Evaluating the Sensitivity of Face Presentation Attack Detection Techniques to Images of Varying Resolutions

Lazaro J. Gonzalez-Soler[1], Marta Gomez-Barrero[2], and Christoph Busch[1]

[1] da/sec - biometrics and internet security research group, Hochschule Darmstadt, Germany
{lazaro-janier.gonzalez-soler;christoph.busch}@h-da.de
[2] Hochschule Ansbach, Germany
marta.gomez-barrero@hs-ansbach.de

Abstract

In the last decades, emerging techniques for face Presentation Attack Detection (PAD) have reported a remarkable performance to detect attack presentations whose attack type and capture conditions are known a priori. However, the generalisation capability of PAD approaches shows a considerable deterioration to detect unknown attacks. In order to tackle those generalisation issues, several PAD techniques have focused on the detection of homogeneous features from known attacks to detect unknown Presentation Attack Instruments without taking into account how some intrinsic image properties such as the image resolution or biometric quality could impact their detection performance. In this work, we carry out a thorough analysis of the sensitivity of several texture descriptors which shows how the use of images with varying resolutions for training leads to a high decrease on the attack detection performance.

1 Introduction

Recently, face-based biometric systems have experienced a large development due to their user convenience and their ability to provide a security level higher than traditional credential-based access control systems. In spite of those advantages, they are still vulnerable to Presentation Attack Instruments (PAIs) which can be easily created by a non-authorized subject by downloading a photo or video of a given person from any social network. The attacker can subsequently use the PAI to gain access to several applications such as bank accounts, to unlock smartphones, and border controls, where face biometric systems are frequently employed.

To address those security threats, numerous software-based Presentation Attack Detection (PAD) approaches have been proposed. Those techniques try to determine whether a face sample stems from a live subject (i.e., it is a bona fide presentation, BP) or from an artificial replica (i.e., it is an attack presentation, AP). Several face PAD techniques have reported a high detection performance to identify attack presentations under the so-called "known attacks" scenario, where AP properties such as PAI species and capture conditions are known a priori. However, most PAD approaches suffer a high accuracy decrease to detect unknown attacks. In particular, PAD techniques based on one-class classifiers [2, 15, 7], AutoEncoders [24], Gaussian Mixture Models [15], tree decision-based models [14], and end-to-end deep learning schemes [20, 9] have reported a high detection performance degradation to detect PAIs when they are trained with a large set of face samples. On the other hand, when they only employ images with a particular quality, an improvement on the performance can be perceived, as shown in Tab. 1. Therefore, one main question still remains unanswered: could the utilisation of images with varying resolutions affect the detection performance of any PAD method? In other words, to which extent are PAD approaches sensitive to training sets containing images of varying quality? And how could this difficulty affect the PAD generalisation capabilities?

Table 1: Benchmark of state-of-the-art approaches in terms of classification accuracy (%) under the Quality Test and Overall Test protocol in [25].

method	**low**	**normal**	**high**	**overall**
LBP+SVM [19]	**83**	78	**90**	80
Network A [21]	**84**	**91**	79	80
Network B [21]	**86**	**93**	80	81
Network C [21]	**94**	**94**	82	87
ShallowCNN [20]	**93**	**92**	84	88

Very few works have addressed these issues so far. Back in 2013, Galbally *et al.* [8] evaluated the potential of general Image Quality Assessment (IQA) as a protection tool against PAIs and showed that a face sample acquired in an attack attempt has different quality than a BP image. Following that idea, Bhogal *et al.* [3] also explored six non-reference IQA metrics to detect attack presentations on iris, fingerprint, and facial characteristics. As a result, the authors found that the best quality measure and classification setting highly depends on the target database, thereby recommending its optimisation for each particular application. More recently, Agarwal *et al.* [1] showed how several image transformations such as gamma correction, log transform, and brightness control can help a non-authorised subject to circumvent a PAD algorithm. In addition, the authors demonstrated that such image transformations decrease the detection performance of handcrafted- and deep learning-based PAD approaches.

In spite of those valuable efforts, the aforementioned questions are still partially unanswered. For that reason, we study in this work both handcrafted and deep learning features and analyse the impact of training with images with varying quality on the PAD performance. In order to allow the reproducibility of our results, the experimental evaluation was conducted over the freely available CASIA Face Antispoofing database [25] following the metrics included in the international standard ISO/IEC 30107-3 on biometric PAD [12].

The remainder of this paper is organised as follows. The PAD methods studied in our work are presented in Sect. 2. Sect. 3 describes the experimental evaluation and results. Finally, conclusions and future work directions are presented in Sect. 4.

2 Presentation Attack Detection Approaches

In order to address the questions posed in Sect. 1, we analyse PAD methods which follow the three-step overview depicted in Fig. 1: i) to remove non-useful information, faces are first detected with the Tensorflow[1] Face Detection method, ii) a global texture descriptor is then extracted per image, and iii) a BP or AP decision is finally taken by a linear Support Vector Machine (SVM). Linear SVMs are popular classifiers since they perform well in high-dimensional spaces, avoid over-fitting, and have good generalisation capabilities. To extract global features from images, we select three well-known texture descriptors: Local Binary Pattern (LBP) [16], Local Phase Quantization (LPQ) [17], and Binarized Statistical Image Features (BSIF) [13], and five deep learning approaches (i.e., MobileNet [10], MobileNetV2 [22], InceptionV3 [23], Xception [4], and DenseNet121 [11]). All of these descriptors have been widely employed for face PAD [2, 15, 5] and are summarised in Tab. 2.

[1] https://github.com/yeephycho/tensorflow-face-detection

Figure 1: General diagram of the PAD approaches analysed.

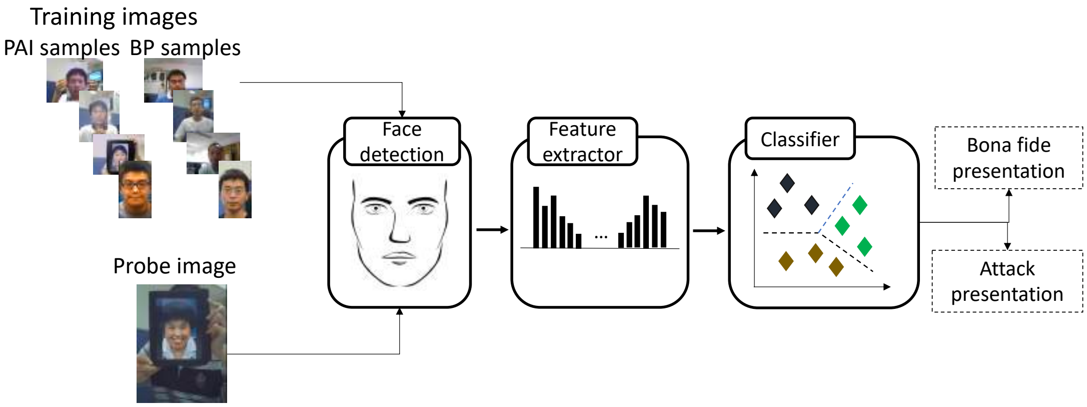


Table 2: Summary of the descriptors studied in this work.

	Handcrafted descriptors		**Deep learning descriptors**	
descriptor	**parameters**	**length**	**architecture**	**length**

LBP [16]	R = {1, 2, 3}
	P = {8, 16, 24}
		59	MobileNet [10]
			MobileNetV2 [22]
				$7 \times 7 \times 1024 = 50176$
				$7 \times 7 \times 1280 = 62720$

LPQ [17]	R = {3, 5} $\alpha = 1$	256	InceptionV3 [23] Xception [4]	$5 \times 5 \times 2048 = 51200$ $7 \times 7 \times 2048 = 100352$

BSIF [13]	N = {5, 6, 7, 8, 9, 10, 11, 12}
	L = {3, 5, 7, 9, 11, 13, 15, 17}
		2^N	DenseNet121 [11]	$7 \times 7 \times 1024 = 50176$

The proposed pipeline was entirely implemented in python. The Scikit-learn and Keras frameworks are employed for the linear SVM and deep learning architectures, respectively. In addition, we use the ImageNet [6] pre-trained deep learning models, and the final descriptor is computed from the last layer after removing the fully connected classification layers.

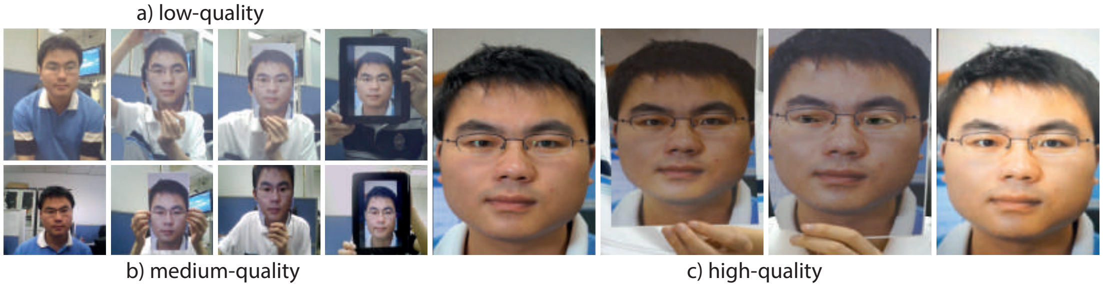
3 Experimental Evaluation

3.1 Experimental Protocol

The experimental evaluation was conducted over the freely available CASIA Face Anti-spoofing database [25], which contains 600 short videos of bona fide and attack presentations stemming from 50 different subjects and acquired with three different devices with varying quality: low (L), medium (M), and high (H). The dataset comprises three PAI species as shown in Fig. 2: *i)* Warped photo attacks, in which the attackers place their face behind the hard copies of high-resolution digital photographs; *ii)* Cut photo attacks, where the face of the attacker is placed behind the hard copies of photos, from which eyes have been cut out; and *iii)* Video replay attacks, where attackers replay face videos using iPads.

In order to evaluate the sensitivity of the current descriptors to images of varying resolutions for the detection of known attacks, we augment the Quality and Overall test protocols in [25] by including new attack-resolution configurations, where a given attack, acquired with a fixed

Figure 2: CASIA Face Antispoofing database, which includes images of varying resolutions, stemming from three PAI species.



resolution capture device, is known on training.

The experimental results are analysed in compliance with the ISO/IEC 30107-3 on biometric PAD [12] using the following two metrics:

- Attack Presentation Classification Error Rate (APCER), or proportion of attack presentations misclassified as bona fide presentations;
- Bona Fide Classification Error Rate (BPCER), or proportion of bona fide presentations misclassified as attack presentations;

We also report the Detection Equal Error Rates (D-EER), which is defined as the error rate value at the operating point the APCER and the BPCER match. In addition, we present the Detection Error Trade-off (DET) curves between APCER and BPCER.

3.2 Results and discussion

3.2.1 Known Attacks: analysis of a single image resolution

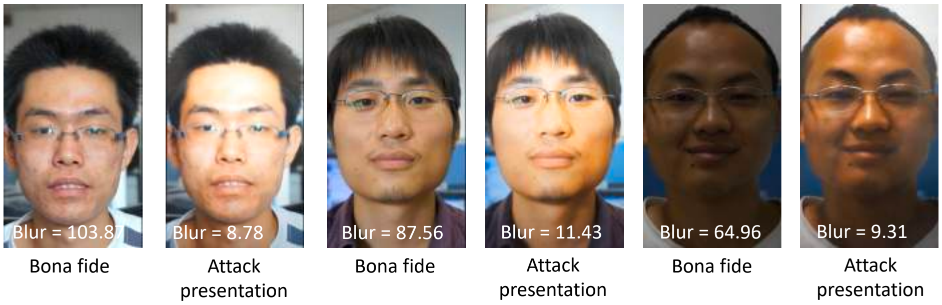
The first set of experiments evaluates the PAD performance under three resolution settings. The corresponding D-EER values are reported in Tab. 3. We can first note that for each particular PAI species the error rates attained depend on the image resolution being evaluated. Specifically, D-EERs of 18.01 ± 1.67, 16.82 ± 2.67, and 15.65 ± 6.08 are reported on average for Cut, Warped, and Video attacks, respectively, hence indicating that PAD approaches widely depend on both the PAI species used (different average D-EER) and the resolution settings employed to acquire them (up to 6% standard deviation in the D-EER).

In addition, it can be observed that the error rates for each PAI species follow different trends depending on the resolution of the images used. Whereas the Cut and Warped photo attacks achieve their best detection performance on average across all descriptors for face images acquired with low-resolution capture devices (i.e., D-EERs of 16.49% and 13.91 for Cut and Warped, respectively), the Video replay shows its best D-EER for high resolution images (i.e., 9.37%). However, taking a closer look we can note that the handcrafted descriptors LBP and BSIF do perform better with high quality images not only for Video but also for Warped attacks. In contrast, the deep learning-based techniques achieve their best detection performance for low to medium quality images for Cut and Warped attacks, and also for Medium quality images for Video attacks. To shed some light into these differences, we investigated some intrinsic image properties and confirmed that the screen projection of the Video replay attacks on a

Table 3: Average D-EER (%) values under the known attack protocol.

Attack	Cut			Warped			Video		
	L	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**
LBP	20.00	**16.67**	17.78	17.78	18.89	**10.00**	18.89	21.11	**4.44**
LPQ	20.00	**15.00**	20.00	**13.33**	21.67	16.67	22.50	20.00	**1.67**
BSIF	**18.61**	23.56	20.56	16.83	26.03	**12.19**	17.28	20.86	**2.17**
MobileNet	**23.33**	26.67	**23.33**	**16.67**	33.33	36.67	23.33	13.33	**10.00**
MobileNetV2	16.67	**13.33**	**13.33**	**10.00**	16.67	**10.00**	23.33	**6.67**	16.67
InceptionV3	**6.67**	20.00	16.67	16.67	**13.33**	16.67	33.33	33.33	**16.67**
Xception	**16.67**	**16.67**	26.67	**10.00**	16.67	26.67	20.00	**10.00**	16.67
DenseNet121	**10.00**	**10.00**	20.00	10.00	**6.67**	10.00	13.33	**3.33**	6.67
avg.	**16.49**	17.74	19.79	**13.91**	19.16	17.39	21.50	16.08	**9.37**

Figure 3: Bona fide and attack presentation samples with their corresponding blurriness values.



high-resolution capture device unveils several blurriness and sharpness artefacts, which are successfully detected by all PAD techniques. In Fig. 3 we show some BP and AP samples with their blurriness values, which were computed as the variation of the Laplacian [18].

3.2.2 Known Attacks: analysis of images of mixed resolutions

We present in Tab. 4 a joint evaluation of the proposed descriptors over several PAI species combinations, which were simultaneously acquired with different resolution capture devices. We can observe that the D-EER values for multiple resolution images are up to seven times higher than the ones achieved for the best single image quality (e.g., 3.00% for high image quality vs. 19.85% for L ∪ M). In contrast, similar D-EER values are reported when several PAI species are employed for training over face images acquired by a single capture device (e.g., 17.94% for Cut photo attacks vs. 16.37% for cut ∪ warped). Thus we can highlight how the utilisation of images of varying resolutions leads to a high PAD performance deterioration across different PAI species.

Finally, it should be also noted that the PAD approaches can be circumvented by launching Cut photo and Warped photo attack samples which were recorded with medium and high-resolution capture devices: a high mean D-EER of 28.40% is attained for that configuration.

Table 4: Average D-EER values under the known attack protocol.

	Quality	**Single**			**Multiple**			
	PAI species	**L**	**M**	**H**	**L ∪ M**	**L ∪ H**	**M ∪ H**	**L ∪ M ∪ H**
Single	**cut**	**17.94**	21.71	19.64	21.82	24.27	27.16	24.22
	warped	16.02	24.03	**12.41**	23.49	19.74	24.92	24.60
	video	17.49	19.64	**3.00**	19.85	14.64	14.67	18.43
Multiple	**cut ∪ video**	19.31	25.06	**16.47**	22.57	23.80	28.19	25.44
	cut ∪ warped	**16.37**	21.19	16.95	21.47	23.62	28.40	26.61
	warped ∪ video	19.43	23.45	**12.73**	22.52	19.42	24.31	24.08
	cut ∪ warped ∪ video	16.25	24.83	**15.11**	22.15	23.01	26.48	25.74

Table 5: D-EER (%) values for single and multiple attack-resolution settings.

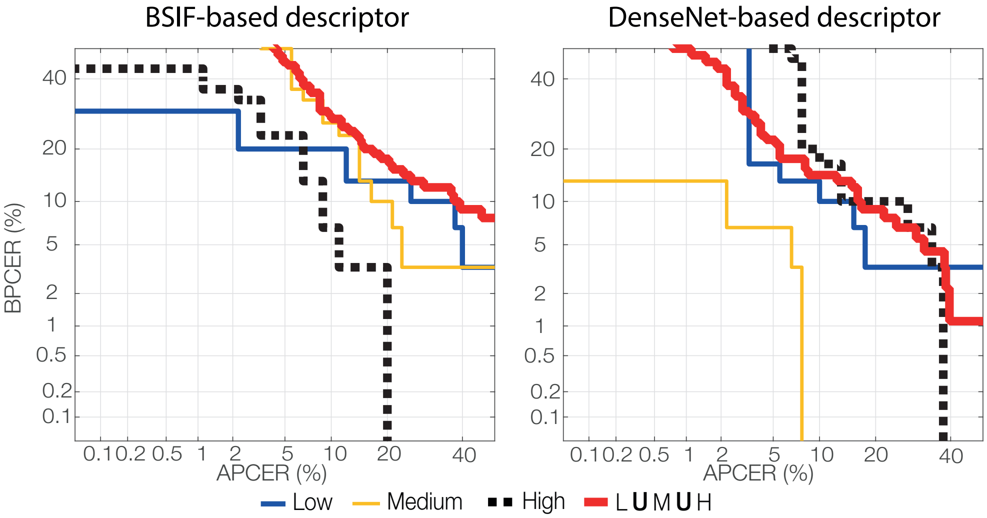
Quality **Attack**	Single	Multiple
Single	16.88 ± 6.17	21.48 ± 4.06
Multiple	18.93 ± 3.99	24.24 ± 2.45

Following those observations, we tried to determine which of these two image properties (i.e., PAI species or image resolution) produces the greatest PAD performance deterioration. To that end, we compute in Tab. 5 the average for each single and multiple combination depicted in Tab. 4. On the one hand, for a configuration where either several PAI species or images of varying resolutions are employed, a high performance decrease can be seen. In particular, a mean D-EER of 21.48% is reported when PAD approaches are trained with a single PAI species which was acquired under several resolution settings. This number is in turn worse than the one achieved when the PAD methods are trained with multiple PAI species with a single image resolution (18.93%). On the other hand, as it could be expected, a high mean D-EER of 24.24% is attained for the worst case where the PAD approaches are optimised utilising several PAI species which were acquired with varying-resolution capture devices. We thus conclude that the utilisation of images with varying quality produces the greatest PAD performance deterioration. In addition, we confirm the PAD performance decrease reported by state-of-the-art algorithms in Tab. 1 and answer the question launched in Sect. 1: PAD techniques based both on handcrafted and deep learning features are sensitive to images with varying resolutions, which lead to a high accuracy decrease in the detection of APs.

3.2.3 Known Attacks: a deeper PAD performance analysis

Finally, we show in Fig. 4 the DET curves for the best handcrafted and deep learning approaches over the Quality test and Overall test protocols from [25]. As it can be observed, the joint training for the analysed PAI species, which were acquired with three varying-resolution capture devices (i.e., thick red line), yields on average a high BCPER of 67.23% for any APCER ≤ 1%. This is in turn higher than the ones attained for each single resolution. In addition, the BSIF descriptor reports its best BPCER value for high-resolution images. In contrast, the deep

Figure 4: DET curves for the best handcrafted and deep learning-based approach over the known attack scenario. For the BSIF computation, we use $N = 10$ filters of size $l = 13$.



learning approach achieves its best BPCER for images of medium quality. In this context, we can conclude that either a down-sampling or up-sampling step performed by the deep learning-based descriptors for fitting the size of a given image into the input layer can lead to an information loss for low and high quality input images. In particular, for a database as CASIA whose cropped face images pose an average size of 180 × 157 and 644 × 545 pixels for low and high resolution settings, respectively, an up-sampling and down-sampling procedure to fit the image size to 224 × 224 pixels (input layer size) can approximately affect on average a 65% of pixels of a given face image. This could in turn remove several artefacts produced in the creation of PAIs. In contrast, this re-sizing procedure affects only 37% of the pixels of medium resolution images, thereby leading to a higher detection performance.

3.2.4 Unknown Attacks: in depth generalisation capability analysis

Once the image resolution issues have been evaluated, we selected the worst case scenario from the previous experiment (i.e., several PAI species acquired under numerous image quality conditions are employed for training) and assessed the generalisation capability of the PAD approaches in Fig. 5. To that end, we follow the leave-one-out protocol in [2] where two PAI species are used for training and the remaining species for testing.

We can observe in Fig. 5 a high D-EER variance for each unknown PAI species evaluated, which confirms the high impact degree of the image resolution on the PAD generalisation capabilities. Unlike the known attack scenario, error rates achieved by handcrafted descriptors tend to increase as the image quality improves, with the exception of Video replay attacks (i.e., D-EER of 21.36% for a high-resolution images). Specifically, a mean D-EER of 24.57% ± 8.64 indicates that the handcraft-based PAD approaches perform better for low-resolution samples stemming from unknown PAI species. In contrast, no such trend can be seen for the deep learning approaches. Depending on the given PAI species, the detection performance attained for low-resolution unknown PAIs outperforms the one reported for medium-resolution samples (e.g., 13.33% ± 4.08 vs. 17.33 ± 6.41 for Warped photo attacks). On the other hand,

Figure 5: Handcrafted vs. Deep Learning performance on the detection of unknown attacks.

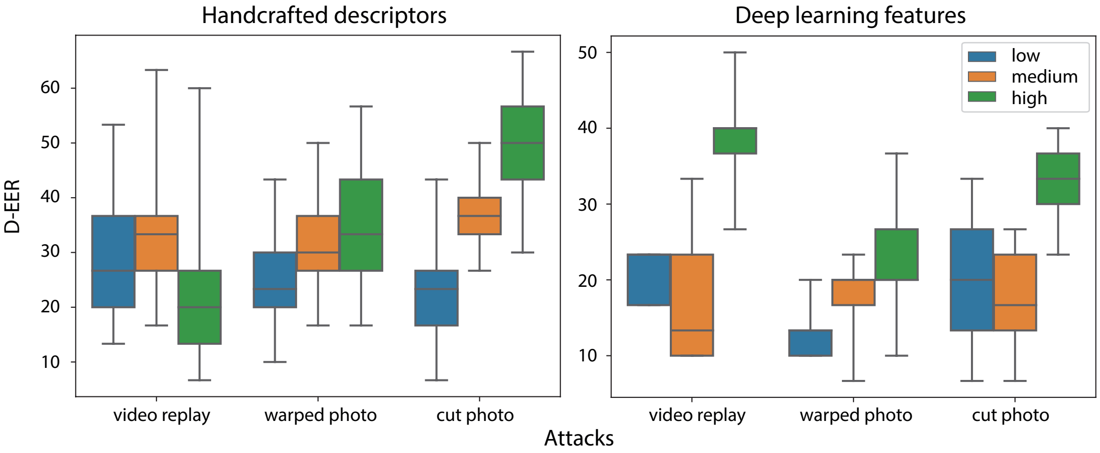
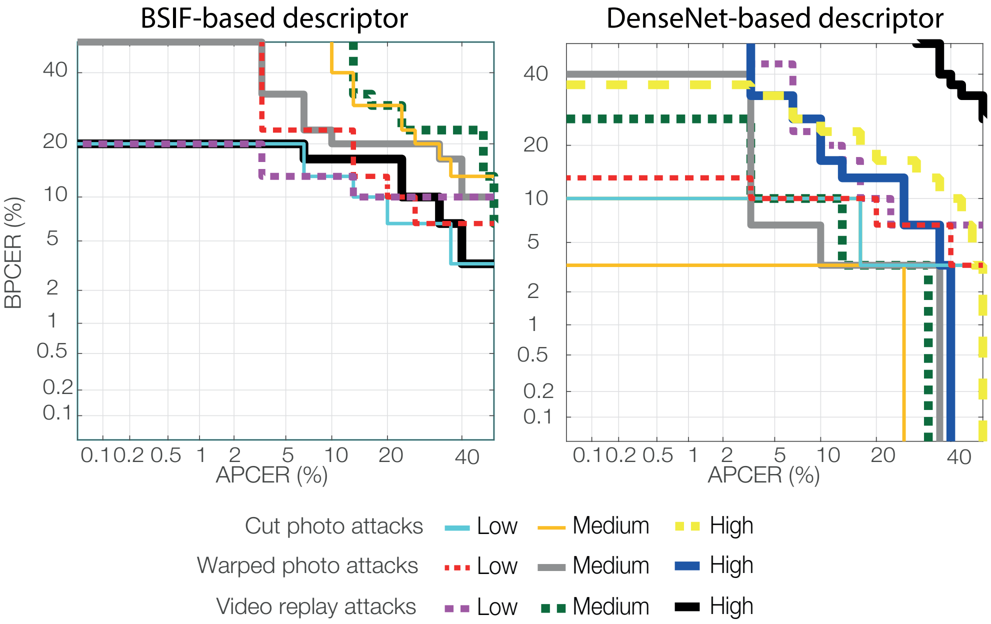


Figure 6: DET curves for the best handcrafted and deep learning-based approaches over the unknown attack scenario. For BSIF computation, we use $N = 10$ filters of size $l = 13$.



the error rates yielded for high-resolution images lead to a considerable detection performance deterioration, thereby resulting in a mean D-EER of 34.14%. Therefore, we conclude that the resolution variation is not the only external factor which affects the detection performance of PAD methods. Other acquisition conditions such as the distance between the PAI and capture device, which differs for different samples on the CASIA database, can also produce a accuracy decrease on facial PAD.

To conclude our analysis, we show in Fig. 6 the DET curves for the best handcrafted and deep learning approaches over the unknown attack scenario. First, we can note that there is a correlation between the error rates represented in Fig. 5 and the detection performance attained by a particular handcrafted and deep learning descriptor: the BSIF descriptor shows its best detection performance for a low-resolution setting, thereby resulting in a BPCER20 of 16.67% for entire set of PAIs. Similarly, the DenseNet descriptor for medium-resolution configuration attains on average a BCPER20 of 6.67% which outperforms the BPCER values achieved for the remaining resolution settings (i.e., a BPCER20 of 13.44% and 35.55% for low and high resolution samples, respectively). These results confirm that the deep learning approaches report an accuracy decrease when the images' size at hand is not close to the size of their input layer. In addition, they reveal that the PAD methods highly depends on the resolution of the capture device and hence should be carefully optimised for each particular application.

4 Conclusions

In this work, a new study on the sensitivity of the eight well-known facial PAD techniques to images with varying resolution is carried out. Experimental results, conducted over the freely available CASIA Face Antispoofing database [25], unveiled that *i*) the utilisation of images of varying quality for the facial PAD produces a high PAD performance decrease, which can be even greater than the use of numerous PAI species; *ii*) the current deep learning-based descriptors report the worse PAD performance deterioration for face images whose size is widely different from the input layer size; *iii*) Video replay attacks, screened on a high-resolution capture device, unveil several blurriness and sharpness artefacts, which can be successfully detected by PAD techniques; and *iv*) training PAD methods with several PAI species which are acquired with varying-resolution capture devices appears to be the worst case for the face PAD task, thereby resulting in a D-EER of 24.24% and a joint BPCER100 of 67.23%. Therefore, we can confirm that the image resolution is a requirement which must carefully be taken into account in order to build a secure and reliable face PAD module.

In addition to the aforementioned findings, we can also point out that PAD methods could suffer a detection performance decrease for a scenario, namely cross-database, where different resolution capture devices might be used for training and testing at the same point in time. This scenario is likely to happen in a long-time deployment where the face capture device might age and eventually stop working. Therefore, we think that the findings in this work must be carefully taken into account for that challenging scenario.

As future work directions, we plan to tackle the impact of the analysed image transformations on the performance of facial PAD methods.

5 Acknowledgments

This research work has been funded by the DFG-ANR RESPECT Project (406880674), and the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] A. Agarwal, A. Sehwag, R. Singh, and M. Vatsa. Deceiving face presentation attack detection via image transforms. In *Proc. Intl. Conf. on Multimedia Big Data (BigMM)*, pages 373–382. IEEE, 2019.

[2] S. R. Arashloo, J. Kittler, and W. Christmas. An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol. *IEEE Access*, 5:13868–13882, 2017.

[3] A. Bhogal, D. Söllinger, P. Trung, and A. Uhl. Non-reference image quality assessment for biometric presentation attack detection. In *Proc. Intl. Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2017.

[4] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, pages 1251–1258, 2017.

[5] T. Chugh, K. Cao, and A. K. Jain. Fingerprint spoof buster: Use of minutiae-centered patches. *IEEE Trans. on Information Forensics and Security*, 13(9):2190–2202, 2018.

[6] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[7] S. Fatemifar, M. Awais, S. Arashloo, and J. Kittler. Combining multiple one-class classifiers for anomaly based face spoofing attack detection. In *International Conference on Biometrics (ICB)*, 2019.

[8] J. Galbally, S. Marcel, and J. Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *IEEE Trans. on image processing*, 23(2):710–724, 2013.

[9] A. George and S. Marcel. Deep pixel-wise binary supervision for face presentation attack detection. *arXiv preprint arXiv:1907.04047*, 2019.

[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[12] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*. International Organization for Standardization, 2017.

[13] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *Intl. Conf. on Pattern Recognition (ICPR)*, pages 1363–1366, 2012.

[14] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu. Deep tree learning for zero-shot face anti-spoofing. In *Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, pages 4680–4689, 2019.

[15] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel. On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing. In *Proc. Intl. Conf. on Biometrics (ICB)*, pages 75–81, 2018.

[16] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[17] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *Proc. Intl. Conf. on Image and Signal Processing*, pages 236–243. Springer, 2008.

[18] S. Pertuz, D. Puig, and M. A. Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, 2013.

[19] L. Po, Y. Li, F. Yuan, and L. Feng. Face liveness detection using shearlet-based feature descriptors. *Electronic Imaging*, 25(4), 2016.

[20] X. Qu, J. Dong, and S. Niu. shallowcnn-le: A shallow cnn with laplacian embedding for face anti-spoofing. In *Intl. Conf. on Automatic Face & Gesture Recognition*, pages 1–8, 2019.

[21] Y. Rehman, L. Po, and M. Liu. Deep learning for face anti-spoofing: an end-to-end approach. In *Proc. Intl. Conf. on Signal Processing: Algorithms, Architectures, Arrangements, and Applications*, pages 195–200, 2017.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[24] F. Xiong and W. AbdAlmageed. Unknown presentation attack detection with face RGB images. In *Proc. Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–9, 2018.

[25] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z Li. A face antispoofing database with diverse attacks. In *Proc. Intl. Conf. on Biometrics (ICB)*, pages 26–31, 2012.