ASPECD: Adaptable Soft-Biometric Privacy-Enhancement Using Centroid Decoding for Face Verification

Peter Rot[1], Philipp Terhörst[2], Peter Peer[1], Vitomir Štruc[1]
[1]University of Ljubljana, Ljubljana, Slovenia; [2]University of Paderborn, Paderborn, Germany

***Abstract*— State-of-the-art face recognition models commonly extract information-rich biometric templates from the input images that are then used for comparison purposes and identity inference. While these templates encode identity information in a highly discriminative manner, they typically also capture other potentially sensitive facial attributes, such as age, gender or ethnicity. To address this issue, Soft-Biometric Privacy-Enhancing Techniques (SB-PETs) were proposed in the literature that aim to suppress such attribute information, and, in turn, alleviate the privacy risks associated with the extracted biometric templates. While various SB-PETs were presented so far, existing approaches do not provide dedicated mechanisms to determine which soft-biometrics to exclude and which to retain. In this paper, we address this gap and introduce ASPECD, a modular framework designed to selectively suppress binary and categorical soft-biometrics based on users' privacy preferences. ASPECD consists of multiple sequentially connected components, each dedicated for privacy-enhancement of an individual soft-biometric attribute. The proposed framework suppresses attribute information using a Moment-based Disentanglement process coupled with a centroid decoding procedure, ensuring that the privacy-enhanced templates are directly comparable to the templates in the original embedding space, regardless of the soft-biometric modality being suppressed. To validate the performance of ASPECD, we conduct experiments on a large-scale face dataset and with five state-of-the-art face recognition models, demonstrating the effectiveness of the proposed approach in suppressing single and multiple soft-biometric attributes. Our approach achieves a competitive privacy-utility trade-off compared to the state-of-the-art methods in scenarios that involve enhancing privacy w.r.t. gender and ethnicity attributes. Source code will be made publicly available.**

I. Introduction

Face verification templates, extracted from input face images through state-of-the-art (SOTA) convolutional neural networks (CNNs) typically encode a wide variety of facial attributes, ranging from identity to different soft-biometrics, such as gender, age or ethnicity [5], [21], [27], [28], [32]. This characteristic represents a significant privacy risk, as potentially sensitive soft-biometric information can easily be extracted and exploited for purposes different from identity recognition, e.g., discrimination, targeted marketing, or user profiling [1], [6], [9], [14], [20]. Subjects, enrolled in a biometric recognition system, should therefore ideally have the ability to control which soft-biometric information can be utilized during the recognition process and which excluded, allowing them to (i) provide explicit consent on the use of specific soft-biometric information, and (ii) minimize the associated privacy risks [17]–[19].

Supported in parts by the ARIS Programmes P2-0250 "Metrology and Biometric Systems" and P2-0214 "Computer Vision", and ARIS Research Project J2-50065 "DeepFake DAD".

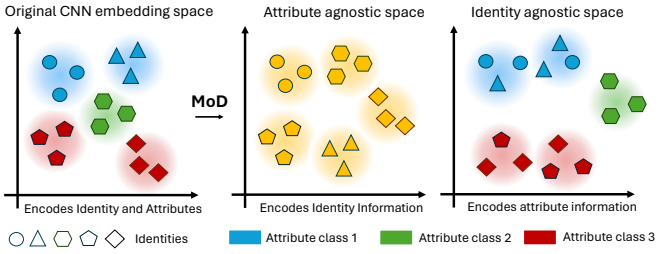

Fig. 1. **Illustration of the Moment-based Disentaglement (MoD) process**. In this paper, we present a novel approach towards suppressing soft-biometric information in face verification templates (i.e., CNN embeddings), called **ASPECD**, that relies on the MoD procedure to split the information encoded in the original CNN embeddings into two distinct components, where the first encodes only identity, but not the targeted soft-biometric attribute, whereas the second encodes only attribute, but not identity information. By manipulating the attribute part, ASPECD effectively suppresses the selected soft-biometric information in the input templates, while making it feasible to suppress multiple attributes one after the other during the privacy-enhancement process.

To address the privacy concerns related to the soft-biometric information encoded in the embedding space of modern CNN-based face recognition models, the research community is increasingly looking into so-called **Soft-Biometric Privacy-Enhancing Techniques** (SB-PETs) that aim to remove (or suppress) sensitive information from face verification templates/embeddings, while maintaining their discriminative power and, in turn, high recognition performance [14]. The interest in SB-PETs has additionally been fueled by stricter privacy regulations, such as GDPR [8], which stipulate that *(i)* the collection and processing of personal data must be confined to what is strictly necessary, and that *(ii)* people must provide explicit consent for each specific purpose for which their data is to be utilized.

To avoid the computational burden of training face recognition models from scratch, modern SB-PETs [2] largely focus on privacy enhancements of face verification templates generated by pretrained face recognition models. While many methods have been proposed for the suppression of specific (individual) soft-biometric attributes [2], [30] or the enhancement of privacy in an unsupervised manner [29], [31], [33], there is a significantly smaller number of solutions dedicated to suppressing multiple soft-biometrics simultaneously [15]. Balancing privacy across multiple attributes poses a unique challenge [4], [15], [19], as individual soft-biometrics make varying contributions to the identity-related information,

each playing a distinct role in the recognition process [12], [13], [35]. Furthermore, the intrinsic correlations among soft-biometric attributes (e.g. beards reveal information about age and gender), add another layer of complexity to this problem by making it challenging to selectively remove or suppress one attribute without affecting others [10], [22].

Motivated by the above discussion, we present in this paper a novel state-of-the-art approach for privacy enhancement, capable of suppressing multiple (selected) soft-biometric attributes in face recognition templates with minimal impact on verification performance. At the core of the approach is an innovative **Mo**ment-based **D**isentaglement (MoD) procedure that splits the original templates/embeddings into two parts, where the first encodes only identity information (in an attribute agnostic manner), while the second part encodes only attribute information (in an identity agnostic manner), as illustrated in Fig. 1. The proposed MoD procedure is general, i.e., applicable to binary as well as categorical attributes, and allows us to separate identity information required for recognition purposes from potentially sensitive attribute information. To suppress attribute information, we replace the identity-agnostic part with an average attribute embedding, i.e., a centroid. Finally, we combine the attribute-agnostic and (modified) identity-agnostic representations and map them back into the initial embedding space using a dedicated decoder, as presented in Fig. 2. Because the attribute information in the decoded templates now corresponds to the centroid of the identity-agnostic embedding space, the original soft-biometric information is no longer inferable. Consequently, privacy w.r.t. the targeted attribute is enhanced in the decoded template.

To facilitate privacy enhancement for multiple attributes, we use a sequential approach and apply the outlined procedure for different attributes one after the other, as also illustrated in Fig. 2. This sequential approach has multiple benefits: (i) it allows suppressing a single attribute at the time, making the required disentanglement process easier to formulate, (ii) it allows processing an arbitrary number of attributes one after the other and adding novel components into the overall pipeline, and (iii) it allows the user to specify, which soft-biometric information to suppress and which to retain, enabling him/her to provide explicit consent on the use of his/her soft-biometrics during the recognition process. The proposed framework, named **A**daptable **S**oft-biometric **P**rivacy-**E**nhancement using **C**entroid **D**ecoding (ASPECD), is evaluated in rigorous experiments on a large-scale face dataset with gender and ethnicity attributes and in comparison to state-of-the-art techniques from the literature. The experimental results show that ASPECD leads to competitive performance w.r.t. privacy protection, while ensuring reasonably minor degradations in recognition results. In summary, we make the following contributions in this paper:

- We propose a novel framework, termed ASPECD, designed to enhance privacy w.r.t. multiple soft-biometrics within a unified and modular approach. Contrary to previous work, ASPECD allows to selectively suppress specific soft-biometric attributes, while retaining others. It also facilitates suppressing multiple attributes, while still ensuring competitive recognition performance.
- We introduce a moment-based disentanglement (MoD) procedure that splits the initial CNN embeddings into two parts by enforcing constraints on the distributions of the data in the two novel embedding spaces through statistical moment-based optimization objectives.
- We report a new state-of-the-art w.r.t. multi-attribute soft-biometric privacy enhancement through comprehensive experimental evaluation with five prominent face recognition CNNs: CosFace [36], FaceNet [25], ArcFace [7], AdaFace [11], and MagFace [16] and two targeted soft-biometric attributes, i.e., gender and ethnicity using a large scale face dataset.

II. Related work

Privacy-enhancing techniques, designed to remove, conceal or suppress soft-biometric information in face recognition templates, have recently gained considerable attention within the biometrics community [2], [14], [21], [30], [33]. In general, existing techniques can be grouped according to whether they target a single attribute, multiple attributes simultaneously, or try to suppress soft-biometric information without explicitly targeting selected attributes, i.e., in an unsupervised manner. Details on the groups are given below.

Single-attribute suppression. In [30], Terhörst *et al.* introduced an elimination technique, named Incremental Variable Elimination (IVE), that removes components from face recognition templates that contribute the most to the prediction of a chosen attribute, e.g., age or gender. Morales *et al.* [21] introduced a supervised model for privacy-enhancement, termed SensitiveNet. The model uses a triplet loss to learn a feature space with suppressed gender or age information and thus, ensures an increased level of soft-biometrics privacy. Most closely related to our work in terms of methodology is the PFRNet approach introduced by Bortolato *et al.* in [2], which also relies on a disentanglement scheme utilizing a statistical-moment based approach. However, PFRNet is only applicable to binary attributes, while our procedure works with categorical variables and includes the PFRNet disentanglement process as a special (simplified) case. Additionally, ASPECD relies on a privacy mechanism build around (ambiguous) centroid decoding, while PFRNet aims to exclude attribute information from the matching procedure of the recognition process. While all the techniques reviewed above may potentially be extended for soft-biometric privacy across multiple target attributes, the authors did not explore the effect of the proposed procedures on multiple attributes simultaneously and the impact of attribute correlations on privacy enhancement, something we address with ASPECD.

Unsupervised attribute suppression. Terhörst *et al.* [31] proposed a Cosine–Sensitive Noise (CSN) transformation, where a specific type of noise is injected into the face representations such that soft–biometric information is masked, while identity information is not. The authors demonstrated the feasibility of their approach on gender and age attributes, with encouraging results. Similarly, in [33], the

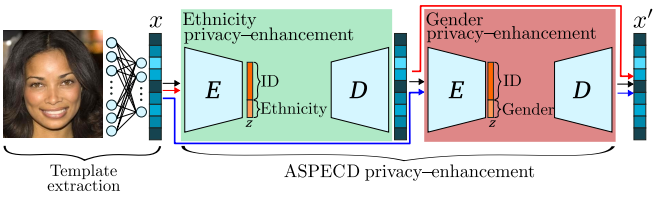

Fig. 2. **High–level overview of ASPECD**, the proposed framework for suppression of multiple soft-biometric attributes in face verification templates. ASPECD is designed in a modular manner and consists of a series of sequential modules that suppress a single attribute each in the input template x and ultimately produces a privacy-enhanced template x', from which the preselected soft-biometrics (i.e., gender and ethnicity in the depicted example) cannot be inferred reliably.

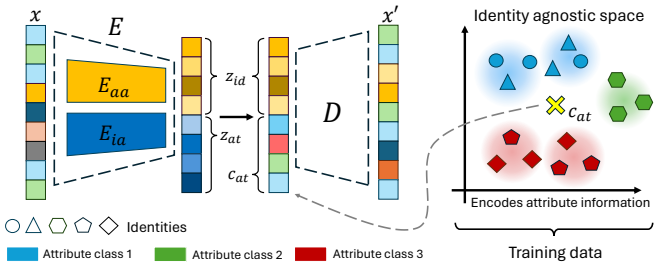

Fig. 3. **Visualization of an ASPECD module.** Each ASPECD module consists of a pair of encoders that maps the input face template to a pair of latent representation that exclusively encode either identity or attribute information. To achieve suppression of a soft-biometric attribute, a centroid representation is swapped into the combined latent representation $z = z_{id} \oplus c_{at}$ and then decoded into the privacy-enhanced template x'.

same authors presented an approach to soft-biometric privacy enhancement that exploited a special type of template coding with minimum information units. Although these approaches might effectively suppress multiple attributes, they lack a controllable mechanism to accommodate varying privacy preferences, which we develop for ASPECD.

Multiple-attribute suppression. As can be seen from the above discussions, various SB-PETs have been proposed and deployed to suppress individual soft-biometric attributes in face verification templates [2], [21], [30]. Nonetheless, only limited work has been done on enhancing privacy across multiple attributes simultaneously in a supervised and controllable manner. One of the few exceptions is the Multi-IVE technique, recently introduced by Melzi *et al.* in [15]. Multi-IVE is an extension of the IVE technique [30] with several enhancements. Instead of directly eliminating components from a feature vector, Multi-IVE transforms templates into a decorrelated subspace using either PCA or ICA. Within this subspace, it estimates the importance of various dimensions to identify the most informative w.r.t. multiple soft-biometric attributes. Subsequently, it sets the identified dimensions to zero and applies the inverse operation (i.e., a reprojection) to obtain privacy-enhanced templates in the original embedding space. While Multi-IVE leads to competitive results, it still removes potentially valuable information from the face templates (due to the entanglement of identity and attributes) that may lead to sub-optimal utility-privacy trade-offs. With ASPECD we address this point through an effective disentanglement process and a sequential treatment of soft-biometrics that can better deal with the correlations among various attributes.

III. Methodology

In this section, we now present the main contributions of this work, i.e.: the Adaptable Soft-biometric Privacy-Enhancement with Centroid Decoding (ASPECD) and the Moment-based Disentaglement (MoD) process that forms the basis for suppression of various soft-biometric attributes.

A. High-level overview

A high-level overview of ASPECD is presented in Fig. 2. As can be seen, the goal of ASPECD is to suppress multiple selected attributes, which are typically encoded in face verification templates, so that a potential attacker is not able to extract sensitive soft-biometric information from the privacy-enhanced templates. The proposed framework consists of multiple sequential modules, each optimized to target one soft-biometric attribute (e.g. gender or ethnicity). Starting with an input image $I \in \mathbb{R}^{w \times h}$ and using a pretrained face-recognition network ψ, a template $x = \psi(I) \in \mathbb{R}^d$ is commonly first extracted and then used for comparison purposes in a typical recognition pipeline. This template is primarily optimized for encoding identity, but, as demonstrated by prior work [14], also implicitly captures various types of soft-biometric information. The goal of ASPECD is, therefore, to transform x into a privacy-enhanced version x', from which selected soft-biometrics (gender and/or ethnicity) cannot be extracted reliably. The adaptable nature of ASPECD allows defining user-specific privacy preferences, so that the face template undergoes selective processing, with users requesting more privacy triggering multiple modules, while those desiring less privacy, activating fewer modules. Note that Fig. 2 shows three distinct pathways that illustrate this characteristic: (i) a red pathway for suppressing gender information only, (ii) a blue pathway for suppressing ethnicity information only, and (iii) a black pathway for suppressing both gender and ethnicity simultaneously. The sequential order of these modules may vary based on the requirements of the application at hand. In the following sections, we elaborate on the technical details behind the architectural design of the ASPECD modules (§III-B) and the MoD procedure (§III-C) that jointly with the centroid decoding process (§III-D) ensures effective attribute suppression.

B. Design of ASPECD modules

ASPECD is a modular framework that can be implemented with an arbitrary number of modules. Without loss of generality, we limit the following discussions on two such modules, one for suppressing gender and one for suppressing ethnicity information, as also shown in Fig. 2.

Each ASPECD module is designed as an autoencoder $D(E(x)) = x'$ that takes the original face template as input and then produces a modified template x' with suppressed attribute information, as shown in Fig. 3. The autoencoder consists of a two-path encoder E with two dedicated sub-encoders: (1) E_{ia} that maps x into identity-agnostic latent

TABLE I

SUMMARY OF TEST DATASET CONFIGURATION[†].

Dataset	#Images	#Subj.	#Mated	#Non-mated	#IDs (Gender)		#IDs (Ethnicity)				
					f	m	A	W	I	B	LH
VGGFace2*	7,760	388	73,720	7,471,200	194	194	84	78	98	78	52

[†]Totals over all 4 test data splits; f – female, m – male, A – asian, W – white, I – indian, B – black, LH – latino-hispanic

space $z_{at} = E_{ia}(x)$ (representing either gender or ethnicity in this context), and (2) E_{aa} that maps the input template into the attribute-agnostic identity space $z_{id} = E_{aa}(x)$. During training, a single-path decoder D is also optimized to reconstruct the concatenated latent representation $z = z_{id} \oplus z_{at}$ into the template x'. The described module design makes it possible to make attribute information in the reconstructed template x' ambiguous by manipulating the latent representation that encodes only the targeted soft-biometrics, i.e., z_{at}. With ASPECD, we achieve this by swapping the attribute representation z_{at} with the centroid of the training data in the identity-agnostic latent space c_{at} before decoding it into $x' = D(z_{id} \oplus c_{at})$, as we discuss in §III-D.

C. Moment-based Disentaglement (MoD) process

The key component of ASPECD is the disentanglement procedure that splits the input face template x into two latent representations, z_{at} and z_{id}. To facilitate the disentanglement and enforce the desired characteristics on the two latent spaces, we design a series of dedicated objective functions/losses for the training procedure. The first is a self-supervised reconstruction loss that ensures that all information from the input side is also present on the output side of the autoencoder. This loss hence allows us to control the information flow through the two latent spaces, i.e.:

$$\mathcal{L}_0 = ||x - D(x) \circ E(x)||_{L_2}\,, \tag{1}$$

where $||\cdot||_{L_2}$ is the L_2 norm and $\circ$ is a composition operator. Inspired by the success of PFRNet [2], we design the optimization objectives for the disentaglement process based on statisticial moments, but keep the process general, so it is applicable to categorical attributes (e.g. ethnicity), and not just binary ones as in [2]. To this end, we take into account all possible binary pairs of probability distributions associated with the given categorical attribute with n classes $c_1, c_2, \ldots, c_n$. During disentaglement, the goal is to prevent information about the categorical soft-biometrics from being encoded in z_{id} and to ensure that the distributions $Q(z_{id}, c_i)$ and $Q(z_{id}, c_j)$ are as similar as possible for all non-identical pairs (c_i, c_j), where $i, j \in 1, 2, \ldots, n$ and $i \neq j$. To achieve this, we define the $\mathcal{L}_\alpha$ loss term as follows:

$$\mathcal{L}_\alpha = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left|\left| \langle z_{id}^{\alpha} \rangle_{c_i} - \langle z_{id}^{\alpha} \rangle_{c_j} \right|\right|_{L_2}\,, \tag{2}$$

where $\langle z_{id}^{\alpha} \rangle_{c_i}$ is the α-order statistical moment of the class-conditional distribution of the identity representation z_{id}. The loss, thus, forces the statistical moments of all identities regardless of their attributes to be as close as possible.

Conversely, for the attribute distributions $Q(z_{at}, c_i)$ and $Q(z_{at}, c_j)$, we aim to make them as discriminative as possible for each i and j, so that all attribute-related information is ultimately encoded in z_{at}, and not z_{id}. During training, we therefore sample batches from each pair of the distributions $Q(z_{at}, c_i)$ and $Q(z_{at}, c_j)$ and compute β-order moments from the sampled data. To maximize the discrimination between the classes c_i and c_j, we then define $\mathcal{L}_\beta$ as follows:

$$\mathcal{L}_\beta = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \exp\left\{ - \frac{\left| \langle z_{at}^{\beta} \rangle_{c_i} - \langle z_{at}^{\beta} \rangle_{c_j} \right|^2}{2\sigma_\beta^2} \right\}, \tag{3}$$

where σ_β is an open hyper-parameter that defines the standard deviation of the Gaussian-shaped loss.

The overall MoD loss is finally defined as follows:

$$\mathcal{L} = \mathcal{L}_0 + \sum_{\alpha=1}^{\alpha_{max}} \lambda_\alpha \mathcal{L}_\alpha + \sum_{\beta=1}^{\beta_{max}} \lambda_\beta \mathcal{L}_\beta, \tag{4}$$

where α_{max} and β_{max} denote the maximum order of the moments used during the disentanglement, and λ_α and λ_β correspond to balancing weights.

D. Privacy Preservation using Centroid Decoding

Once the input template x is split into the latent representations z_{id} and z_{at}, we manipulate the attribute representation z_{at} to suppress soft-biometric information. Specifically, we replace the attribute representation z_{at} with the centroid of the training data c_{at}, as also shown in Fig. 3, and then decode the identity representation and centroid into a reconstructed template $x' = D(z_{id} \oplus c_{at}) \in \mathbb{R}^d$ with enhanced privacy. Because all templates x' are reconstructed with the same attribute information, the original attributes are effectively suppressed. With the decoding procedure, we maintain a consistent length of the templates, ensuring $|x| = |x'|$, where $|\cdot|$ is the cardinality operator. This enables us to directly compare privacy-enhanced templates with different privacy-preferences, as they all closely resemble the original space. The entire procedure is visually illustrated in Fig. 3.

IV. EXPERIMENTS

A. Datasets and experimental setup

Datasets. For the experiments, we use the VGGFace2 [3] dataset, beacuse it includes a variety of diverse demographic groups. Specifically, we subsample the original VGGFace2 dataset to roughly balance the data across demographic factors for both training and testing purposes and refer to the data as VGGFace2*. To train all considered privacy-enhancing techniques, we use the VGGFace2* training set consisting of $64,032$ images across 717 identities, balanced

with respect to gender and representing 5 ethnic groups (Asian, White, Indian, Black, and Latino-Hispanic). Details on the testing data are provided in Table I. In the Supplementary material, we provide a list of images from the VGGFace2* dataset to ensure reproducibility and also present results on an additional test dataset.

Verification templates. We select five SOTA pretrained and publicly available face recognition models for the extraction of face templates in the experiments. The models were optimized using different loss functions, namely, CosFace [36], ArcFace [7], AdaFace [11], MagFace [16], and FaceNet [25]. Prior to template extraction, face images were cropped and aligned using MTCNN [37], and finally resized to 224×224 pixels to fit the models' architectures. Additional details on the models can be found in the Supplementary material.

B. Performance measures

Verification performance. In the experiments, we first assess the baseline verification performance on the unmodified original test images. Here, we follow the international ISO standard [26] and report the EER and FNMR at specific FMR values, specifically at 10^{-2} and 10^{-3}. To assess privacy-enhancement, we recompute these performance indicators on the templates with suppressed attributes and report their relative change (RC), as advocated in [14].

Extractable Soft-Biometrics. To quantify the maximum potential information leakage, i.e., the maximum amount of attribute information that is inferable from the face templates, we train different soft-biometric classification models (SVM, MLP, and LR), which typically extract varying levels of soft-biometric information from the original templates and their privacy-enhanced versions. We then select the best performing model as the worst-case scenario in terms of information leakage. Thus, for a chosen soft-biometric attribute a, we report the highest AUC_a, calculated as:

$$AUC_a = \max(AUC_{a,\text{SVM}}, AUC_{a,\text{MLP}}, AUC_{a,\text{LR}}). \quad (5)$$

In our specific case, $a \in \{\text{gender, ethnicity}\}$. We use the notation AUC_g for the maximal gender AUC and AUC_e for the maximal ethnicity AUC (Area Under the ROC Curve).

Suppression of soft-biometric information. In the context of SB-PETs, the suppression rate (SR) is commonly used to evaluate the effect of privacy-enhancement on soft-biometric classifiers [2], [31]. To calculate the SR with respect to a single attribute, we utilize the equation from [24]:

$$SR_a = \frac{-1}{(AUC_{ao} - 0.5)}(AUC_{ap} - AUC_{ao}) \in [0, 1], \quad (6)$$

where AUC_{ao} measures the performance of the attribute classifier on the original (unmodified) templates, and AUC_{ap} measures the performance after privacy enhancement. In the case of non-binary attributes (i.e., ethnicity, in our experiments), AUC scores are calculated using the one-versus-rest approach using macro-averaging.

When considering the privacy enhancement of both gender (g) and ethnicity (e), we compute the total SR, treating both attributes as equally important, using the following equation: $SR = 0.5(SR_g + SR_e)$.

C. Experimental setup

We trained ASPCED and the competing models using the VGGFace2* training set. To evaluate performance, we partition the testing data of VGGFace2* into four folds, ensuring that there is no overlap in IDs among the folds. These folds are also designed to maintain a rough balance with respect to ethnicity and gender. Finally, we conduct the following investigations in the experimental part of the paper:

1) **Baseline Performances:** In this set of experiments, we evaluate the verification performance on unmodified face templates and assess the extent, to which soft-biometrics can be extracted from these templates.
2) **Privacy Enhancement of Individual Soft-Biometrics:** In the first set of experiments, we evaluate the efficacy of ASPECD in scenarios, where a single attribute (gender or ethnicity) is selected for suppression. We compare ASPECD with methods utilizing single-attribute suppression (IVE and PFRNet) and an unsupervised approach (CSN).
3) **Privacy Enhancement of Multiple Soft-Biometrics:** Next, we enhance privacy w.r.t. two soft-biometric attributes by sequentially applying two ASPECD components. We evaluate this process in two different configurations: (i) by first suppressing gender and then ethnicity information, and (ii) vice versa, by first suppressing ethnicity and then gender information.
4) **Comparison to the State-of-the-Art:** In this experiment, we compare ASPECD with the state-of-the-art in multiple-attribute suppression, i.e., Multi-IVE. To facilitate a direct comparison, we set the operating points of Multi-IVE to match ours. We individually evaluate three scenarios - when matching the EER, when matching AUC_g and when matching AUC_e.

D. Implementation details

Optimization of ASPECD components. To optimize both ASPECD components, we use the same set of training parameters across all template extractors for both targeted soft-biometric attributes. We employ the Adam optimizer with a learning rate of 0.001 and betas set to 0.9 and 0.999, along with an epsilon value of $1e-8$. The model is trained using batches of $10,000$ verification templates until convergence, typically occurring at around 500 epochs. We use $\alpha_{max} = 2$ and $\beta_{max} = 2$, as advocated in [2].

Compared methods. We conduct a comparative analysis of ASPECD against IVE [30], PFRNet [2], CSN [31], and Multi-IVE [15] ensuring that the parameters are appropriately aligned with those used in the original publications.

V. Results

A. Baseline results on unmodified templates

Baseline results are essential for evaluating the performance of soft-biometric privacy-enhancing techniques and

TABLE II

BASELINE VERIFICATION PERFORMANCE ON VGGFACE2*.

Extractor	EER (↓)	FNMR@FMR10^{-2}(↓)	FNMR@FMR10^{-3}(↓)
CosFace	0.010 ± 0.001	0.011 ± 0.002	0.027 ± 0.003
FaceNet	0.024 ± 0.001	0.048 ± 0.004	0.180 ± 0.011
ArcFace	0.034 ± 0.002	0.069 ± 0.008	0.198 ± 0.024
AdaFace	0.013 ± 0.000	0.014 ± 0.001	0.019 ± 0.002
MagFace	0.013 ± 0.002	0.013 ± 0.002	0.018 ± 0.002

TABLE III

BASELINE ATTRIBUTE-CLASSIFIER PERFORMANCE ON VGGFACE2*

Extractor	Gender (AUC$_g$)	Ethnicity (AUC$_e$)
CosFace	0.991 ± 0.006	0.947 ± 0.009
FaceNet	0.991 ± 0.003	0.938 ± 0.008
ArcFace	0.991 ± 0.005	0.928 ± 0.010
AdaFace	0.724 ± 0.058	0.738 ± 0.018
MagFace	0.912 ± 0.026	0.875 ± 0.019

evaluate the privacy-utility trade-off. In the first set of experiments, we therefore explore the verification performance and amount of extractable soft-biometric information using the selected five face recognition models.

Baseline verification performance. Table II presents verification baselines for the five considered template extractors. The best performance in terms of EER and FNMR@FMR10^{-2} is achieved with CosFace. AdaFace and MagFace achieve similar performance, while MagFace is the best performer in terms of FNMR@FMR10^{-3}. ArcFace exhibits the worst performance.

Baseline soft-biometric extraction. To assess the amount of soft-biometric information in the initial templates, we train gender and ethnicity classifiers on the training part of VGGFace2* and then evaluate their performance on the corresponding test set. We report gender– and ethnicity-related AUC values in Table III. AdaFace features consistently demonstrate the lowest capacity for extracting gender and ethnicity information, while the remaining models perform similarly. Overall, these results show that all models are able to extract a significant amount of soft-biometric information from the raw, unprotected templates.

B. Evaluation of Individual Privacy-Enhancing Components

In this section, we examine ASPECD's performance when focusing on the suppression of a single attribute, specifically either gender or ethnicity. We compare it with three competing state-of-the-art methods: PFRNet, IVE, and CSN, and results are presented in Table IV. When targeting only gender, we observe that ASPECD's performance is comparable to that of PFRNet. ASPECD outperforms CSN on CosFace, ArcFace and MagFace. CSN achieves better SR_g on AdaFace with comparable verification performance. In all verification scenarios, ASPECD outperforms IVE. IVE outperforms ASPECD in terms of SR$_g$ in certain scenarios, such as on CosFace and ArcFace; however, this leads to significant degradation in verification performance, consequently reducing IVE's practical utility.

Shifting focus to experiments targeting only ethnicity, first

TABLE IV

COMPARISON OF SINGLE-ATTRIBUTE SUPPRESSION WITH STATE-OF-THE-ART APPROACHES.

Extractor	Approach	Targeting Gender		Targeting Ethnicity	
		†(↓)	SR$_g$(↑)	†(↓)	SR$_e$(↑)
CosFace	ASPECD (ours)	**0.047**	0.235	**0.054**	0.333
	PFRNet [2]	0.048	0.186	n/a	n/a
	IVE [30]	0.292	**0.459**	0.176	**0.385**
	CSN [31]	0.275	0.099	0.275	0.096
FaceNet	ASPECD (ours)	**0.181**	0.028	**0.177**	**0.221**
	PFRNet [2]	0.203	0.042	n/a	n/a
	IVE [30]	0.302	0.030	0.239	0.055
	CSN [31]	0.393	**0.078**	0.393	0.128
ArcFace	ASPECD (ours)	**0.265**	0.161	**0.253**	0.325
	PFRNet [2]	0.278	0.154	n/a	n/a
	IVE [30]	0.421	**0.287**	0.338	**0.384**
	CSN [31]	0.449	0.153	0.449	0.258
AdaFace	ASPECD (ours)	**0.033**	0.093	**0.031**	0.424
	PFRNet [2]	0.036	**0.050**	n/a	n/a
	IVE [30]	0.050	0.301	**0.031**	**0.583**
	CSN [31]	0.045	0.419	0.045	0.530
MagFace	ASPECD (ours)	**0.024**	0.497	**0.023**	0.604
	PFRNet [2]	**0.024**	**0.472**	n/a	n/a
	IVE [30]	0.034	0.568	0.024	**0.659**
	CSN [31]	0.033	0.060	0.045	0.530

† = FNMR@FMR10^{-3}

note that PFRNet cannot be evaluated as it does not support categorically defined attributes. In this set of experiments, ASPECD generally outperforms IVE and CSN in verification performance across all scenarios. Exceptions are AdaFace and MagFace, where it performs comparably to IVE. Note however that this often comes at the expense of slightly lower SR_e. On FaceNet, ASPECD outperforms IVE and CSN in terms of both, verification performance and SR$_e$.

C. Sequential Execution Testing

When enhancing multiple soft-biometrics, different sequences of ASPECD modules are theoretically available. In this section, we investigate whether and to what extent the order of ASPECD modules influences the overall performance. Considering gender and ethnicity, we evaluate both available sequential executions: E→G, where the ethnicity component E precedes the gender module G, and G→E, where the G precedes the E module. From the results in Table V, we can draw the following conclusions. First, we observe that the attribute that is targeted first is more substantially suppressed than the second attribute in the sequence, a pattern that remains consistent across all face recognition models considered. Furthermore, variations are noted among extractors in their encoding of gender and ethnicity. For example, MagFace shows a higher correlation in this regard compared to CosFace. Additionally, in all instances, the sequence G→E consistently leads to greater degradations in verification performance, suggesting that with certain sequences it is more challenging to effectively disentangle identity information from attribute information using our MoD procedure.

D. Visual analysis

To get an in-depth understanding of the privacy enhancement process, we visualize the different feature spaces

TABLE V

COMPARISON OF SEQUENTIAL EXECUTIONS E→G AND G→E IN TERMS OF SUPPRESSION RATE (SR) AND RELATIVE CHANGES OF GENDER AUC, ETHNICITY AUC, AND FNMR@FMR10^{-2}.

Sequence	Extractor	Gender△(↓)	Ethnicity△(↓)	SR(↑)	▽(↓)
E→G	CosFace	−9.7%	−21.2%	0.323	141.8%
	FaceNet	−3.0%	−14.9%	0.189	59.4%
	ArcFace	−5.6%	−20.0%	0.273	107.0%
	AdaFace	−4.4%	−23.0%	0.432	97.9%
	MagFace	−26.9%	−27.6%	0.619	56.2%
G→E	CosFace	−21.5%	−9.6%	0.319	1454.5%
	FaceNet	−10.5%	−9.4%	0.207	662.3%
	ArcFace	−17.0%	−15.6%	0.341	631.6%
	AdaFace	−23.7%	−19.0%	0.701	1128.6%
	MagFace	−31.0%	−27.7%	0.665	474.6%

△ – Relative Change in AUC
▽ – Relative Change in FNMR@FMR10^{-2}

involved in ASPCED using t-distributed Stochastic Neighbor Embedding (t-SNE). In Fig. 4, we present t-SNE projections [34] of E→G, offering insights into the clustering of demographic groups in *(i)* the unmodified templates x, *(ii)* the privacy-enhanced templates x', *(iii)* the identity-agnostic attributes embedding spaces belonging to either z_{at_e} or z_{at_g}.

Focusing on x in the first row with respect to gender, we observe distinct patterns. In the case of CosFace and ArcFace, the gender attribute can be visually separated quite easily, while for AdaFace, the classes are mixed, indicating more challenging separability. This aligns with the results when training soft-biometric classifiers. On the other hand, MagFace shows a more clustered female group in the center, with mixed features around. In the last row, where we examine z_{at_e} and ideally expect to see separated clusters, the observations differ. CosFace and FaceNet exhibit well-separated clusters, with red and blue mixed in the middle. For ArcFace, yellow and green are somewhat separated at the borders, while other colors mix in the center. For AdaFace and MagFace, we hardly discern any distinct patterns.

In general, we observe that the identity-agnostic embedding spaces offer solid attribute separation across all models, suggesting that modest of the attribute information is indeed encoded in this space as intended by the MoD procedure. Similarly, we see that the attributes in the privacy-enhanced spaces x' overlap more and are, therefore, more difficult to infer than from the initial representations x.

E. Comparison to SOTA

In this section, we conduct a comparative analysis between ASPECD and the state-of-the-art Multi-IVE method for enhancing privacy w.r.t. both gender and ethnicity. To ensure a fair comparison, we evaluate it under three scenarios by setting matching operating points.

Matching EER. From the results in Table VI, we see that in terms of suppressing attributes, ASPECD outperforms Multi-IVE on all extractors, except on CosFace. On CosFace at a matched EER of 0.062, Multi-IVE outperforms ASPECD with a SR = 0.483, while ASPECD scores a SR of 0.323. This disparity in performance is mainly due to the substantial impact on gender AUC_g, where Multi-IVE exhibits a -31.2% relative change (RC) reduction compared

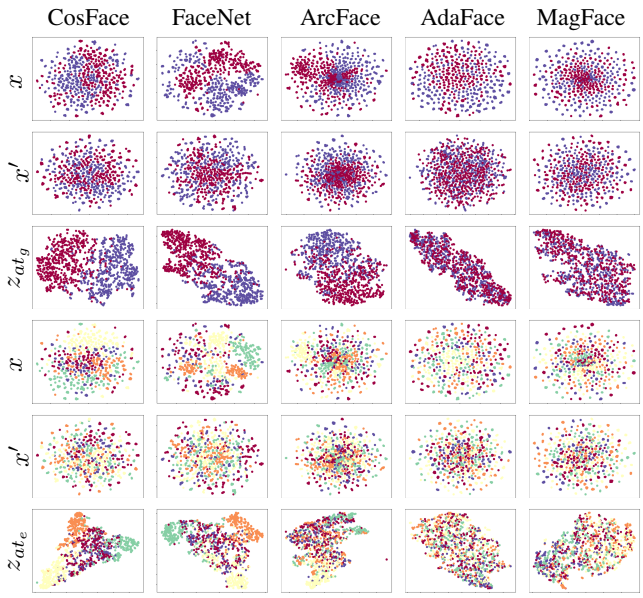

Fig. 4. **Visual analysis using t-SNE projections.** The first three rows focus on the gender attribute, where blue represents male and red female samples. The last three rows focus on ethnicity, with each color representing a different ethnicity. The red color corresponds to White, orange to Black, yellow to Indian, green to Asian, and blue to Latino-Hispanic. The first and fourth (and second and fifth) rows represent the same data, color-coded based on the considered soft-biometric. The third and sixth row show the embedding space for encoding gender and ethnicity, respectively.

to our method's -9.7% RC. However, the verification process is significantly stronger affected by Multi-IVE than by ASPECD. Additionally, ASPECD outperforms Multi-IVE in terms of ethnicity suppression, achieving a -21% RC compared to Multi-IVE's -17.7% RC. On FaceNet, ASPECD outperforms Multi-IVE in both, gender and ethnicity suppression. For AdaFace, which initially exhibited a considerably lower starting ethnicity AUC_e, ASPECD reduces it to near random performance of 0.568, lower than Multi-IVE's value of 0.635. ASPECD also achieves a higher suppression of gender with a lower AUC_g. For MagFace ASPECD achieves a higher suppression of ethnicity (-26.9% RC compared to Multi-IVE's -9.7% RC) and exhibits a higher suppression of gender (-27.6% RC compared to MultiIVE's -22.8%).

Matching gender AUC_g. From the results in Table VII, we observe that for CosFace at a fixed gender AUC of 0.888, ASPECD significantly outperforms Multi-IVE in terms of ethnicity suppression (-21.2% RC vs. -8.3% RC). However, this comes at the expense of a much higher drop in verification performance (141.8% increase in FNMR@FMR10^{-2} for ASPECD compared to 65.5% for Multi-IVE). For all the other extractors, ASPECD consistently achieves lower degradations in verification performance and higher ethnicity suppression rates. For MagFace, both gender and ethnicity are suppressed to a similar extent.

Matching ethnicity AUC_e. The results reported in Table VIII show that ASPECD outperforms Multi-IVE in terms of FNMR@FMR10^{-2} with all face recognition models, primarily due to ASPECD's retention of more identity-

TABLE VI

ASPECD VS. MULTI-IVE AT MATCHED EER ON VGGFACE2*

Extractor	EER	Approach	Gender AUC$_g$ (↓)			Ethnicity AUC$_e$ (↓)			FNMR@FMR10^{-2} (↓)			SR (↑)
			Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	
CosFace	0.062	ASPECD (ours)	0.991	0.895	−9.7%	0.947	0.746	−21.2%	0.011	**0.027**	141.8%	0.323
		Multi-IVE		0.682	−31.2%		0.780	−17.7%		0.206	1776.4%	**0.483**
FaceNet	0.035	ASPECD (ours)	0.991	0.962	−3.0%	0.938	0.798	−14.9%	0.048	**0.076**	59.4%	**0.189**
		Multi-IVE		0.984	−0.7%		0.838	−10.7%		0.090	87.5%	0.062
ArcFace	0.059	ASPECD (ours)	0.991	0.936	−5.6%	0.928	0.743	−20.0%	0.069	**0.143**	107.0%	**0.273**
		Multi-IVE		0.943	−4.9%		0.784	−15.5%		0.167	142.5%	0.149
AdaFace	0.024	ASPECD (ours)	0.705	0.674	−4.4%	0.738	0.568	−23.0%	0.014	**0.028**	97.9%	**0.432**
		Multi-IVE		0.718	1.8%		0.635	−14.0%		0.031	122.1%	0.211
MagFace	0.018	ASPECD (ours)	0.912	0.667	−26.9%	0.875	0.634	−27.6%	0.013	**0.020**	53.8%	**0.619**
		Multi-IVE		0.824	−9.7%		0.675	−22.8%		**0.020**	53.8%	0.047

TABLE VII

ASPECD VS. MULTI-IVE AT MATCHED GENDER AUC$_g$ ON VGGFACE2*.

Extractor	Gender AUC$_g$	Approach	EER (↓)			Ethnicity AUC$_e$ (↓)			FNMR@FMR10^{-2} (↓)			SR (↑)
			Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	
CosFace	0.888	ASPECD (ours)	0.010	0.020	98.0%	0.947	0.746	−21.2%	0.011	0.027	141.8%	**0.323**
		Multi-IVE		0.015	51.0%		0.868	−8.3%		**0.018**	65.5%	0.167
FaceNet	0.958	ASPECD (ours)	0.024	0.036	52.1%	0.938	0.798	−14.9%	0.048	**0.076**	59.4%	**0.189**
		Multi-IVE		0.067	177.9%		0.794	−15.3%		0.243	406.5%	0.145
ArcFace	0.927	ASPECD (ours)	0.034	0.060	75.6%	0.928	0.743	−20.0%	0.069	**0.143**	107.0%	**0.273**
		Multi-IVE		0.063	84.4%		0.775	−16.4%		0.172	149.4%	0.178
AdaFace	0.673	ASPECD (ours)	0.013	0.023	80.0%	0.738	0.568	−23.0%	0.014	**0.028**	97.9%	**0.432**
		Multi-IVE		0.048	268.5%		0.61	−17.3%		0.102	632.1%	0.365
MagFace	0.663	ASPECD (ours)	0.013	0.018	40.0%	0.875	0.634	−27.6%	0.013	**0.020**	56.2%	**0.619**
		Multi-IVE		0.129	893.8%		0.648	−25.9%		0.498	3729.2%	0.400

TABLE VIII

ASPECD VS. MULTI-IVE AT MATCHED ETHNICITY AUC$_e$ ON VGGFACE2*.

Extractor	Ethnicity AUC$_e$	Approach	EER (↓)			Gender AUC$_g$ (↓)			FNMR@FMR10^{-2} (↓)			SR (↑)
			Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	
CosFace	0.769	ASPECD (ours)	0.010	0.020	98.0%	0.991	0.895	−9.7%	0.011	**0.027**	141.8%	0.323
		Multi-IVE		0.072	623.0%		0.682	−31.2%		0.265	2310.0%	**0.496**
FaceNet	0.794	ASPECD (ours)	0.024	0.036	52.1%	0.991	0.962	−3.0%	0.048	**0.076**	59.4%	**0.189**
		Multi-IVE		0.067	177.9%		0.958	−3.3%		0.243	406.5%	0.145
ArcFace	0.744	ASPECD (ours)	0.034	0.06	75.6%	0.991	0.936	−5.6%	0.069	**0.143**	107.0%	**0.273**
		Multi-IVE		0.142	319.1%		0.887	−10.5%		0.613	787.8%	0.262
AdaFace	0.562	ASPECD (ours)	0.013	0.023	80.0%	0.705	0.674	−4.4%	0.014	**0.028**	97.9%	0.432
		Multi-IVE		0.114	780.0%		0.549	−22.2%		0.416	2874.3%	**0.739**
MagFace	0.633	ASPECD (ours)	0.013	0.018	40.0%	0.912	0.667	−26.9%	0.013	**0.020**	56.2%	**0.619**
		Multi-IVE		0.210	1519.2%		0.612	−32.9%		0.879	6660.8%	0.525

related information. On the other hand, Multi-IVE consistently excels in gender suppression. Nevertheless, this advantage in gender suppression is offset by a significant loss in verification performance.

VI. CONCLUSIONS

In this paper, we introduced ASPECD, a novel framework for the suppression of multiple soft-biometric attributes in face verification templates that includes a control mechanism to determine which soft-biometrics in the templates to exclude and which to retain. ASPECD was evaluated in comprehensive experiments with five state-of-the-art template extractors, achieving competitive results in both single- and multi-attribute privacy-enhancement settings. The experimental results showed that ASPECD exhibits significant capacity for suppressing multiple-attributes in face templates across a variety of face recognition models, and that it compares favorably against the state-of-the-art. As part of our future work, we plan to explore unsupervised versions of our disentanglement procedure to account for different types of sensitive information in one single step, where only identity supervision is used to drive the disentanglement process.

REFERENCES

[1] N. Y. Almudhahka, M. S. Nixon, and J. S. Hare. *Comparative Face Soft Biometrics for Human Identification*, pages 25–50. 2018.

[2] B. Bortolato, M. Ivanovska, P. Rot, J. Križaj, P. Terhörst, N. Damer, P. Peer, and V. Štruc. Learning privacy-enhancing face representations through feature disentanglement. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 495–502, 2020.

[3] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 67–74, 2018.

[4] S. Chhabra, R. Singh, M. Vatsa, and G. Gupta. Anonymizing k-facial attributes via adversarial perturbations. In *International Joint Conferences on Artificial Intelligence*, pages 656–662, 2018.

[5] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? a survey on soft biometrics. *IEEE Transactions on Information Forensics and Security*, 11(3):441–467, 2015.

[6] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2):739–777, 2011.

[7] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[8] A. Goldsteen, G. Ezov, R. Shmelkin, M. Moffie, and A. Farkash. Data minimization for gdpr compliance in machine learning models. *AI and Ethics*, 2(3):477–491, 2022.

[9] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez. Facial soft biometrics for recognition in the wild: Recent works, annotation, and cots evaluation. *IEEE Transactions on Information Forensics and Security*, 13(8):2001–2014, 2018.

[10] B. Hassan, E. Izquierdo, and T. Piatrik. Soft biometrics: a survey: Benchmark analysis, open challenges and recommendations. *Multimedia Tools and Applications*, pages 1–44, 2021.

[11] M. Kim, A. K. Jain, and X. Liu. Adaface: Quality adaptive margin for face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18750–18759, 2022.

[12] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain. Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security*, 7(6):1789–1801, 2012.

[13] Y. Lin and H. Xie. Face gender recognition based on face recognition feature vectors. In *IEEE International Conference on Information Systems and Computer Aided Education*, pages 162–166, 2020.

[14] B. Meden, P. Rot, P. Terhörst, N. Damer, A. Kuijper, W. Scheirer, A. Ross, P. Peer, and V. Štruc. Privacy–Enhancing Face Biometrics: A Comprehensive Survey. *IEEE Transactions on Information Forensics and Security*, 16:4147–4183, 2021.

[15] P. Melzi, H. O. Shahreza, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, J. Fierrez, S. Marcel, and C. Busch. Multi-ive: Privacy enhancement of multiple soft-biometrics in face embeddings. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 323–331, 2023.

[16] Q. Meng, S. Zhao, Z. Huang, and F. Zhou. Magface: A universal representation for face recognition and quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14225–14234, 2021.

[17] V. Mirjalili, S. Raschka, A. Namboodiri, and A. Ross. Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images. In *IEEE International Conference on Biometrics*, pages 82–89, 2018.

[18] V. Mirjalili, S. Raschka, and A. Ross. Flowsan: Privacy-enhancing semi-adversarial networks to confound arbitrary face-based gender classifiers. *IEEE Access*, 7:99735–99745, 2019.

[19] V. Mirjalili, S. Raschka, and A. Ross. Privacynet: Semi-adversarial networks for multi-attribute face privacy. *IEEE Transactions on Image Processing*, 29:9400–9412, 2020.

[20] V. Mirjalili and A. Ross. Soft biometric privacy: Retaining biometric utility of face images while perturbing gender. In *IEEE International Joint Conference on Biometrics*, pages 564–573, 10 2017.

[21] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana. Sensitivenets: Learning agnostic representations with application to face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2158–2164, 2020.

[22] D. A. Reid, S. Samangooei, C. Chen, M. S. Nixon, and A. Ross. Soft biometrics for surveillance: an overview. *Handbook of statistics*, 31:327–352, 2013.

[23] J. P. Robinson, C. Qin, Y. Henon, S. Timoner, and Y. Fu. Balancing biases and preserving privacy on balanced faces in the wild. *IEEE Transactions on Image Processing*, 32:4365–4377, 2023.

[24] P. Rot, K. Grm, P. Peer, and V. Štruc. PrivacyProber: Assessment and detection of soft–biometric privacy–enhancing techniques. In *IEEE Transactions on Dependable and Secure Computing*, pages 1–18, 2023.

[25] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.

[26] I. technology-biometric performance testing, reporting-part 1: Principles, and framework. Standard. *International Organization for Standardization*, 2006.

[27] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper. Beyond identity: What information is stored in biometric face templates? In *IEEE International Joint Conference on Biometrics*, pages 1–10, 2020.

[28] P. Terhörst, D. Fährmann, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper. Maad-face: A massively annotated attribute dataset for face images. *IEEE Transactions on Information Forensics and Security*, 16:3942–3957, 2021.

[29] P. Terhörst, M. Huber, N. Damer, F. Kirchbuchner, and A. Kuijper. Unsupervised enhancement of soft-biometric privacy with negative face recognition. In *arXiv preprint arXiv:2002.09181*, 2020.

[30] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper. Suppressing gender and age in face templates using incremental variable elimination. In *IEEE International Conference on Biometrics*, pages 1–8, 2019.

[31] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper. Unsupervised privacy-enhancement of face representations using similarity-sensitive noise transformations. *Applied Intelligence*, pages 1–18, 2019.

[32] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper. On soft-biometric information stored in biometric face embeddings. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4):519–534, 2021.

[33] P. Terhörst, K. Riehl, N. Damer, P. Rot, B. Bortolato, F. Kirchbuchner, V. Štruc, and A. Kuijper. Pe-miu: A training-free privacy-enhancing face recognition approach based on minimum information units. *IEEE Access*, 8:93635–93647, 2020.

[34] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[35] R. Vera-Rodriguez, M. Blazquez, A. Morales, E. Gonzalez-Sosa, J. C. Neves, and H. Proenca. Facegenderid: Exploiting gender information in dcnns face recognition systems. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2254–2260, 2019.

[36] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu. Cosface: Large margin cosine loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

[37] J. Xiang and G. Zhu. Joint face detection and facial expression recognition with mtcnn. In *IEEE International Conference on Information Science and Control Engineering*, pages 424–427, 2017.

APPENDIX

In the main part of the paper, we conducted comprehensive experiments with ASPECD to demonstrate its effectiveness in privacy enhancement. In this *Supplementary material*, we provide additional results that further showcase the capabilities of ASPECD. Specifically, we: *(i)* evaluate individual components for suppressing either ethnicity or gender, *(ii)* evaluate ASPECD on the additional BFW dataset, *(iii)* investigate the types of errors caused by ASPECD using confusion matrices, and *(iv)* provide links to the used models and repositories for reproducibility reasons.

A. In-detail analysis of ASPECD components

In the main paper, we reported results from tests on both individual and sequential configurations of ASPECD. Here, we provide additional results for individual components (G and E), evaluated separately.

Module E. When using only module E, which focuses solely on privacy-enhancement of ethnicity, we obtain the results reported in Table IX. In Table X we provide corresponding, in-depth results for the verification experiments.

TABLE IX

ASPECD TARGETING ONLY ETHNICITY (COMPONENT E).

Extractor	Gender AUC (↓)			Ethnicity AUC (↓)			Verification AUC (↑)			SR (↑)
	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	
CosFace	0.991	0.983	−0.8%	0.947	0.798	−15.7%	0.999	0.997	−0.2%	0.175
FaceNet	0.991	0.986	−0.5%	0.938	0.841	−10.3%	0.996	0.994	−0.2%	0.116
ArcFace	0.991	0.965	−2.7%	0.928	0.789	−15.0%	0.993	0.988	−0.6%	0.189
AdaFace	0.724	0.705	−2.7%	0.738	0.637	−13.7%	0.996	0.994	−0.3%	0.166
MagFace	0.912	0.665	−27.0%	0.875	0.648	−25.9%	0.996	0.995	−0.1%	0.601

TABLE X

ASPECD TARGETING ONLY ETHNICITY (COMPONENT E) WITH THE FOCUS ON VERIFICATION PERFORMANCE.

Extractor	EER (↓)			FNMR@FMR10^{-2} (↓)			FNMR@FMR10^{-3} (↓)		
	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC
CosFace	0.010	0.016	60.0%	0.011	0.019	77.3%	0.027	0.054	99.6%
FaceNet	0.024	0.029	22.1%	0.048	0.057	18.8%	0.180	0.177	−1.4%
ArcFace	0.034	0.048	42.4%	0.069	0.109	58.0%	0.198	0.253	27.6%
AdaFace	0.013	0.020	56.9%	0.014	0.022	58.6%	0.019	0.031	63.7%
MagFace	0.013	0.017	27.7%	0.013	0.018	37.7%	0.018	0.024	30.6%

The impact of privacy-enhancement on ethnicity AUC varies depending on the chosen template extractor. The RC ranges from a decrease of -10.3% for FaceNet features to a decrease of up to -25.9% for MagFace features. Note that, as desired, the non targeted gender exhibits proportionally smaller changes in AUC. For example, there is a minor decrease of only 0.5% in RC for FaceNet and a decrease of -0.9% for CosFace. MagFace stands as an exception in this context, where the relative drop in gender AUC is even greater than that of ethnicity AUC, amounting to -27% in RC. The trend of correlated drops of gender AUC and ethnicity AUC in scores concerning MagFace templates is consistent and emerges in later analyses as well. This suggests that gender and ethnicity attributes are highly correlated and hard to separate, making MagFace suitable for cases where both, gender and ethnicity need to be privacy-enhanced. Turning our attention to Table X, the highest relative drop in verification performance occurs on CosFace features (e.g. 60% RC in EER), however the absolute EER of 0.016 and FNMR@FMR10^{-2} of 0.019 are still among the lowest. While ArcFace achieved ethnicity AUC RC of -15%, a value similar to CosFace's ethnicity AUC RC of -15.7, both absolute FNMR@FMR scores are approximately 5 times higher in comparison to CosFace. On AdaFace, a relatively low starting ethnicity AUC of 0.738 was additionally reduced to 0.637 without significant effect on gender AUC and promising verification performance in terms of absolute FNMR@FMR scores (e.g. 0.031 FNMR@FMR10^{-3} is second-lowest score). In summary, when considering extractors other than MagFace, we can effectively target only ethnicity with less significant impact on gender, while the effect on verification performance varies among template extractors.

Module G. The results of targeting only gender using ASPECD are presented in Table XI, while Table XII provides corresponding in-detail verification performance. When focusing solely on gender, the drops in gender AUC are higher than drops in ethnicity AUC. As desired, they are also notably higher in comparison to gender AUC values presented in Table IX, where only ethnicity was targeted. This holds true for all compared template extractors. On MagFace templates, we again observe substantial reductions in both gender AUC and ethnicity AUC. Regarding templates other than MagFace, targeting gender causes varying levels of reduction in ethnicity AUC. We observe that reductions of ethnicity AUC in this experiment are higher compared to reductions of gender AUC in the previous experiment, where only ethnicity was targeted. Moreover, the verification performance is more impacted when targeting gender (e.g. EERs for FaceNet and ArcFace become higher than 0.1).

TABLE XI

ASPECD TARGETING ONLY GENDER (COMPONENT G).

Extractor	AUC_g (↓)			AUC_e (↓)			Verification AUC (↑)			SR (↑)
	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	
CosFace	0.991	0.778	−21.5%	0.947	0.890	−6.1%	0.999	0.988	−1.1%	0.281
FaceNet	0.991	0.914	−7.8%	0.938	0.873	−6.9%	0.996	0.960	−3.6%	0.153
ArcFace	0.991	0.851	−14.1%	0.928	0.812	−12.5%	0.993	0.928	−6.6%	0.279
AdaFace	0.705	0.554	−21.4%	0.738	0.610	−17.3%	0.996	0.983	−1.3%	0.636
MagFace	0.912	0.635	−30.4%	0.875	0.657	−24.9%	0.996	0.990	−0.6%	0.627

TABLE XII

ASPECD TARGETING ONLY GENDER (VERIFICATION).

Extractor	EER (↓)			FNMR@FMR10^{-2} (↓)			FNMR@FMR10^{-3} (↓)		
	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC	Orig.	Priv. enh.	RC
CosFace	0.010	0.051	408.0%	0.011	0.126	1041.8%	0.027	0.278	930.7%
FaceNet	0.024	0.105	335.8%	0.048	0.324	574.4%	0.180	0.557	209.6%
ArcFace	0.034	0.147	331.8%	0.069	0.459	565.8%	0.198	0.676	241.5%
AdaFace	0.013	0.054	314.6%	0.014	0.116	728.6%	0.019	0.240	1162.1%
MagFace	0.013	0.033	151.5%	0.013	0.049	277.7%	0.018	0.106	488.9%

B. Evaluation on the BFW dataset

In this section, we present an extended evaluation of ASPECD, applying it to the BFW dataset [23], following the same methodology as used for the VGGFace2* dataset in the main paper. Table XIII details the baseline verification performance, while Table XIV outlines the baseline

performance for soft-biometric attribute extraction. It's worth reiterating that VGGFace2* considers 5 different ethnic groups, while BFW only considers 4, making it a somewhat easier classification task due to fewer classes.

TABLE XIII

BASELINE VERIFICATION PERFORMANCE ON THE BFW DATASET.

Extractor	Verification AUC (↑)	EER (↓)	FNMR@FMR10^{-2}(↓)	FNMR@FMR10^{-3}(↓)
CosFace	0.992 ± 0.001	0.040 ± 0.004	0.085 ± 0.009	0.209 ± 0.010
FaceNet	0.975 ± 0.001	0.078 ± 0.001	0.311 ± 0.008	0.605 ± 0.010
ArcFace	0.955 ± 0.002	0.108 ± 0.004	0.402 ± 0.013	0.920 ± 0.032
AdaFace	0.975 ± 0.004	0.059 ± 0.008	0.079 ± 0.012	0.102 ± 0.014
MagFace	0.977 ± 0.004	0.057 ± 0.008	0.077 ± 0.013	0.099 ± 0.015

TABLE XIV

BASELINE PERFORMANCE OF SOFT-BIOMETRIC CLASSIFIERS ON THE BFW DATASET IN TERMS OF AUC.

Extractor	**Gender AUC**	**Ethnicity AUC**
CosFace	0.987 ± 0.003	0.989 ± 0.003
FaceNet	0.967 ± 0.005	0.978 ± 0.004
ArcFace	0.975 ± 0.003	0.972 ± 0.004
AdaFace	0.633 ± 0.029	0.666 ± 0.025
MagFace	0.852 ± 0.021	0.878 ± 0.006

C. Examining Privacy-Enhancement Induced Errors

The objective of privacy enhancement is to reduce the effectiveness of soft-biometric classifiers. In this section, we investigate the types of misclassifications that occur as a consequence of privacy enhancement. In Fig. 5, we examine the impact of E→G by analyzing the resulting misclassifications through confusion matrices for both considered testsets, VGGFace2* and BFW.

Specifically, our analysis examines CosFace (rows 1 and 2) with less correlated gender and ethnicity, and MagFace (rows 3 and 4) with a high correlation. Note that the VGGFace2* train and test sets involve 5 ethnicities, whereas BFW only encompasses 4, rendering it a comparatively simpler classification problem. For both datasets, we observe that while diagonals are clearly discernible for original templates, they are notably less discernible for privacy-enhanced templates, demonstrating the effectiveness of privacy enhancement.

D. Reproducibility

All of our experiments are reproducible, as we used publicly available datasets and official repositories for all models used in the experiments. We also note at this point that further details on the models, such as the choice of backbone, the selection of training datasets and hyperparameter settings can be found from the links posted below.

Template extractors:

- AdaFace: https://github.com/mk-minchul/AdaFace
- ArcFace & FaceNet-512: https://github.com/serengil/deepface/

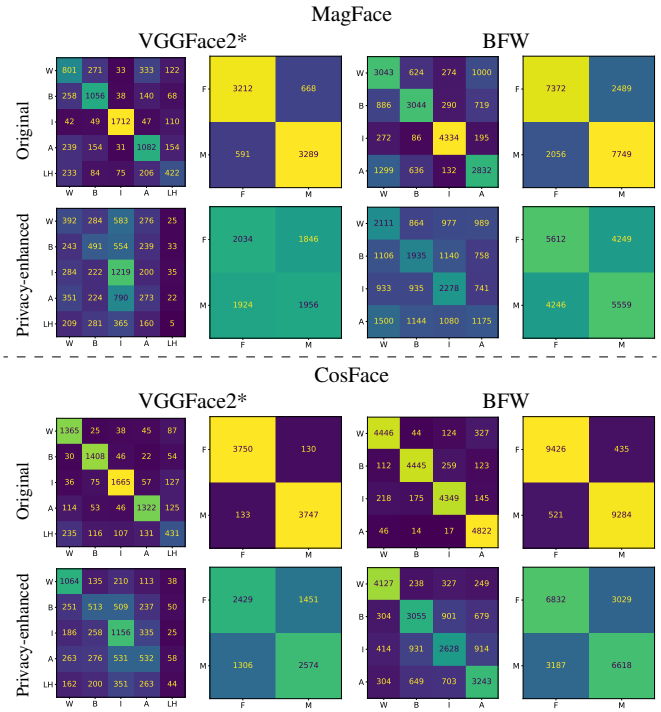

Fig. 5. **Confusion matrices computed over the two test datasets.** The confusion matrices show the effect of ASPECD's privacy enhancement, specifically of model E→G, on ethnicity (columns 1 and 3) and gender (columns 2 and 4) classification. We consider two extractors: MagFace in rows 1 and 2, and CosFace in rows 3 and 4, on two datasets, VGGFace2* (columns 1 and 2) and BFW (columns 3 and 4).

- CosFace: https://github.com/MuggleWang/CosFace_pytorch
- MagFace: https://github.com/IrvingMeng/MagFace

Compared state-of-the-art methods:

- Multi-IVE: https://github.com/otroshi/multi-ive
- IVE and CSN: https://github.com/pterhoer/PrivacyPreservingFaceRecognition

List of VGGFace2* dataset images:
https://github.com/to_come_after_review

We also plan to make the code for ASPECD publicly available after the review.