Can personalised hygienic masks be used to attack face recognition systems?

Alain Komaty akomaty@idiap.ch
Vedrana Krivokuća Hahn vkrivokuca@idiap.ch
Christophe Ecabert cecabert@idiap.ch
Sébastien Marcel marcel@idiap.ch
Idiap Research Institute (Martigny, Switzerland)

Abstract

The proliferation of automated face recognition (FR) necessitates increasingly accurate person identification. The COVID-19 pandemic has exposed the limitations of FR systems when presented with faces occluded by hygienic masks. However, the security risks of personalised hygienic mask attacks, whereby an attacker wears the mask on which the bottom part of an enrolled user's face is printed, have not yet been studied. To address this research gap, we introduce a novel face dataset consisting of smartphone-recorded videos of real (bona-fide) faces and personalised hygienic mask attacks. We also analyse the vulnerability of two state-of-the-art FR systems to this type of attack, using our dataset. Our results indicate that personalised hygienic mask attacks have the potential to compromise system security, particularly for FR systems that are tuned towards optimising user convenience. These findings underscore the importance of developing suitable Presentation Attack Detection (PAD) algorithms. Our dataset will help researchers and practitioners work towards this goal, thereby enhancing the security and reliability of FR systems.

1. Introduction

Although automated face recognition (FR) technologies have been investigated since the 1960s, the past decade has seen unprecedented gains in the recognition accuracy thanks to the adoption of deep neural networks. For this reason, FR has gained significant attention as a reliable means of human identity management, with applications ranging from verification in personal devices (e.g., smartphones) to large-scale identification (e.g., surveillance). Despite considerable progress in improving the accuracy of these technologies, FR systems continue to struggle against many difficult challenges, such as the ability to recognize occluded faces. A recent example is the difficulty of recognising faces covered by hygienic masks, studied in [1, 2]. However, the effect of *personalised* hygienic masks on FR systems has thus far been overlooked.

While the primary purpose of hygienic masks is to protect individuals from viruses such as COVID-19, they have also become a fashion statement and a means of personal expression. As such, *personalised* masks have gained popularity in recent years, where people customise their hygienic masks with various designs, patterns, and images. Some individuals have even taken mask personalisation a step further, by printing the *bottom half of their face* onto their mask [3]. This trend has gained attention in the media and has raised concerns about its potential impact on FR technologies, particularly in security-sensitive areas such as airports, banks, and other high-security facilities [4, 5].

This paper studies the vulnerability of FR systems to personalised hygienic mask attacks. This type of attack involves printing the bottom part of an enrolled user's face onto a hygienic mask, then placing the mask onto an attacker's face to try to fool the FR system into accepting the attacker as the genuine user they are attempting to impersonate. Our two main contributions are the following:

1. A new public dataset [1] of face videos, including bona-fide (i.e., normal, non-attack) videos, videos of personalised hygienic mask attacks, and videos of printed photograph and phone replay attacks for comparison (e.g., see Figure 1). All face videos were captured using the 'selfie' cameras of five different smartphones.

[1] https://www.idiap.ch/en/dataset/phymatt

Bona-fide Session 1 | Bona-fide Session 2 | Hygienic mask attack | Print attack | Replay attack

Figure 1. Video frames of bona-fide recordings for the same data subject, captured during two data acquisition sessions, and the three types of attacks present in our new dataset (captured using iPhone 12's 'selfie' camera).

2. An evaluation of how easy it is to fool FR systems into accepting an attacker as the genuine user they are attempting to impersonate via a personalised hygienic mask attack. We present an assessment of the vulnerability of two state-of-the-art FR systems to this type of attack, and we compare this to the systems' vulnerability against printed photograph and phone replay attacks. The code for this analysis is publicly available [2] to enable other researchers to reproduce our results.

As far as we are aware, neither a study nor a public dataset of personalised hygienic mask attacks exists in the literature. Our aforementioned contributions aim to fill these gaps. This work will provide insights into the dangers of personalised hygienic mask attacks, in terms of their ability to fool FR systems and thereby jeopardise their security.

The remainder of this paper is structured as follows. Section 2 considers related work from the literature, Section 3 describes our new (public) face dataset, Section 4 analyses the vulnerability of two state-of-the-art FR systems to personalised hygienic mask attacks, and Section 5 presents concluding remarks and plans for future work.

2. Related work

This section considers face presentation attack (PA) datasets and literature on FR system vulnerability analysis.

Face PA datasets: As the face PAD domain has gained increasingly more attention over the past 10 years, several datasets have been acquired and shared in the scientific literature. Our analysis of the literature allowed us to identify at least 23 relevant datasets, grouped by PA in Figure 2.

To the best of our knowledge, hygienic masks are only present in CRMA [8] (*not publicly available*), where the bona-fide data subjects were recorded both with and without masks; however, the hygienic masks themselves were not considered as a PA. Instead, this dataset contains print and replay attacks, which were generated for the data subjects with and without hygienic masks. This allowed the authors to examine the performance of several advanced PAD algorithms on print and replay attacks of masked faces under various experimental conditions, as well as to conduct a vulnerability analysis of FR systems to such attacks. Partial face attacks (top of the face printed on paper and placed over the attacker's face) were presented in Rose-Youtu [22], but no specific studies on the effect of this occlusion have been reported. Other face occlusions, such as partial eyes, partial faces and paper crafts were presented in SiW-M-v2 [25], WMCA [26] and HQ-WMCA [28].

Most partial face attacks (printed half-faces, printed eyes, etc.), as well as the other attacks in the existing face

[2] https://gitlab.idiap.ch/bob/bob.paper.ijcb2023_vuln_analysis_hyg_mask_attack

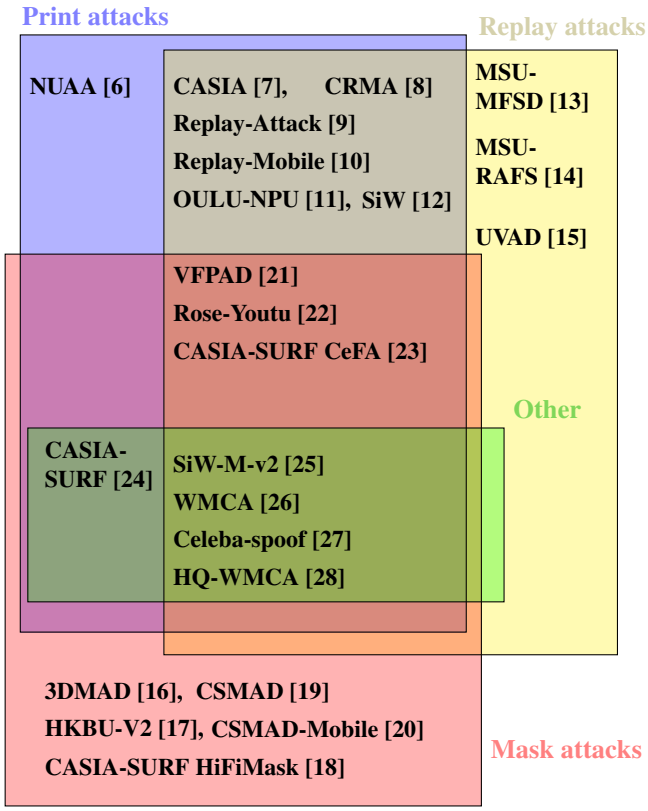


Figure 2. List of PA datasets, grouped by PA type.

datasets (e.g., print, replay, 3D masks) are unlikely to pose a real security threat in public places such as airports, since the attack would usually be evident to human observers. On the other hand, personalised hygienic mask attacks are more subtle and thus have the potential to go unnoticed, particularly if hygienic masks are allowed to be worn when passing through the FR system. For this reason, the focus of this work is on evaluating the vulnerability of state-of-the-art FR systems to personalised hygienic mask attacks, where the "personalisation" aspect involves printing the bottom part of a different person's face onto the mask. We additionally contribute a *publicly available* dataset of both bona-fide and personalised hygienic mask attack face videos from 70 data subjects, covering a wide range of ages and ethnic backgrounds, with an almost equal male/female split.

Vulnerability analysis: Although PAD algorithms have been tested on partially occluded faces in much of the aforementioned research, vulnerability analysis for state-of-the-art FR systems has been performed only in [8]. In fact, the effect of hygienic masks on FR systems has been an open research problem since the beginning of the COVID-19 pandemic. Although [29] found that FR systems experience a drop in performance when the subjects wear hygienic masks, [8] showed that FR systems are significantly *less*

vulnerable to print and replay PAs when the subject wears a mask compared to when full-face attacks are launched.

However, the vulnerability of FR systems to *personalised* hygienic mask PAs, where the bottom part of an enrolled user's face is printed on the mask, has not yet been investigated. This is the aim of our paper, which will assess the vulnerability of two state-of-the-art FR systems, IResNet100 and IResnet50 [30], to personalised hygienic mask attacks. The results are presented in Section 4.4.

3. New Dataset

We believe that personalised hygienic mask attacks have the potential to pose a significant challenge to FR technologies, in terms of providing a relatively easy avenue for impersonation attacks. So, studying the vulnerability of FR systems to this type of attack is important, but there does not exist a public dataset of such attacks to perform this type of analysis. So, we collected a new face dataset, including personalised hygienic mask attacks, which will be released to the public to encourage further research in this domain.

Our new dataset consists of face videos of both bona-fide (non-attack) captures, as well as personalised hygienic mask attacks. Specifically, the dataset contains 1400 bona-fide videos with variations in pose and background, along with 345 videos of personalised hygienic mask attacks. For comparison, we also generated 1400 videos of printed photograph attacks, and 2800 videos of phone replay attacks.

The face data was obtained from 70 volunteers, each of whom participated in two data capture sessions. The data was acquired using the front (i.e., selfie) cameras of five smartphones: Apple iPhone 12, Apple iPhone 6s, Xiaomi Redmi 6 Pro, Xiaomi Redmi 9A and Samsung Galaxy S9. Each recorded video was 10 seconds long, where for the first 5 seconds the data subject was required to stay still and look at the camera, then for the last 5 seconds the subject was asked to turn their head from one side to the other (such that profile views could be captured). The videos were acquired indoors, under normal office lighting conditions.

Sections 3.1, 3.2, and 3.3 present more details on the capture of the bona-fide face videos, personalised hygienic mask attacks, and print and replay attacks, respectively.

3.1. Bona-fide face captures

Each of our 70 data subjects was asked to participate only in the bona-fide face captures. The volunteers were required to be present during two recording sessions, which on average were separated by about three weeks. The idea was to incorporate some natural (uncontrolled) variability in the acquired face videos. Figure 1 shows an example of two bona-fide video frames captured from the same data subject during the two different recording sessions – some evident natural variabilities include different hairstyles and image backgrounds (due to the use of different recording rooms, or the same room but different locations within that room).

In each recording session, the volunteers were asked to record a video of their own face using the front (i.e., selfie) camera of each of the five smartphones mentioned earlier. The face data was additionally captured while the data subjects wore plain (not personalised) hygienic masks, to simulate the scenario where face recognition might need to be performed on a masked face (e.g., during a pandemic like COVID-19). Figure 3 shows examples of video frames from the same subject, both without and with a plain hygienic mask, acquired using the five different smartphones.

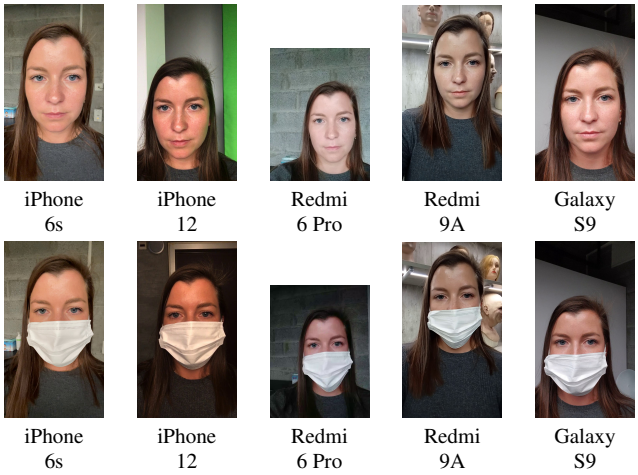


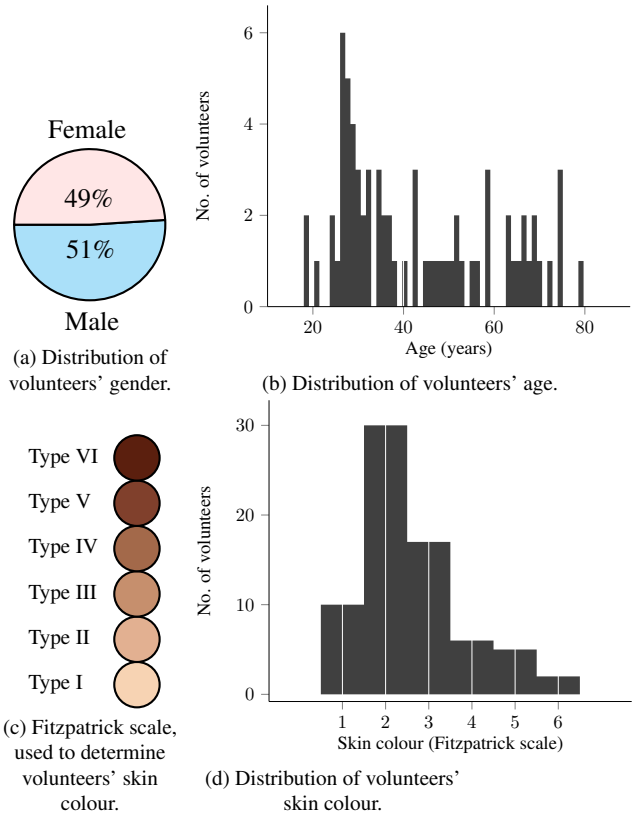
Figure 3. Bona-fide video frames from the same data subject, captured using different smartphones. The top row shows the normal face capture and the bottom row shows the face capture when the subject is wearing a plain (non-personalised) hygienic mask.

In selecting the volunteers for our data collection, we tried our best to have a balanced gender distribution, a wide age range, and a uniform distribution of skin colours. Figure 4 shows the distribution of genders, ages, and skin colours across our 70 data subjects. The skin colours were based on the Fitzpatrick scale (see Figure 4c), and each data subject was consulted on their opinion of their own skin colour.

From Figure 4, it is evident that our dataset consists of an almost perfectly balanced gender split, which is a notable achievement. Concerning the age distribution, we did manage to gather volunteers spanning a wide age range (from about 20-80 years old); however, most subjects were in the 20-30 age range. We also succeeded in collecting volunteers across the whole Fitzpatrick-scale skin colour spectrum; however, the majority of our data subjects were Types II and III, while skin colours on either end of the spectrum (particularly dark) were difficult to acquire.

3.2. Personalised hygienic mask attacks

For each of the 70 data subjects, a personalised hygienic mask was created from one of their best quality video



(a) Distribution of volunteers' gender.

(b) Distribution of volunteers' age.

(c) Fitzpatrick scale, used to determine volunteers' skin colour.

(d) Distribution of volunteers' skin colour.

Figure 4. Distribution of volunteer demographics.

frames. The selected frames (face images) were sent to the Cadeaux Folies [3] company in Switzerland, which generated the corresponding personalised hygienic masks.

To simulate attacks, the personalised hygienic mask for each of 69 data subjects was placed in turn onto the face of a human attacker (who was actually the 70th data subject). Figure 5 shows examples of two subjects (a man and a woman) being impersonated by the same attacker, using each of the five smartphones in turn to capture the attack.

At this stage, only one human attacker was used, so the gender, age, and skin colour of the attacker did not necessarily match those of the person being attacked; however, in a real-life attack, it is normal to expect a mismatch between the demographics of the attacker and the person being impersonated. Nevertheless, in the future we plan to extend this investigation to involve multiple human attackers with different demographic attributes. Furthermore, most of the hygienic masks exhibit a somewhat orange hue (e.g., see Figure 5), which does not correspond perfectly to the skin colour of the underlying person. So, another idea for an extension to this investigation is to generate the personalised hygienic masks using multiple suppliers (besides Cadeaux Folies), in case the quality of the printed masks differs.

[3] https://www.cadeauxfolies.ch/

Figure 5. Personalised hygienic mask attacks carried out by the same attacker for two different genuine users: a man (top row) and a woman (bottom row). The attacks were captured using five different smartphones.

Having made these observations and suggestions for future work, we must emphasize that our current dataset of personalised hygienic mask attacks is nevertheless novel and provides fertile ground for investigating the vulnerability of FR systems to this type of attack. Our vulnerability investigation is presented in Section 4.

3.3. Print and replay attacks

Our dataset also includes printed photograph and phone replay attacks. The print attacks involved printing the target's face on a matte A4 paper, then showing this paper to each of the five smartphones in turn to record videos of the attack. For the replay attack, different phones were paired such that one of the pair was used to replay the bona-fide videos while the second (attacked) phone recorded the videos using its front camera. Figure 1 shows examples of a print and replay attack for the same data subject.

4. Vulnerability Analysis

This section analyses the vulnerability of two state-of-the-art FR systems to a personalised hygienic mask attack (based on our new dataset, described in Section 3). Section 4.1 presents our chosen FR systems, Sections 4.2 and 4.3 outline the bona-fide and PA evaluation protocols, and Section 4.4 discusses the results of our vulnerability analysis.

4.1. Face recognition systems

Modern FR systems rely on trained neural networks, CNNs or Vision Transformers, to extract representations of face images. These networks are trained to generate discriminative face features with respect to identity. The resulting feature vectors are in turn used as templates for the input face images. Then, to determine whether or not two

face images match (i.e., originate from the same person), the corresponding face templates are compared using distance or similarity measures.

The vulnerability analysis carried out in the work presented in this paper, uses an off-the-shelf neural network trained with an additive angular margin-based loss function, namely ArcFace [31]. Two backbones [4], IResNet100 and IResNet50, derived from a modified ResNet [30] architecture, have been selected. Both networks were trained on the MS1MV2 dataset, a semi-automatically refined version of the MS-Celeb-1M dataset [32]. Models trained with an additive angular-margin framework generate highly discriminative feature vectors for face recognition, thus reaching state-of-the-art performance on various datasets. For example, IResNet100 achieves 99.83% verification accuracy on the Labeled Faces in the Wild (LFW) [33] dataset and 98.02% on the YouTube Faces (YTF) [34] dataset. Consequently, IResNet100 seems to be a good candidate for analysing the vulnerability of state-of-the-art FR systems to PAs, such as personalised hygienic mask attacks (the focus of this paper). The trained IResNet100 model attained a verification accuracy of 98.5% on our new face dataset. We include the more lightweight architecture, IResNet50, in our investigation as well. On the same dataset, the IResNet50 model achieved a verification accuracy of 97.6%.

4.2. Bona-fide evaluation protocols

This section describes the two protocols used to evaluate the accuracy of the FR systems outlined in Section 4.1. For both protocols, face images (video frames) from the first recording session were used as references, and images from the second session were used as probes. The images were 20 equally-spaced frames extracted from the corresponding face videos. Thus, the following two protocols were used:

Protocol 1 (p1): Both references and probes were images of full faces, without hygienic masks.

Protocol 2 (p2): The probes were images of faces covered by plain (not personalised) hygienic masks. The references were full face images, as in p1.

The similarities between reference and probe images were calculated in terms of cosine distance [5]. Based on the *zero-effort impostor* (probe ID does *not* match reference ID) and *genuine* (probe ID *matches* reference ID) scores, the accept/reject decision threshold was set. Using this threshold, the Impostor Attack Presentation Match Rate (IAPMR) was calculated to evaluate the vulnerability of the face recognition systems to personalised hygienic mask attacks. The evaluation protocol for these attacks (as well as print and replay attacks) is explained in Section 4.3.

4.3. Presentation attack evaluation protocols

For each attack video, we extracted 20 equally-spaced frames, all of which were used for the PA evaluations. For both Genuine and Zero-Effort Impostor (ZEI) references and probes, we used bona-fide images (frames) of full faces, and for PAs we used the videos from all five smartphones. The whole process is illustrated in Figure 6.

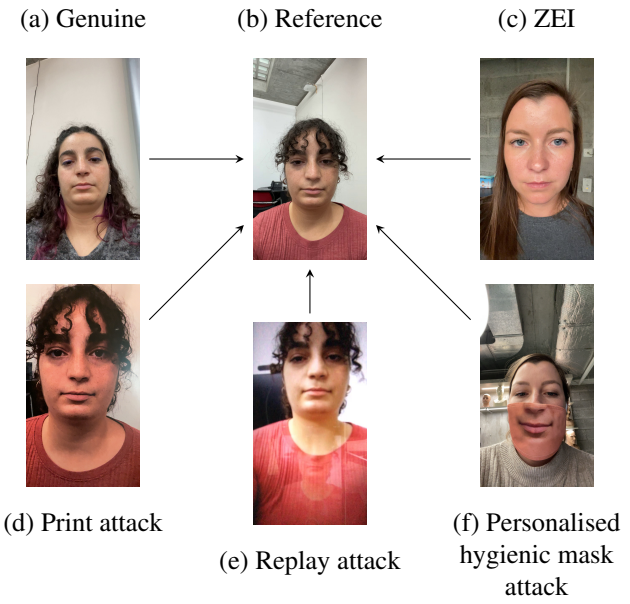


Figure 6. A reference is enrolled using the full bona-fide face (b). This reference is then compared to genuine (a) and ZEI (c) probes. It is also compared to three PAs: print attack (d), replay attack (e) and personalised hygienic mask attack (f).

4.4. Results and discussion

Table 1 shows the performance of our two FR systems, IResNet100 and IResNet50, on our new face dataset, when the accept/reject decision threshold is set at FMR = 0.1% on the FRGC dataset [6]. A different dataset was used to set the threshold to simulate the scenario where the FR systems are tuned prior to deployment (in this case, the deployment scenario is represented by our new dataset). The recognition accuracy is reported in terms of the False Match Rate (FMR) and False Non-Match Rate (FNMR), which were computed using the bona-fide protocols p1 and p2 (see Section 4.2). The same threshold was used for p1 and p2, because the FRGC dataset does not contain images of faces covered by hygienic masks, so we could not tune the threshold separately for p2.

[4] https://github.com/deepinsight/insightface

[5] The distances were multiplied by -1 to turn them into similarity scores.

[6] https://www.nist.gov/programs-projects/face-recognition-grand-challenge-frgc

	IResNet100		IResNet50	
	p1	p2	p1	p2
FMR	0.0%	0.0%	0.0%	0.0%
FNMR	2.4%	27.0%	5.4%	47.1%
IAPMR	1.4%		1.3%	

Table 1. False Match Rate (FMR), False Non-Match Rate (FNMR), and Impostor Attack Presentation Match Rate (IAPMR) for the IResNet100 and IResNet50 FR systems, evaluated on our new dataset using a decision threshold set at FMR = 0.1% on the FRGC dataset. The FMR and FNMR were evaluated using the bona-fide evaluation protocols p1 and p2, while IAPMR was evaluated using the PA protocol described in Section 4.3.

Table 1 also shows the Impostor Attack Presentation Match Rate (IAPMR) for each FR system. The IAPMR quantifies the vulnerability of the system to a personalised hygienic mask attack, in terms of the percentage of these attacks that are falsely "accepted" as genuine users of the FR system. So, the higher the IAPMR, the more vulnerable the system is to the attack. Ideally, the IAPMR should be 0, which would indicate that the FR system never mistakes the attacker for the enrolled user whom they are impersonating. Note that the IAPMR for each FR system in Table 1 was based on the same decision threshold used to compute the FMR and FNMR. We report a single IAPMR value per FR system, since the same threshold was used for p1 and p2, and the attacks are independent of these protocols.

The main observation from Table 1 is that the IAPMR for both IResNet100 and IResNet50 is quite low, which suggests that neither FR system is particularly vulnerable to a personalised hygienic mask attack. However, these results are based on a single (system-specific) decision threshold, which was tuned on a different dataset (FRGC) to that used for the evaluation (our new face dataset). While this threshold seems excellent for maintaining a low (zero) FMR for both FR systems, the FNMR may be considered unacceptably high, particularly the FNMRs of 27.0% and 47.1% obtained under protocol p2. These results suggest that, if users of the IResNet100 and IResNet50 FR systems were to be enrolled using full face images, but then the subjects wore hygienic masks during the recognition stage, about one-third of the users would be rejected as impostors in the IResNet100 system and close to half would be rejected in the IResNet50 system. These high FNMRs for the p2 evaluation protocol may be attributed to the fact that FRGC, on which the accept/reject decision threshold was set, does not contain images of faces covered by hygienic masks. So, this threshold is not appropriate for the scenario where masked faces are presented to the FR systems during the recognition stage. This finding prompted us to experiment with setting the decision threshold on our own dataset, which includes both full-face images and faces covered by hygienic masks, instead of on the FRGC dataset.

Table 2 shows the IAPMR of our two FR systems when the decision threshold was set on our own face dataset. As for the evaluation in Table 1, the threshold was once again set at FMR = 0.1% (which is typical when evaluating FR systems in the literature). This time, however, the threshold was not only system-specific, as in Table 1, but also protocol-specific. In other words, the threshold used to compute the IAPMR was set separately for evaluation protocols p1 and p2. This is because our results in Table 1 indicate that the recognition accuracy of the FR systems suffers if a single threshold is used for both the scenario where face recognition is based on full face images (p1) and the scenario where the faces are occluded by a hygienic mask (p2). Note that, although these decision thresholds are required for the IAPMR calculation, the PAs themselves are independent of the p1 and p2 protocols (i.e., the attacks remain the same). Table 2 compares the IAPMR of the IResNet100 and IResNet50 FR systems for personalised hygienic mask attacks (described in Section 3.2), to the IAPMR for the more common printed photograph and phone replay attacks (described in Section 3.3).

Presentation attack (PA)	IResNet100		IResNet50	
	p1	p2	p1	p2
Printed photograph	99.9%	99.1%	99.8%	98.8%
Phone replay	95.5%	95.7%	93.3%	93.6%
Personalised hygienic mask	1.9%	3.3%	1.8%	2.9%

Table 2. Impostor Attack Presentation Match Rate (IAPMR) for different PAs on the IResNet100 and IResNet50 FR systems. The decision threshold used to calculate IAPMR was set at FMR = 0.1% (separately for each system and each evaluation protocol).

From Table 2, it is clear that, for both IResNet100 and IResNet50, the IAPMR of the personalised hygienic mask attack is much lower than the IAPMR corresponding to printed photograph and phone replay attacks. This suggests that the two FR systems are much less vulnerable to personalised hygienic mask attacks than to print or replay attacks. However, a personalised hygienic mask attack may be easier to carry out in a subtler way than printed photograph or replay attacks, since wearing a hygienic mask in public is perfectly acceptable. So, this type of attack should still be defended against in practice.

Thus far, our evaluation of the vulnerability of FR systems to personalised hygienic mask attacks has been based on decision thresholds set at FMR = 0.1%. In this scenario, which favours system security over user convenience, we found that both the IResNet100 and IResNet50 FR systems are unlikely to be fooled by this type of attack (based on the low IAPMR values in Tables 1 and 2). However, we were interested in finding out whether the vulnerability of our FR systems to personalised hygienic mask attacks would increase if the systems were set to operate at different decision thresholds. Table 3 shows the IAPMR evaluated at three different thresholds (once again set on our own face dataset), to provide insight into the vulnerability of our IResNet100

and IResNet50 FR systems to personalised hygienic mask attacks when the FR systems are tuned to operate in higher-system-security versus higher-user-convenience scenarios.

There are several important observations from Table 3. Firstly, it is clear that the vulnerability of both FR systems to personalised hygienic mask attacks increases when the decision threshold is set to favour user convenience over system security. For example, under evaluation protocol p1, when the threshold is set at FMR = 0.1% (higher security) the IAPMR for both IResNet100 and IResNet50 is less than 2%. On the other hand, when the threshold is set at FNMR = 1% (higher convenience), the IAPMR jumps to 44.2% for IResNet100 and 66.8% for IResNet50. These findings indicate that personalised hygienic mask attacks may pose a serious threat to FR systems that are tuned towards optimising user convenience. As for systems that are tuned to value security and convenience equally, for example by setting the threshold at the Equal Error Rate (EER), our results show an IAPMR of 12.9% for IResNet100 and 17.4% for IResNet50. These figures suggest that, in this type of general face recognition scenario, where there is an equal trade-off between security and convenience, personalised hygienic mask attacks present a non-negligible threat to the FR systems. For example, an IAPMR of 12.9% (17.4%) implies that about 13 (17) out of every 100 attacks would be falsely accepted as a real authentication attempt by a genuine user of the FR system.

Another important observation from Table 3 is that, for all three threshold settings, both IResNet100 and IResNet50 are significantly more vulnerable to personalised hygienic mask attacks under evaluation protocol p2 than p1 [7]. This is especially the case when the thresholds are set to favour user convenience (e.g., at FNMR = 1%) or to strike an equal compromise between system security and user convenience (at EER). For example, when the threshold is set at EER using evaluation protocol p2, for both FR systems the IAPMR is close to 50%, suggesting that half of all personalised hygienic mask attacks would succeed in fooling the FR systems. When the threshold is set at FNMR = 1%, the vulnerability jumps to over 90%. Since the attacks in p1 and p2 are the same, this difference in the IAPMRs between the two protocols can be attributed mainly to the worse recognition accuracy of the two FR systems under p2, which is when the probe faces are occluded by a (plain) hygienic mask.

For example, Figure 7 illustrates the IAPMRs for IResNet100 at the three different decision thresholds, for evaluation protocols p1 and p2. The thresholds and corresponding IAPMR values are plotted along with the score distributions of the bona-fide "genuine" and "zero-effort impostor (ZEI)"

[7]Recall that p2 represents the scenario where users enroll into the FR system using full face images, but then they attempt to be recognised when wearing a plain (non-personalised) hygienic mask. In contrast, for p1, both enrollment and recognition are performed using full face images.

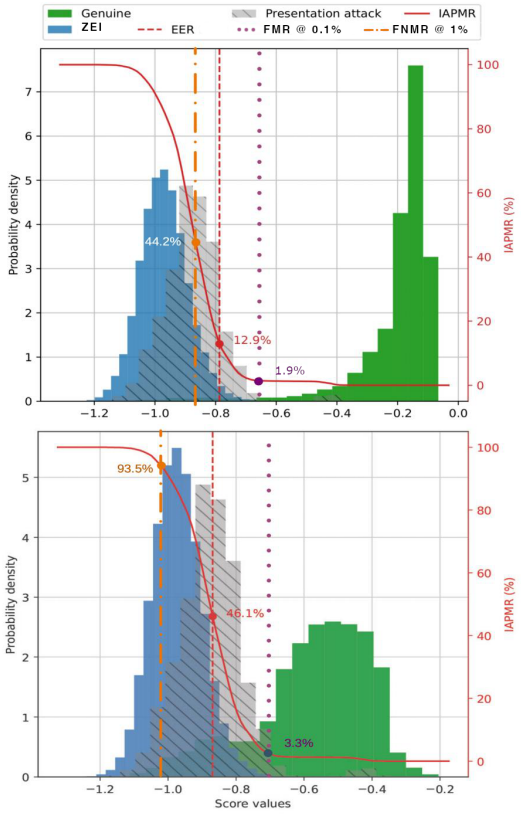


Figure 7. Similarity score distributions and IAPMR values for the IResNet100 FR system, for evaluation protocols p1 and p2. The further the "presentation attack" distribution is from the "genuine" distribution, the lower the IAPMR and the less vulnerable the FR system is to a personalised hygienic mask attack.

face presentations, as well as the "presentation attack" score distribution corresponding to personalised hygienic mask attacks. Note that, for an IAPMR of 0, we would expect to see full separation between the "genuine" and "presentation attack" score distributions. A visual comparison of the p1 and p2 distributions in Figure 7 shows that there is a much greater separation between the "ZEI" and "genuine" distributions in p1 than in p2. In fact, the "genuine" distribution for p2 is much more spread out than for p1. Consequently, the "presentation attack" distribution overlaps much more with the "genuine" distribution in p2 than in p1, resulting in a higher IAPMR. This is because a higher IAPMR indicates a greater likelihood of mistaking a PA for a genuine face, which increases as the threshold moves from FMR = 0.1% to FNMR = 1%.

A final observation from Table 3 is that, in general, the IResNet50 FR system seems slightly more vulnerable than IResNet100 to personalised hygienic mask attacks. To fully understand the reason for this, as part of our future work we intend to perform an in-depth investigation into the vulner-

	IResNet100			IResNet50		
	@ FMR = 0.1%	**@ FNMR = 1%**	**@ EER**	**@ FMR = 0.1%**	**@ FNMR = 1%**	**@ EER**
p1	1.9%	44.2%	12.9%	1.8%	66.8%	17.4%
p2	3.3%	93.5%	46.1%	2.9%	97.5%	50.3%

Table 3. IAPMR for personalised hygienic mask attacks on the IResNet100 and IResNet50 FR systems. Results are reported for the p1 and p2 evaluation protocols, for which the attacks are the same, but where the bona-fide probes are full face images in p1 and faces occluded by a plain hygienic mask in p2. The following accept/reject decision thresholds were used to calculate IAPMR: @ FMR = 0.1% (representing higher system security), @ FNMR = 1% (representing higher user convenience), and @ EER (representing equal security/convenience).

abilities of several different types of FR systems.

Overall, our vulnerability analysis shows that state-of-the-art FR systems, such as IResNet100 and IResNet50, are vulnerable to personalised hygienic mask attacks to some extent, depending on the threshold that is used for accept/reject decisions in the underlying FR system. The more this threshold is tuned towards improving user convenience (i.e., reducing the FNMR), the more vulnerable the FR system becomes. This is particularly the case for FR systems that are set up to perform face recognition on faces occluded by (standard, non-personalised) hygienic masks. Our results show that, in this scenario, close to 50% of the attacks would be accepted as genuine authentication attempts when the decision threshold is set at the EER, and over 90% of the attacks would fool the system when the threshold is set at an FNMR of 1%. So, our findings indicate that it is important to consider ways of mitigating this type of attack in practice, especially considering the ease with which such attacks can be launched by practically anyone.

5. Conclusion

This paper made two important contributions towards improving the security of FR systems. Firstly, we presented a new, publicly available dataset of face videos captured using the 'selfie' cameras of five smartphones. The videos were acquired from 70 data subjects, spanning a wide range of ages and skin colours, with an almost equal male/female split. Both full and partially masked (via a plain hygienic mask) faces were captured. The dataset additionally contains videos of three presentation attacks: printed photograph, phone replay, and personalised hygienic mask. To the best of our knowledge, ours is the first public dataset to include personalised hygienic mask attacks. This attack involves printing the bottom part of a genuine (enrolled) user's face image onto a hygienic mask, then placing the mask on an attacker's face. Therefore, our new dataset will enable researchers to study the vulnerability of FR systems to these types of attacks and to test the efficacy of their Presentation Attack Detection (PAD) algorithms, as well as to evaluate the accuracy of FR systems in general.

Our second main contribution was an analysis of the vulnerability of two state-of-the-art FR systems, IResNet100 and IResNet50, to personalised hygienic mask attacks. We found that, although these systems are significantly more likely to be fooled by a printed photo or phone replay attack, they are still vulnerable to personalised hygienic mask attacks to some extent, depending on the decision threshold of the underlying FR system. In particular, as the threshold is tuned more towards optimising user convenience (i.e., decreasing FNMR), the system becomes more vulnerable to the attack. For a threshold set at the EER, to balance user convenience and system security, our experiments showed an IAPMR of 12.9% (17.4%) for IResNet100 (IResNet50). The IAPMR was found to increase significantly, to approximately 50%, when the FR systems were tuned to perform recognition on faces partially occluded by plain (standard) hygienic masks. Therefore, our findings indicate that personalised hygienic mask attacks may present a threat to FR systems in practice, particularly considering the ease with which such attacks can be launched by the average person.

Our plans for future work include extending our dataset and vulnerability investigation to include a larger number of attackers, different manufacturers for the personalised hygienic masks, a wider range of lighting conditions for the face data acquisition, and several different types of FR systems. We also intend to explore suitable PAD algorithms for thwarting this type of attack.

6. Acknowledgments

We would like to thank the EU project SOTERIA for funding this work, our project partner IDnow and Idiap colleague Philip Abbet for creating the mobile data capture application, our Idiap colleagues Yannick Dayer and Laurent Colbois for their help in developing the software package for our experiments, and our former colleague Karine Vaucher for performing the bulk of the data collection.

References

[1] N. Damer, F. Boutros, M. Süßmilch, M. Fang, F. Kirchbuchner, and A. Kuijper, "Masked face recognition: Human versus machine," *IET Biometrics*, vol. 11, no. 5, pp. 512–528, 2022.

[2] A. V. L.R. Rabbitt, Y.B. Sirotin, "Human-Algorithm Teaming: An investigation on masks and algorithm accuracy on human decisions," 2022. In the Unfamiliar Face Identification Group (UFIG), virtual meeting.

[3] N. Morris, "You can get your own face printed on a face mask – and it is creepy." https://metro.co.uk/2020/05/27/can-get-face-printed-face-masks-creepy-12764385, May 2020.

[4] C. Smith, "This company will print your face on a mask so you can unlock your phone." https://nypost.com/2020/02/18/this-company-will-print-your-face-on-a-mask-so-you-can-unlock-your-phone, Feb 2020.

[5] S. Liberatore, "Bizarre surgical mask printed with your face protects from 'viral epidemics' and claims to unlock your smartphone." https://www.dailymail.co.uk/sciencetech/article-8020889/Surgical-mask-printed-face-protects-viral-epidemics-claims-unlock-phone.html, Feb 2020.

[6] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *European Conference on Computer Vision*, pp. 504–517, Springer, 2010.

[7] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *2012 5th IAPR International Conference on Biometrics (ICB)*, pp. 26–31, 2012.

[8] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper, "Real masks and spoof faces: On the masked face presentation attack detection," *Pattern Recognition*, vol. 123, p. 108398, 2022.

[9] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)*, pp. 1–7, IEEE, 2012.

[10] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–7, IEEE, 2016.

[11] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A Mobile Face Presentation Attack Database with Real-World Variations," in *12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pp. 612–618, 2017.

[12] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 389–398, 2018.

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

[14] 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 *2015 International Conference on Biometrics (ICB)*, pp. 98–105, IEEE, 2015.

[15] A. Pinto, W. Robson Schwartz, H. Pedrini, and A. De Rezende Rocha, "Using Visual Rhythms for Detecting Video-Based Facial Spoof Attacks," *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 1025–1038, May 2015.

[16] E. Nesli and S. Marcel, "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect," in *IEEE 6th International Conference on Biometrics: Theory, Applications and Systems (BTAS'13)*, pp. 1–8, 2013.

[17] S. Liu, B. Yang, P. C. Yuen, and G. Zhao, "A 3D mask face anti-spoofing database with real world variations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 100–106, 2016.

[18] A. Liu, C. Zhao, Z. Yu, J. Wan, A. Su, X. Liu, Z. Tan, S. Escalera, J. Xing, Y. Liang, G. Guo, Z. Lei, S. Z. Li, and D. Zhang, "Contrastive Context-Aware Learning for 3D High-Fidelity Mask Face Presentation Attack Detection," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2497–2507, 2022.

[19] S. Bhattacharjee, A. Mohammadi, and S. Marcel, "Spoofing deep face recognition with custom silicone masks," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–7, IEEE, 2018.

[20] R. Ramachandra, S. Venkatesh, K. B. Raja, S. Bhattacharjee, P. Wasnik, S. Marcel, and C. Busch, "Custom silicone face masks: Vulnerability of commercial face recognition systems & presentation attack detection," in *2019 7th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6, IEEE, 2019.

[21] K. Kotwal, S. Bhattacharjee, P. Abbet, Z. Mostaani, H. Wei, X. Wenkang, Z. Yaxi, and S. Marcel, "Domain-Specific Adaptation of CNN for Detecting Face Presentation Attacks in NIR," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 135–147, 2022.

[22] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1794–1809, 2018.

[23] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, and S. Z. Li, "CASIA-SURF CeFA: A Benchmark for Multi-modal Cross-ethnicity Face Anti-spoofing," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1179–1187, 2021.

[24] S. Zhang, A. Liu, J. Wan, Y. Liang, G. Guo, S. Escalera, H. J. Escalante, and S. Z. Li, "CASIA-SURF: A Large-Scale Multi-Modal Benchmark for Face Anti-Spoofing," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 2, pp. 182–193, 2020.

[25] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4680–4689, 2019.

[26] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 42–55, 2019.

[27] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, "CelebA-Spoof: Large-Scale Face Anti-Spoofing Dataset with Rich Annotations," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, pp. 70–85, Springer, 2020.

[28] G. Heusch, A. George, D. Geissbühler, Z. Mostaani, and S. Marcel, "Deep Models and Shortwave Infrared Information to Detect Face Presentation Attacks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2020.

[29] M. Ngan, P. Grother, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face recognition accuracy with face masks using post-COVID-19 algorithms," 2020-11-30 2020.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[31] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Long Beach, CA, USA), pp. 4685–4694, IEEE, June 2019.

[32] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), vol. 9907, pp. 87–102, Cham: Springer International Publishing, 2016.

[33] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.

[34] L. Wolf, T. Hassner, and I. Maoz, "Face Recognition in Unconstrained Videos with Matched Background Similarity," in *CVPR 2011*, (Colorado Springs, CO, USA), pp. 529–534, IEEE, June 2011.