Open access • Proceedings Article • DOI:10.1109/BTAS.2018.8698550

Spoofing Deep Face Recognition with Custom Silicone Masks — Source link

Sushil Bhattacharjee, Amir H. Mohammadi, Sébastien Marcel

Institutions: Idiap Research Institute

Published on: 01 Oct 2018 - International Conference on Biometrics: Theory, Applications and Systems

Spoofing Deep Face Recognition with Custom Silicone Masks

Sushil Bhattacharjee Amir Mohammadi Sébastien Marcel
Idiap Research Institute. Centre du Parc, Rue Marconi 19, Martigny (VS), Switzerland
{sushil.bhattacharjee; amir.mohammadi; sebastien.marcel}@idiap.ch

Abstract

We investigate the vulnerability of convolutional neural network (CNN) based face-recognition (FR) systems to presentation attacks (PA) performed using custom-made silicone masks*. Previous works have studied the vulnerability of CNN-FR systems to 2D PAs such as print-attacks, or digital-video replay attacks, and to rigid 3D masks. This is the first study to consider PAs performed using custom-made flexible silicone masks. Before embarking on research on detecting a new variety of PA, it is important to estimate the seriousness of the threat posed by the type of PA. In this work we demonstrate that PAs using custom silicone masks do pose a serious threat to state-of-the-art FR systems.*

Using a new dataset based on six custom silicone masks, we show that the vulnerability of each FR system in this study is at least 10 times higher than its false match rate. We also propose a simple but effective presentation attack detection method, based on a low-cost thermal camera.

1. Introduction

The high accuracy of state-of-the-art face recognition (FR) methods comes from their ability to handle the inevitable variability in the input data. In recent years, convolutional neural network (CNN) based FR systems have demonstrated an extraordinary capacity to compensate for such variability, as evidenced by their near-perfect recognition performance on difficult datasets such as the Labeled Faces in the Wild (LFW) dataset [8]. This very ability to discount the inter-session variability, however, also makes an FR system vulnerable to *presentation attacks* (PA) [6, 7, 12] (also called spoofing attacks).

(a) Rigid 3D mask (b) Custom silicone mask

Figure 1: Illustration of presentation attacks based on 3D masks: rigid mask [7] and flexible custom silicone mask (**this study**).

PAs fall into two categories: *impersonation*, where the attacker attempts to impersonate the victim, and *obfuscation*, where the attacker wishes to avoid being recognized. Mohammadi *et al*. [12] have shown that state-of-the-art CNN-FR systems are highly, and at times, completely vulnerable to 2D impersonation-PAs, that is, to print-attacks, and to replay-attacks performed using electronic *presentation attack instruments* (PAI).

In this work we consider PAs using custom flexible masks made of silicone. At present, a large body of research exists on the subject of face *presentation attack detection* (PAD) [11, 15]. Most face-PAD countermeasures, however, have been developed to detect 2D PAs. Very few researchers [2, 7, 16] have considered the threat from 3D-mask based PAs (two examples of such attacks are illustrated in Fig. 1). The PAs considered in these studies have been created using rigid-masks [7, 16], and generic latex masks [2]. This is the first systematic study on the vulnerability of FR systems to impersonation attacks based on custom-made silicone masks. One reason for the lack of research on this specific kind of PA is that constructing a custom-silicone mask requires a high level of expertise, and therefore, is an expensive process. The cost of manufacturing such custom silicone masks, however, is dropping, and custom-masks are expected to become fairly affordable in the near future.

In this paper we primarily address the question: "How vulnerable are CNN-FR systems to PAs using custom silicone masks?" Specifically, we investigate the vulnerability of three CNN-FR systems – VGG-Face [13], LightCNN [23] and Facenet [17] – to mask-based PAs. The reason these specific CNN-FRs have been chosen for this study is that they were shown to have very high FR performance, as well as the highest vulnerability to 2D PAs [12].

The main contributions of this study are as follows.

1. This is the first FR-vulnerability study involving impersonation PAs made using custom silicone masks. Our experiments clearly show that the three CNN-FR methods are all significantly vulnerable to custom-mask based PAs.
2. We also present preliminary results of PAD for such PAs, based on thermal imagery. Thermal images of live face-images are starkly distinct from those of PAs, including silicone mask attacks. Simple PAD methods, not requiring elaborate statistical modelling, may be used [5] with thermal imagery for face-PAD.
3. A new Custom Silicone Mask Attack Dataset (CS-MAD) (containing *bona fide* presentations, as well as PAs made using custom silicone masks for six identities), as well as the source-code for our experiments.

In Section 2 we provide brief reviews of relevant works by other researchers. The three CNN-FR methods evaluated here are briefly described in Section 3. The data and protocols used in this study are described in Section 4, and experimental results are discussed in Section 5. Finally, Section

6 provides a summary of this study, the conclusions drawn from the results, and an outlook on future work.

2. Related Works

Several strands of related works are relevant to this paper. Here we provide brief overviews of relevant background.

CNNs for Face Recognition

Several CNN-FR methods have been proposed in the recent scientific literature. Taigman *et al*. [21] have proposed DeepFace, a 9-layer CNN for FR. They use 3D modeling and piecewise affine transforms to first align the input faces to an upright position, before feeding the corrected face-image to the DeepFace network. This network achieves a recognition accuracy of 97.25% on the LFW dataset. Schroff *et al*. [18] report an accuracy of 99.63% on the LFW dataset using FaceNet, a CNN with 7.5 million parameters, trained using a novel *triplet* loss function. One of the most popular CNN-FR systems today is the VGG-Face CNN [13].

In this work we analyze the vulnerability to PAs of three CNN-FR methods: the popular VGG-Face [13], LightCNN [23], and FaceNet [17]. These networks are discussed in more detail in Section 3.

3D-Mask PAD

Although 3D-mask based attacks are quickly evolving into significant threats, methods to detect such attacks have only recently attracted research interest.

The first publicly available experimental dataset, 3DMAD [7], was created in 2013, using a set of rigid custom masks manufactured on a 3D-printer. Other researchers [1] have used the 3DMAD dataset for 2D face-PAD experiments. Liu *et al.* [9] have published the more recent HKBU-MARs dataset containing images of 3D-rigid-mask based PAs. They have successfully used remote photo-plethysmography (rPPG) to detect 3D-mask PAs. Both works ([7, 9]) have demonstrated the vulnerability of FR methods to the PAs using rigid masks.

Previous works involving flexible masks, in particular [2, 10], have proposed PAD methods for obfuscation PAs, based on generic silicone and latex masks, not impersonation attacks using custom-made masks. The observational study by Manjani *et al*. [10] is based on images collected from the Internet (the SMAD dataset), that is, the data used in the study was not collected under controlled conditions. Hence, we cannot rule out the influence of factors beyond the control of the authors over their PAD results. The work of Agarwal *et al.* [2] (based on the MLFP dataset) also addresses obfuscation attacks based on generic latex masks. Unlike impersonation-attacks (studied in the present work) obfuscation attacks, by their very nature, do not involve custom-made masks.

Extended-Range Imagery for PAD

Raghavendra *et al*. [14] have published a detailed study on the vulnerability of FR systems in multispectral imaging domain, involving seven-band imagery covering the visible light and NIR illumination. Steiner *et al*. [19] have recently demonstrated the use of multispectral SWIR imagery to reliably distinguish human skin from other materials, and have shown that such multispectral devices can be used for PAD. Both these works have used custom-built imaging systems. Bhattacharjee and Marcel [5] have used a new crop of off-the-shelf extended-range imaging devices to detect 2D as well as mask-based PAs.

3. The Studied CNN-FR Methods

A typical FR system functions in three phases: training, enrollment, and probe. In the training phase a background model, assumed to broadly represent a sub-space of face-images, is constructed using training data. In the enrollment phase, the FR system uses the background model to generate templates for the given enrollment samples, which are then stored in the gallery. In the probe phase, the FR system is presented with a probe-image and a claimed identity. The template created for the probe-image is compared with the set of enrolled templates for the claimed identity. The match-score is thresholded to reach a decision (accept/reject).

Sub-images representing individual faces are first extracted from the raw input image. Geometric and color transforms may also be applied to the extracted face-images, depending on the requirements of the specific FR method. The result of this pre-processing stage is a *normalized* face image, of predefined size and scale, that may form an input to a FR system. Before describing the different CNN-FR methods, we explain the pre-processing steps applied to normalize the input face images.

3.1. Face Image Normalization

For the FaceNet and VGG-Face networks, the input images are normalized color (RGB) images, whereas the LightCNN expects normalized gray-level face-images of fixed 2D-shape. Our first pre-processing step is to convert the input RGB image to a gray-level image. The face region in the input image is localized as follows. First, a face-detector [4] is used to localize the face-region. Next, facial-landmarks are extracted from the face-region using the *flandmarks* method [22]. This method returns pixel coordinates for seven facial landmarks, including the two corners of each eye, from which we estimate the locations of the two eye-centres. Imposing the constraints that the straight-line joining the two eye-centres should be horizontal, and should have a predefined length, an affine transform is used to extract a normalized face-image of fixed size from the given input image.

The three CNN-FR methods are discussed in the following sub-sections. Note that, for all three CNNs, we have used the same parameter-settings as used in [12].

3.2. CNN-Based Systems

For FR applications, CNNs are usually trained for face identification, using face-images as input, and the set of identities to be recognized as output. For face-verification applications, CNNs are typically used as feature-extractors. The terms *representation* and *embedding* are used interchangeably, to denote the outputs of the various layers of a deep network. Representations generated by a specific layer of a pre-trained CNN may be used as templates (feature-vectors) representing the corresponding input images. Such templates may be subsequently be compared to each other using appropriate similarity measures. Brief descriptions of the selected CNNs are provided below.

VGG-Face CNN: VGG-Face is a CNN consisting of 16 hidden layers [13]. The input to this network is an appropriately normalized color face-image of pre-specified dimensions. We use the representation produced by the penultimate fully-connected layer ('fc7') of the VGG-Face CNN as a template for the input image. When enrolling a client, the template produced by the VGG-Face network for each enrollment-sample is recorded.

LightCNN: Even though LightCNN [23] involves a much smaller number of parameters than the VGG-Face network, it shows marginally better FR performance on the LFW dataset than VGG-Face. Here we use as templates the 256-D representation produced by the 'eltwise_fc1' layer of LightCNN.

FaceNet CNN: We have used the implementation of FaceNet, and associated models, published by David Sandberg [17]. This is the closest open-source implementation of the FaceNet CNN proposed by Schroff *et al*. [18]. Sandberg's FaceNet implements an Inception-ResNet V1 DNN architecture [20]. In our tests, we have used the 20170512-110547 model from Sandberg. We use the 128-D representation at the output of the 'embeddings:0' layer of FaceNet to construct enrollment and probe templates.

During verification, the CNN generates a template for the probe face-image, which is then compared to the enrollment templates of the claimed identity using the Cosine-similarity measure. The score assigned to the probe is the average Cosine-similarity of the probe-template to all the enrollment-templates of the claimed identity. If this score exceeds a preset threshold, the probe is accepted as a match for the claimed identity.

4. Description of Masks and Dataset

The masks used in this study are described in this section. We also explain here our data-collection process.

4.1. Custom-made Silicone Masks

Of the 14 participants in this study (denoted as subjects **A** – **N** hereafter), custom silicone masks have been manufactured for six subjects, **A** – **F**. These masks have been manufactured by a professional special-effects company, at a cost of approximately USD 4,000 per mask. For each of the six subjects, the manufacturer was provided with the following data: (1) 3D-scan of the face, collected using an Intel Realsense SR300 camera, (2) color photographs of subject's face (frontal, lateral and diagonal views), and, (3) physical measurements of facial features. For each subject, the manufacturer first made a cast of the subject's face and used that to create the silicone mask. Superficial features of each mask, such as eye-brows and facial make-up, were finished manually. The manufacturer has also provided a bespoke matching support for each mask.

Silicone mask examples are shown in Figure 3. The inner surface of each mask is coated with an adhesive substance, which helps to hold the mask in position when worn. The masks are manufactured with holes in eye-locations. In Figure 3 the masks are shown mounted on their bespoke supports provided by the manufacturer, with glass eyes and silicone eye-sockets, also provided by the manufacturer.

4.2. Data Collection

Two cameras – the Realsense SR300 from Intel, and the Compact Pro from Seek-Thermal – have been used for recording *bona fide* and attack presentations in this work.
SR300: this camera, priced at about USD 150, captures three kinds of data: color (RGB) images, near-infrared (NIR, 860nm) images, and depth-data. We have captured color images at full-HD resolution, and the corresponding NIR and depth images at VGA resolution, at 30 frames per sec. (fps).
Compact Pro: this camera, costing about USD 500, collects thermal (long-wave infra-red (LWIR)) images at QVGA resolution, at approximately 10 fps.
In our experiments the two cameras have been used in a fixed spatial configuration, such that it is straightforward to extract face-regions in the thermal image (from Compact Pro), based on computations on the color image (from SR300). Samples images in the different wavelengths are shown in Figure 4. Images in the top row of the figure correspond to *bona fide* presentations and images in the bottom row correspond to mask-PAs. Note that the mask is clearly distinguishable in the thermal image.

4.3. Dataset and Experimental Protocols

The dataset collected for this study is named the CS-MAD (Custom Silicone Mask Attack Dataset)[1]. The dataset consists of videos as well as still photographs.

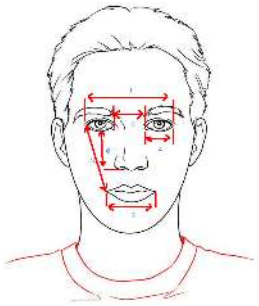
As mentioned before, masks for subjects **A** – **F** have been used in this work. *Bona fide* presentations have been captured for each of the six subjects, as well as for remaining eight (**G** – **N**) subjects. Videos of *bona fide* presentations have been recorded under four different illumination conditions, namely: (1) *i1*: ceiling lights, (2) *i2*: floor-standing incandescent lamp illuminating the subject from the left side only, (3) *i3*: similar floor-standing illumination from the right side only, and (4) *i4*: floor-standing illumination from both sides of the subject. Thus, four short (5 – 10 sec.) *bona fide* presentation-videos have been recorded for each subject. For certain subjects, four additional *bona fide* videos have been recorded (corresponding to the four illumination conditions described above), where the subjects are wearing spectacles. The spectacles as well as the different illumination conditions introduce substantial variability in the dataset.

High quality still color photographs have also been captured using a Nikon Coolpix P520 camera, for subjects **A** – **F**. Several photographs, from varying angles, have been collected for each subject. These photographs were initially collected to aid the mask-manufacturer but have also been used in the vulnerability analysis experiments.

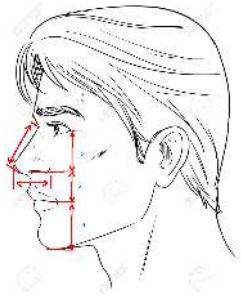
Mask PA videos using both cameras have also been captured under the same four illumination conditions. Two kinds of PA videos have been recorded: (1) where each mask is worn by a person, and (2) where the masks are mounted on their corresponding supports provided by the manufacturer. In the first case, the masks have each been worn by two subjects, in turn. In total, the dataset used in this work consists of 88 *bona fide* videos, 160 mask-PA videos, and 60 high-quality still color *bona fide* photographs.

[1] www.idiap.ch/dataset/csmad

(a) 3D-scan



(b) Frontal face measurements (6)



(c) Profile face measurements (5)

(d) Right 90°

(e) Right 45°

(f) Frontal

(g) Left 45°

(h) Left 90°

Figure 2: Examples of facial data collected for each subject. The red double-headed arrows in (b) and (c) indicate the facial measurements taken for subjects A – F.

Figure 3: Custom silicone half-masks for six subjects. **Rows 1 & 3:** *bona fide* presentations of the six subjects for whom mask-attacks have been prepared; **Rows 2 & 4:** the masks corresponding to the subjects above.

The vulnerability of each CNN-FR system has been assessed for the six identities **A** – **F**. The face-verification protocol for the vulnerability analysis is as follows:
Enrollment: one frontal photograph and 3 videos (illuminations *i1*, *i2*, and *i3*), for subjects **A** – **F**.
Probe: one *bona fide* video (illumination *i4*) and all the remaining still images for subjects **A** – **F**, all *bona fide* videos for subjects **G** – **N**, as well as all mask-attack videos.
No training set is necessary in these experiments, as we have used pre-trained CNNs provided by the creators of the respective networks. Given the small size of the CS-MAD (custom-masks for only 6 identities), we have used the entire dataset as *test* data. The vulnerability analysis results presented in Section 5 are based on *a posteriori* selection of classification thresholds.

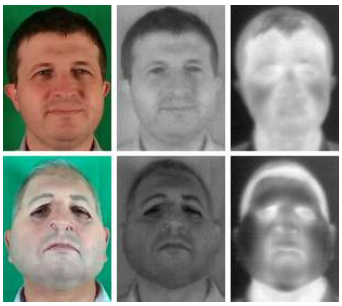


Figure 4: Samples images. **Top:** images of a *bona fide* presentation. **Bottom:** Silicone mask attack presentation. **Columns (left to right):** color- (RGB), NIR-, and Thermal images. The RGB and NIR images are captured with the SR300 camera, and the thermal images are captured using the Compact-Pro.

5. Experiments

Vulnerability analysis results for the three CNN-FR methods, followed by PAD results are summarized in this section.

5.1. Vulnerability Analysis

Vulnerability of each CNN-FR method to custom silicone masks is estimated as follows: identities are enrolled in the FR system using *bona fide* enrollment samples and a gallery is created for each identity. During evaluation, the gallery-templates are compared against genuine samples (*bona fide* samples compared against gallery templates of the true identities respectively) as well as ZEI samples (*bona fide* samples compared against the gallery of another identity). Based on the comparison scores obtained so far, the operating score-threshold is chosen so as to minimize the misclassification rate between genuine and ZEI presentations. The probe-template of each mask-PA sample

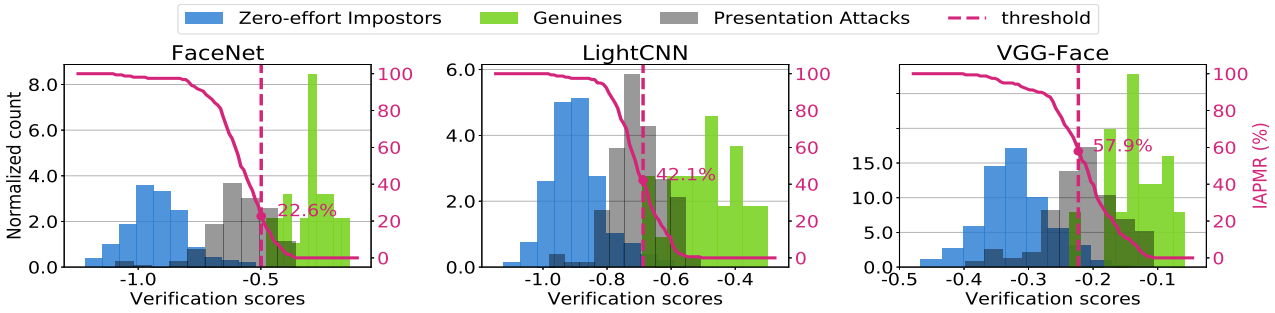


Figure 5: FR score-distributions. Each plot shows three histograms of scores – for genuine (green), ZEI (blue) and attack (gray) presentations. The red vertical dashed line marks the score-threshold $\mathcal{T}_{EER}$, for the FR method in question. Samples with scores lower than $\mathcal{T}_{EER}$ are rejected. The solid red line shows the evolution of the IAPMR as the operating threshold is varied from left to right.

Table 1: FR accuracy and vulnerability of each FR system are shown for two test-sets: CS-MAD and '2D Attacks'. The '2D Attacks' here refer to the results of the *Combined* dataset as published in [12]. The IAPMR values have been computed based on the $\mathcal{T}_{EER}$ score-threshold. 95% confidence intervals for the IAPMR values are shown in brackets.

FR Method	**Datasets**	**Score-Threshold @ EER** ($\mathcal{T}_{EER}$)			
		FMR	**FNMR**	**HTER**	**IAPMR**
FaceNet	CS-MAD	0.0	0.0	0.0	22.6 [16.4, 29.9]
	2D Attacks	0.1	0.5	0.3	99.9 [99.7, 100]
LightCNN	CS-MAD	3.0	3.6	3.3	42.1 [34.4, 50.2]
	2D Attacks	1.0	1.4	1.2	99.8 [99.6, 99.9]
VGG-Face	CS-MAD	4.2	3.6	3.9	57.9 [49.8, 65.6]
	2D Attacks	0.3	2.0	1.2	97.0 [96.4, 97.6]

is then compared against gallery-templates of the identity corresponding to the mask in question, and the performance-metrics are reported using the previously chosen threshold. The metrics used in this work are:

False match rate (FMR): the proportion of ZEI that are wrongly accepted as genuine presentations (Type-I error).

False non-match rate (FNMR): the proportion of wrongly rejected genuine samples (Type-II error).

Half-total error rate (HTER): average of FMR and FNMR.

Impostor attack presentation match rate (IAPMR): the proportion of PAs accepted by the FR system as genuine presentations.

Table 1 summarizes the performance of each CNN-FR system. The accuracy of each FR method is reported in terms of FMR and FNMR. For convenience, the HTER for each experiment is also provided. The table also shows the vulnerability of each FR method to presentation attacks (IAPMR). To facilitate comparison, the vulnerability of each FR method to 2D PAs, as published in [12] (for the *Combined* dataset), is also reported in this table. Please note that in Table 1, for 2D Attacks [12] the $\mathcal{T}_{EER}$ score-threshold has been computed *a priori* (using the development set of the *Combined* dataset), whereas for the CS-MAD, the same threshold ($\mathcal{T}_{EER}$) has been determined *a posteriori*.

The VGG-Face network shows the highest vulnerability (IAPMR = 57.9%). The HTER of the VGG-Face network is also higher than that of the other networks. In general, the IAPMR values are roughly 10 times higher than the FMR values in all three FR methods using the CS-MAD. However, comparing the vulnerability analysis results on the CS-MAD and those on the Combined dataset of [12], we note that the IAPMR values of the custom-masks are significantly lower than IAPMR values of 2D attacks. This may be due to the fact that the mask attacks in the CS-MAD are not as realistic as the 2D attacks in the Combined dataset.

Imaging conditions can also affect the success rate of the PAs. Table 2 shows the IAPMR values for six attack categories. First we group the attacks into two categories depending on whether the mask is worn by the attacker (category: 'Worn') or not (category: 'Stand'). As the first two rows of the table show, there is no significant difference between IAMPR values the two categories of mask-attacks. Next we partition the PAs into four categories (*i1* $\cdots$ *i4*) according to the illumination conditions of the attack. The table shows that PAs in categories *i1* and *i4*, where the illumination is relatively uniform, are more successful compared to the attacks that were recorded under the *i2* lighting conditions.

Table 2: IAPMR (%) values of the FR methods for custom silicone mask PAs. The IAPMR value is shown separately for attacks performed under different conditions (see text for details). 95% confidence intervals are shown for the IAPMR values in brackets.

Category	**FaceNet**	**LightCNN**	**VGG-Face**
Stand	27.5 [15.9, 41.7]	37.3 [24.1, 51.9]	60.8 [46.1, 74.2]
Worn	20.4 [13.2, 29.2]	44.4 [34.9, 54.3]	56.5 [46.6, 66.0]
i1	25.6 [13.0, 42.1]	48.7 [32.4, 65.2]	64.1 [47.2, 78.8]
i2	20.0 [9.1, 35.6]	27.5 [14.6, 43.9]	50.0 [33.8, 66.2]
i3	17.1 [7.2, 32.1]	46.3 [30.7, 62.6]	61.0 [44.5, 75.8]
i4	28.2 [15.0, 44.9]	46.2 [30.1, 62.8]	56.4 [39.6, 72.2]

The histograms in Figure 5 show score-distributions of the three kinds of presentations – genuine (green), ZEI (blue) and PAs (gray). Ideally, there should be no overlap between the green histogram and the other two histograms. The distribution of PA-scores lies between genuine and ZEI scores for all three CNN-FR systems, that is, in general, the PAs are scored higher than the ZEIs by all three FR methods. Even for FaceNet, which has the lowest IAPMR of the three FR systems, the score distribution of PAs are considerably to the right of the ZEI-scores.

Examples of successful PAs are shown in Figure 6. The least successful attacks all occur under the *i2* illumination.

5.2. PAD Using Thermal Imagery

A simple PAD method relies on the mean-intensity over the face-regions in the thermal images in the CS-MAD. We use the color image from the SR300 camera for face-localization. The face coordinates determined on the color

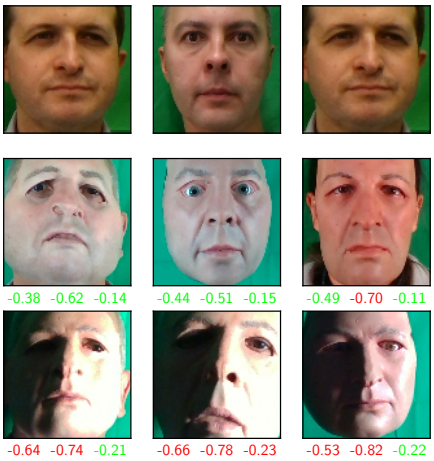
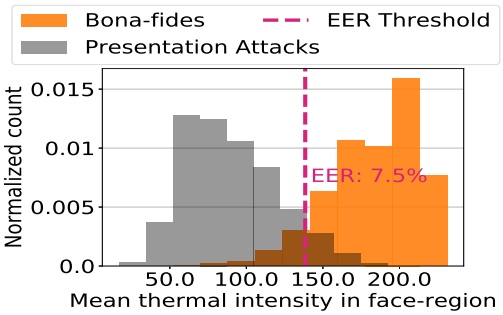
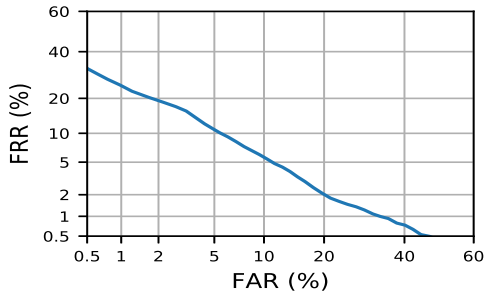


Figure 6: Examples of PAs most/least successful in attacking the FR methods. **Top:** enrollment sample of each column. **Middle:** most successful PA when attacking, from left to right, FaceNet, LightCNN, and VGG-Face networks respectively. **Bottom:** least successful PA in attacking, from left to right, FaceNet, LightCNN, and VGG-Face networks of the same identity. Below each PA, 3 scores from the FaceNet, LightCNN, and VGG-Face FR methods are shown, respectively, from left to right. Scores higher than $\mathcal{T}_{EER}$ of the corresponding FR method are shown in green. Otherwise, the score is shown in red.



(a) Score distribution



(b) DET curve

Figure 7: Results of presentation attack detection using thermal images captured using the Seek-Thermal Compact-Pro camera.

image are projected onto the thermal image via an affine transform, to locate the corresponding face-region in the thermal image. For *bona fide* presentations the face regions are quite bright in thermal images (see Fig. 4). For mask-PAs, we expect the face-region in the thermal image to be fairly dark, because the mask is at a relatively lower temperature than the human body. When worn by an attacker, the mask may warm up with time. However, it is still reasonable to expect the mask-region to be darker than live-presentations. Therefore, mean facial brightness in thermal images is an effective indicator of the presence of a mask.

PAD results are shown in Figure 7. For a given presentation, the mean facial brightness statistic is directly used as the PAD score. Figure 7(a) shows the score-distributions for *bona fide* presentations and for PAs. The overlap between the two score-distributions may be attributed to several reasons:

1. Incorrect face-localization may lead to some background pixels also being included when computing the face-brightness statistic.
2. Spectacles in *bona fide* presentations also significantly lower the mean face-brightness in the thermal image.
3. Mask temperature rises significantly if the attacker wears the mask for a relatively long period of time.

Ignoring these challenges for the time being, we note that the simple temperature based threshold performs quite well in detecting mask based PAs. Fig. 7(b) shows the detection error tradeoff (DET) curve for this PAD experiment, indicating an EER of 7.5%. If we permit a FAR of 5%, then the corresponding FRR is about 10%. The challenges listed above are not insurmountable. Indeed, a Principal Component Analysis based PAD method, to be investigated in future work, should improve over this PAD result significantly.

6. Conclusions

This is the first study to evaluate the vulnerability of these CNN-FR systems to impersonation-PAs using custom-made silicone masks. The selected CNN-FR systems are considered state-of-the-art FR methods. We also present a new dataset, CS-MAD[1], containing *bona fide* presentations and custom-silicone mask PAs, captured using two low-cost cameras – the Intel Realsense SR300 (color and NIR images) and the Seek Thermal Compact Pro (thermal images).

Our experiments show that the vulnerability of all three CNN-FR methods is at least 10 times higher than the corresponding FMR. In other words, each FR system in this study is highly effective at detecting ZEI attacks, but is less than 10 times as effective in detecting PAs based on custom silicone masks. FR systems with IAPMR to FMR ratio > 10 are graded as highly vulnerable [3].

We also propose a simple but effective mask-PAD method based on thermal imagery. In future work we will develop more accurate mask-PA countermeasures based on multi-spectral data, using a larger set of PAs. This will also lead to experiments using fused (FR+PAD) systems.

Source code for all our experiments is made freely available for research purposes[2].

Acknowledgements

We take this opportunity to thank Mr. Tom Lauten of Nimba Creations (UK) for his meticulous efforts in creating the custom silicone masks. This work has been supported by the European H2020-ICT project TeSLA (grant agreement no. 688520), the Norwegian Research Council project on Secure Access Control over Wide Area Networks (SWAN, grant no. IKTPLUSS 248030/O70), and by the Swiss Center for Biometrics Research and Testing.

[2] `gitlab.idiap.ch/bob/bob.paper.btas2018_siliconemask`

References

[1] A. Agarwal et al. Face Anti-Spoofing using Haralick Features. In *Proc. IEEE Intl. Conf. BTAS*, Niagara Falls, NY, 2016.

[2] A. Agarwal et al. Face Presentation Attack with Latex Masks in Multispectral Videos. In *2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 275–283, July 2017.

[3] Z. Akhtar et al. Biometrics: In Search of Identity and Security (Q & A). *IEEE MultiMedia*, PP, 2017.

[4] C. Atanasoaei. *Multivariate Boosting with Look-up Tables for Face Processing*. PhD thesis, EPFL, 2012.

[5] S. Bhattacharjee and S. Marcel. What You Can't See Can Help You – Extended Range Imaging for 3D-Mask Presentation Attacks. In *Proc. Intl. Conf. of Biometrics Special Interest Group (BIOSIG)*, pages 1 – 8, 2017.

[6] N. M. Duc and B. Q. Minh. Your Face Is Not Your Password Face Authentication Bypassing Lenovo–asus–toshiba. *Black Hat Briefings*, 2009.

[7] N. Erdogmus and S. Marcel. Spoofing in 2D Face Recognition With 3D Masks and Anti-spoofing With Kinect. In *Proc. IEEE Intl. Conf. BTAS*, Washington D.C., 2013.

[8] G. B. Huang et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, Univ. Massachusetts, Amherst, October 2007.

[9] S. Liu et al. A 3D Mask Face Anti-spoofing Database with Real World Variations. In *Proc. IEEE Conf. on Comp. Vision and Patt. Rec. Workshop (CVPRW)*, Las Vegas, 2016.

[10] I. Manjani et al. Detecting Silicone Mask-Based Presentation Attack via Deep Dictionary Learning. *IEEE Trans. Information Forensics and Security*, 12(7):1713 – 1723, 2017.

[11] S. Marcel et al., editors. *Handbook of Biometric Anti-Spoofing*. Springer-Verlag, 2014.

[12] A. Mohammadi et al. Deeply Vulnerable – A Study of The Robustness of Face Recognition Methods to Presentation Attacks. *IET Biometrics*, 2017. Accepted on 29-Sep-2017.

[13] O. M. Parkhi et al. Deep Face Recognition. In *British Machine Vision Conference*, volume 1, pages 41.1 – 41.12. BMVA Press, 09 2015.

[14] R. Raghavendra et al. On the Vulnerability of Extended Multispectral Face Recognition Systems Towards Presentation Attacks. In *Proc. IEEE Intl. Conf. on Identity, Security and Behavior Analysis (ISBA)*, New Delhi, 2017.

[15] R. Ramachandra and C. Busch. Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey. *ACM Computing Surveys*, 50(1), 2017.

[16] S. S. Liu et al. A 3D Mask Face Anti-Spoofing Database with Real World Variations. In *2016 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1551–1557, June 2016.

[17] D. Sandberg. facenet: Face Recognition Using Tensorflow. `https://github.com/davidsandberg/facenet`. Accessed: 2017-08-01.

[18] F. Schroff et al. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 815 – 823, 2015. 00297.

[19] H. Steiner et al. Design of an Active Multispectral SWIR Camera System for Skin Detection and Face Verification. *Journal of Sensors*, (1):1 – 8, 2016. Article ID 9682453, Special Issue on Multispectral, Hyperspectral, and Polarimetric Imaging Technology.

[20] C. Szegedy et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv preprint arXiv:1602.07261v1*, 2016.

[21] Y. Taigman et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *IEEE Conf. On Computer Vision and Pattern Recognition (CVPR)*, 2014.

[22] M. Uřičář et al. Detector of Facial Landmarks Learned by the Structured Output SVM. In *Proc. 7th Intl. Conf. on Computer Vision Theory and Applications (VISAPP)*, volume 1, pages 547 – 556, February 2012.

[23] X. Wu et al. A Light CNN for Deep Face Representation with Noisy Labels. *arXiv preprint arXiv:1511.02683*, 2015.