Fun Selfie Filters in Face Recognition: Impact Assessment and Removal

Cristian Botezatu, Mathias Ibsen, *Graduate Student Member, IEEE*, Christian Rathgeb, and Christoph Busch, *Senior Member, IEEE*

***Abstract*—This work investigates the impact of fun selfie filters, which are frequently used to modify selfies, on face recognition systems. Based on a qualitative assessment and classification of freely available mobile applications, ten relevant fun selfie filters are selected to create a database. To this end, the selected filters are automatically applied to face images of public face image databases. Different state-of-the-art methods are used to evaluate the influence of fun selfie filters on the performance of face detection using dlib, RetinaFace, and a COTS method, sample quality estimated by FaceQNet and MagFace, and recognition accuracy employing ArcFace and a COTS algorithm. The obtained results indicate that selfie filters negatively affect face recognition modules, especially if fun selfie filters cover a large region of the face, where the mouth, nose, and eyes are covered. To mitigate such unwanted effects, a GAN-based selfie filter removal algorithm is proposed which consists of a segmentation module, a perceptual network, and a generation module. In a cross-database experiment the application of the presented selfie filter removal technique has shown to significantly improve the biometric performance of the underlying face recognition systems.**

***Index Terms*—Biometrics, face recognition, selfie filter, face occlusion, generative adversarial network, inpainting.**

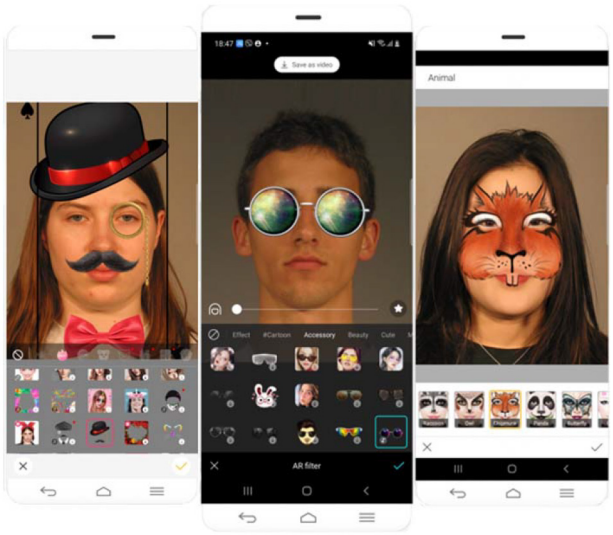


Fig. 1. Example of fun selfie filters [11], [12], [13].

I. Introduction

In THE recent past, the use of deep convolutional neural networks has achieved remarkable improvements in face recognition (FR) accuracy, surpassing human-level performance [1], [2]. Due to these breakthrough advances FR technologies have become an essential tool for identity management systems and forensic investigations worldwide. In the latter application scenario, public content plays an important role, especially facial images from social media [3], [4], [5]. However, before sharing their face images on social media platforms, e.g., Facebook or Instagram, users frequently edit them to achieve a desired impact. Common editing tools include beautification filters which may apply significant alterations to the facial shape and texture, e.g., by enlarging the eyes or smoothing of the skin. Furthermore, so-called *fun selfie filters* (FSF) are frequently applied by users to add to the amusement. FSF applications automatically edit facial images in a funny and conspicuous way. In Fig. 1, it can be seen that FSFs may induce severe alterations and occlusions to face images.

In many forensic investigations, public authorities are frequently confronted with cases where only edited facial images of a subject are available, e.g., in searches for missing persons. This is particularly the case for teenagers which often apply FSFs to their face images [6], [7]. In such cases, an automatic reconstruction of the original face images can facilitate forensic investigations.

In a FR system, FSFs applied to face images are expected to represent a challenge for various processing stages [8], [9]. For instance, a large coverage of the facial region by a FSF may hamper face detection or face sample quality estimation. In addition, biometric comparison scenarios where one of the face images to be compared has been altered with a FSF are expected to be challenging. So far, a first attempt towards evaluating the impact of FSFs on certain facial analysis tasks

Manuscript received 11 February 2022; revised 14 April 2022 and 13 June 2022; accepted 20 June 2022. Date of publication 6 July 2022; date of current version 22 December 2022. This work was supported in part by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE and in part by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Grant under Agreement 860813—TReSPAsS-ETN. This article was recommended for publication by Associate Editor A. O'Toole upon evaluation of the reviewers' comments. *(Corresponding author: Cristian Botezatu.)*

Cristian Botezatu was with the Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany. He is now with the Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, 2815 Gjøvik, Norway (e-mail: cristian.botezatu@ntnu.no).

Mathias Ibsen, Christian Rathgeb, and Christoph Busch are with the Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, 2815 Gjøvik, Norway, and also with the Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany.

Digital Object Identifier 10.1109/TBIOM.2022.3185884

has been made in [10]. However, the impact of FSFs on the performance of state-of-the-art FR systems has not been investigated in a detailed manner yet.

Recently, deep learning techniques have been applied for the purpose of image inpainting. In particular methods based on generative adversarial network (GAN) have shown impressive results for removing facial occlusion [14], e.g., caused by medical masks [15]. In order to perform well, such techniques usually require a large amount of realistic training data containing face image with and without targeted occlusions. To the best of the authors' knowledge, so far the feasibility of GAN-based removal of FSFs has not been investigated in the scientific literature.

In order to assess the impact of FSFs on FR systems and mitigate their potentially negative effects, this work makes the following contributions:

- A qualitative assessment of FSFs, available in mobile application stores, is conducted. Based on this assessment, ten highly relevant FSFs are identified and classified w.r.t. the face image alterations.
- The automated creation of a dataset generated from images of more than 1,000 subjects of the public FRGCv2 [16] and FERET [17] face image databases.[1]
- A comprehensive evaluation of the impact of FSFs on the performance of face detection, face sample quality, and FR, along with a detailed discussion of obtained results.
- A FSF removal algorithm which is created by adapting existing network architectures for segmentation and inpainting [15], [18], along with a detailed evaluation of FR performance before and after the FSF removal.

The remainder of this paper is organised as follows: related works are briefly summarised in Section II. The creation of the FSF face image dataset is described in detail in Section III. Subsequently, the impact of FSFs on different FR sub-systems is evaluated and discussed in Section IV. Section V introduces the novel GAN-based selfie removal which is applied to the created dataset to evaluate to which extent it mitigates the effects of FSFs. Section VI concludes the work.

II. Related Work

Facial occlusions challenge FR systems which are able to cope with the occlusion problem in three main ways [14]:

- *Occlusion Robustness:* apply a patch-based or learning-based feature extraction strategy to describe the feature space that is less affected by facial occlusions (e.g., [19], [20], [21], [22], [23]).
- *Occlusion Awareness:* feature extraction methods which detect occluded facial regions and subsequently only employ visible face parts for FR (e.g., [24], [25], [26], [27]).
- *Occlusion Recovery:* techniques which aim at reconstructing the occluded face parts prior to applying the feature extraction of the FR systems (e.g., [28], [29], [30], [31], [32]).

Many works have reported negative impacts of facial occlusions on FR, e.g., caused by sunglasses or face masks [14].

[1]Due to license restrictions the database created in this work is only made available to visiting researchers.

TABLE I
Categorization of Occlusion Types [14]

Occlusion Scenario	Examples
Facial accessories	eyeglasses, sunglasses, scarves, mask, hat, hair
External occlusions	occluded by hands and random objects
Limited field of view	partial faces
Self-occlusions	non-frontal pose
Extreme illumination	part of face highlighted, other parts in shadow
Artificial occlusions	random black or white rectangles.

The common facial occlusions that challenge the current state-of-the-art FR systems are listed in Tab. I.

Similar to occlusions, strong makeup [33], [34], [35] or even facial tattoos and painting [36] have been shown to negatively influence FR systems, especially in cases where significant parts of a face are covered with tattoos or paint.

During the COVID-19 pandemic, which at the time of writing is still ongoing, many countries made it mandatory to wear facial masks in public and indoor spaces as a preventive measure to help stop the spread of the virus [37]. However, occlusions from facial masks significantly change the operational condition of face recognition systems and, as a result, can make automated face recognition more challenging [38]. Additionally, it has been shown that facial masks can be used to successfully launch attacks against commercial FR systems [39]. Hence, many research activities focus on algorithms to increase FR performance when dealing with masks that cover a large area of an individual's face [40], [41].

Focusing on selfie filters, the impact of beautification filters has firstly been analysed in [42]. In contrast to FSFs, beautification filters aim at beautifying the facial appearance in a discreet way. It was shown that the performance of a FR system might significantly drop in case a beautification filter drastically alters the facial appearance. In more recent works it has been shown that FR systems can be robust to moderate alterations resulting from the use of beautification applications, e.g., in [43], [44], [45]. With respect to FSFs, the Specs on Faces (SoF) dataset was introduced in [10] for the purpose of evaluating various tasks, e.g., face detection and gender prediction, in challenging environmental scenarios. This face database contains face images of 112 subjects to which two FSFs have been applied. However, its small size and the fact that facial images of the said database were mostly captured in a single session makes the SoF dataset less suitable for FR performance evaluations. The large amount of facial occlusion variations, as well as their possible random placement on the face, makes FR under occlusion a yet unresolved issue.

So far, FSFs have not been considered as potential occlusions a FR system has to deal with. Hence, the impact of FSFs on FR performance is assessed both in the direct face comparison but also following an occlusion recovery approach, using inpainting techniques to treat the occluded face as an image repairing problem.

Fig. 2. Fun selfie filter dataset creation workflow.

A. Occlusion Detection and Segmentation

For an effective occlusion removal, it is important to accurately detect and segment the occluded facial regions. In the past years, various deep learning-based object detection and segmentation methods have been shown to obtain outstanding performance [46], [47].

Regions with convolutional neural networks (R-CNNs) [48], Fast R-CNNs [49], or Faster R-CNNs [50] are well-known for their state-of-the-art object detection performance. These approaches use selective search algorithms to extract regions from an image, feeding them to a CNN to produce a feature vector for each processed region. Subsequently, machine learning-based classifiers, e.g., support vector machines (SVMs), analyse the features extracted from each candidate region, to determine the presence of the object. Despite their competitive detection performance, such approaches are computationally expensive [51].

While R-CNNs have also been used for semantic image segmentation tasks [48], Long *et al.* showed in [52] an improved performance for segmentation tasks using a fully convolution network. However, the need for more precise segmentation that works for relatively few training images led to the development of U-Net [53], being one of the most used end-to-end FCNs in image segmentation. The U-Net encoder uses a series of convolutions with max pooling layers, while the decoder uses transformed convolutions to upsample the encoded information. The encoder and decoder feature maps are concatenated to better learn the contextual information. For an accurate FSF segmentation, the current work adopts the idea presented in [15], using U-Net architecture supplemented with a squeeze and excitation (SE) [54] block at the output of the first three layers of the encoder.

B. Occlusion Removal

Deep learning-based algorithms have been effectively used to reconstruct occluded facial regions, e.g., [30]. On the other hand, inpainting techniques focus on reconstructing the occluded elements of the image, leaving FR out of consideration. It is a challenging task to recover details of facial features on high-level image semantics, being used in many FR scenarios, such as when a subject wears sunglasses [32], a facial mask [15], or when there are other external facial occlusions [18], [55]. The purpose of inpainting is to reconstruct missing information in an image.

Inpainting methods usually consider information from the whole image (i.e., low-level texture information and high-level semantic information). Traditional inpainting methods rely on low level information to find best corresponding patches from the unaltered regions in the same image [56], [57]. These methods work well for background completions and repetitive texture patterns. However, as the face image consists of many unique components, low level features are limited for face inpainting tasks. Thus, the inpainting process needs to be carried out with a high semantic confidence [58].

Facial inpainting (also referred to as face completion) methods have been found to improve FR performance on occluded face images [59]. Rapid progress in deep learning, in particular GANs, inspired studies [15], [60] on facial inpainting. Here, GANs are proposed to deal with both low-level textural features and high-level semantic features utilised for removing facial occlusions. In [15] several GAN-based image inpainting models, i.e., [55], [61], [62], [63], are benchmarked on real world images showcasing significant reconstruction capability.

III. Fun Selfie Filter Database

To create the FSF database used in this work, a qualitative assessment of popular mobile applications for adding FSFs were conducted. In addition, various styles that focus on occluding different facial regions was considered. Subsequently, the selected FSFs were applied to 1,441 face images of the FRGCv2 [16] dataset. For this purpose, an automated software that emulates the chosen mobile applications was used, as illustrated in Fig. 2. The used subset of the FRGCv2 dataset has good-quality face images which allows to analyse the sole impact of FSFs on FR modules in the absence of quality-related factors [64], e.g., variations in pose or illumination.

A. Face Image Selection

The used database consists of constrained reference images and unconstrained probe images. For reference images frontal faces with neutral expression have been manually chosen. Further, probe images were selected which exhibit variations in pose, expression, focus and illumination. If possible, probe images were preferably chosen from different acquisition session in order to obtain a realistic scenario. Examples of probe and reference images of both face image subsets are depicted in Fig. 3. FSFs are applied to the constrained

Fig. 3. Examples of reference (top row) and probe (bottom row) images selected from the FRGCv2 database.

TABLE II
RANKING CRITERIA FOR FUN SELFIE FILTER APPLICATIONS

Criteria	Score		
	1	2	3
Gallery support	black-box	possible	supported
Ease of use	hard	medium	easy
Daily usability	no	relative	yes
Variety	no	relative	yes
Popularity	low	medium	high
Development	no	relative	yes
Complexity	easy	medium	diverse
Cost	high	medium	low

TABLE III
SELECTED FUN SELFIE FILTER APPLICATIONS WITH THE NUMBER OF DOWNLOADS AND SCORE BASED ON THE USER REVIEWS IN 2022

Mobile Application	Popularity	
Sweet Face Camera [65]	500M+ downloads	4.3/5.0 score
B612 [11]	100M+ downloads	4.5/5.0 score
Snow [66]	100M+ downloads	4.2/5.0 score
YouCam Fun [12]	10M+ downloads	4.5/5.0 score
Bloom Camera [13]	1M+ downloads	4.3/5.0 score

reference images. This facilitates the automated application of FSF which requires a detection of faces.

B. Fun Selfie Filter Selection

To create an appropriate database of facial images with FSFs, a total of ten FSFs were selected from five different FSF mobile applications. The mobile applications were selected by performing a ranking based on the criteria in Tab. II. The scores have been assigned based on the available reviews from users, as well as the authors' experience while using the applications. Tab. III shows the five FSF mobile applications that received the highest rankings and which are the mobile applications used in this work.

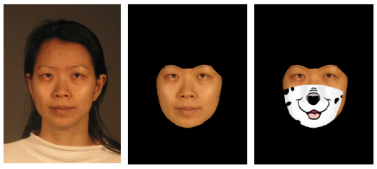


Fig. 4. Cropping of facial region polygon.

TABLE IV
FUN SELFIE FILTER GROUPS BASED ON THE FACIAL COVERAGE METRIC

Scenario	$Coverage_{intensity}$
Low Coverage	less than 15% difference
Medium Coverage	between 15% and 40% difference
High Coverage	more than 40% difference

In Tab. II, *gallery support* refers to the possibility of applying selfie filter to existing photos, rather than live captures; *variety* provides an estimate of how different selfie filters are relative to each other; *development* captures the continuous work on the application in terms of new feature or selfie filter addition.

When investigating the impact of FSFs on FR systems, it is interesting to see how the FSF coverage and placement affect the performance of the tested systems. The chosen FSFs are depicted as part of Fig. 2.

C. Categorisation

According to the criteria presented in Tab. II, a categorisation of FSFs was performed based on facial coverage and placement of the FSF.

Coverage: The FSF coverage is quantified by focusing on the facial region polygon that is used as a mask to the original image, cropping the facial part of the image as illustrated in Fig. 4.

This information is used further on to investigate the impact of the FSF based on its facial alteration, as well as focusing on specific elements that drive the eventual decrease in facial recognition performance.

$$Coverage_{intensity} = \frac{\Delta\text{Facial Pixel Intensity}}{\text{Number of Facial Pixels}} \quad (1)$$

Eq. (1) reports on the average pixel intensity variation due to the FSF, being a stable and accurate way of estimating the significance of the FSF. In the equation, Δ*Facial Pixel Intensity* is the absolute difference between pixels in the cropped facial regions of the masked and original face. Moreover, *Number of Facial Pixels* is the total pixel amount in the cropped facial region. Using this metric, transparent FSFs will achieve a lower score in comparison to corresponding solid color FSFs that cover the same area. Additionally, smoothing, compression, and other effects which do not occlude facial attributes, will not have a big impact on the calculated coverage intensity score. After trying a wide variety of FSFs provided by various mobile application, the visual complexity of recognising the identity behind the FSF has been defined following the thresholds presented in Tab. IV.

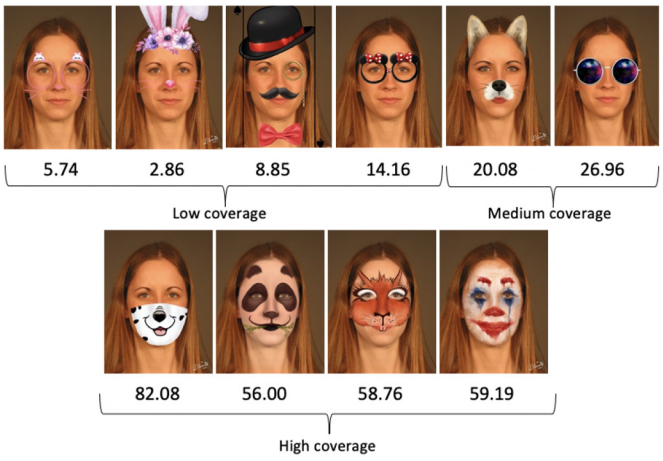


Fig. 5. Fun selfie filters with facial coverage score (see Eq. (1)).

Fig. 6. Fun selfie filter groups by the most affected facial region.

It is expected that the coverage score will correlate with the actual difficulty of recognising the original face once the FSF is applied. Tab. IV highlights the main FSF groups and Fig. 5 presents the scores for each of the ten samples.

Placement: Each of the FSFs presented in Fig. 6 focuses on altering certain facial features. This information will be used in concluding whether certain regions of the face are more important for recognition purposes and if there is any difference across the state-of-the-art FR systems.

IV. Impact on Face Recognition

The impact of FSFs on face detection and sample quality is estimated for scenarios with different facial coverage measures. In experiments on recognition performance, the most relevant scenario where either one of the face images to be compared has been modified using a FSF is considered. For evaluations on recognition performance, the placement of FSFs is additionally considered. In all evaluations a comparison with a baseline of unaltered face images is made.

Fig. 7 presents the min-max normalized face detection score, while Tab. V refers to the actual score, where the detection score ranges differ across the used algorithms (e.g., in our case, the range for detection scores on dlib is [0; 4], RetinaFace is [0; 1], and COTS is [0; 5.4]).

TABLE V
Average Face Detection Scores, Standard Deviation, and Errors (in %) by the Fun Selfie Filter Facial Coverage

Scenario	dlib			RetinaFace			COTS		
	μ	σ	ϵ	μ	σ	ϵ	μ	σ	ϵ
Baseline	2.031	0.529	0	0.9993	0.0002	0	3.058	0.820	0
Low	1.343	0.462	0.681	0.9995	0.0002	0.005	1.646	0.877	0.319
Medium	1.380	0.568	1.158	0.9993	0.0005	0.015	1.265	0.964	30.55
High	0.874	0.388	5.658	0.9992	0.0003	2.838	0.408	0.502	21.36

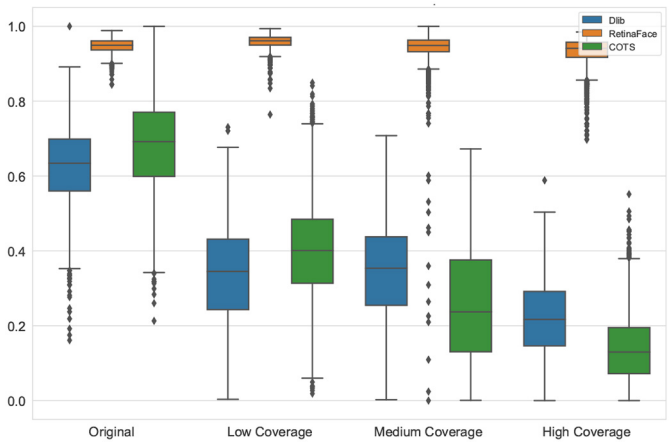


Fig. 7. Min-max normalized (eq. (2)) face detection confidence scores by the fun selfie filter facial coverage.

The min-max normalization is done to ensure an equal scale when comparing the performance of various algorithms as defined in Eq. (2).

$$X_i^{\text{normalized}} = \frac{X_i - X_{max}}{X_{max} - X_{min}} \cdot (R_{max} - R_{min}) + R_{min} \qquad (2)$$

where R_{min} and R_{max} cover the desired range of normalized data (i.e., in this case $R_{min} = 0$ and $R_{max} = 1$) and X_i refers to the detection score of sample i.

A. Face Detection

As shown in Tab. V and Fig. 7, using dlib [67], RetinaFace [68], and a COTS method, the confidence scores of detected faces for the FSF-filtered images, in general, degrade as the FSF coverage increases.

The used COTS is particularly prone to detection errors (ϵ) when the face is covered at a higher intensity or if the eye region is occluded. Such recognition systems are designed based on constraint environments, where the subject is required to follow some rules for an increased FR pipeline accuracy (e.g., in the border control scenario, every person follows a well-defined FR protocol and no unnecessary object is allowed to occlude the face).

B. Sample Quality

For face quality assessment, on the basis of FaceQNet [69] and MagFace [70], results, ranging in the interval [0; 1], are shown in Fig. 8 and Tab. VI.

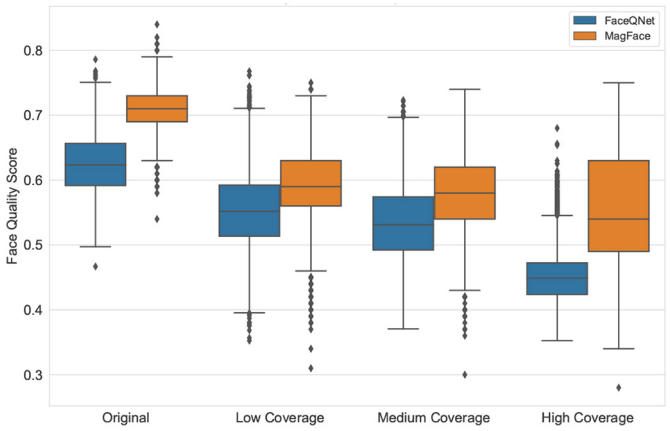


Fig. 8. Face quality scores by the fun selfie filter facial coverage.

TABLE VI
AVERAGE FACE QUALITY SCORES BY THE FUN SELFIE FILTER FACIAL COVERAGE

Scenario	FaceQNet		MagFace	
	μ	σ	μ	σ
Baseline	0.625	0.048	0.705	0.037
Low Coverage	0.554	0.059	0.592	0.056
Medium Coverage	0.534	0.060	0.576	0.056
High Coverage	0.450	0.044	0.556	0.081

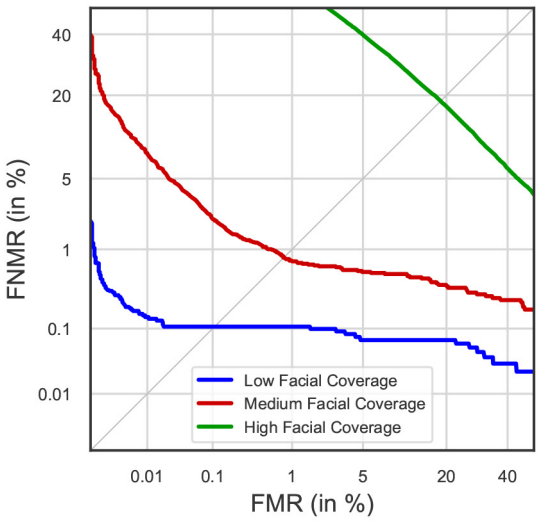
In addition to the case when the face is fully altered by the FSF, a significant effect on face sample quality is shown by medium coverage FSFs. FaceQNet as well as MagFace return a consistently lower image quality score as the FSF coverage increases. For MagFace, the comparatively high variance for high coverage FSFs is mainly caused by attributing relatively good face quality scores to face images where the joker mask is applied. Hence, if the FSF highlights certain facial characteristics, the magnitude of the facial embedding increases while the face may still be significantly occluded.

C. Recognition Performance

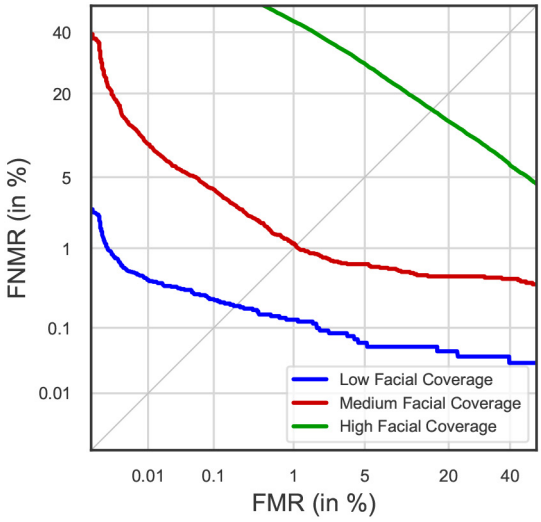
Biometric recognition performance is measured in terms of false non-match rate (FNMR) at certain false match rate (FMR) [71], [72]. In addition, the failure-to-enrol rate (FTE) and the equal error rate (EER) are reported. For the ArcFace [73] and COTS system, a higher FSF facial coverage results in a higher FNMR and EER, see detection error trade-off (DET) curves in Fig. 9 and Tab. VII. With respect to the placement of FSFs, the impact of the altered facial region differs but is especially challenging for mouth. Due to its more constraint target environment, COTS performs poorly on FTE when the eyes are covered.

V. FUN SELFIE FILTER REMOVAL

The facial coverage resulting from the application of FSFs can range from being almost non-existent to the extreme case of full coverage (*cf.* Fig. 5). As shown in the previous section, FR systems recognize well on low facial coverage scenarios,



(a) ArcFace



(b) COTS

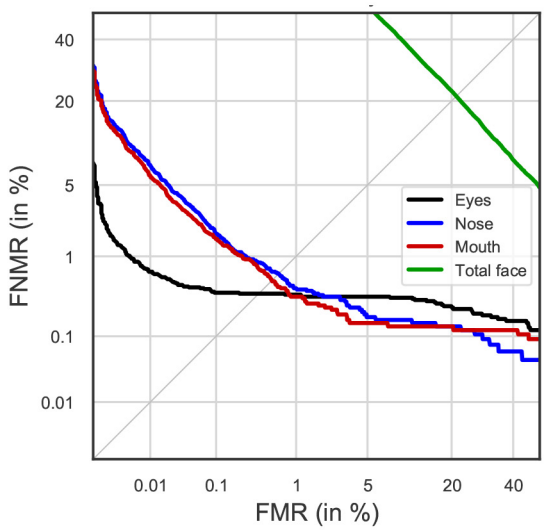
Fig. 9. DET curves by fun selfie filter coverage.

TABLE VII
FACE RECOGNITION PERFORMANCE (IN %) BY FUN SELFIE FILTER FACIAL COVERAGE

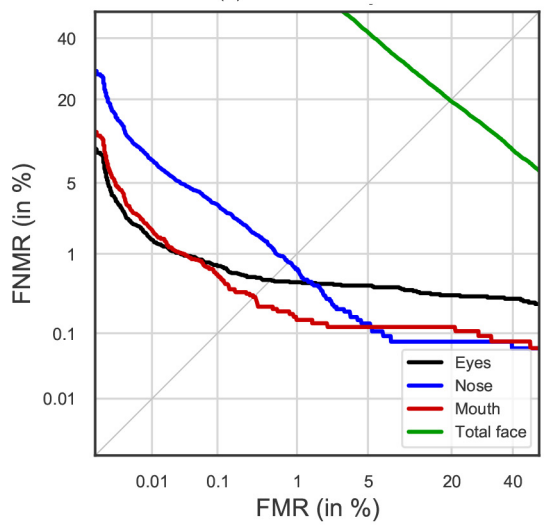
System	Scenario	FTE	EER	FNMR		
				FMR=0.01	FMR=0.1	FMR=1
ArcFace	Baseline	0	0	0	0	0
	Low Coverage	0.005	0.100	0.210	0.100	0.100
	Medium Coverage	0.015	0.850	9.400	3.500	0.830
	High Coverage	2.838	19.80	73.00	59.20	48.80
COTS	Baseline	0	0	0	0	0
	Low Coverage	0.319	0.250	0.650	0.400	0.175
	Medium Coverage	30.55	1.050	10.70	4.100	1.010
	High Coverage	21.36	18.20	66.00	50.60	44.70

while they are rather challenged by FSFs that produce high facial coverage.

This section introduces the proposed FSF removal method which represents an adaptation of the inpainting technique of [18]. As suggested in [15], the architecture is supplemented with a perceptual network [74], in a form of a pre-trained VGG-19 fixed network, to stimulate the generator output, to have similar feature representation to the ground truth ones.



(a) ArcFace



(b) COTS

Fig. 10. DET curves by fun selfie filter placement.

TABLE VIII
FACE RECOGNITION PERFORMANCE (IN %) BY THE FACIAL REGION THAT IS ALTERED BY THE FUN SELFIE FILTER

System	Scenario	FTE	EER	FNMR		
				FMR=0.01	FMR=0.1	FMR=1
ArcFace	Baseline	0	0	0	0	0
	Mouth	0.244	0.820	6.200	2.200	0.550
	Eyes	0	0.650	0.870	0.550	0.500
	Nose	0.139	0.850	8.800	2.780	0.620
	Total face	0.418	22.90	94.80	78.60	56.80
COTS	Baseline	0	0	0	0	0
	Mouth	0.662	0.500	2.200	0.700	0.210
	Eyes	20.66	0.760	1.920	0.910	0.710
	Nose	0.069	0.900	9.700	3.750	0.730
	Total face	28.97	19.80	82.90	69.70	58.30

The segmentation module and perceptual network are inspired from [58]. As proposed in [58], a good way of improving the GAN-based image inpainting accuracy is to split the occluded image segmentation from the actual image inpainting.

It is acknowledged that there are more possible solutions to the face occlusion removal (e.g., sunglasses, mask). However, without aiming to benchmark various algorithms, the aforementioned approach is chosen and tailored to the FSF case, proving the feasibility of selfie filter removal for an enhanced FR performance.

A. Segmentation

An overview of the architecture of the segmentation algorithm is depicted in Fig. 11. The output of the segmentation module is a binary map indicating the pixels covered by the FSF. The generator of the segmentation map is a modified version of the U-Net architecture [53] consisting of a CNN-based encoder and decoder:

Encoder consists of five blocks comprising of a convolution and a 'Squeeze-and-Excitation' layer, followed by a down-sampling of the input along its spatial dimensions by applying a *MaxPool* of *kernel size* 2 and *stride* 2;

Decoder resembles the encoder architecture except that the *MaxPool* is replaced by the up-sampling layer and instead of the convolution layers, deconvolution layers are applied, where the last layer of the decoder uses *sigmoid* activation function.

The local information is combined with the global one by concatenating the result of the deconvolution layers with the feature maps from the encoder at the same level. As a loss function, cross-entropy is used between the predicted binary map and corresponding target map, adding a post processing step to handle image processing operations of erosion and dilation.

B. Inpainting

The goal of this module is to remove the FSF and reconstruct the representation of the facial characteristics that have been covered by the FSF in a way that is both structural and appearance wise consistent with the ground truth image. The main building blocks for the image inpainting module are the generator, the discriminator and the perceptual network.

Generator: The generator has the same encoder and decoder architecture as the generator of the segmentation map, with the addition of gated convolution for the image inpainting network, accounting for a dynamic feature selection mechanism for each channel and spacial location.

Fig. 14 depicts the used architecture, where each convolution is distinctively marked based on its type (e.g., gated, dilated gated, or normal convolution). Overall, the GAN model takes as input the original or FSF-filtered image together with the corresponding FSF binary segmentation and passes it to the first generator network. Once the coarse output is derived, it is passed through a refinement network for an improved inpainting. The refined inpainting together with the FSF binary segmentation is the input to the fully convolutional discriminator and to the perceptual network. As a result, the GAN loss is computed and the training proceeds until it has reached the targeted number of iterations.

Discriminator: For training free-form image inpainting networks, the fully convolutional discriminator architecture is

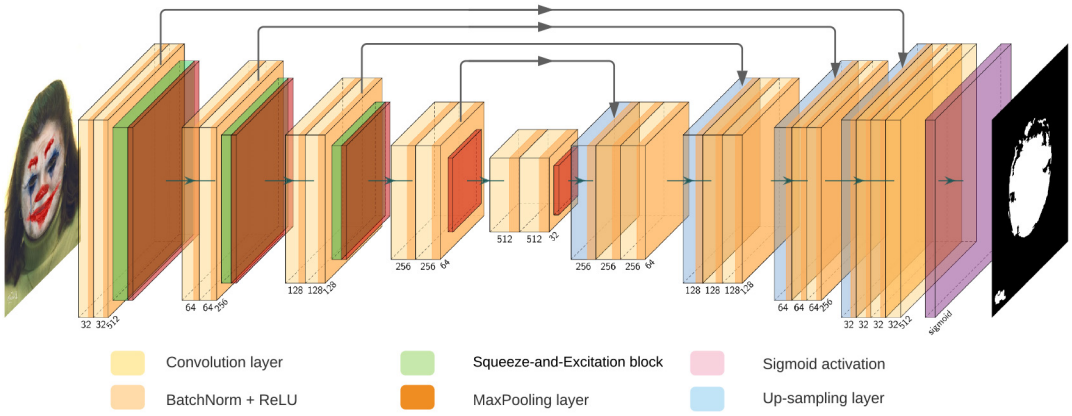


Fig. 11. Architecture of the fun selfie filter segmentation network.

used. As indicated earlier, the network is inspired by global and local GANs [61], MarkovianGANs [75], and perceptual loss [76].

A six strided convolutional network with kernel size 5 and stride 2 is used as the discriminator. Additionally, GANs are applied for each feature element in this feature map, formulating $h \times w \times c$ number of GANs focusing on different locations and different semantics of the input image.

Perceptual network: The third module of the image inpainting module presented in Fig. 14 is a perceptual network, in the form of a pre-trained VGG-19 fixed network [74] with a perceptual loss [76] that is applied to penalize the outputs that are perceptually not reasonable by defining a feature level distance measure between the intermediate feature maps of the reconstructed image and its original counterpart. The purpose of this network is to encourage the generator's output to be similar to the original image.

As a optimization, [15] suggests exploiting the intermediate convolution layer feature maps of the VGG-19 network to get rich structural information. This is expected to help in recovering plausible structure of the face semantics.

The overall generator loss function, $\mathcal{L}$, is defined as:

$$\mathcal{L} = \lambda_{\text{rc}_{\text{coarse}}} \cdot \mathcal{L}_{\text{rc}_{\text{coarse}}} + \lambda_{\text{rc}_{\text{refined}}} \cdot \mathcal{L}_{\text{rc}_{\text{refined}}} + \lambda_{\text{perc}} \cdot \mathcal{L}_{\text{perc}} + \lambda_G \cdot \mathcal{L}_G \tag{3}$$

where $\lambda_{\text{rc}_{\text{coarse}}} = 30$, $\lambda_{\text{rc}_{\text{refined}}} = 70$, $\lambda_{\text{perc}} = 50$, and $\lambda_G = 0.7$.

The aforementioned constants are determined by incrementally optimizing the parameter values, so that the loss function would converge to low values [15]. $\mathcal{L}_{\text{rc}} = \mathcal{L}_{\text{H}} + \mathcal{L}_{\text{SSIM}}$ (calculated based on the coarse and refined outputs, as presented in Fig. 14). $\mathcal{L}_{\text{H}}$ uses the mean squared error (MSE) if the absolute element-wise error falls below one and the l_1-distance, otherwise. Its combination with the $\mathcal{L}_{\text{SSIM}}$ ensures that the resulting image resembles its target, being also similar in terms of structural similarity index (SSIM). $\mathcal{L}_{\text{perc}}$ refers to the perceptual loss, while $\mathcal{L}_G$, the generator loss, captures the MSE loss given the discriminator's refined output and the target image.

Additionally, the discriminator loss function, $\mathcal{L}_{\mathcal{D}}$, is:

$$\mathcal{L}_{\mathcal{D}} = 0.5 \cdot \left(\mathcal{L}_{\text{MSE}_{\text{fake}}} + \mathcal{L}_{\text{MSE}_{\text{real}}}\right) \tag{4}$$

where $\mathcal{L}_{\text{MSE}}$ is the MSE between the input and target tensor.

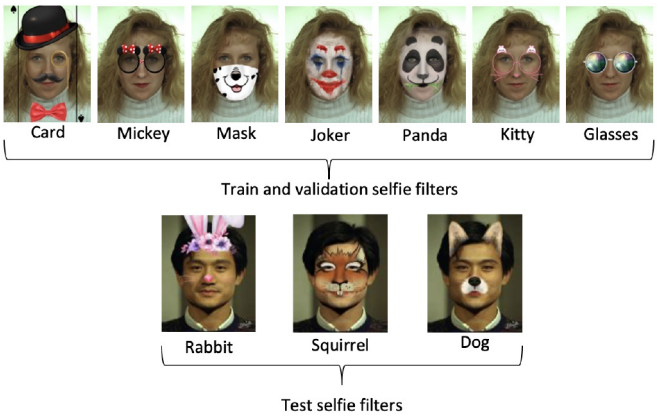


Fig. 12. Fun selfie filters used for training, validation and testing.

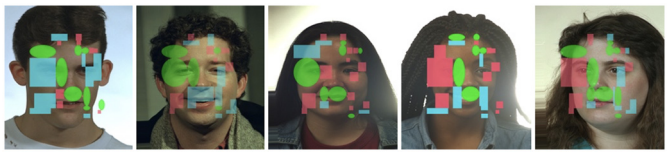


Fig. 13. Data augmentation: Overlay with multiple shapes of random color, size, and transparency.

For both the generator and the discriminator, the Adam optimizer is applied [77] with an initial learning rate of 0.001 that is adjusted every 50,000 training iterations by 0.1.

C. Training

To assess the GAN model's generalizability, seven FSFs are chosen for training and validation, while the remaining three are used for testing, as shown in Fig. 12.

The overall image inpainting process is directly related to the quality of the FSF segmentation. Hence, for enhanced accuracy and generalizability, the training of the segmentation module should be done on a wide variety of FSFs. Given the limited number of considered popular FSFs (Fig. 12), it becomes rather challenging to generalize well on unknown FSFs occluding unknown facial images.

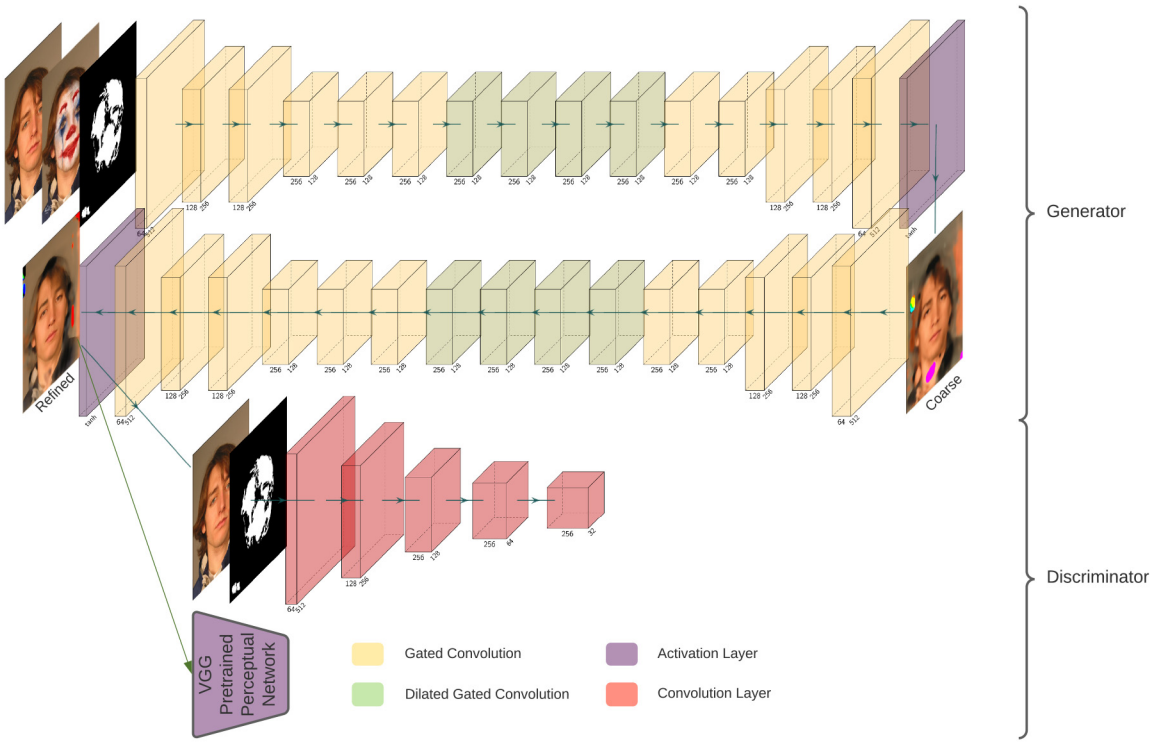


Fig. 14. Architecture of the inpainting network.

To improve robustness towards unknown FSFs, data augmentation was applied. The employed data augmentation is motivated by observed FSF properties: on the one hand, FSFs can contain various colours and be transparent to a certain degree: on the other hand, FSFs can be of different sizes and cover different facial parts. The implemented data augmentation method, for which examples are shown in Fig. 13, consists of three steps:

1) Identify the facial region based on its landmarks;
2) Divide the identified region in a the desired number of subregions that do not overlap;
3) Place random shapes (of random colour and intensity) on a subset of subregions such that only one shape is attributed to a subregion and its size is a perfect fit for the target subregion.

In a cross-database experiment, the training is performed on the FERET dataset [17], while the FSF removal is applied to the FRGC dataset [16]. As a result, the training and test images are captured from different subjects and affected by different environmental factors, e.g., different illumination.

The training of the GAN based model for FSF removal was done on a Tesla M10 GPU over 12,225 training samples of size 512 × 512, where 4,355 are FSF-filtered images (Fig. 12) based on the FERET references and 7,870 are semi-synthetically created images with shapes (Fig. 13) based on the FERET probes. The model is trained for 70 epochs (i.e., as a good time and performance trade-off), using an Adam optimizer as a replacement optimization algorithm for stochastic gradient descent.

Fig. 15 presents the evolution of the GAN based FSF removal training, where samples exceeding 500,000 iterations do not differ much from the performance attained just after 500,000 iterations. However, fine-tuning the generator and the discriminator over more iterations was seen to be very important when testing the FSF removal. Fig. 16 highlights the FSF removal performance on unseen FSFs over a set of pre-trained weights.

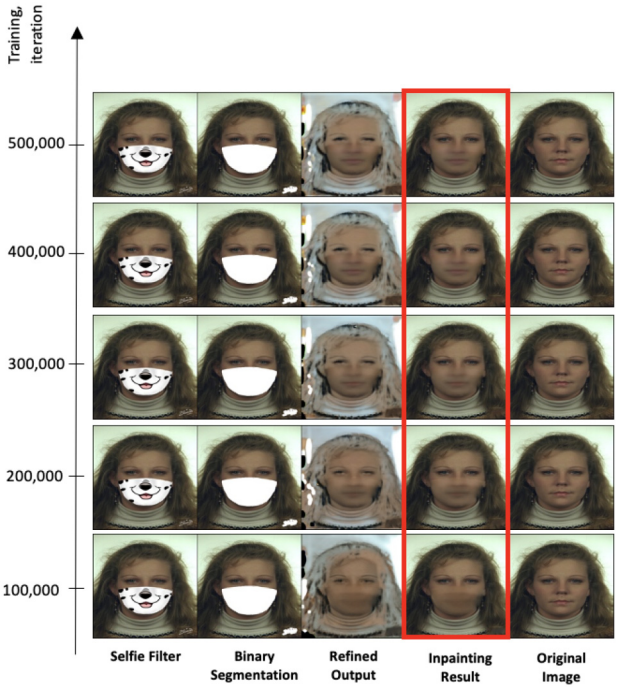


Fig. 15. Incremental GAN based fun selfie filter removal training.

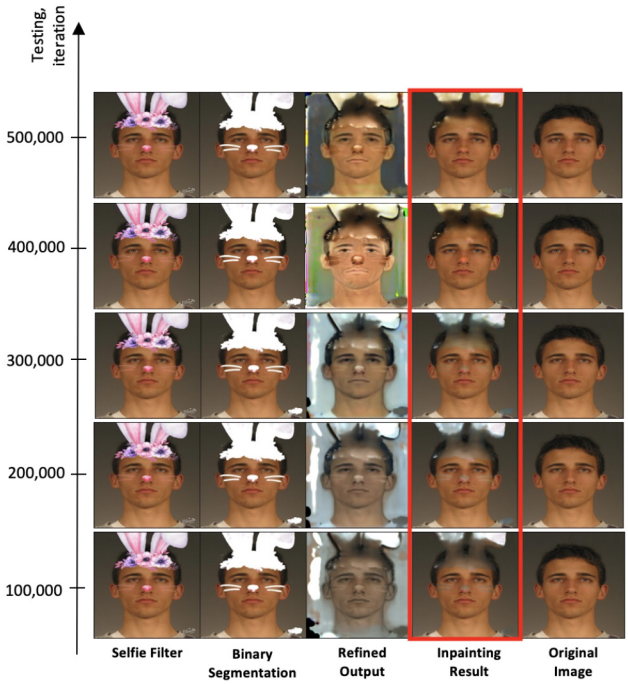


Fig. 16. GAN testing on removing unseen fun selfie filters.

TABLE IX
THE PSNR AND MSSIM SCORES FOR THE FUN SELFIE FILTERED IMAGE (FIG. 12) AND ITS RECONSTRUCTED COUNTERPART, AVERAGED OVER 1,441 IMAGES

FSF	Coverage	Placement	FSF-Filtered		Reconstructed Image (FSF Removal)	
			PSNR	MSSIM	PSNR	MSSIM
Card	Low	mouth	22.42	0.967	**25.32**	**0.970**
Kitty	Low	eyes	22.12	0.965	**27.88**	**0.967**
Bunny	Low	nose	21.55	0.966	**34.71**	**0.976**
Mickey	Low	eyes	22.08	0.966	**23.85**	**0.968**
Glasses	Medium	eyes	19.44	0.927	**22.11**	**0.968**
Dog	Medium	nose	21.27	0.955	**23.96**	**0.964**
Mask	High	mouth	16.26	0.930	**22.42**	**0.959**
Panda	High	total	18.87	0.915	**19.10**	**0.943**
Squirrel	High	total	20.60	0.920	**22.52**	**0.964**
Joker	High	total	19.60	0.902	**26.82**	**0.945**

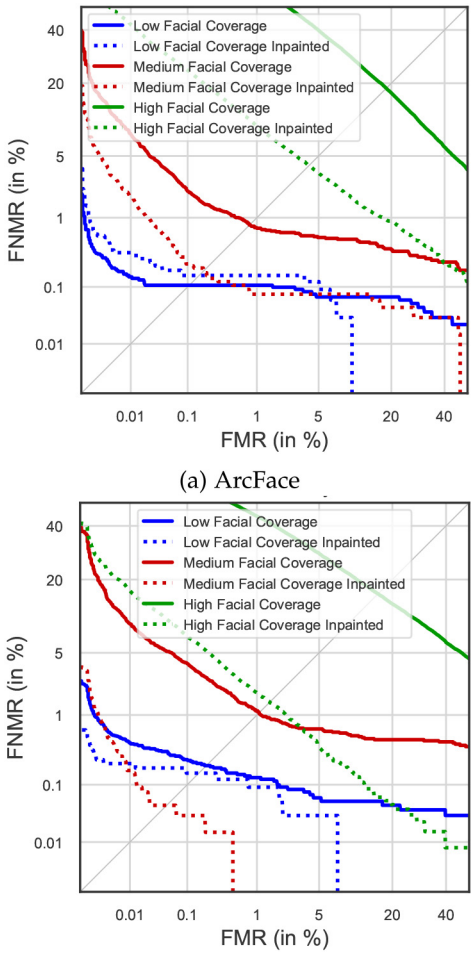
D. Face Recognition Performance With Fun Selfie Filter Removal

Having trained the GAN-based FSF removal model for at least 500,000 iterations, following the train and test split presented in Fig. 12, a comparison with the original unaltered counterparts is performed.

Following the Peak Signal-to-Noise Ratio (PSNR) [78] and Mean Structural Similarity Index (MSSIM) [79] scores, Tab. IX indicates an overall higher similarity between the FRGC reference and the image where the FSF has been removed relative to the FSF-filtered counterpart.

PSNR captures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation.

MSSIM is a perception-based model that considers image degradation as perceived change in structural information, where pixels have strong inter-dependencies especially when they are spatially close. A higher *PSNR* or *MSSIM* score relate to an increased degree of similarity between the compared images.



(a) ArcFace

(b) COTS

Fig. 17. DET curves for FRGC based fun selfie filter test dataset by facial coverage.

TABLE X
EER COMPARISON BETWEEN THE FUN SELFIE SELFIE-FILTERED IMAGES AND THEIR FUN SELFIE FILTER REMOVAL COUNTERPART BY FACIAL COVERAGE

FR System	Coverage	EER %	FSF Removal EER %	Δ
ArcFace	None	0	n.a.	n.a.
	Low	0.100	0.200	0.100
	Medium	0.850	0.250	**-0.600**
	High	19.80	4.500	**-15.30**
COTS	None	0	n.a.	n.a.
	Low	0.250	0.210	**-0.040**
	Medium	1.050	0.060	**-0.990**
	High	18.20	2.100	**-16.10**

To quantify the benefit of applying the proposed removal method, the FR performance is re-evaluated after removing the FSFs. The corresponding DET curves are plotted for FSFs of various facial coverage (Fig. 5) and affected facial regions (Fig. 6).

Comparing the results in Fig. 17 and Tab. X with their FSF-filtered counterparts, the EER has mostly decreased, improving

(a) FSF-filtered (b) Removal (c) Original

Fig. 18. Examples of total facial coverage FSFs.

(a) FSF-filtered (b) FSF Removal (c) Original

Fig. 19. Examples of medium facial coverage fun selfie filters.

(a) FSF-filtered (b) FSF Removal (c) Original

Fig. 20. Examples of low facial coverage fun selfie filters.

the FR performance across facial coverage and placement. The only exception is proposed by the low coverage FSFs, where, due to some facial artifacts after the removal, the metric is slightly higher than the FSF-filtered counterpart, while still keeping the superior FR performance.

The high coverage FSFs have benefited the most from the proposed FSF removal approach, achieving a 15.30% percentage point lower EER for ArcFace (Fig. 17a) and a 16.10% percentage point lower EER for COTS (Fig. 17b) compared to the FSF-filtered variant (Fig. 19).

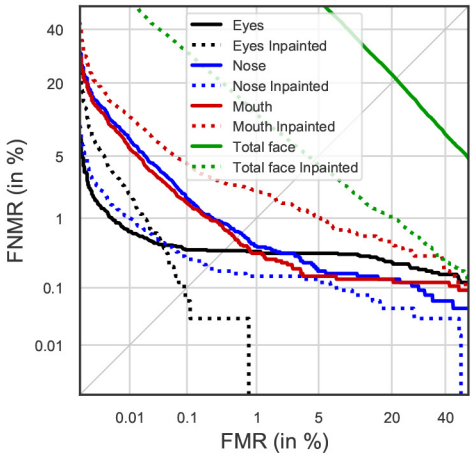
For medium facial coverage FSFs, despite of a realistic reconstruction of the face, the challenge of approximating facial elements of the mouth region in particular has not allowed for a significant growth in FR performance (Fig. 20).

In the case of low coverage FSFs, facial images before and after FSF removal maintain a high visibility of the original facial characteristics (Fig. 20).

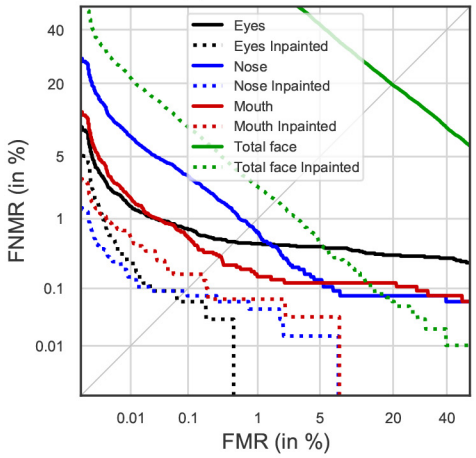
Despite of the relatively accurate FSF removal, the resulting images might vary in brightness or might still contain FSF related artefacts. In the case of the total facial coverage only some facial elements are approximated, while still leading to an enhanced FR performance.

The application of the FSF removal when the entire face is occluded sees the highest FR performance enhancement but it can also improve performance in other scenarios as, for instance, illustrated in Fig. 21 and Tab. XIII.

In addition to reducing the EER, the FSF removal improves the FTE. Given that COTS has shown to be vulnerable to FSFs covering the eye region, the selfie removal has reduced the corresponding FTE by 19.87% points. Furthermore, the FSF removal has reduced the FTE on COTS for the total coverage FSFs by 16.37% points. The FTE metric on ArcFace is stable before and after the use of the FSF removal method. Tab. XII



(a) ArcFace



(b) COTS

Fig. 21. DET curves for FRGC based fun selfie filter test dataset by placement.

TABLE XI
EER COMPARISON BETWEEN THE FUN SELFIE-FILTERED AND ITS FUN SELFIE FILTER REMOVAL COUNTERPART BY FACIAL PLACEMENT

FR System	FSF Coverage	EER %	FSF Removal EER %	Δ
ArcFace	None	0	n.a.	n.a.
	Mouth	0.820	3.100	2.280
	Eyes	0.650	0.100	**-0.550**
	Nose	0.850	0.360	**-0.490**
	Total face	22.90	4.700	**-18.20**
COTS	None	0	n.a.	n.a.
	Mouth	0.500	0.250	**-0.250**
	Eyes	0.760	0.095	**-0.665**
	Nose	0.900	0.100	**-0.800**
	Total face	19.80	2.400	**-17.40**

and Tab. XIII highlight the benefits of the FSF removal method in terms of FTE across facial coverage and placement, where the FRGC test dataset is considered.

VI. CONCLUSION AND FUTURE WORK

FSFs are frequently used to modify selfies, e.g., prior to sharing them on social media. Alterations and occlusions that are added to face images by applying such FSFs represent

TABLE XII
FTE COMPARISON BETWEEN THE FUN SELFIE-FILTERED AND ITS FUN SELFIE FILTER REMOVAL COUNTERPART BY FACIAL COVERAGE

FR System	Coverage	FTE %	FSF Removal FTE %	Δ
ArcFace	None	0	n.a.	n.a.
	Low	0.005	0.190	0.185
	Medium	0.015	0.278	0.263
	High	2.838	1.528	**-1.310**
COTS	None	0	n.a.	n.a.
	Low	0.319	0.502	0.183
	Medium	30.55	0.328	**-30.22**
	High	21.36	12.34	**-9.02**

TABLE XIII
FTE COMPARISON BETWEEN THE FUN SELFIE-FILTERED AND ITS FUN SELFIE FILTER REMOVAL COUNTERPART BY FACIAL PLACEMENT

FR System	FSF Coverage	FTE %	FSF Removal FTE %	Δ
ArcFace	None	0	n.a.	n.a.
	Mouth	0.244	1.917	1.673
	Eyes	0	0.208	0.208
	Nose	0.139	0	**-0.139**
	Total face	0.418	0.162	**-0.256**
COTS	None	0	n.a.	n.a.
	Mouth	0.662	1.439	0.777
	Eyes	20.66	0.786	**-19.87**
	Nose	0.069	0.494	0.425
	Total face	28.97	12.60	**-16.37**

a challenge for FR systems. The results obtained during this work have shown that FSFs may negatively impact commercial and open-source FR modules. Across face detection, sample quality estimation, and FR, this is especially the case for facial images with high FSF facial coverage and for FSFs that cover the mouth and nose. Furthermore, for the used COTS system, eye coverage has a high correlation with an increased FTE.

To tackle the above challenge, a FSF removal algorithm has been proposed. The proposed GAN-based method was shown to reduce the negative effects caused by the FSF when removing it prior to FR.

Nevertheless, it is important to note that the presented work represents an initial study that has certain limitations. The following should be addressed in future work to confirm the effectiveness of the proposal and to show the applicability of the proposed approach in a realistic scenario:

- The experimental evaluation should be extended, introducing comparisons with related and alternative facial inpainting techniques applicable to selfie filter removal.
- More unconstrained images containing common variations should be considered to supplement the images used in the work and prove selfie filter removal generalizability.

REFERENCES

[1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 1701–1708.

[2] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102805.

[3] *Police Embrace Social Media as Crime-Fighting Tool*, CNN, Atlanta, GA, USA, 2012. Accessed: Aug. 13, 2021.

[4] *Detroit Police Department Weekly Report on Facial Recognition*, Detroit Police Dept., Detroit, MI, USA, May 2021. Accessed: Apr. 4, 2022.

[5] *How Police Monitor Social Media to Find Crime and Track Suspects*, MLive, Grand Rapids, MI, USA, 2021. Accessed: Aug. 13, 2021.

[6] *Warum Sucht die Polizei mit Instagram-Foto Nach Rebecca?* Berliner Tageszeitung, Berlin, Germany, 2019. Accessed: Apr. 8, 2022.

[7] *How Instagram, Facebook Photo Filters are Creating Challenges for Police When People Go Missing*, iHeartRadio, New York, NY, USA, 2019. Accessed: Apr. 8, 2022.

[8] C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks* (Advances in Computer Vision and Pattern Recognition), 1st ed. Cham, Switzerland: Springer-Verlag, 2022.

[9] M. Ibsen, C. Rathgeb, D. Fischer, P. Drozdowski, and C. Busch, "Digital face manipulation inäbiometric systems," in *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks* (Advances in Computer Vision and Pattern Recognition). Cham, Switzerland: Springer Verlag, 2022, pp. 27–43.

[10] M. Afifi and A. Abdelhamed, "AFIF^4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces," *J. Vis. Commun. Image Represent.*, vol. 62, pp. 77–86, Jul. 2019.

[11] *B612*, SNOW Corp., Seongnam-si, South Korea. Accessed: Aug. 13, 2021.

[12] *YouCam Fun*, Perfect Corp., New York, NY, USA. Accessed: Aug. 13, 2021.

[13] *Bloom Camera*, Nox Future Corp., Hong Kong. Accessed: Aug. 13, 2021.

[14] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," *IET Biometrics*, vol. 10, pp. 581–606, Nov. 2021.

[15] N.-U. Din, K. Javed, S. Bae, and J. Yi, "A novel GAN-based network for unmasking of masked face," *IEEE Access*, vol. 8, pp. 44276–44287, 2020.

[16] P.-J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2005, pp. 947–954.

[17] J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput. J.*, vol. 16, no. 5, pp. 295–306, 1998.

[18] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 4470–4479.

[19] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer Vision (ECCV)*. Berlin, Germany: Springer, 2004, pp. 469–481.

[20] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S.-Z. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*. Berlin, Germany: Springer, 2007, pp. 828–837.

[21] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, 2005, pp. 786–791.

[22] G. Hua and A. Akbarzadeh, "A robust elastic and partial matching metric for face recognition," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 2082–2089.

[23] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *Proc. 31st SIBGRAPI Conf. Graph. Patterns Images (SIBGRAPI)*, 2018, pp. 471–478.

[24] Y. Xia, B. Zhang, and F. Coenen, "Face occlusion detection based on multi-task convolution neural network," in *Proc. 12th Int. Conf. Fuzzy Syst. Knowl. Discov. (FSKD)*, 2015, pp. 375–379.

[25] H.-F. Neo, C.-C. Teo, and A. B.-J. Teoh, "Development of partial face recognition framework," in *Proc. 7th Int. Conf. Comput. Graph. Imag. Vis.*, 2010, pp. 142–146.

[26] L. He, H. Li, Q. Zhang, Z. Sun, and Z. He, "Multiscale representation for partial face recognition under near infrared illumination," in *Proc. IEEE 8th Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, 2016, pp. 1–7.

[27] S. Liao, A.-K. Jain, and S.-Z. Li, "Partial face recognition: Alignment-free approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1193–1205, May 2013.

[28] J.-S. Park, Y.-H. Oh, S.-C. Ahn, and S.-W. Lee, "Glasses removal from facial image using recursive error compensation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 805–811, May 2005.

[29] J. Wright, A.-Y. Yang, A. Ganesh, S.-S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[30] F. Zhao, J. Feng, J. Zhao, W. Yang, and S. Yan, "Robust LSTM-Autoencoders for face de-occlusion in the wild," *IEEE Trans. Image Process.*, vol. 27, pp. 778–790, 2018.

[31] A. Vijayalakshmi, "Recognizing faces with partial occlusion using inpainting," *Int. J. Comput. Appl.*, vol. vol. 168, no. 13, pp. 20–24, 2017.

[32] B. Hu, Z. Zheng, P. Liu, W. Yang, and M. Ren, "Unsupervised eyeglasses removal in the wild," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4373–4385, Sep. 2021.

[33] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?" in *Proc. IEEE Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, 2012, pp. 391–398.

[34] C. Chen, A. Dantcheva, and A. Ross, "Impact of facial cosmetics on automatic gender and age estimation algorithms," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, vol. 2, 2014, pp. 182–190.

[35] C. Rathgeb, A. Dantcheva, and C. Busch, "Impact and detection of facial beautification in face recognition: An overview," *IEEE Access*, vol. 7, pp. 152667–152678, 2019.

[36] M. Ibsen, C. Rathgeb, T. Fink, P. Drozdowski, and C. Busch, "Impact of facial tattoos and paintings on face recognition systems," *IET Biometrics*, vol. 10, pp. 706–719, Apr. 2021.

[37] *What Countries Require Masks in Public or Recommend Masks?* #Mask4All, Dec. 2020. Accessed: Apr. 4, 2022.

[38] M. Gomez-Barrero *et al.*, "Biometrics in the era of COVID-19: Challenges and opportunities," 2021, *arXiv:2102.09258*.

[39] R. Raghavendra *et al.*, "Custom silicone Face Masks: Vulnerability of commercial face recognition systems & presentation attack detection," in *Proc. 7th Int. Workshop Biometrics Forensics (IWBF)*, May 2019, pp. 1–6.

[40] J. Deng, J. Guo, X. An, Z. Zhu, and S. Zafeiriou, "Masked face recognition challenge: The InsightFace track report," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021, pp. 1437–1444.

[41] F. Boutros *et al.*, "MFR 2021: Masked face recognition competition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, 2021, pp. 1–10.

[42] M. Ferrara, A. Franco, and D. Maltoni, "On the effects of image alterations on face recognition accuracy," in *Face Recognition Across the Imaging Spectrum*. Cham, Switzerland: Springer, 2016, pp. 195–222.

[43] A. Bharati, R. Singh, M. Vatsa, and K.-W. Bowyer, "Detecting facial retouching using supervised deep learning," *IEEE Trans. Inf. Forensics Security*, vol. 11, pp. 1903–1913, 2016.

[44] C. Rathgeb *et al.*, "PRNU-based detection of facial retouching," *IET Biometrics*, vol. 9, no. 4, pp. 154–164, Jul. 2020.

[45] C. Rathgeb, C.-I. Satnoianu, N. E. Haryanto, K. Bernardo, and C. Busch, "Differential detection of facial retouching: A multi-biometric approach," *IEEE Access*, vol. 8, pp. 106373–106385, 2020.

[46] L. Jiao *et al.*, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019.

[47] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022. [Online]. Available: https://www.computer.org/csdl/journal/tp/2022/07/09356353/1rigXK0s5Ak

[48] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[49] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1440–1448.

[50] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[51] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 5325–5334.

[52] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3431–3440.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Munich, Germany, 2015, pp. 234–241.

[54] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[55] M.-J. Khan, N.-U. Din, S. Bae, and J. Yi, "Interactive removal of microphone object in facial images," *Electronics*, vol. 8, no. 10, p. 1115, 2019.

[56] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. Image Process.*, vol. 12, pp. 882–889, 2003.

[57] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, p. 24, 2009.

[58] Y. Chen and H. Hu, "An improved method for semantic image inpainting with GANs: Progressive inpainting," *Neural Process. Lett.*, vol. 49, pp. 1355–1367, Jun. 2019.

[59] J. Mathai, I. Masi, and W. AbdAlmageed, "Does generative face completion help face recognition?" in *Proc. Int. Conf. Biometrics (ICB)*, 2019, pp. 1–8.

[60] Y.-A. Chen, W.-C. Chen, C.-P. Wei, and Y.-C. F. Wang, "Occlusion-aware face inpainting via generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 1202–1206.

[61] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Jul. 2017.

[62] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5505–5514.

[63] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "EdgeConnect: Structure guided image inpainting using edge prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV) Workshops*, Oct. 2019, pp. 3265–3274.

[64] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *ACM Comput. Surv.*, to be published.

[65] *Sweet Face Camera*, Sweet Snap Studio, Singapore. Accessed: Aug. 13, 2021.

[66] *SNOW*, SNOW Corp., Seongnam-si, South Korea. Accessed: Aug. 13, 2021.

[67] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Dec. 2009.

[68] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot multi-level face localisation in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 5202–5211.

[69] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "FaceQnet: Quality assessment for face recognition based on deep learning," in *Proc. Int. Conf. Biometrics (ICB)*, 2019, pp. 1–8.

[70] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 14220–14229.

[71] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 2382-37:2012 Information Technology—Vocabulary—Part 37: Biometrics*, ISO/IEC Standard 2382-37:2012, 2012.

[72] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 19795-1:2021. Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework*, ISO/IEC Standard 19795-1:2021, Jun. 2021.

[73] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 4685–4694.

[74] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1–14.

[75] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5967–5976.

[76] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694–711.

[77] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–15.

[78] L. C. Chan and P. Whiteman, "Hardware-constrained hybrid coding of video imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, no. 1, pp. 71–84, Jan. 1983.

[79] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, 2004.

Cristian Botezatu received the M.Sc. degree in computer science and engineering, with specialization in artificial intelligence and algorithms, from the Technical University of Denmark in 2021. He is currently pursuing the Ph.D. degree with the Norwegian University of Science and Technology, being a member with the Norwegian National Laboratory. He is working for the EU-funded Image Manipulation Attack Resolving Solutions Project bringing his contribution to new face quality standards. His main research interests focus on pattern recognition, biometrics, and face manipulations.

Mathias Ibsen (Graduate Student Member, IEEE) received the M.Sc. degree in computer science and engineering, with specialization in artificial intelligence and algorithms, from the Technical University of Denmark in 2020. He is currently pursuing the Ph.D. degree with the EU-Funded TReSPAsS-ETN Project and the Faculty of Computer Science, Hochschule Darmstadt, Germany. He is a member of the da/sec—Biometrics and Internet Security Research Group affiliated with the National Research Center for Applied Cybersecurity (ATHENE) and a member of the European Association for Biometrics. He has published several peer-reviewed papers and won a Best Paper Award (WIFS'21). His principal research interests are focused on pattern recognition, machine learning, biometrics, and face manipulations.

Christian Rathgeb is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt, Germany. He is a Principal Investigator with the National Research Center for Applied Cybersecurity ATHENE. He coauthored over 100 technical papers in the field of biometrics. His research includes pattern recognition, iris and face recognition, security aspects of biometric systems, secure process design, and privacy enhancing technologies for biometric systems. He is a winner of the EAB—European Biometrics Research Award 2012, the Austrian Award of Excellence 2012, the Best Poster Paper Awards (IJCB'11, IJCB'14, and ICB'15), the Best Paper Award Bronze (ICB'18), and the Best Paper Award (WIFS'21). He is a member of the European Association for Biometrics (EAB), a Program Chair of the International Conference of the Biometrics Special Interest Group (BIOSIG), and an Editorial Board Member of *IET Biometrics* (IET BMT). He has served for various program committees and conferences, e.g., ICB, IJCB, BIOSIG, and IWBF and journals as a Reviewer, e.g., IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE, and *IET Biometrics*.

Christoph Busch (Senior Member, IEEE) is a member of the Norwegian University of Science and Technology, Norway. He holds a joint appointment with Hochschule Darmstadt, Germany. Further, he lectures Biometric Systems with DTU, Denmark, since 2007. On behalf of the German BSI, he has been the Coordinator for the project series BioIS, BioFace, BioFinger, BioKeyS Pilot-DB, KBEinweg, and NFIQ2.0. He was/is the partner of the EU projects 3D-Face, FIDELITY, TURBINE, SOTAMD, RESPECT, TReSPsS, iMARS, and others. He is also a Principal Investigator with the German National Research Center for Applied Cybersecurity (ATHENE) and is the Co-Founder of the European Association for Biometrics. He coauthored more than 500 technical papers and has been a speaker at international conferences. He is a member of the editorial board of the IET journal on Biometrics and formerly of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. Furthermore, he chairs the TeleTrusT Biometrics Working Group as well as the German standardization body on Biometrics and is the Convenor of WG3 in ISO/IEC JTC1 SC37.