An Attack on Facial Soft-Biometric Privacy Enhancement

Dailé Osorio-Roig, *Graduate Student Member, IEEE*, Christian Rathgeb, Pawel Drozdowski, Philipp Terhörst, Vitomir Štruc, *Senior Member, IEEE*, and Christoph Busch, *Senior Member, IEEE*

***Abstract*—In the recent past, different researchers have proposed privacy-enhancing face recognition systems designed to conceal soft-biometric attributes at feature level. These works have reported impressive results, but generally did not consider specific attacks in their analysis of privacy protection. We introduce an attack on said schemes based on two observations: (1) highly similar facial representations usually originate from face images with similar soft-biometric attributes; (2) to achieve high recognition accuracy, robustness against intra-class variations within facial representations has to be retained in their privacy-enhanced versions. The presented attack only requires the privacy-enhancing algorithm as a black-box and a relatively small database of face images with annotated soft-biometric attributes. Firstly, an intercepted privacy-enhanced face representation is compared against the attacker's database. Subsequently, the unknown attribute is inferred from the attributes associated with the highest obtained similarity scores. In the experiments, the attack is applied against two state-of-the-art approaches. The attack is shown to circumvent the privacy enhancement to a considerable degree and is able to correctly classify gender with an accuracy of up to approximately 90%. Future works on privacy-enhancing face recognition are encouraged to include the proposed attack in evaluations on the privacy protection.**

***Index Terms*—Biometrics, face recognition, privacy protection, privacy enhancement, soft-biometrics, attack.**

Manuscript received November 24, 2021; revised January 21, 2022 and March 14, 2022; accepted April 23, 2022. Date of publication May 9, 2022; date of current version June 21, 2022. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie under Grant 860813 (TReSPAsS-ETN); in part by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE; and in part by the ARRS Project J2–1734 "Face deidentification with Generative Deep Models" (FaceGEN). The work of Philipp Terhörst was supported in part by the European Research Consortium for Informatics and Mathematics. This article was recommended for publication by Associate Editor T. Hassner upon evaluation of the reviewers' comments. *(Corresponding author: Dailé Osorio-Roig.)*

Dailé Osorio-Roig, Christian Rathgeb, Pawel Drozdowski, and Christoph Busch are with the Department of Computer Science, Hochschule Darmstadt, 64295 Darmstadt, Germany (e-mail: daile.osorio-roig@h-da.de).

Philipp Terhörst is with the Smart Living and Biometric Technologies Department, Fraunhofer Institute for Computer Graphics Research, 64283 Darmstadt, Germany, and also with the Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway.

Vitomir Štruc is with the Faculty of Electrical Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia.

Digital Object Identifier 10.1109/TBIOM.2022.3172724

I. Introduction

FACE recognition technologies are deployed in many personal, commercial, and governmental identity management systems around the world. Current state-of-the-art face recognition technologies utilise deep learning and massive training datasets to embed face images as discriminative representations in the latent space [1], [2]. Similar kinds of deep learning techniques, e.g., deconvolutional neural networks, have shown impressive results for reconstructing facial images from their corresponding embeddings [3]. Further, it has been demonstrated that, sensitive soft-biometric information, e.g., gender, race, or age, can be directly derived from facial embeddings [4], [5].

In response to these privacy issues, a considerable amount of research has been conducted over the past years. In order to protect individuals' privacy, *biometric template protection* schemes have been proposed for various biometric characteristics, including the face. Biometric template protection methods are commonly categorized as *cancelable biometrics* and *biometric cryptosystems*. Cancelable biometrics employ transforms in the signal or feature domain which enable a biometric comparison in the transformed domain [6]. In contrast, the majority of biometric cryptosystems binds a key to a biometric feature vector resulting in a protected template. Biometric authentication is then performed indirectly by verifying the correctness of a retrieved key [7]. For comprehensive surveys on this topic, the interested reader is referred to [8], [9]. Alternatively, homomorphic encryption has frequently been suggested for biometric template protection [10]. Homomorphic encryption makes it possible to compute operations in the encrypted domain which are functionally equivalent to those in the plaintext domain and thus enables the estimation of certain distances between protected biometric templates. Biometric template protection are designed to fulfill the major requirements of irreversibility and unlinkability which are defined in ISO/IEC IS 24745 [11].

In addition to face-based biometric template protection, methods that attempt to remove (or conceal) certain sensitive information from facial biometric data (while leaving other useful information unchanged) have been proposed by various research laboratories. Said schemes have recently been summarized under the umbrella term *privacy–enhancing face biometrics*, a comprehensive survey can be found in [12]. A large amount of published methods which are referred

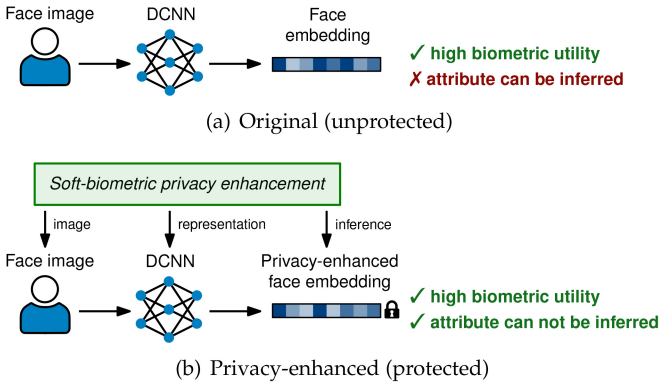


Fig. 1. Extraction of facial embeddings: (a) original face embeddings exhibit high biometric utility, but sensitive attributes can be derived from them; (b) application of soft-biometric privacy enhancement at image, representation or inference level is usually claimed to result in face embeddings with high biometric utility of which sensitive attributes can not be inferred.

to as *soft-biometric privacy enhancement* aim at removing or suppressing sensitive attributes in facial data. In the context of a face recognition system, this group of techniques can be applied at either image level, representation level, or at inference level [12]. Approaches applied on image level, e.g., obfuscation, have been shown to enhance privacy at the cost of biometric utility. In other words, privacy-enhanced face images obtained by said techniques become less usable for facial recognition tasks. Further, different methods have been applied at representation level or inference level, i.e., these methods operate at feature level. Interestingly, the latter schemes have been reported to retain biometric utility and, at the same time, provide strong privacy protection [12], see Figure 1. This clearly contradicts with the assumption that a removal or suppression of facial information yield less discriminative face embeddings which results in a decrease in biometric performance, analogous to methods applied on image level. This necessitates a closer examination of soft-biometric privacy enhancement methods. In particular since published approaches often lack a rigorous analysis with respect to privacy protection [13].

The main contribution of this work is the proposal of a novel attack on privacy-enhancing face recognition systems. Here, we mainly focus on methods operating at representation or inference level while the attack is generally applicable to any soft-biometric privacy enhancement method (including image level-based methods). The attack builds upon the following observations: it has recently been shown that facial recognition algorithms produce higher similarity scores and, hence, significantly more false matches for subjects with similar soft-biometric attributes – in particular gender and race. This effect is referred to as *broad homogeneity* [20]. Further, it has been shown that it is possible that face recognition algorithms operate on facial features that are unrelated to soft-biometric attributes, albeit with somewhat lower recognition accuracy [21].

We show that the aforementioned properties also hold for privacy-enhancing face recognition systems. This can be exploited to attack these schemes, i.e., infer soft-biometric attributes from privacy-enhanced face embeddings. In the proposed attack, a face database with known soft-biometric attributes is used to generate a set of privacy-enhanced face representations against which a privacy-enhanced face representation with unknown soft-biometric attributes is compared. The best obtained similarity scores are then analyzed to derive the unknown attributes of the attacked privacy-enhanced face representation. The attack can be performed offline and only requires the privacy-enhancing algorithm as black box and an arbitrary set of facial images with known soft-biometric attributes. In experimental evaluations, the attack is applied to privacy-enhanced face representation obtained by two recently published algorithms, i.e., privacy-enhancing face-representation learning network (PFRNet) [17] and privacy-enhancing face recognition based on minimum information units (PE-MIU) [18]. High success rates of up to 90% with respect to gender prediction are obtained for attacking both state-of-the-art algorithms.

The results reported in this work indicate that privacy protection capabilities of facial soft–biometric privacy enhancement methods are commonly over-estimated in the current scientific literature. Towards the creation of privacy-preserving biometric systems various attacks have been proposed against different types of popular biometric cryptosystems and cancelable biometrics, e.g., in [22], [23]. Uncovered gaps in privacy protection have in turn led to (continuous) improvements of such schemes. Therefore, we believe that the developments of facial soft–biometric privacy enhancement can benefit from considering the proposed attack. In particular, to advance developments of facial soft–biometric privacy enhancement, it is strongly suggested to employ the proposed kind of attack in evaluations of privacy protection capabilities of future methods.

This work is organized as follows: Section II briefly summarises most relevant works on soft-biometric privacy-enhancing techniques applied at feature level. Section III describes the proposed attack in detail. The experimental setup and results are reported in Sections IV and V, respectively. They are subsequently discussed in Section VI, while Section VII contains a summary and concluding remarks.

II. Related Works

Several efforts have been made in recent years to introduce different soft-biometric privacy-enhancing techniques at feature level, i.e., approaches operating at representation or inference level. Table I lists the most relevant works in this research area. The performance metrics are reported in the table exactly as in the cited papers. Note that differently named metrics often correspond to the same underlying concept, e.g., ADA is expected to be the same as COCR.

Terhörst *et al.* [14] proposed a Cosine–Sensitive Noise (CSN) transformation applied to face embeddings to enhance privacy in terms of gender and age attributes. To this end, the authors introduced an specific type of noise over the face representation which hides the soft–biometric information. Morales *et al.* [15] proposed SensitiveNets, a privacy-preserving learning method. By incorporating soft-biometric

TABLE I
OVERVIEW OF RELEVANT FACIAL SOFT-BIOMETRIC PRIVACY ENHANCEMENT APPROACHES OPERATING AT FEATURE LEVEL (RESULTS REPORTED FOR BEST CONFIGURATIONS; NOTE THE DIFFERENCES IN THE USED EVALUATION DATASETS AND PERFORMANCE METRICS)

Authors	Method	Level	Protected Attribute	Datasets	Original		Privacy-Enhanced	
					Biometric Perf.	Classification Perf.	Biometric Perf.	Classification Perf.
Terhörst *et al.* [14]	CSN-Transformation	Representation	Gender Age	ColorFeret	$\sim 0.09\%$ EER	$\sim 90.0\%$ ADA –	$\sim 0.18\%$ EER	$\sim 65.00\%$ ADA 10.6% ASR
Morales *et al.* [15]	SensitiveNets	Representation	Gender Ethnicity	LFW	98.4% VA	97.70% ADA 98.8% ADA	95.8% VA	54.6% ADA 53.5% ADA
Terhörst *et al.* [16]	IVE	Representation	Gender Age	ColorFeret	3.1% EER	94.8% COCR 68.7% COCR	3.8% EER	77.9% COCR 50.6% COCR
Bortolato *et al.* [17]	PFRNet	Representation	Gender	CelebA Adience LFW	5.9% EER 5.6% EER 1.8% EER	1.8% fic 14.5% fic 4.9% fic	8.6% EER 6.4% EER 2.8% EER	43.5% fic 50.2% fic 41.4% fic
Terhörst *et al.* [18]	PE-MIU	Inference	Gender	LFW Adience ColorFeret	0.49% EER 3.27% EER 2.15% EER	89.50% ADA 89.81% ADA 97.62% ADA	0.56% EER 3.63% EER 3.11% EER	50.23% ADA 44.71% ADA 51.87 ADA
Terhörst *et al.* [19]	NFR	Inference	Gender Age Ethnicity Gender Age	ColorFeret Adience 	1.97% EER 3.83% EER 	97.30% ADA 57.40% ADA 88.73% ADA 84.91% ADA 60.36% ADA	3.18% EER 4.43% EER 	22.2% ASR 30.2% ASR 14.6% ASR 26.1% ASR 28.7% ASR

EER: Equal Error Rate, VA: Verification Accuracy, ADA: Attribute Decision Accuracy, ASR: Attribute Suppression Rate, COCR: Correct Overall Classification Rate, fic: Fraction of Incorrectly Classified Images

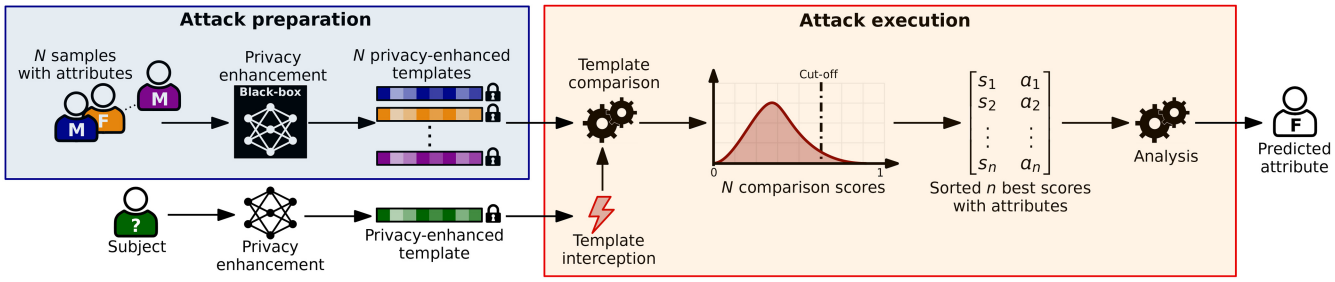


Fig. 2. Overview of the attack: an attacker is in possession of the soft-biometric privacy enhancement method and applies it to a database of images with known labels (collected in the preparation phase); then, an intercepted privacy enhanced face embedding is compared against the database and the best scores are analyzed to predict the soft-biometric attribute.

classifiers in the loss function of during algorithm training, this approach learns new feature representations suppressing gender and ethnicity information. Terhörst *et al.* [16] proposed a strategy called Incremental Variable Elimination (IVE) to eliminate (or remove) components related to soft-biometric information from the face feature representation. Bortolato *et al.* [17] managed to learn a disentangled feature representation in their so-called Privacy-Enhancing Face-Representation learning Network (PFRNet). PFRNet is an autoencoder which learns to separate gender attributes from the identity information.

A few works operating at the inference level have been proposed recently. These methods apply transformations and adapt the biometric comparator accordingly. Terhörst *et al.* [19] proposed such a method based on Negative Face Recognition (NFR). So-called negative embeddings are obtained by introducing features to them that are intentionally different from the original (positive) embeddings, thereby concealing soft-biometric attributes. Further, Terhörst *et al.* [18] proposed the Privacy-Enhancing face recognition approach based on Minimum Information Units (PE-MIU). This method allows the creation privacy-enhanced face template by partitioning the original feature vector into smaller parts (called minimum information units). Then, these blocks are randomly shuffled to obtain a privacy-enhanced template.

Whereas several authors have explored the development of novel techniques for removal of information on soft-biometrics with promising results, there still exists a need for deeper analysis of the achieved privacy protection. Terhörst *et al.* [13] recently argued that the absence of a standardized evaluation protocol hampers a meaningful comparison of proposed approaches to facial soft-biometric privacy enhancement. They propose a framework to evaluate the trade-off between suppressing an attribute and maintaining the recognition performance. However, their framework does not consider specific attacks.

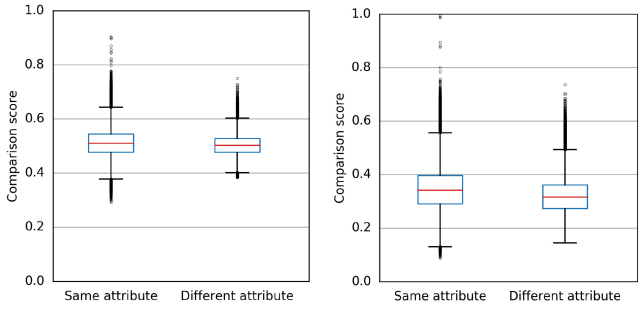
III. PROPOSED ATTACK

This section presents the proposed attack. Section III-A provides background information and theoretical foundations of the attack. Figure 2 shows an overview of the proposed attack; a detailed description of the attack execution is given in Section III-B.

A. Background

The proposed attack relies on several observations about:

1) The effects of broad homogeneity and demographic differentials in face recognition.



(a) FaceNet (b) VGGFace2

Fig. 3. Boxplots of similarity scores for original (unprotected) non-mated comparison trials with same and different soft-biometric attributes for two face recognition systems on the LFW database. Comparison trials for the same attribute (gender) yield slightly higher similarity scores and more outliers compared to those for different attributes.

2) Properties and general operating principles of the privacy-enhancing methods the attack is aimed at.

Regarding the first of the above, let P denote a probability measure and s a similarity scoring function between two non-mated samples with given soft-biometric attributes a_1 and a_2, which can be identical (e.g., female vs female) or different (e.g., female vs male). For the purpose of this example, let s return similarity scores in the range [0, 1], where 0 represents a complete dissimilarity and 1 a perfect similarity.

Many works have shown that comparisons between non-mated samples of same/similar soft-biometric attributes tend to generally yield higher similarity scores and consequently more frequent false matches, e.g., in [20], [24], [25]. This property is especially pertinent in face recognition, but does not necessarily hold for all other biometric characteristics (e.g., iris) [21]. In many face recognition systems, the relation,

$$P\big(\mathrm{s}_{a_1=a_2}(\cdot) > \mathrm{s}_{a_1\neq a_2}(\cdot)\big) \gg P\big(\mathrm{s}_{a_1=a_2}(\cdot) < \mathrm{s}_{a_1\neq a_2}(\cdot)\big) \quad (1)$$

generally holds true, where $\mathrm{s}_{a_1=a_2}(\cdot)$ and $\mathrm{s}_{a_1\neq a_2}(\cdot)$ denote similarity scores obtained from comparisons of non-mated samples with same and different soft-biometric attributes, respectively.

Additionally, beyond the general shift in the non-mated similarity scores distributions, the highest non-mated similarity scores (i.e., those at the tail of the score distribution) tend to stem from comparisons of two non-mated samples with identical, rather than different soft-biometric attributes. In other words, as the similarity score increases, the probability of the contributing samples being associated with the same soft-biometric attribute also increases,

$$\mathrm{s}_{a_1,a_2}(\cdot) \to 1 \Leftrightarrow P(a_1 = a_2) \to 1 \quad (2)$$

where s_{a_1,a_2} denotes the similarity score between two samples with soft-biometric attributes a_1 and a_2.

Figure 3 illustrates the above two propositions empirically. It can be seen, that the body of the boxplot for the "same attribute" similarity scores is shifted towards higher similarity scores; furthermore, its whisker and outliers are likewise shifted w.r.t. the boxplot for the "different attribute" similarity scores.

The general goal of the privacy-enhancing methods is to maintain the biometric performance and to simultaneously make infeasible inferring the soft-biometric attributes of the protected template. In other words, it is assumed that the methods retain sufficient identity information, while the information about the soft-biometric attributes is somehow disentangled/removed, e.g., in [14], [15], [17], [18]. Intuitively, such a process appears challenging. It would be surprising if this was possible, i.e., that not even the slightest overlap between identity and, e.g., gender or ethnicity information existed. Thus far, this assertion has neither been theoretically proven nor rigorously tested empirically. While the privacy-enhancing methods may change the feature space to be no longer separable (i.e., prevent classification by, e.g., SVMs), this does not necessarily guarantee security from other types of attacks, e.g., as described below.

In order to reach a decision based on a computed similarity score, biometric systems typically operate using a fixed decision threshold. Let t denote such a decision threshold; if $\mathrm{s}(\cdot) > t$, the compared samples are deemed to be mated by the system. In case the samples are actually non-mated, this means a false match. Although the feature representation and/or the comparator may operate completely different in the protected and unprotected domain, the basic principles regarding similarity scores and decision threshold remain unchanged. Hence, if biometric performance is to be maintained by the privacy-enhancing method, then the relations,

$$P\big(\mathrm{s}_{\mathrm{unprotected}}(\cdot) > t\big) \equiv P\big(\mathrm{s}_{\mathrm{protected}}(\cdot) > t\big) \quad (3)$$

must hold, where $\mathrm{s}_{\mathrm{unprotected}}(\cdot)$ and $\mathrm{s}_{\mathrm{protected}}(\cdot)$ denote similarity scores of an original and a privacy-enhancing face recognition system, respectively.

To satisfy these relations, the mapping performed by the privacy-enhancing method must be done in such a way, that sample pairs which would have achieved a high similarity score in the unprotected domain also do so in the protected domain. Due to the nearly inevitable overlap between the mated and non-mated score distributions, this implies that some non-mated sample pairs will be clustered closely together in the latent space generated by the privacy-enhancement method (i.e., at least those scoring above t corresponding to a certain false-match rate), i.e., *equation 2 likely holds true also in the protected domain*, thus opening an attack vector.

Bringing together the above points enables an attack aimed at inferring the soft-biometric attributes of templates protected by the aforementioned privacy-enhancing methods. The prerequisites for the attack are modest:

1) The attacker intercepts a privacy-enhanced template.
2) The attacker knows which algorithm was used to protect the template and can operate it as a black-box to generate new templates from own data.
3) The attacker possesses or can synthesize a dataset of arbitrary facial image, with approximate balanced distribution of the target attribute.

The attack, described in detail in the next subsection, takes advantage of the demographic differentials exhibited by most facial recognition systems, the imperfect separation between

mated and non-mated distributions in the vast majority of biometric recognition systems, and other circumstances which prevent the privacy-enhancing methods to fully disentangle identity and soft-biometric information.

B. Attack Execution

In the first step of the attack, an intercepted template is compared against the attacker's own database of privacy-enhanced facial templates. Let N represent the number of samples in the attacker's database. Further, let $A = [a_1 \dots a_k]$ represent the list of distinct soft-biometric attributes (e.g., male and female for gender) in the attacker's database, and k the count thereof. Thus, a list of $S = [s_1 \dots s_N]$ similarity scores is created; furthermore, a list of same length containing the soft-biometric attributes of the samples from attacker's database is maintained.

Instead of considering the entire list of scores, only a subset of highest similarity scores is considered. Depending on the selected analysis method (described further below), the attacker selects *one of the following*:

1) A single list, denoted S_n, representing similarity scores taken from S_N, sorted in descending order, and cut-off after n first entries.
2) k lists S_{a_n}, each representing similarity scores taken from S_N for each specific attribute a present in the attacker's database. The lists are sorted by similarity score in descending order, and cut-off after n first entries.

In the analysis step, the attacker applies simple algorithms or calculations to quantify the aforementioned behaviors and predict a soft-biometric attribute from a privacy-enhanced template. Let $c(a)$ represent a function which computes a loosely defined "strength of evidence" or a probability (not in a strict mathematical sense) of the intercepted template having a given soft-biometric attribute a. Further, let $C_{attack\_type} = [c(a_1) \dots c(a_k)]$ represent a list containing such probabilities for all (k) considered soft-biometric attributes for a given attack type:

Majority vote C_{vote} contains the count for all k possible attributes.

Averaging C_{av} is a list of averages for all k possible attributes.

Weighted averaging C_{av_lin} and C_{av_log} contain average similarity scores which are linearly and logarithmically weighted, respectively. Weights are assigned according to their position i in the list of n highest scores. Precisely, the linear weight $1 - i/n+1$ and the logarithmic weight $-\log i/n+1$ are applied.

To reach a decision denoted $P(x)$ (i.e., to predict the unknown attribute x of the intercepted template), the maximum value for the chosen attack type is found, i.e., $p = \underset{x}{\text{argmax}}\, C_{attack\_type}$. Finally, the corresponding soft-biometric attribute is selected accordingly, i.e., $P(x) = a_p$.

IV. Experimental Setup

This section describes the setup of the conducted experiments. Specifically, Section IV-A describes the experimental protocol, the used datasets are summarized in Section IV-B, while Section IV-C outlines the metrics used in the evaluations.

TABLE II
Overview of the Analyzed Soft–Biometric Privacy Enhancement Approaches

Approach	Recognition Model	Protected Attribute	Training	Test
PFRNet [17]	VGGFace2 [26]	Gender	CelebA [30]	Adience [28] LFW [29] CelebA [30]
PE-MIU [18]	FaceNet [27]	Gender	*none*	Adience [28] LFW [29] ColorFeret [31]

A. Choice of Algorithms and Protocol

Two soft–biometric privacy enhancement approaches, i.e., PFRNet and PE-MIU were selected. PFRNet and PE-MIU are based on a model [26] trained on VGGFace2 database (hereafter referred to as VGGFace2) and the FaceNet [27] face recognition system. Accordingly, these face recognition systems are used in experiments representing the original unprotected systems. While these face recognition system may not represent the current state-of-the-art, these are used to reproduce the results reported in previous works. Additionally, the effect of broad homogeneity has recently also been confirmed for various state-of-the-art systems [21].

The selection of the algorithm is based on several observations. Firstly, it is noteworthy that these methods are publicly available, i.e., the experiments in this work are reproducible. Secondly, like the chosen methods, most soft–biometric privacy enhancement approaches are designed to conceal gender information, see Table II. In fact, it is worth noting that there are hardly any available implementations of soft–biometric privacy enhancement methods protecting attributes other than gender. Thirdly, the two methods represent conceptually different soft-biometric privacy enhancement approaches, i.e., applied on representation level (PFRNet) and inference level (PE-MIU). Fourthly, these approaches achieved a promising trade-off between soft-biometric privacy protection and biometric performance over challenging databases such as Adience [28] and LFW [29]. Although other methods do exist in the literature, they were either superseded by the aforementioned methods or their authors were not able to provide the generated templates and/or the code/models for generating them.

The evaluation consists of following parts, organized accordingly in Section V:

Performance analysis in a baseline evaluation, the biometric performance and gender prediction accuracy are computed using the original (unprotected) and privacy-enhanced (protected) templates, similar to the protocol described in the respective publications [17], [18].

Vulnerability analysis the attacks described in Section III are carried out and their efficacy is evaluated.

B. Datasets

The experiments were conducted using the facial image databases with soft-biometric attribute annotations and face recognition models used by the authors of each of the considered soft–biometric privacy-enhancement approach, see

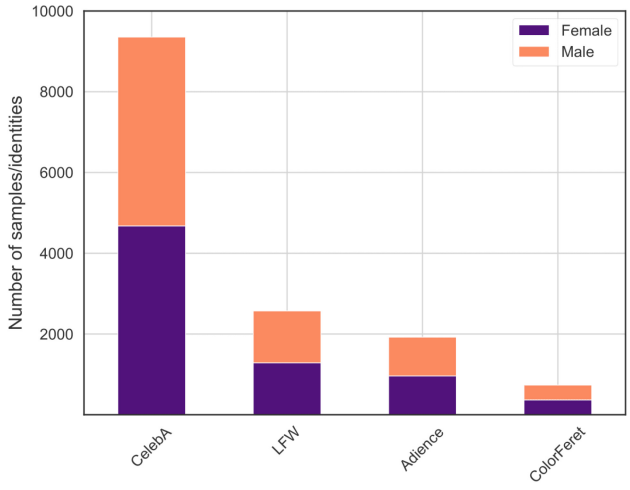


Fig. 4. Gender-balanced attacker databases sorted by size.

Table II. The privacy-enhanced templates generated by PFRNet for each dataset were provided directly by their authors [17]. The method was trained and applied on disjoint subsets of the CelebA database. For PE-MIU, the templates were generated using the publicly available PE-MIU software.[1] This method does not require any training. The underlying face recognition models VGGFace2 and FaceNet are trained with the VGGFace2 and MS-Celeb-1M databases.

To simulate an attacker possessing their own dataset, subsets of said databases were created by selecting one sample (with highest quality) per identity. These subsets are then balanced w.r.t. the protected soft-biometric attribute (i.e., gender) resulting in an approximate equal number of male and female subjects in the database. This is done to avoid a higher false match probability for one of the genders. In order to avoid duplicate identities in cross-database evaluations, high similarity scores obtained from cross-database comparisons were analyzed. To that end, face embeddings were extracted using a face recognition system, i.e., FaceNet [27], and cosine distance was used for comparison. Then, potential duplicate identities were identified by visual inspection. As a result, cross-database evaluations (e.g., CelebA against LFW) containing duplicated identities were removed from our evaluations. Figure 4 depicts an overview of the number of identities and the distribution of the gender attribute used for each dataset. Note that numerous subjects/samples have to be removed in order to obtain gender-balanced attacker databases. In the cross-database evaluations, the dataset possessed by the attacker and the dataset from which the targeted privacy-enhanced templates stem from are always different (e.g., attacker is in possession of FERET database, while the target stems from LFW database).

Scenarios in which the identity of an attacked privacy-enhanced template is contained in the training database of the face recognition model or soft-biometric privacy enhancement method represent a clear disadvantage for the attacker. On the one hand, if an image from the attacker's database has been seen by the recognition model during training, it is expected that it is more easily separable from other identities and, hence, it is less likely to produce a high similarity score. On the other hand, in case an image from the attacker's database has been seen by the privacy enhancing technique during training, i.e., PFRNet, it can be assumed that gender information will be suppressed more effectively for this identity. Thus, this identity has less chance to produce a high similarity score with an attacked template of the same gender.

[1] https://github.com/pterhoer/PrivacyPreservingFaceRecognition

TABLE III
BIOMETRIC PERFORMANCE FOR ORIGINAL (UNPROTECTED) AND PRIVACY-ENHANCED (PROTECTED) SYSTEMS (IN %)

Method	Dataset	Original			Privacy-enhanced		
		EER	FMR	FNMR	EER	FMR	FNMR
PFRNet	LFW	0.80	0.001 0.01 0.1	3.562 0.665 0.066	1.38	0.001 0.01 0.1	7.098 1.762 0.231
	Adience	6.70	0.001 0.01 0.1	74.445 40.022 4.332	6.72	0.001 0.01 0.1	80.191 43.501 4.632
	CelebA	6.47	0.001 0.01 0.1	30.786 15.533 4.962	9.37	0.001 0.01 0.1	34.326 20.108 9.034
PE-MIU	LFW	0.55	0.001 0.01 0.1	2.116 0.347 0.049	0.64	0.001 0.01 0.1	2.234 0.512 0.165
	Adience	4.72	0.001 0.01 0.1	63.675 22.640 2.057	4.72	0.001 0.01 0.1	63.675 22.638 2.347
	ColorFeret	2.16	0.001 0.01 0.1	16.721 4.083 0.419	2.70	0.001 0.01 0.1	16.724 4.613 1.246

C. Metrics

The experimental evaluation is conducted according to ISO/IEC 19795-1 [32] standard methods. The standard and additional metrics used in the experimental evaluation are as follows:

Biometric performance the False Non-Match Rate (FNMR) and False Match Rate (FMR) denote the proportion of falsely classified mated and non-mated attempts in a biometric verification scenario, respectively. Additionally, the equal error rate (EER), which is the point where FMR and FNMR are equal, is reported.

Attack success rate percentage of samples correctly classified in terms of soft-biometric attribute by an attack. This rate can also be seen as gender prediction accuracy.

V. RESULTS

In this section, Section V-A presents an performance analysis of the used soft-biometric privacy-enhancing approaches. Subsequently, a vulnerability analysis of said methods to the proposed attack is conducted in Section V-B.

A. Performance Analysis

As a first step, the biometric performance of the unprotected systems, i.e., original system, is estimated and compared against that of the corresponding privacy-enhanced systems. In Table III, the face verification performance is reported on different databases for each method. PFRNet and PE-MIU

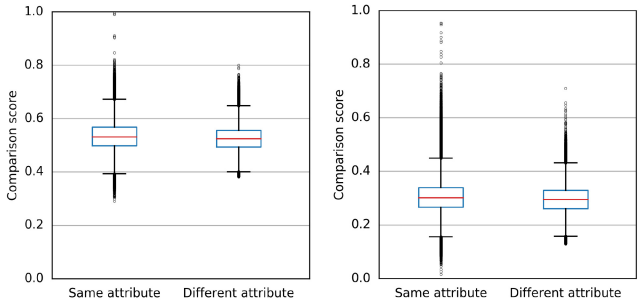
TABLE IV
GENDER PREDICTION PERFORMANCE OF BASIC MACHINE LEARNING-BASED CLASSIFIERS ON ORGINAL (UNPROTECTED) AND PRIVACY-ENHANCED (PROTECTED) TEMPLATES IN CROSS-DATABASE SCENARIOS (IN %)

Method	Training	Testing	Original SVM				Privacy-enhanced SVM			
			kNN	Poly	RBF	Sigmoid	kNN	Poly	RBF	Sigmoid
PFRNet	LFW	Adience	80.80	**91.00**	90.95	87.36	69.82	79.60	79.70	**62.64**
	Adience	CelebA	77.40	91.24	**94.54**	87.84	63.16	59.33	58.42	**51.19**
		LFW	83.36	94.56	**96.93**	87.91	73.79	76.39	75.42	**67.80**
	CelebA	Adience	80.75	89.65	**91.42**	86.73	73.20	83.71	86.00	**70.00**
PE-MIU	LFW	Adience	90.42	**91.31**	87.04	60.21	58.55	56.20	64.06	**50.00**
		ColorFeret	95.38	89.13	**96.20**	68.34	61.41	**55.43**	70.92	59.78
	Adience	LFW	95.18	89.81	**96.73**	72.86	59.18	57.93	63.06	**54.74**
		ColorFeret	**88.99**	84.24	88.18	80.98	62.09	60.46	61.68	**57.20**
	ColorFeret	LFW	93.51	84.56	**98.21**	75.00	57.62	50.27	**50.00**	51.63
		Adience	85.17	80.84	**89.54**	77.39	59.22	**50.00**	50.36	51.88

have both been applied on LFW and Adience. In addition PFRNet has been applied to CelebA and PE-MIU to ColorFeret, respectively. Based on the obtained results we can observe that the verification performance on privacy-enhanced system is slightly degraded compared to the original system. This confirms that privacy enhancement defines a trade-off between identity information and suppression of privacy-sensitive attributes, as it is shown in [18]. It can be observed that for FMRs$<$ 0.1%, the PE-MIU biometric performance is up to seven times lower than the PFRNet performance over similar databases, i.e., LFW and Adience. In particular, for a practical scenario (FMR$=$ 0.1%), PE-MIU rejects only approximately 0.17% and 2.35% of the mated samples over these two challenging databases, respectively. Overall, both systems obtain impressive performance rates across different databases which are generally retained in their privacy-enhanced versions.

In the second experiment, the gender prediction performance of both both approaches, i.e., PRF-Net and PE-MIU, is explored. To that end, machine learning-based gender classifiers are trained on original face embeddings and privacy-enhanced templates obtained by both approaches in cross-database scenarios, e.g., training on LFW and gender prediction on Adience, where the number of subjects for each gender attribute is also balanced (see Figure 4). In Table IV, the gender prediction performance is reported for different classic classifiers, e.g., kNN and SVM. Here, SVM is employed by training different kernels (Poly, RBF, and Sigmoid). Note that hyper-parameters of both classifiers were set to basic configurations without optimization.

A significant degradation of the gender prediction performance is observable for privacy-enhanced templates compared to unprotected templates. Lowest average gender prediction accuracy of 52.37% is obtained by PE-MIU, in contrast to 65.22% for PFRNet, over similar cross-database scenarios (i.e., training on LFW and Adience to predict on Adience and LFW respectively). These results indicate that machine learning-based classifiers are not able to reliably predict gender from privacy-enhanced templates. This is further supported by looking at visualizations obtained by dimensionality reduction tools. Examples using t-distributed stochastic neighbor embeddings (t-SNE) [33] are depicted in Figure 6. It can be observed that in their original embeddings, faces are clustered with respect to gender, which is not the case for the privacy-enhanced templates. At this point, it is important to repeat that such observations are the basis for reporting high level of soft-biometric privacy in some published works, e.g., in [14], [15], [17], [18].



(a) PFRNet (b) PE-MIU

Fig. 5. Boxplots of similarity scores for non-mated comparison trials of privacy-enhanced templates with same and different soft-biometric attributes for both algorithms on the LFW database. Comparison trials for the same attribute (gender) yield slightly higher similarity scores and more outliers compared to those for different attributes.

B. Vulnerability Analysis

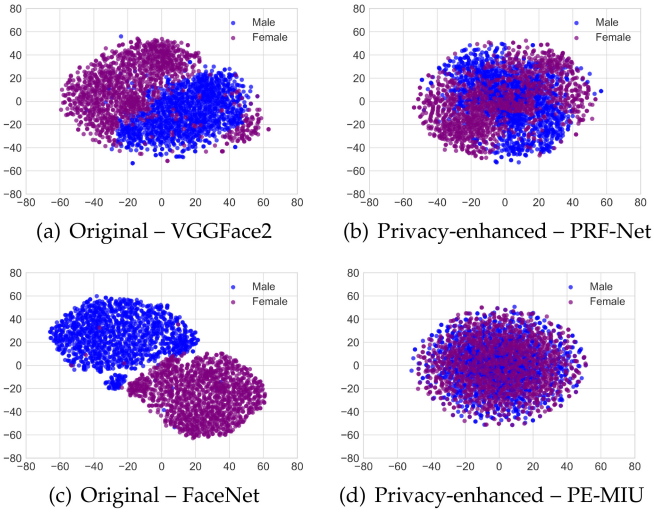
In a first experiment, it is analyzed whether the propositions about the properties of privacy-enhanced templates hold. Figure 5 depicts examples of similarity scores for non-mated comparison trails of privacy-enhanced templates with same and different soft-biometric attributes for both used methods (analogous to Figure 3). Like in the original unprotected systems, "same attribute" boxplots are shifted towards higher similarity score. In addition, facial image pairs which produce high similarity scores when comparing their corresponding privacy-enhanced templates have been visually inspected. Examples of samples and top-ranked samples that obtain high similarity scores in non-mated comparison of privacy-enhanced template are depicted in Figure 7. It can be seen that with high probability the gender of top-ranked

TABLE V
ATTACK SUCCESS RATES OF THE ATTACK EMPLOYING THE MAJORITY VOTING STRATEGY (IN %)

Method	Attacker	Target	Attack success rate					
			$n=1$	$n=5$	$n=11$	$n=51$	$n=101$	$n=201$
PFRNet	LFW	Adience	77.14	82.15	**82.50**	80.17	78.62	76.13
	Adience	CelebA	76.69	79.71	**80.91**	80.12	79.08	77.63
		LFW	75.29	78.30	**78.36**	76.48	74.56	70.14
	CelebA	Adience	70.56	**71.83**	71.06	66.32	63.75	60.35
PE-MIU	LFW	Adience	86.51	**87.99**	87.67	85.03	82.58	78.58
		ColorFeret	83.44	84.18	**84.33**	81.53	78.73	72.74
	Adience	LFW	86.05	87.40	**87.92**	84.59	82.88	79.96
		ColorFeret	79.54	81.57	**81.73**	77.98	74.75	70.01
	ColorFeret	LFW	91.03	92.39	**92.80**	91.98	89.13	86.01
		Adience	84.78	**87.91**	87.77	86.96	84.92	84.38

TABLE VI
ATTACK SUCCESS RATES OF THE ATTACK EMPLOYING THE AVERAGING STRATEGY (IN %)

Method	Attacker	Target	Attack success rate					
			$n=1$	$n=5$	$n=10$	$n=50$	$n=100$	$n=200$
PFRNet	LFW	Adience	77.14	83.36	**83.59**	81.61	79.74	77.41
	Adience	CelebA	76.69	80.49	**81.37**	81.32	80.70	78.77
		LFW	75.29	79.45	**79.97**	77.47	75.81	72.06
	CelebA	Adience	70.56	**72.71**	72.35	67.84	64.98	61.80
PE-MIU	LFW	Adience	86.51	89.74	90.40	88.65	86.16	83.16
		ColorFeret	83.44	86.82	**86.98**	84.56	81.77	78.38
	Adience	LFW	86.05	88.60	**89.02**	87.82	86.05	83.71
		ColorFeret	79.54	83.08	**83.13**	81.05	78.87	76.58
	ColorFeret	LFW	91.03	92.39	**93.89**	93.21	91.58	89.40
		Adience	84.78	87.91	87.77	**88.59**	86.96	85.73



(a) Original – VGGFace2 (b) Privacy-enhanced – PRF-Net

(c) Original – FaceNet (d) Privacy-enhanced – PE-MIU

Fig. 6. Visualization of original (unprotected) and privacy-enhanced (protected) face representations over the LFW database using t-SNE.

samples is the same as that of the leftmost sample. This suggests that the effect of broad homogeneity still exists in the protected domain.

In the second experiment, the different types of attacks are launched to derive the gender attribute from privacy-enhanced templates. It is important to note that all the attack strategies are analyzed in cross-database scenarios. In the first step, the attack is applied using the majority-based voting strategy to derive gender from privacy-enhanced templates. Obtained results are summarized in Table V where best obtained results for each cross-database scenario are marked bold. Scenarios in which a Web-collected face image database (Adience, LFW, or CelebA) are used in the attack is considered most relevant since an attacker could effortlessly access and collect such images. Employing the majority voting-based strategy, the attacker obtains the gender attribute from the n odd best scores. Highest attack success rates are achieved for employing a small number of $n = 11$ best scores. The average obtained attack success rates for this attack strategy lies around 85% which is clearly above that achieved by machine learning-based classifiers (*c.f.* Table IV on the right hand side for privacy-enhanced templates).

Table VI, Table VII, and Table VIII list the attack success rates for the averaging strategy. Again, best attack success rates for each cross-database experiment are marked bold. In this attack strategies n best scores against male and female subjects form the attacker database are averaged and compared to obtain the gender attribute from privacy-enhanced templates. For averaging without weights, competitive attack success rates are achieved for considering $n = 10$ best males and females scores. Overall, slight improvements (up to approximately two percent points) are observable when comparing the averaging strategies to the majority voting-based strategy. Further, in case scores are weighted w.r.t. their rank, higher

Fig. 7. Ranked examples of samples that reach a high similarity score in a non-mated comparison of privacy-enhanced templates of PFRNet (first and second row) and PE-MIU (third and fourth row); images taken from the LFW database.

TABLE VII
ATTACK SUCCESS RATES OF THE ATTACK EMPLOYING THE LINEARLY WEIGHTED AVERAGING STRATEGY (IN %)

Method	Attacker	Target	Attack success rate					
			$n=1$	$n=5$	$n=10$	$n=50$	$n=100$	$n=200$
PFRNet	LFW	Adience	77.14	82.08	**83.67**	82.78	81.18	79.12
	Adience	CelebA	76.69	79.81	81.48	**81.79**	81.32	80.12
		LFW	75.29	78.82	**79.60**	78.67	76.80	74.92
	CelebA	Adience	70.56	**72.84**	72.54	69.45	66.97	64.06
PE-MIU	LFW	Adience	86.51	89.93	90.16	**90.32**	88.37	85.73
		ColorFeret	83.44	86.20	**87.40**	85.93	84.21	81.30
	Adience	LFW	86.05	88.44	**88.91**	88.55	87.61	85.74
		ColorFeret	79.54	82.04	**83.29**	82.72	81.10	78.81
	ColorFeret	LFW	91.03	91.85	93.34	**93.75**	92.93	91.30
		Adience	84.78	87.50	87.91	**89.67**	88.45	87.09

TABLE VIII
ATTACK SUCCESS RATES OF THE ATTACK EMPLOYING THE LOGARITHMICALLY WEIGHTED AVERAGING STRATEGY (IN %)

Method	Attacker	Target	Attack success rate					
			$n=1$	$n=5$	$n=10$	$n=50$	$n=100$	$n=200$
PFRNet	LFW	Adience	77.14	81.18	83.24	**83.44**	82.19	81.07
	Adience	CelebA	76.69	79.19	80.39	**82.15**	81.32	81.11
		LFW	75.29	78.67	78.98	**79.45**	78.04	76.27
	CelebA	Adience	70.56	72.48	**72.89**	70.79	68.62	66.31
PE-MIU	LFW	Adience	86.51	89.39	90.12	**90.75**	90.05	88.34
		ColorFeret	83.44	85.46	**87.05**	86.90	85.65	83.59
	Adience	LFW	86.05	88.18	88.81	**88.86**	88.34	87.30
		ColorFeret	79.54	81.31	82.82	**83.08**	82.61	80.95
	ColorFeret	LFW	91.03	91.71	92.26	**94.43**	94.02	92.93
		Adience	84.78	87.50	87.91	**89.54**	88.86	88.32

values of $n \geq 50$ can reveal improved attack success rates (around one percent point) compared to the simple averaging. Moreover, by weighting the scores the attack is expected to become more robust, i.e., less sensitive to n.

The mentioned disadvantage for the attacker (overlapping identities in the training database of the privacy enhancement method and the attacker's database) becomes clear for the scenario where the privacy-enhanced templates produced by PFRNet are attacked using the CelebA as attacker database. Here, the attack success chances are generally lower compared to the other evaluated scenarios.

In summary, the obtained results confirm that both analyzed schemes, i.e., PE-MIU and PFRNet, are highly vulnerable to the proposed attack. For a better overview, Table IX

TABLE IX
SUMMARY OF THE BEST AVERAGE ATTACK SUCCESS RATES ACROSS ALL CROSS-DATABASE SCENARIOS (IN %)

Method	Attack success rate					
	$n = 1$	$n = 5$	$n = 10$	$n = 50$	$n = 100$	$n = 200$
PFRNet	74.92 ± 4.79	79.04 ± 7.08	79.50 ± 7.42	78.96 ± 9.06	77.54 ± 9.88	75.63 ± 12.79
PE-MIU	85.23 ± 3.97	88.12 ± 3.28	88.65 ± 3.68	88.95 ± 3.99	88.26 ± 4.08	86.91 ± 4.38

summarises the best average attack success rates across all cross-database scenarios for different values of *n* with a 95% confidence interval.

VI. DISCUSSION

This section discusses different relevant aspects of the attack. Section VI-A describes potential countermeasures against the attack. Alternative attack methods are briefly discussed in Section VI-B. The application of the proposed attack to systems based on other biometric characteristics is discussed in Section VI-C. Finally, different attack models are described in Section VI-D.

A. Attack Prevention

The proposed attack may be prevented by other techniques which meet the goal of protecting soft-biometric information by protecting biometric data entirely, i.e., biometric template protection scheme, as elaborated below:

Cancelable biometrics [6] obscure biometric signals by applying irreversible transformations to them. To achieve unlinkability, application- or subject-specific transformation parameters, i.e., keys, are employed. In case an attacker would be in possession of the key that was used to protect the biometric data, the presented attack could be performed offline. Note that key possession usually does not suffice to revert the protected biometric signal. If the attacker does not have the key, the proposed attack would only be applicable online, provided that a sufficiently large set of face images can be presented to the cancelable biometric system.

Biometric cryptosystems [7] do not return biometric comparison scores. In contrast, biometric cryptosystems retrieve keys which are validated and usually only released if these are correct, otherwise a failure message is returned. This means, to perform the proposed attack to a biometric cryptosystem, a certain amount of false matches would need to be achieved when presenting the set of biometric probe images to the system. Obviously, this would depend on the size of the image set the attacker is using and the false match rate the system is operated at. For the conducted experiments, Table X lists the average proportion of false matches for the best obtained score for decision thresholds corresponding to relevant false match rates in verification mode. It can be observed that for the conducted experiments only extremely low false match rates would considerably reduce the probability of false matches. However, if a biometric cryptosystem would return erroneous or random keys in case the key validation fail, the attacker may not be able to correctly identify false matches which in turn would prevent from the proposed attack.

TABLE X
RELATIVE AMOUNT OF FALSE MATCHES OBTAINED BY THE ATTACK IN RELATION TO VERIFICATION-BASED FALSE MATCH RATES (FMRS) (IN %)

Method	Attacker	Target	FMR	Threshold	FMs in Attack
PFRNet	LFW	Adience	0.001	0.84	0.08
			0.01	0.76	1.17
			0.1	0.61	99.22
	Adience	CelebA	0.001	0.70	60.93
			0.01	0.66	97.66
			0.1	0.60	100.00
		LFW	0.001	0.72	19.98
			0.01	0.67	79.97
			0.1	0.61	99.90
	CelebA	Adience	0.001	0.84	0.01
			0.01	0.76	0.73
			0.1	0.61	99.51
PE-MIU	LFW	Adience	0.001	0.83	0.16
			0.01	0.70	0.97
			0.1	0.42	100.00
		ColorFeret	0.001	0.62	3.42
			0.01	0.47	68.23
			0.1	0.37	100.00
	Adience	LFW	0.001	0.53	56.51
			0.01	0.44	99.95
			0.1	0.37	100.00
		ColorFeret	0.001	0.62	11.20
			0.01	0.47	78.49
			0.1	0.37	100.00
	ColorFeret	LFW	0.001	0.53	56.66
			0.01	0.44	100.00
			0.1	0.37	100.00
		Adience	0.001	0.83	0.54
			0.01	0.70	3.80
			0.1	0.42	100.00

Homomorphic encryption [10] requires a probe to be encrypted with a public key prior to comparing it to the reference in the encrypted domain. Subsequently, the comparison score is decrypted using the private key. Hence, an attacker would require the private key of the system in order to obtain comparison scores, which would be a prerequisite to launch the proposed attack. Under the assumption that an attacker has full access to private keys, a direct decryption of encrypted references could be performed. Subsequently, soft-biometric attributes could be reliably extracted from unprotected references. That is, if the secrecy of the private keys can be guaranteed in homomorphic encryption schemes, the presented attack can not be applied.

In summary, it can be argued that certain template protection mechanisms, in particular biometric cryptosystems and homomorphic encryption, prevent the presented attack while

under specific circumstances cancelable biometric systems are expected to be vulnerable to the attack. However, the latter assumption would require further investigations which are beyond the scope of this work.

B. Alternative Attacks

Apart from the proposed attack, facial soft–biometric privacy enhancement techniques may be vulnerable to further attacks. As previously mentioned, analyses on privacy protection capabilities of these methods have mostly been conducted by employing well-known machine learning-based classifiers, e.g., SVM. However, alternative classification methods based on different (machine learning-based) classifier might be capable of inferring soft-biometric information from privacy-enhanced templates. In addition, classifiers could be trained to retrieve unprotected soft-biometric attributes which are interrelated with a protected soft-biometric attribute. For instance, soft–biometric privacy enhancement methods may be circumvented by deriving gender from another unprotected attribute such as hairstyle or makeup.

C. Application to Other Characteristics

It is worth mentioning that the attack may only be applicable to biometric systems based on characteristics for which the effect of broad homogeneity is observable. It has been shown that biometric attributes can be derived from various popular biometric characteristics [34], e.g., fingerprints, iris, or voice. However, this does not necessarily mean that a biometric system based on such characteristics utilises these soft-biometric attributes for recognition purposes. For instance, it has recently been shown that the effect of broad homogeneity can not be observed for commercial iris recognition systems [21], while many researchers reported high accuracies for predicting soft-biometric attributes such as gender from iris images [34].

D. Attack Models

Different models exist for describing scenarios and assumptions of attacks on biometric information protection schemes which are standardized in [35]. The most restrictive model is referred to as *naïve model* in which an adversary has neither information of the underlying algorithm, nor owns a large biometric database. However, it has recently been argued that privacy-enhancing face recognition system should be analyzed under Kerckhoffs's principle [13]. In this *general model*, an adversary is assumed to possess full knowledge of the underlying algorithm. In addition, the adversary may have access to one or more privacy-enhanced templates from one or more databases. The adversary may also possess knowledge of the statistical properties of biometric features. In contrast, in the proposed attack, full knowledge of the underlying algorithm is not required, i.e., merely applying it as a black-box is sufficient. More precisely, the attack only requires the privacy-enhancing method as black-box and a small database. It is noteworthy that such a scenario is identical to a scenario in which a machine learning-based classifier would be trained to extract soft-biometric attributes from privacy-enhanced templates. The latter scenario is usually considered in the scientific literature for analysing privacy protection capabilities of soft-biometric privacy enhancement methods [12].

VII. Conclusion

We showed that in order to maintain biometric performance, privacy-enhancing face recognition methods have to retain certain properties of the original face recognition systems. This includes the well-documented effect of broad homogeneity [20], i.e., face recognition systems produced higher similarity scores for subjects which share certain soft-biometric attributes such as gender or race. Based on these observations an attack was proposed which can be performed offline with the minimal requirements that the algorithm is available as black box along with a small set of arbitrary face images. In experiments, high success rates were achieved for attacking two state-of-the-art algorithms for facial soft-biometric privacy enhancement. Such an attack may also be applicable to other schemes which are conceptually similar to the ones used in the experiments of this work. While the proposed attack is applied to infer gender information in this work, it can theoretically be applied to further protected attribute, e.g., age or race. It is concluded that the privacy protection capabilities of some facial soft–biometric privacy enhancement techniques are currently over-estimated in published works. Future research on this topic, therefore, needs to focus on more rigorous evaluations when assessing privacy protection capabilities of soft-biometric privacy-enhancing techniques and consider potential attacks, such as the one introduced in this work.

References

[1] R. Ranjan *et al.*, "Deep learning for understanding faces: Machines may be just as good, or better, than humans," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 66–83, Jan. 2018.

[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] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 5, pp. 1188–1202, May 2019.

[4] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper, "Beyond identity: What information is stored in biometric face templates?" in *Proc. Int. Joint Conf. Biometrics (IJCB)*, 2020, pp. 1–10.

[5] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper, "On soft-biometric information stored in biometric face embeddings," *IEEE Trans. Biom., Behav., Ident. Sci.*, vol. 3, no. 4, pp. 519–534, Oct. 2021.

[6] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 54–65, Sep. 2015.

[7] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: Issues and challenges," *Proc. IEEE*, vol. 92, no. 6, pp. 948–960, Jun. 2004.

[8] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–17, Jan. 2008.

[9] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," *EURASIP J. Inf. Security*, vol. 3, p. 3, Sep. 2011.

[10] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey, "Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 108–117, Mar. 2013.

[11] *ISO/IEC 24745:2011. Information Technology–Security Techniques–Biometric Information Protection*, ISO/IEC Standard JTC 1/SC 27 IT Security techniques, Jun. 2011.

[12] B. Meden *et al.*, "Privacy–enhancing face biometrics: A comprehensive survey," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4147–4183, 2021.
[13] P. Terhörst *et al.*, "Privacy evaluation protocols for the evaluation of soft-biometric privacy-enhancing technologies," in *Proc. Int. Conf. Biometrics Spec. Interest Group (BIOSIG)*, 2020, pp. 215–222.
[14] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised privacy-enhancement of face representations using similarity-sensitive noise transformations," *Appl. Intell.*, vol. 49, no. 8, pp. 3043–3060, 2019.
[15] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "SensitiveNets: Learning agnostic representations with application to face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2158–2164, Jun. 2021.
[16] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Suppressing gender and age in face templates using incremental variable elimination," in *Proc. Int. Conf. Biometrics (ICB)*, 2019, pp. 1–8.
[17] B. Bortolato *et al.*, "Learning privacy-enhancing face representations through feature disentanglement," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, 2020, pp. 495–502.
[18] P. Terhörst *et al.*, "PE-MIU: A training-free privacy-enhancing face recognition approach based on minimum information units," *IEEE Access*, vol. 8, pp. 93635–93647, 2020.
[19] P. Terhörst, M. Huber, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised enhancement of soft-biometric privacy with negative face recognition," 2020, *arXiv:2002.09181*.
[20] J. J. Howard, Y. B. Sirotin, and A. R. Vemury, "The effect of broad and specific demographic homogeneity on the imposter distributions and false match rates in face recognition algorithm performance," in *Proc. Int. Conf. Biometric Theory Appl. Syst. (BTAS)*, Sep. 2019, pp. 1–8.
[21] J. J. Howard, Y. B. Sirotin, J. L. Tipton, and A. R. Vemury, "Quantifying the extent to which race and gender features determine identity in commercial face recognition algorithms," 2020, *arXiv:2010.07979*.
[22] A. Kong, K.-H. Cheung, D. Zhang, M. Kamel, and J. You, "An analysis of biohashing and its variants," *Pattern Recognit.*, vol. 39, no. 7, pp. 1359–1368, 2006.
[23] W. J. Scheirer and T. E. Boult, "Cracking fuzzy vaults and biometric encryption," in *Proc. Biometrics Symp.*, 2007, pp. 1–6.
[24] P. Grother, M. Ngan, and K. Hanaoka, "Ongoing face recognition vendor test (FRVT) part 3: Demographic effects," Nat. Inst. Stand. Technol., Gaithersburg, MD, USA, Rep. NISTIR 8280, Dec. 2019.
[25] Y. B. Sirotin and A. R. Vemury (EAB, Amsterdam, The Netherlands). *Demographic Variation in the Performance of Biometric Systems: Insights Gained from Large-Scale Scenario Testing (Virtual Events Series–Demographic Fairness in Biometric Systems)*. (Mar. 2021). [Online]. Available: https://mdtf.org/publications/EAB2021-Demographics.pdf
[26] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, 2018, pp. 67–74.
[27] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 815–823.
[28] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Trans. Inf. Forensics Security*, vol. 9, pp. 2170–2179, 2014.
[29] 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 *Proc. Workshop Faces Real-Life Images Detection Alignment Recognit.*, Sep. 2008. [Online]. Available: https //cased-dms.fbi.h-da.de/literature/Huang-LabeledFaces InTheWild-HAL-Inria-2008.pdf
[30] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3730–3738.
[31] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[32] *ISO/IEC 19795-1:2021. Information Technology–Biometric Performance Testing and Reporting–Part 1: Principles and Framework*, ISO/IEC Standard JTC1 SC37 Biometrics, Jun. 2021.
[33] L. V. der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
[34] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? a survey on soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, pp. 441–467, 2016.
[35] *ISO/IEC 30136:2018 Information Technology–Performance Testing of Biometric Template Protection Schemes*, ISO/IEC Standard JTC 1/SC 37 Biometrics, 2018.

Dailé Osorio-Roig (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from the Technological University of Havana, in 2014. She is currently pursuing the Ph.D. degree with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. She joined the Advanced Technologies Application Center (CENATAV), Havana, Cuba, for computer science graduate training. She is a member of the da/sec—Biometrics and Internet Security Research Group and the National Research Center for Applied Cybersecurity (ATHENE), Germany. Her principal research interests are focused in areas of pattern recognition, biometrics and machine learning, specifically, biometric indexing, and privacy-enhancing technologies.

Christian Rathgeb is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He is a Principal Investigator with the National Research Center for Applied Cybersecurity ATHENE. He has 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 European Association for Biometrics (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 EAB, the Program Chair of the International Conference of the Biometrics Special Interest Group (BIOSIG), and an Editorial Board Member of *IET Biometrics*. 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*).

Pawel Drozdowski worked as a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He has coauthored over 35 technical publications in the field of biometrics. His research interests include biometrics, information security and privacy, pattern recognition, and algorithmic fairness. He won the European Biometric Industry Award of the European Association for Biometrics in 2021, the Best Student Paper Runner-Up Award (WIFS'18), the Best Poster Award (BIOSIG'19), and the Best Paper Award (WIFS'21).

Philipp Terhörst received the Master of Science degree in physics from the Technical University of Darmstadt in 2017, and the Ph.D. degree in computer science in 2021 for his work on "Mitigating Soft-Biometric Driven Bias and Privacy Concerns in Face Recognition Systems". He is currently working as an ERCIM "Alain Bensoussan" Fellow with the Norwegian Institute for Science and Technology. He has been working with the Smart Living and Biometric Technologies Department, Fraunhofer Institute for Computer Graphics Research (IGD) as a Research Scientist and as a Ph.D. student with the Technical University of Darmstadt. He has authored several publications in conferences and journals, such as CVPR and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and regularly works as a reviewer for e.g., IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, CVPR, PR, BTAS, and ICB. His areas of specialization include topics in machine learning as well as biometric face recognition with a focus on quality assessment, privacy, and fairness. For his scientific work, he received several awards, such as the 2020 EAB Biometrics Industry Award from the European Association for Biometrics for his dissertation or the IJCB 2020 Qualcomm PC Chairs Choice Best Student Paper Award. He furthermore participated in the "Software Campus" Program, a management program of the German Federal Ministry of Education and Research (BMBF).

Vitomir Štruc (Senior Member, IEEE) received the Doctoral degree from the Faculty of Electrical Engineering, Ljubljana, in 2010. He is a Member of the University of Ljubljana, Slovenia. His research interests include problems related to biometrics, computer vision, image processing, pattern recognition, and machine learning. He has (co-)authored more than 150 research papers for leading international peer reviewed journals and conferences in these and related areas. He served in different capacities on the organizing committees of several top-tier vision conferences, including IEEE Face and Gesture, ICB, WACV, and IJCB. He is a Senior Area Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, a Subject Editor of *Signal Processing* (Elsevier), and an Associate Editor of *Pattern Recognition* and *IET Biometrics*. He served as an Area Chair for WACV 2018, 2019, 2020, ICPR 2018, Eusipco 2019, and FG 2020, 2021 and as the Program Cahir for IJCB 2020. He is currently the Program Co-Chair of IWBF 2022, the Workshop Co-Chair of WACV 2022, and the Competition Co-Chair of IJCB 2022. He is a member of IAPR, EURASIP, and Slovenia's National Contact Point for the EAB and the Former President and Current Executive Committee Member of the Slovenian Pattern Recognition Society and the Slovenian branch of IAPR. He also serves as the VP Technical Activities for the IEEE Biometrics Council from 2022 to 2024.

Christoph Busch (Senior Member, IEEE) is a Member of the Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with Hochschule Darmstadt (HDA), Germany. He has been lectures Biometric Systems with Denmark's DTU 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 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 the Co-Founder of the European Association for Biometrics. He has 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 convenor of WG3 in ISO/IEC JTC1 SC37.