protocol-2. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-ConvA Protocol 2													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	CASIA 3DMASK	WMCA	WFFD	CASIA -SURF	CSMAD	OULU NPU	HKBU MarV2	MSU MFSD	CASIA FASD	REPLAY ATTACK	ROSE YOUTU	CeFA
0	99.64%	-	-	-	-	-	-	-	-	-	-	92.67%	68.68%
1	98.27%	97.71%	-	-	-	-	-	-	-	-	-	85.86%	60.09%
2	98.51%	94.85%	96.83%	-	-	-	-	-	-	-	-	87.14%	66.69%
3	97.10%	92.78%	96.26%	84.56%	-	-	-	-	-	-	-	80.23%	54.04%
4	97.39%	93.13%	96.79%	81.81%	93.04%	-	-	-	-	-	-	82.45%	52.28%
5	95.89%	83.46%	96.29%	83.25%	93.55%	99.06%	-	-	-	-	-	76.19%	52.14%
6	96.17%	80.52%	96.96%	83.62%	91.86%	97.62%	95.74%	-	-	-	-	80.26%	61.46%
7	97.23%	84.25%	96.06%	83.25%	94.14%	99.11%	93.77%	100.00%	-	-	-	80.79%	54.90%
8	97.85%	90.70%	95.16%	74.63%	91.43%	88.40%	97.19%	97.37%	98.06%	-	-	91.94%	68.24%
9	97.29%	90.09%	94.99%	75.39%	91.76%	89.71%	97.17%	98.54%	98.32%	99.61%	-	91.11%	70.18%
10	95.24%	88.00%	95.96%	72.23%	91.73%	93.41%	94.56%	96.87%	98.81%	98.94%	99.66%	89.42%	76.58%

Table a14. Detail experimental results of each session of ViT-CDCA in Protocol-2. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-CDCA in Protocol 2													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	CASIA 3DMASK	WMCA	WFFD	CASIA -SURF	CSMAD	OULU NPU	HKBU MarV2	MSU MFSD	CASIA FASD	REPLAY ATTACK	ROSE YOUTU	CeFA
0	99.64%	-	-	-	-	-	-	-	-	-	-	92.67%	68.68%
1	98.59%	96.66%	-	-	-	-	-	-	-	-	-	92.72%	67.13%
2	99.28%	92.72%	96.16%	-	-	-	-	-	-	-	-	93.52%	63.76%
3	98.02%	83.72%	96.14%	82.59%	-	-	-	-	-	-	-	88.28%	51.10%
4	97.38%	83.61%	96.77%	83.42%	92.19%	-	-	-	-	-	-	85.49%	48.53%
5	95.91%	77.01%	96.55%	84.37%	92.05%	99.07%	-	-	-	-	-	81.27%	46.94%
6	95.77%	75.51%	96.30%	84.12%	91.54%	98.09%	94.60%	-	-	-	-	84.08%	55.46%
7	97.60%	83.64%	97.81%	82.45%	93.29%	98.87%	94.06%	99.97%	-	-	-	86.30%	52.81%
8	98.95%	88.65%	97.66%	78.26%	93.19%	96.49%	96.73%	99.31%	97.70%	-	-	91.18%	60.10%
9	99.11%	89.31%	95.13%	73.18%	92.68%	93.21%	97.29%	97.31%	98.60%	99.80%	-	94.22%	66.70%
10	96.95%	85.37%	95.33%	70.74%	90.63%	92.28%	96.58%	93.48%	98.00%	99.46%	99.10%	93.95%	73.56%

Table a15. Detail experimental results of each session of ViT-DCDCA in Protocol-2. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset. The horizontal direction represents the training dataset, and the vertical direction represents the testing dataset.

ViT-DCDCA in Protocol 2													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	CASIA 3DMASK	WMCA	WFFD	CASIA -SURF	CSMAD	OULU NPU	HKBU MarV2	MSU MFSD	CASIA FASD	REPLAY ATTACK	ROSE YOUTU	CeFA
0	99.37%	-	-	-	-	-	-	-	-	-	-	90.02%	64.62%
1	98.74%	95.83%	-	-	-	-	-	-	-	-	-	94.02%	69.25%
2	99.39%	95.02%	95.60%	-	-	-	-	-	-	-	-	94.08%	68.97%
3	98.95%	93.19%	95.24%	78.98%	-	-	-	-	-	-	-	90.37%	58.27%
4	98.93%	91.61%	95.16%	77.44%	91.35%	-	-	-	-	-	-	91.12%	61.33%
5	98.00%	84.71%	95.25%	78.57%	92.44%	99.49%	-	-	-	-	-	85.70%	54.76%
6	97.78%	79.23%	95.11%	77.74%	89.97%	98.24%	95.21%	-	-	-	-	89.84%	64.60%
7	97.77%	79.03%	94.96%	79.50%	92.37%	99.52%	89.99%	99.99%	-	-	-	88.23%	55.52%
8	98.70%	83.30%	95.33%	74.48%	91.39%	97.23%	95.65%	99.26%	96.62%	-	-	92.03%	65.91%
9	98.79%	87.05%	96.18%	70.76%	91.54%	95.32%	95.29%	96.60%	97.56%	99.44%	-	94.49%	72.54%
10	96.72%	84.96%	97.06%	69.00%	89.66%	93.25%	94.97%	93.01%	97.91%	99.13%	97.98%	94.62%	77.69%

variations. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 100–106, 2016. 2, 3, 7

[10] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 389–398, Salt Lake City, UT, 2018. 2, 3, 4

[11] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015. 2, 3, 6

[12] Zitong Yu, Jun Wan, Yunxiao Qin, Xiaobai Li, Stan Z. Li, and Guoying Zhao. NAS-FAS: Static-Dynamic Central Difference Network Search for Face Anti-Spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(9):3005–3023, 2021. 2, 3, 10

[13] Shifeng Zhang, Ajian Liu, Jun Wan, Yanyan Liang, Guodong Guo, Sergio Escalera, Hugo Jair Escalante, and Stan Z Li. CASIA-SURF: A Large-scale Multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics,*

Table a16. Detail experimental results of each session of ViT-DCDCA-PPCR in Protocol-2.

ViT-DCDCA-PPCR in Protocol 2													
Session ID	0	1	2	3	4	5	6	7	8	9	10	Unseen	
Test	Base	CASIA 3DMASK	WMCA	WFFD	CASIA -SURF	CSMAD	OULU NPU	HKBU MarV2	MSU MFSD	CASIA FASD	REPLAY ATTACK	ROSE YOUTU	CeFA
0	99.63%	-	-	-	-	-	-	-	-	-	-	92.71%	70.37%
1	98.17%	96.50%	-	-	-	-	-	-	-	-	-	93.55%	76.04%
2	98.34%	91.25%	95.33%	-	-	-	-	-	-	-	-	92.33%	76.98%
3	98.33%	82.09%	97.44%	77.77%	-	-	-	-	-	-	-	88.34%	65.06%
4	98.20%	85.93%	97.95%	71.58%	91.65%	-	-	-	-	-	-	87.22%	63.01%
5	96.91%	76.05%	97.82%	76.60%	93.37%	99.29%	-	-	-	-	-	85.76%	58.73%
6	97.64%	72.33%	97.91%	78.32%	93.42%	98.48%	95.83%	-	-	-	-	87.97%	71.16%
7	97.05%	72.18%	98.34%	80.59%	93.70%	99.57%	92.54%	99.92%	-	-	-	87.00%	60.03%
8	98.93%	80.98%	97.89%	74.98%	93.89%	98.45%	96.53%	99.25%	97.06%	-	-	91.21%	73.15%
9	98.83%	78.26%	97.28%	75.61%	92.32%	98.50%	96.92%	99.82%	97.15%	99.79%	-	91.34%	75.50%
10	98.68%	80.30%	96.75%	70.51%	91.45%	97.05%	95.34%	96.29%	98.17%	99.23%	98.73%	93.33%	80.90%

Behavior, and Identity Science, 2(2):182–193, 2020. 2, 3, 8

[14] Yuanhan Zhang, Zhenfei Yin, Yidong Li, Guojun Yin, Junjie Yan, Jing Shao, and Ziwei Liu. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In *European Conference on Computer Vision (ECCV)*, 2020. 2, 3, 4

[15] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and S. Z. Li. A face anti-spoofing database with diverse attacks. In *IAPR International Conference on Biometrics*, pages 26–31, 2012. 2, 3, 6