DECEIVING FACE PRESENTATION ATTACK DETECTION VIA IMAGE TRANSFORMS

Akshay Agarwal, Akarsha Sehwag, Richa Singh, and Mayank Vatsa
IIIT-Delhi, India

{akshaya, akarsha15010, rsingh, mayank}@iiitd.ac.in

ABSTRACT

Presentation attacks can provide unauthorized access to the users and fool face recognition systems for both small scale and large scale applications. Among all the presentation attacks, 2D print and replay attacks are very popular due to their ease and cost-effectiveness in attacking face recognition systems. However, over the years, there are several successful presentation attack detection algorithms developed to detect 2D print and replay attacks. Generally, 2D presentation attacks are detected using the presence or absence of *micro patterns* which distinguish a real input from an attacked input. However, if a smart attacker digitally "pre-processes" the image using intensity transforms and then performs 2D presentation attack, differences between real and attacked samples due to the micro-patterns would be minimized. In this paper, for the first time, we show that simple intensity transforms such as Gamma correction, log transform, and brightness control can help an attacker to deceive face presentation attack detection algorithms. Experimental results demonstrate that the smart attacker can increase the error rate of the hand-crafted as well as deep learning based presentation attack detectors.

***Index Terms*—** Face recognition, Presentation attack detection, Image transforms

1. INTRODUCTION

Biometrics is now considered as one of the widely used technologies for identity management, including at large scale national level projects. There are numerous benefits of using biometrics [1] over traditional identity management systems such as PINs and passwords. However, this technology also has some vulnerabilities such as presentation attacks by an attacker. As defined by recent ISO/IEC standard[1], *'presentation attack'* aims to either (i) impersonate an identity (where the attacker wants to access the system through the identity of another person) or (ii) obfuscate the identity (where an attacker wants to hide his/her own identity being caught by the surveillance systems).

There are different kinds of presentation attacks: (i) printed photo display, (ii) display the photo on an electronic device, (iii) replaying the face video on electronic mediums, and (iv) 3D mask attacks such as hard resin masks, latex masks, or silicone masks. [2, 3, 4, 5, 6, 7, 8]. Among different attacks, print and replay attacks are the most cost-effective and easy way to attack a biometric system, particularly for a face recognition system. Fig. 1 shows the difference between samples of real and replay attacked images acquired in existing research. The difference between real and attack images are high due to factors such as screen brightness, the skin tone of the subject, and luminance property of the attacking and capturing medium. This difference leads to the successful detection of 2D based presentation attack (i.e., replay). In the face presentation attack detection literature [9, 10], several algorithms are presented for both 2D and 3D attacks. Generally, these algorithms attempt to highlight the micro pattern variations to

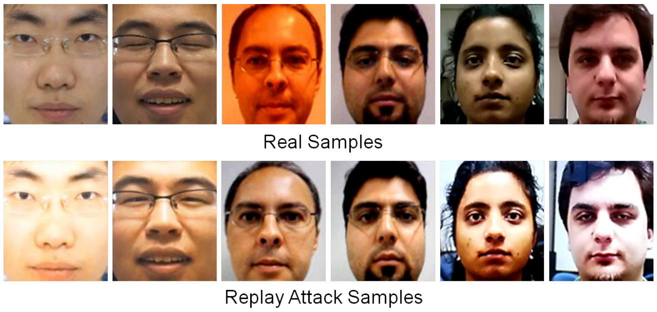


Fig. 1: The real and attack samples: images are taken from three popular face spoofing databases: CASIA-FASD [2], Replay-Attack [3], and MSU-MFSD [4].

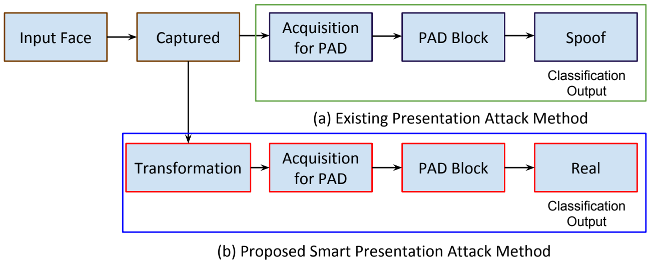


Fig. 2: A smart attacker can attack face recognition pipeline such that the video is pre-processed and then replayed on a screen: the image transform techniques can reduce the difference between real and attacked samples.

[1] https://www.iso.org/standard/53227.html

discriminate the differences between real and attacked input. While research in 3D presentation attacks, particularly with silicon masks, is still in early stages, Presentation Attack Detection (PAD) for 2D attacks is mature and almost all existing databases have shown very high PAD accuracy.

As mentioned previously, 2D presentation attacks are easy to perform as well as easy to detect. While detecting, micro patterns are "enhanced" to distinguish between real and attacked samples. In a similar fashion, a smart attacker can use image pre-processing techniques to "transform" the attacked samples such that the difference between real and attacked samples are less. As shown in Fig. 2, we propose that the effectiveness of 2D replay attack can be enhanced using image intensity transformations to fool PAD algorithms.

The key contribution of this research is: transform real high-quality videos to perform effective replay based face presentation attack. Image transforms such as log transform, Gamma correction, and intensity/brightness correction are used to neutralize the effect of factors such as environment, attacking medium, and capturing device. The extensive experiments on CASIA-FASD database [2] showcase that **'image transformation'** on 2D replay based presentation attack is able to fool face PAD algorithms using both handcrafted and deep CNN features.

2. RELATED WORK

In this section, we present existing work done towards creating the presentation attacks followed by algorithms developed to counter these attacks.

Presentation Attacks: In the literature, face presentation attacks are performed using either digital means such as morphing/swapping [11] and retouching [12] or by physical mediums such as 2D photo/replay and 3D silicone mask [10]. One of the first physical face presentation attack database namely NUAA [13] is prepared using the printed color photo of real users. Following that, several 2D photo and video attack databases are prepared such as CASIA-FASD [2], Replay-Attack [3], MSU-USSA [14] and UVAD [15]. Another medium of physical presentation attack is 3D masks. The advantage of mask attacks over photo attacks is the face like structure and texture of these masks. The first popular 3D mask attack database, namely 3DMAD [6] is prepared from 17 subjects using the Kinect sensor. Other popular 3D mask attack databases are prepared using latex and silicone masks [5, 7, 8]. These 3D mask attack database are more effective in comparison to the 2D photo and video-based attacks but require sophisticated devices and are costly to build the mask.

Agarwal et al. [16] have presented the first work to fool the face PAD algorithms using tampering of PAD features. The focus of this research is on 2D video based attacks because of its advantages in terms of cost effectiveness and easy availability. While capturing the video attack database, various factors are considered in existing databases such as resolution of the capturing and attacking device, background, and illumination. Other than these factors, as shown in Fig. 1, the skin tone of the person whose face image is being displayed and the brightness of the electronic medium on which the images/videos are displayed play important roles. This research aims to minimize the difference between real and attacked samples so that when the videos are recaptured back from the screen of electronic device, it must be close to the images/videos acquired from a real face.

PAD Algorithms: The images either captured from 2D attack or 3D mask attack generally suffers artifacts in terms of texture, quality, and natural face motions. To counter the presentation attacks, these artifacts are explored using various image texture, quality, and motion features. The most popular algorithms developed to detect the spoofing samples are either based on texture measure, motion features, and hybrid [17, 18, 19]. The popular texture measure explored in literature are local binary patterns (LBP) [20, 21, 22] and its variants such as binarized statistical image features (BSIF) [23], Gabor features [24], Haralick features [25], Moiŕe patterns [26], and image quality [27]. Similarly, motion countermeasures are based on the measurement on optical flow [19], Gaussian mixture model (GMM) [28], and dynamic frequency [15]. The boom in the hardware and software technology rise the completely different era of machine learning algorithms referred to as deep learning. The deep features computed using the multiple layers of CNN networks show a huge success in person authentication and autonomous driving and therefore leads its use for face PAD algorithm. The PAD algorithms using deep features proposed in [29, 30, 31, 8, 32, 33, 34] are based on 2D CNN, 3D CNN, deep dictionary, and deep textures. Based on the popularity of hand crafted texture and deep CNN features in detecting face presentation attacks, in this research, we have utilized LBP [35], BSIF [23], SURF [36], and VGG-16 [37] CNN features based PAD algorithms.

3. IMAGE TRANSFORMS FOR IMPROVING PRESENTATION ATTACKS

This section presents the proposed database prepared for performing *smart* 2D video based face presentation attack (i.e., replay attack) using image intensity transforms. The database contains the real and attack videos of 50 subjects. The real part of the database is taken from a high-quality subset of CASIA-FASD database[2] [2]. The real videos are captured using a high-quality camera with $1,280 \times 720$ resolution. First column of Fig. 3 shows the sample images of the real dataset. The attack dataset consists of five different subsets: one being the normal attack set and four are grouped under *Image Trans-*

[2] CASIA-FASD contains both real and attacked samples, however, we have not used attacked samples.

formation attack. In order to perform the attack, we have used the android phone with full HD display and high quality USB camera with resolution $1,920 \times 1,080$. In place of using the attack videos of CASIA-FASD database, we have collected the attack videos to keep the device and other environmental factors consistent while capturing different sets of attack videos. The attack videos in this research, can be broadly divided into two categories: (i) input transformation based and (ii) screen brightness based. Each type of image transformation attack videos are described next.

3.1. Normal Attack (Set A)

We refer the normal attack where the real videos are replayed on an electronic device without 'any' pre-processing and captured using high quality camera. Second column of Fig. 3 shows the normal attack samples. As explained earlier the attack images are of high brightness and contrast. This clear difference between the real and attack samples makes it easier to detect the 2D replay based attack.

3.2. Attack with Image Transformation - 1 (Set B)

Based on the contrast difference between real and spoof face images, one simple solution is to adjust the contrast of the images before displaying them on the screen. As the first intensity transform, we have performed *Gamma* correction before displaying the image on the screen. Gamma correction can be expressed as,

$$I_{out} = \alpha \cdot I_{in}^{\gamma} \tag{1}$$

where, α is the constant value, set to 1 and $\gamma = 0.5$ (default) is used for the experimentation. Gamma correction applies the non-linear operation to map each input pixel I_{in} to the corresponding I_{out} to increase the recordable dynamic range. The dark input values of narrow range are mapped to wider output range when γ is set to value less than 1. Third column of Fig. 3 shows the attack images recaptured after performing the Gamma correction on input images. Post Gamma correction, the intensity values are all mapped in same range and therefore, as shown in these examples, the contrast is very similar.

3.3. Attack with Image Transformation - 2 (Set C)

Log transformation is applied to decrease/remove the skewness from image data. Similar to Gamma correction, when this is applied on the input image before displaying on the screen, it expands the dark pixels present in the image. Mathematically, the transformation can be defined as:

$$I_{out} = c \cdot log(1 + I_{in}) \tag{2}$$

where, c represents the amount of enhancement, and 1 is added in the pixel values to counter $log(0)$. Fourth column of Fig. 3 shows the attack samples after log transformation with $c = 2$. The spoof images are darker as compared to Gamma correction but visually are close to real images.

3.4. Attack with Image Transformation - 3 (Set D)

In this set, we have increased the enhancement magnitude (c) of *log* transformation by two times. The log transformation might be helpful in improving the contrast of dark skin face images by increasing the details of lower intensities. The higher magnitude enhanced images as shown in the fifth column of Fig. 3. These images are much clearer as compared to Gamma correction (Set B) and brighter as compared to lower magnitude enhanced images (Set C) .

3.5. Attack with Image Transformation - 4 (Set E)

As discussed earlier, the screen brightness of the attacking medium such as mobile phone or iPad plays an important role in the spoof images. In this research, we have studied the effect by capturing the images at high brightness and low brightness of the attacking medium. All other factors such as capturing device and environmental conditions are kept fixed. Last two columns of Fig. 3 shows the effect of brightness of attacking medium. We hypothesize that similar to paper quality of print attack, various factors of screen medium must be considered while evaluating or performing the attack.

To the best of our knowledge, this is the first work where face presentation attacks are improved (made more effective) using image intensity transforms. The proposed attack approach can also help in evaluating the robustness of face PAD algorithms. Overall, the proposed database contains 50 real videos and 250 presentation attack videos. To perform the experiments, the database is divided into subject independent training-testing sets. Similar to the original protocol of CASIA-FASD, the videos corresponding to 20 subjects are used for training and 30 subjects' videos are used for evaluation. The database will be released for others at http://iab-rubric.org/resources.html.

4. FACE PRESENTATION ATTACK DETECTION: ALGORITHMS AND PERFORMANCE METRICS

In this section, different feature extraction algorithms used to develop the face PAD algorithm are described. Each feature descriptor computed over both real and spoof sets are given to the linear support vector machine (SVM) [38] classifier for binary classification. The features used for PAD algorithm development are: (i) Uniform LBP (ULBP) [35], (ii), BSIF [39], (iii) SURF proposed by Boulkenafet et al. [36], and (iv) pretrained VGG-16 [37]. The intuition of using these features is the robustness in capturing the discriminatory edge artifacts, moiŕe patterns, and image texture present in real and spoof images.

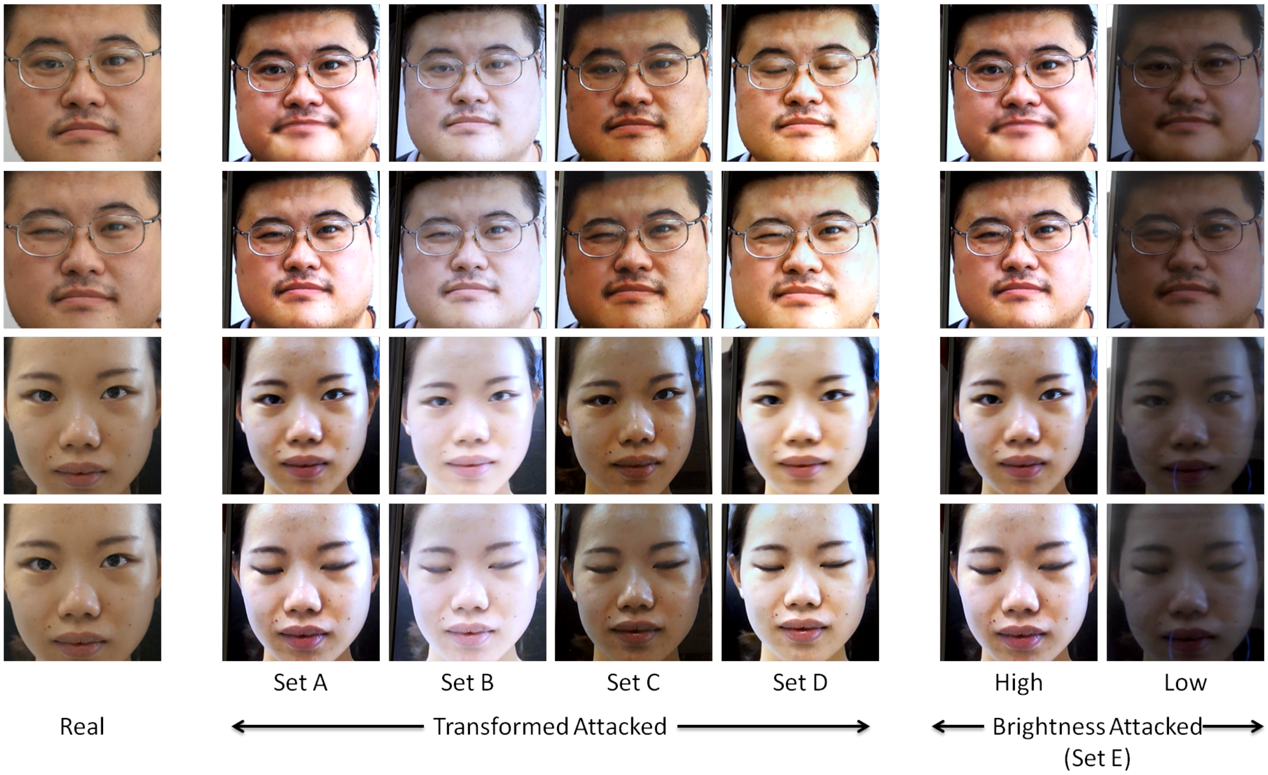


Fig. 3: Real and the proposed set of attacked images.

- **ULBP:** ULBP is one of the robust measures of local image texture. Each neighborhood pixel of a 3×3 patch is compared with the center pixel and thresholded based on the sign of the difference. In ULBP, "uniform" patterns are placed into separate bins, and other non-uniform patterns are put into single bin.

- **BSIF:** BSIF feature extraction works on the principle of LBP but in-place of using identity filter or no filter, the image is convolved with a set of learned filters. The convolved output is then thresholded based on its sign and formed binary pattern is then converted into decimal value. Finally, the histogram vector is calculated for evaluation.

- **SURF:** SURF feature is one of the scale and rotation invariant feature descriptors, which is efficient in handling variations that might be present in spoof data. The feature vector is computed by dividing the region into 4×4 cells followed by the 'Haar' wavelet decomposition. The code provided with the original paper [36] is used in default setting for feature calculation.

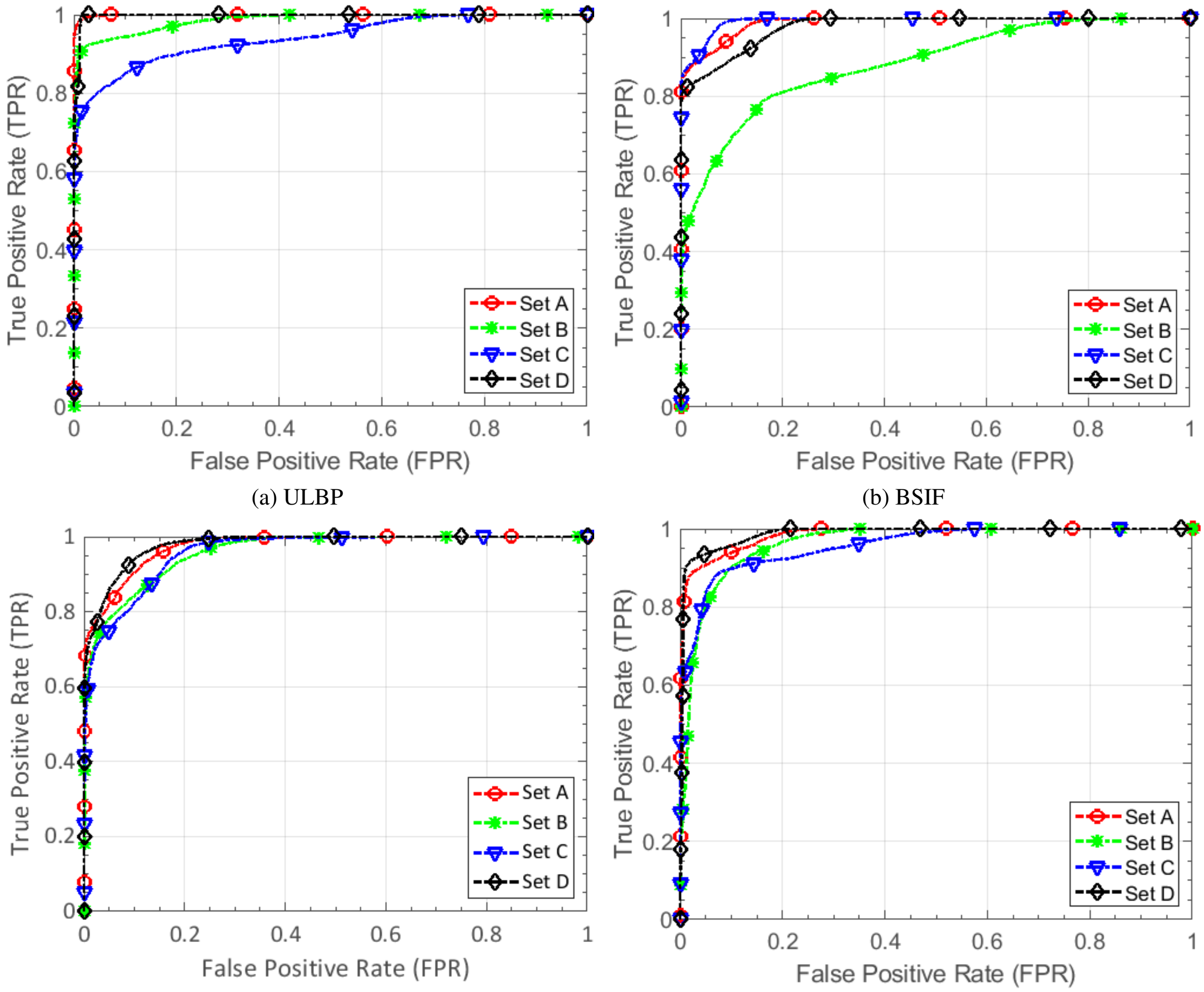
- **VGG-16:** Feature descriptors defined above represent the hand-crafted category, whereas VGG-16 defines the class of automatic feature learning from the images themselves. The VGG-16 model used in this research contains 16 layers deep CNN architecture. The pre-trained VGG-16 model is used for feature extraction from the last dense layer.

Performance Metrics: The performance of the face PAD algorithms using each feature descriptor is reported in terms of equal error rate (EER) and average classification error rate (ACER). EER is defined as the point where false accept rate (FAR) is equal to false reject rate (FRR) of receiver operating characteristic (ROC) curve. ACER is the average of bonafide presentation attack classifier error rate (BPCER) and attack presentation attack classifier error rate (APCER).

5. FACE PRESENTATION ATTACK DETECTION: RESULTS AND ANALYSIS

In this research, we have performed image intensity transforms to enhance presentation attacks and highlight the vulnerabilities of face PAD algorithms. The results corresponding to the traditional presentation attack and proposed presentation attacks in terms of EER and ACER are reported in Tables 1 and 2, respectively. The ROC curves using four feature descriptors across each type of presentation attack images are shown in Fig. 4. The results are evaluated under frame based detection where each frame is classified as real or fake.

Further in this section, the results corresponding to normal (i.e., traditional) attack set is discussed followed by the findings related to the proposed transformation and brightness based attacks are described.



(a) ULBP

(b) BSIF

(c) SURF

(d) VGG16

Fig. 4: ROC curves for face PAD algorithm used for both normal and smart attack videos.

Table 1: Frame based face presentation detection results in terms of EER% on the normal and image transformation based attacks on the proposed database.

Features	Normal Attack	Image Transformation Attack		
	Set A	Set B	Set C	Set D
ULBP	**1.41**	6.49	**13.01**	1.26
BSIF	**7.62**	**19.25**	5.44	10.22
SURF	**9.83**	12.78	**13.02**	8.31
VGG-16	**7.65**	9.96	**10.17**	5.96

Table 2: Frame based face presentation detection results in terms of ACER% on the normal and image transformation based attacks on the proposed database.

Features	Normal Attack	Image Transformation Attack		
	Set A	Set B	Set C	Set D
ULBP	**9.08**	9.31	**12.44**	9.46
BSIF	**7.29**	**25.11**	7.05	10.04
SURF	**12.14**	13.29	**14.29**	11.83
VGG-16	**9.03**	11.30	**18.45**	7.09

Normal attack set: Among the face PAD algorithms, the algorithm developed with ULBP texture features yields lowest EER value of 1.41%. In terms of ACER, the texture based BSIF descriptor yields the lowest error value of 7.29%. VGG-16 deep learning based PAD shows ACER of 9.03% and EER of 7.65%.

Image transformation sets: The ULBP feature, which yields the lowest EER value of 1.41% on normal (traditional) attack set, shows an increment of more than 11% when log enhanced images are used to perform the attack. Similarly, BSIF texture

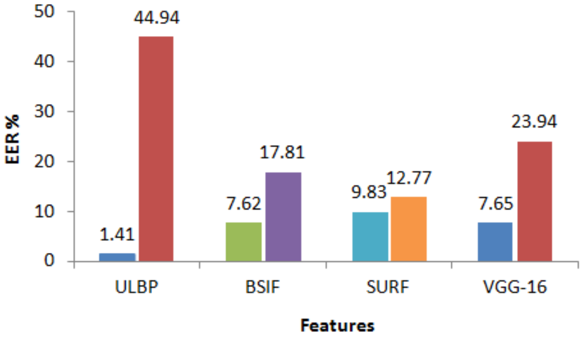


Fig. 5: Face presentation attack detection results in terms of EER % corresponding to different brightness of the attacking medium. First and second bar in each features represents the error corresponding to high and low brightness (i.e., set E), respectively.

descriptor shows the highest vulnerability towards Gamma corrected presentation attack images, and EER is increased for more than 2.5 times compared to normal attack set. In terms of EER, the VGG-16 descriptor shows the lowest sensitivity. The EER increases by 2.52% when log transform with 2 times enhancement factor (i.e., set C) is used for face presentation attack. Similarly, for other combination of image transformations, face PAD algorithms based on texture and CNN descriptor show the sensitivity towards these transformations.

In terms of ACER evaluation metric, BSIF texture feature shows the lowest value but at the same time indicates the highest sensitivity towards proposed attacks. In the case of gamma corrected attack images, the ACER value of BSIF descriptor increases by more than 3 times. The error value of ULBP texture features increases by 3.36% for log enhanced images. Similarly, the error value of deep CNN features increases by more than 2 times for log transformed images. In brief, the analysis can be summarized as follows:

- In terms of EER, ULBP descriptor yields the lowest error value whereas, in terms of ACER, BSIF shows the lowest ACER value. However, the proposed attack successfully highlights the vulnerabilities of both texture and CNN based face PAD algorithms;
- Both ULBP and VGG-16 feature descriptors are highly sensitive towards log enhanced (i.e., set C) presentation attack images. Similarly, BSIF texture descriptor shows highest vulnerabilities towards Gamma corrected (i.e., set B) attack images;
- It is interesting to note that the EER of ULBP and VGG-16 improves with higher contrast images by applying larger factor in log correction. On the other hand, the performance of BSIF improves with lower factor log enhanced images;

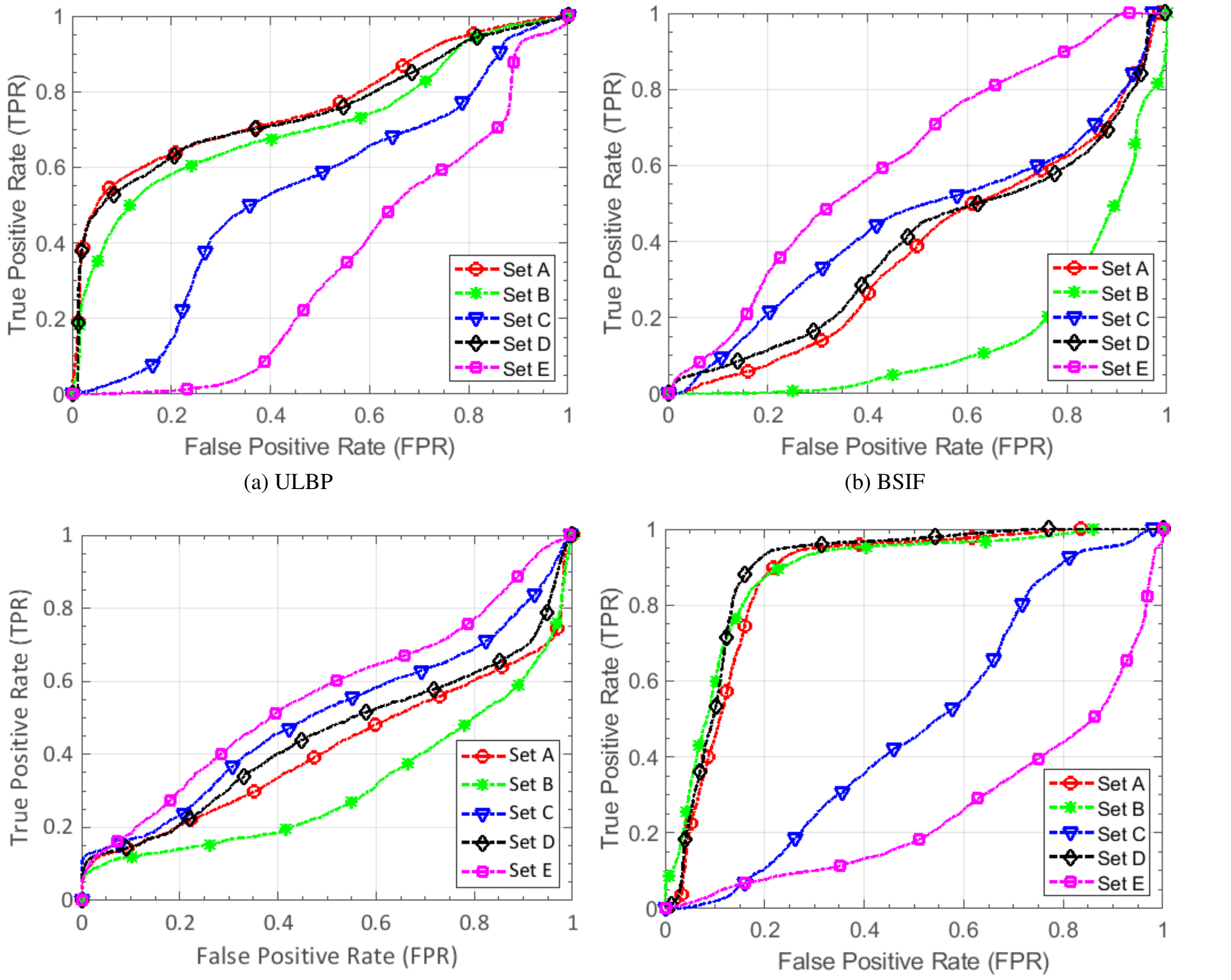
Analysis regarding brightness of attacking medium: The screen brightness of the attacking medium is also a very important factor while displaying the images for an attack. The brightness factor can largely affect the luminance component of the images. In this research, we have also performed the analysis regarding screen brightness. The sample attack images (set E) collected with high and low brightness are shown in Fig. 3. Fig. 5 shows the EER performance on the set E. When high brightness is used to perform the attack, the ULBP, BSIF, and VGG-16 feature descriptors yield EER of 1.41%, 7.62%, and 7.65%, respectively. The ULBP feature which yields the lowest EER value shows the highest sensitivity towards screen brightness and EER increases by 43.53% when low brightness attack is performed. Similarly, the performance of BSIF and VGG-16 is reduced by more than $2-3$ times on low brightness attack as compared to high brightness attack.

6. CROSS-DATABASE/ATTACK EXPERIMENTS

We have performed the experiments to model the real world scenario where the face presentation attack detector is trained on one database but tested on another [20]. For that purpose, we have used the original training set of CASIA-FASD database to develop the face presentation attack detector. The detector is trained on real and high-quality replay attack videos of 20 subjects as specified by the protocol of the CASIA-FASD database. Since, our database is built using CASIA-FASD real videos and have also followed the same subject-wise split of training and testing; training on train-set of CASIA-FASD and test on test-set of the proposed database can be considered as cross-database/attack evaluations[3]. We have evaluated ULPB, BSIF, SURF, and VGG-16 based face PAD algorithms. The performance, both in terms of EER and ACER, are reported in Table 3 and 4, respectively. Fig. 6 shows the ROC curves corresponding to these experiments. Overall, the results show the sensitivities of the PAD algorithm regarding brightness of the attacking medium, luminance correction through Gamma and log-transformed processed videos. Two key observations are summarized below:

- **Analysis regarding transformations:** Similar to brightness factor, the detector is vulnerable to Gamma and log transformed videos/images. The EER of ULBP on Gamma and log transformed images ranges from 31.89% to 44.51%. Similarly, the BSIF detector yields high ACER values, in the range of 55.07% to 68.08%.
- **Analysis regarding screen brightness:** ULBP and VGG-16 feature based PAD algorithms have shown high sensitivity towards low brightness screen attack videos. On the other hand, BSIF based face PAD algorithm has

[3] Same number of subjects' videos are used for the results reported in Sections 5 and 6.



(a) ULBP

(b) BSIF

(c) SURF

(d) VGG16

Fig. 6: ROC curves for face PAD algorithm under real-world evaluation. In all the experiments, we have observed that generally, the PAD algorithms are fooled by image transform based approaches.

Table 3: Frame based PAD results in terms of EER% on image transformation and screen brightness based attacks on the proposed database. The detector is trained using original CASIA-FASD training set for cross database/attack evaluation.

Features	High Brightness	Low Brightness	Image Transformation Attack		
	Set A	Set E	Set B	Set C	Set D
ULBP	31.49	59.37	34.47	44.51	31.89
BSIF	54.85	41.80	77.31	50.55	53.66
SURF	55.10	44.79	64.26	48.39	51.81
VGG-16	17.86	67.22	16.83	52.51	14.71

Table 4: Frame based PAD results in terms of ACER% on image transformation and screen brightness based attacks on the proposed database. The detector is trained using original CASIA-FASD training set for cross database/attack evaluation.

Features	High Brightness	Low Brightness	Image Transformation Attack		
	Set A	Set E	Set B	Set C	Set D
ULBP	35.50	57.40	41.88	50.45	36.74
BSIF	57.53	41.73	68.08	55.07	59.32
SURF	60.82	47.20	64.00	53.45	59.27
VGG-16	16.48	51.92	19.32	45.32	13.82

shown high EER and ACER when the attack videos are captured using high brightness of the attacking medium. EER and ACER values of VGG-16 PAD algorithm on high brightness attack are 17.86% and 16.48% which is increased to 67.22% and 51.92%, respectively under low brightness attack.

7. CONCLUSION

In this paper, for the first time, we have shown that, for 2D video replay attacks, image intensity transformations and brightness of the display screen can increase PAD error rates. We have highlighted the vulnerabilities of the face PAD algorithms based on handcrafted and deep CNN features. In future, the efforts can be made to further analyze the image processing operations or learning based algorithms to counter the face PAD algorithms. Moreover, digital and adversarial attacks [40, 41, 42, 43] can also be coupled with replay attacks to fool PAD algorithms. Such attacks can impede the applicability of biometrics/face recognition systems. Therefore, the research should be extended to increase the robustness of PAD algorithms against such smartly crafted presentation attacks.

8. ACKNOWLEDGEMENTS

M. Vatsa and R. Singh are partly supported by the Infosys Center for AI, IIIT Delhi. M. Vatsa is also supported through the Swarnajayanti Fellowship by Government of India. A. Agarwal is partially supported by Visvesvaraya fellowship by Government of India.

A. APPENDIX: FACE PAD WITH STATE-OF-THE-ART CNN MODEL

In the literature, several algorithms have utilized the discriminating power of CNNs to propose efficient face PAD algorithms [8, 29, 30, 31, 32, 33, 34, 44]. While in the paper (Section 5), we have already included the results of VGG-16 based approach, we have performed additional experiments using ResNet [45]. We have used four variants of ResNet ranging from 18 layers deep model to 101 layers deep model. In these experiments, pre-trained (trained on ImageNet) ResNet models are fine-tuned for face PAD. As shown in Table 5 and Fig. 7 below, the error rate of the ResNet-18 model increases from 8.33% to 26.98% (for Gamma transformation i.e., set B). Similar observations against each transformed attack have been observed across different ResNet architectures.

Table 5: PAD ACER% using various finetuned ResNet models and SURF+Softmax classifier on normal and input transformation attack videos.

PAD Classifier	Normal Attack	Transformation Attack		
	Set A	Set B	Set C	Set D
ResNet18	**8.33**	26.98	**30.41**	5.69
ResNet34	**5.76**	27.61	**33.72**	4.26
ResNet50	**2.92**	**17.06**	15.86	5.24
ResNet101	**2.93**	8.30	**16.41**	2.55

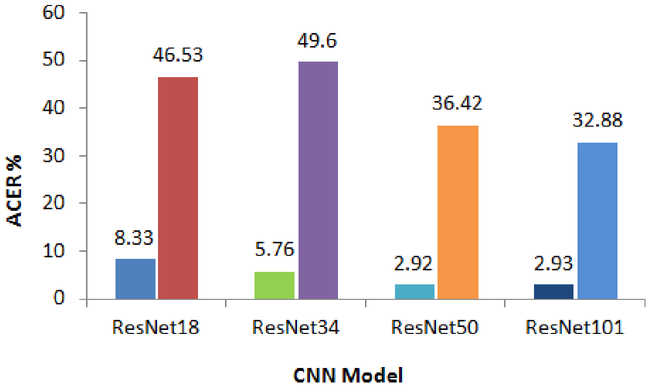


Fig. 7: PAD ACER% using various finetuned ResNet models and SURF+Softmax classifier. First and second bars in each features represent the error corresponding to high and low brightness (i.e., set E), respectively.

B. REFERENCES

[1] A. K Jain, A. A. Ross, and K. Nandakumar, *Introduction to biometrics*, Springer Science & Business Media, 2011.

[2] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *IAPR ICB*, 2012, pp. 26–31.

[3] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *BIOSIG*, 2012, number EPFL-CONF-192369.

[4] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE TIFS*, vol. 10, no. 4, pp. 746–761, 2015.

[5] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore, "Face presentation attack with latex masks in multispectral videos," in *IEEE CVPRW*, 2017, pp. 275–283.

[6] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3d masks," *IEEE TIFS*, vol. 9, no. 7, pp. 1084–1097, 2014.

[7] S. Liu, B. Yang, P. C. Yuen, and G. Zhao, "A 3d mask face anti-spoofing database with real world variations," in *IEEE CVPRW*, 2016, pp. 100–106.

[8] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar, “Detecting silicone mask-based presentation attack via deep dictionary learning,” *IEEE TIFS*, vol. 12, no. 7, pp. 1713–1723, 2017.

[9] J. Galbally, S. Marcel, J. Fierrez, et al., “Biometric antispoofing methods: A survey in face recognition.,” *IEEE Access*, vol. 2, no. 1530-1552, pp. 1, 2014.

[10] R. Ramachandra and C. Busch, “Presentation attack detection methods for face recognition systems: a comprehensive survey,” *ACM CSUR*, vol. 50, no. 1, pp. 8, 2017.

[11] A. Agarwal, R. Singh, M. Vatsa, and A. Noore, “Swapped! digital face presentation attack detection via weighted local magnitude pattern,” in *IEEE IJCB*, 2017, pp. 659–665.

[12] A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, “Detecting facial retouching using supervised deep learning,” *IEEE TIFS*, vol. 11, no. 9, pp. 1903–1913, 2016.

[13] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model,” in *ECCV*. Springer, 2010, pp. 504–517.

[14] K. Patel, HK Han, and A. K. Jain, “Secure face unlock: Spoof detection on smartphones,” *IEEE TIFS*, vol. 11, no. 10, pp. 2268–2283, 2016.

[15] A. Pinto, W. R. Schwartz, H. Pedrini, and A. d. R. Rocha, “Using visual rhythms for detecting video-based facial spoof attacks,” *IEEE TIFS*, vol. 10, no. 5, pp. 1025–1038, 2015.

[16] A. Agarwal, A. Sehwag, M. Vatsa, and R. Singh, “Deceiving the protector: Fooling face presentation attack detection algorithms,” in *IEEE/IAPR ICB*, 2019.

[17] J. Komulainen, A. Hadid, M. Pietikäinen, A. Anjos, and S. Marcel, “Complementary countermeasures for detecting scenic face spoofing attacks,” in *IAPR ICB*, 2013, pp. 1–7.

[18] J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection with component dependent descriptor.,” *IAPR ICB*, vol. 1, pp. 2, 2013.

[19] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, “Computationally efficient face spoofing detection with motion magnification,” in *IEEE CVPRW*, 2013, pp. 105–110.

[20] T. de Freitas Pereira, A. Anjos, J. Mario De Martino, and S. Marcel, “Can face anti-spoofing countermeasures work in a real world scenario?,” in *IAPR ICB*, 2013, pp. 1–8.

[21] J. Mtt, A. Hadid, and M. Pietikinen, “Face spoofing detection from single images using micro-texture analysis,” in *IEEE IJCB*, 2011, pp. 1–7.

[22] T. A. Siddiqui, S. Bharadwaj, T. I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, and N. Ratha, “Face antispoofing with multifeature videolet aggregation,” in *IAPR ICPR*, 2016, pp. 1035–1040.

[23] R. Raghavendra, K. B. Raja, and C. Busch, “Presentation attack detection for face recognition using light field camera,” *IEEE TIP*, vol. 24, no. 3, pp. 1060–1075, 2015.

[24] M. Waris, Honglei Zhang, I. Ahmad, S. Kiranyaz, and M. Gabbouj, “Analysis of textural features for face biometric anti-spoofing,” in *IEEE EUSIPCO*, 2013, pp. 1–5.

[25] A. Agarwal, R. Singh, and M. Vatsa, “Face antispoofing using haralick features,” in *IEEE BTAS*, 2016, pp. 1–6.

[26] K. Patel, H. Han, A. K. Jain, and G. Ott, “Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks,” in *IAPR ICB*, 2015, pp. 98–105.

[27] J. Galbally, S. Marcel, and J. Fierrez, “Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition,” *IEEE TIP*, vol. 23, no. 2, pp. 710–724, 2014.

[28] J. Yan, Z. Zhang, Z. Lei, D. Yi, and S. Z. Li, “Face liveness detection by exploring multiple scenic clues,” in *IEEE ICARCV*, 2012, pp. 188–193.

[29] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, “Face antispoofing using patch and depth-based cnns,” in *IEEE IJCB*, 2017, pp. 319–328.

[30] N. N. Lakshminarayana, N. Narayan, N. Napp, S. Setlur, and V. Govindaraju, “A discriminative spatio-temporal mapping of face for liveness detection,” in *IEEE ISBA*, 2017, pp. 1–7.

[31] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot, “Learning generalized deep feature representation for face anti-spoofing,” *IEEE TIFS*, vol. 13, no. 10, pp. 2639–2652, 2018.

[32] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falco, and A. Rocha, “Deep representations for iris, face, and fingerprint spoofing detection,” *IEEE TIFS*, vol. 10, no. 4, pp. 864–879, 2015.

[33] R. Shao, X. Lan, and P. C. Yuen, “Joint discriminative learning of deep dynamic textures for 3d mask face antispoofing,” *IEEE TIFS*, vol. 14, no. 4, pp. 923–938, 2019.

[34] X. Tu and Y. Fang, "Ultra-deep neural network for face anti-spoofing," in *ICNIP*. Springer, 2017, pp. 686–695.

[35] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, vol. 24, no. 7, pp. 971–987, 2002.

[36] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face antispoofing using speeded-up robust features and fisher vector encoding," *IEEE SPL*, vol. 24, no. 2, pp. 141–145, 2017.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[38] C. Cortes and V. Vapnik, "Support vector machine," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[39] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *IAPR ICPR*, 2012, pp. 1363–1366.

[40] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14410–14430, 2018.

[41] G. Goswami, N. Ratha, A. Agarwal, R. Singh, and M. Vatsa, "Unravelling robustness of deep learning based face recognition against adversarial attacks," *AAAI*, pp. 6829–6836, 2018.

[42] G. Goswami, A. Agarwal, N. K. Ratha, R. Singh, and M. Vatsa, "Detecting and mitigating adversarial perturbations for robust face recognition," *IJCV*, vol. 127, no. 6-7, pp. 719–742, 2019.

[43] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *ICLR*, 2014.

[44] D. Yadav, N. Kohli, A. Agarwal, M. Vatsa, R. Singh, and A. Noore, "Fusion of handcrafted and deep learning features for large-scale multiple iris presentation attack detection," in *IEEE CVPRW*, 2018, pp. 572–579.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.