Facial Biometrics in the Social Media Era: An in-Depth Analysis of the Challenge Posed by Beautification Filters.

Nelida Mirabet-Herranz, Chiara Galdi, and Jean-Luc Dugelay

Abstract—Automatic beautification through social media filters has gained popularity in recent years. Users apply face filters to adhere to beauty standards, posing challenges to the reliability of facial images and complicating tasks like automatic face recognition. In this work, the impact of digital beautification is assessed, focusing on the most popular social media filters from three different platforms, on a range of AI-based face analysis technologies: face recognition, gender classification, apparent age estimation, weight estimation, and heart rate assessment. Tests are performed on our extended Facial Features Modification Filters dataset, containing a total of 24312 images and 260 videos. An extensive set of experiments is carried out to show through quantitative metrics the impact of beautification filters on the performance of the different face analysis tasks. The results reveal that employing filters significantly disrupts soft biometric estimation, resulting in a pronounced impact on the performance of weight and heart rate networks. Nevertheless, we observe that certain less aggressive filters do not adversely affect face recognition and gender estimation networks, in some instances enhancing their performances. Scripts and more information are available at https://github.com/nmirabeth/filters_biometrics

Index Terms—Face analysis, Beautification, Social media filters, Face Recognition, Soft biometrics, Hidden biometrics

1 Introduction

Human face images encode different biometric information whose estimation is of great utility in security and health applications, including but not limited to people identification and health assessment. Face is considered as a *hard* biometric trait, meaning that face analysis technologies can infer explicit information from it, such as a person's identity, in addition to extracting different *soft* biometric traits. Soft biometrics are not able to univocally authenticate a person due to a lack of distinctiveness and permanence but give additional information about the subject for instance gender, age, height, or weight [1]. Moreover, in the medical domain, the term *hidden* biometrics (concerning human vision) refers to methods used to quantify and/or measure parameters extracted from medical data enabling the use of biosignals for tasks such as individual identification [2]. Parameters including Heart Rate (HR) or Blood Pressure can be inferred from face images or videos for people's daily quality of life assessment in addition to giving complementary information for a person's identification.

The act of uploading facial pictures to the internet has become more common since the use social media platforms, also known as Social Networks (SN), has grown. More than 18.2 million text messages are transmitted in a minute. And a big portion of them are multimedia content [3]. Nowadays these popular SN platforms offer many different tools to automatically modify and embellish images to mask or enhance certain facial traits. Filtered images have been proven to be among the most heavily engaged photos on SNs [4], augmenting the popularity of these techniques. About 600 million people use filters each month on Instagram or Facebook and 76% of Snapchat users use them every day [5].

Beautification is the modification of the appearance of a subject to make it more visually attractive. In the past years, *digital* beautification has gained popularity due to the accessibility of photo editing software (which often requires an expert user) and more recently with the prominence of facial beautification filters. Filters available on SNs allow for the automatic beautification of a person's image, requiring very little or no user expertise at all. Due to the rapid growth of beautification filter usage, some concerns have been raised in the research community. Retouched face images on SNs may be considered idealized images and could thus negatively impact viewers' body image [6]. Moreover, racial biases in beautification filters have been found in recent studies [7]. Consequently, researchers in the field of biometrics have begun to wonder whether beautification can be considered a compromising factor for facial processing techniques since they modify distinctive facial traits. Filters are not by definition used to compromise Face Recognition (FR) systems. Still, they might as well affect those models in different stages of their pipeline e.g: detecting the face, extracting features, and successfully classifying individuals [8] in addition to impacting other facial processing tasks such as gender classifiers and weight estimators [9]. However, to the best of the authors' knowledge, the impact of social media filters on the estimation of other biometric traits, for instance, apparent age or HR, has not been investigated yet, neither the comparison between filters from different SNs.

Consequently, in our work, we target the study of the impact of the most popular SN filters on the most relevant AI-based face analysis tasks. Our focus is set on filters that alter facial characteristics such as eyes and face contour being those modifications subtle and not explicitly notice-

- *N. Mirabet-Herranz, C. Galdi and J.L. Dugelay are with the Department of Digital Security, EURECOM, Biot, 06410, France.*
  E-mail: {mirabet, galdi, dugelay}@eurecom.fr

TABLE 1: Works addressing the impact of beautification in different manners. We select a representative set of papers based on the date of publication and number of citations.

Beaut. Type	Work	Analysis performed				
		FR	Gender	Age	Weight	HR
Makeup	[10]	x	x			
	[11]		x	x		
	[12]	x				
Plastic surgery	[13]	x				
	[14]	x				
	[15]	x				
Digital	[16]	x				
	[17]	x				
	[18]	x				
Filters	[19]	x				
	[8]	x				
	[9]	x	x		x	
	Ours	x	x	x	x	x

able by the human eye. A face image refashioned using these techniques could be taken as unprocessed for official documents, e.g. passport ID picture.

The contributions of this work are the following:

1) The impact of facial beautification is extensively tested in up-to-date face analysis tasks namely identity verification, gender recognition, apparent age, weight and heart rate estimation;
2) The FFMF original database composed of 10130 images has been extended to 24312 face pictures and 260 videos and includes now 6 new popular filters from Instagram, Snapchat and TikTok;
3) We use structural similarity (SSIM) as a metric to objectively measure the *agressivity* of a filter.

The rest of this paper is organized as follows. In Section 2, relevant state-of-the-art works that address the problem of digital beautification are described. Section 3 presents the methodology adopted including the FFMF database extension and the technologies evaluated in our experiments; while in Section 4, the implementation details and metrics used in our study are presented. The impact of beautification filters on different facial hard, soft, and hidden biometrics is extensively analyzed and in Section 5. Finally, conclusions and future directions are reported in Section 6.

2 Related work

Beautification can be applied in the physical (makeup, plastic surgery) or digital domain (image editing software, social media filters).

Makeup: Facial cosmetics have the ability to enhance or disguise facial traits. The use of makeup can visually modify the proportions of different facial characteristics such as eyes and cheekbones. In 2010, Ueda *et al.* [10] observed that light make-up makes it easier to recognize a face while heavy make-up makes it more challenging. Moreover, makeup has been proven as an effective attack for FR [12]. In this type of attack, the attacker might apply a substantially high amount of makeup with the purpose of emulating the facial appearance of a target user. Chen *et al.* [11] explored the impact that such modifications have on other traits namely gender and age. Other works evaluated integrating makeup detection schemes into biometric systems to improve FR models [20] by proposing a dynamic weighting of the extracted traits for FR according to the makeup classification result. More recently, Albeiro *et al.* demonstrated that the use of makeup can degrade the performance of face recognition networks, especially for genuine female subjects [21].

Plastic surgery: Facial plastic surgery is commonly used for correcting feature imperfections or improving the visual appearance of a subject by removing birthmarks and scars, enhancing desired traits and correcting asymmetric features. When a subject gets plastic surgery, both the shape and texture of facial features are changed to varying degrees altering their appearance. Consequently, existing identities may become unknown to the already existing FR systems and their reference templates. Facial plastic surgery is usually employed benignly to improve a person's appearance but researches have pointed out the usage of plastic surgery by criminals to "manipulate" their facial identity with the intent to deceive FR systems [13], [14], [15].

Image editing: Digital face retouching is a widespread application available across a large spectrum of devices such as mobile phones, tablets, and personal computers. Rathgeb *et al.* [16] presented in their survey facial retouching in the *digital* domain with different software. In [17], high error rates were reported in FR SotA systems when faces were digitally modified. Kose *et al.* [22], simulated nose alterations traditionally achieved with makeup and/or plastic surgery. The results reported a loss of precision in FR models for both 2D and 3D faces. In the past years, some works have focused on automatically beautifying faces without interfering with the identity of the subject. In [18], [23], novel methods for digital face beautification are presented to increase the predicted attractiveness of a subject, maintaining a strong similarity between the pre- and post-modified face.

Social media filters: In this work we are interested in analyzing automatic filtering via social media tools. In the past years, SNs have included a feature allowing users to easily modify their published content by using filters. Our focus is set on filters that are commonly applied to faces. In [9], a categorization of these filters is presented: **a)** Colour Adjustment Filters (CAF): Apply changes to the color and luminance channels or convert the image/video from RGB to black-and-white modality; **b)** Smoothing Filters (SF): Smooth and blurry face skin, as if a layer of foundation was applied. They focus on skin pixels, leaving the rest of the face regions untouched; **c)** Facial Features Modification Filters (FFMF): Modify the biometric features of the user's face. Some of the most common effects are the enlarging, shrinking, and sharpening of different facial lines (e.g: augmenting the eyes or the lips size, defining the nose lines and modifying the eyebrows position); **d)** Augmented Reality Filters (ARF): An Augmented Reality (AR) filter is a mask-like filter that adds unexisting elements to the scene. Allow the user to incorporate different gadgets on their faces such as animal ears, but also to see how a specific product might look on their faces (e.g. eyeglasses); **e)** Immersive AR Face Filter (IARFF): Place in real-time, the users' face or some elements of it (e.g: eyes, mouth) into a virtual scene.

In [8], Hedman *et al.* assessed the impact of CAF and ARF on FR and propose a counter-filter based on a modified version of the U-NET segmentation network. To improve the FR system, deep learning algorithms and distance measures are applied to the features extracted using a ResNet-34

TABLE 2: Overview of categories and face features modified by filters from Instagram, Snapchat, and TikTok. All filters except "Thinner_face" include Colour Adjustment (CA).

SN	Filter Name	F. Category		Modified Face Feature by F.			
		SF*	ARF**	Contour	Eyes	Nose	Lips
Ig	Thinner_face			x	x		
	Relax! You Pretty!	x				x	x
	Hawaii Grain	x	x		x	x	x
	Glam Grain	x	x		x	x	x
Sc	Fresh vibes	x	x	x	x	x	x
	Fresh light	x	x	x	x	x	x
	Mellow glow	x	x	x	x	x	x
Tk	Belle	x	x				
	Spring glow	x		x		x	

*SF = Smoothing Filters; **ARF = Augmented Reality Filters

Fig. 1: Example of original and beautified images, one filter for each SN: Thinner_face (Ig), Mellow Glow (Sc) and Belle (Tk).

network. Botezatu *et al.* [24] studied the impact of ARF, also called "fun" selfie filters. The authors evaluated their impact on different face detector and recognition models and propose a GAN-based filter removal algorithm. Although ARF add artificial elements that act as occlusion therefore removing important information from the face image, their risk to biometric systems is minor since a face of this kind is easily detected as processed via visual inspection in e.g: border control scenario. In a social-oriented approach, Riccio *et al.* [19] drew key insights such as the discovery of a general homogenization of the beautified faces when compared to the original ones. Mirabet-Herranz *et al.* [9] conducted a preliminary first study to assess the impact of Instagram FFMF on different facial processing tasks and Libourel *et al.* [25] quantify the impact of such filters on deepfake detection.

In Table 1, a compendium of relevant works assessing the impact of all types of beautification is presented. We provide conclusions extracted from these works as well as the facial processing technique(s) analyzed.

In this article, we aim to study the impact of filters that modify or distort the shape of facial biometric traits. These filters apply modifications to facial features that are not easily noticeable to the naked eye but play a role in facial processing tasks. More specifically, we study filters that are a result of combinations between SF, FFMF, and virtually added makeup as ARF.

3 METHODOLOGY

The objective of our study is to observe the impact of beauty filters on certain face analysis tasks. We first extended the FFMF database to contain more filters and at the same time to include the data needed to enable the study of new tasks. In this section, we describe the extension of the database, the methods used, and the evaluation protocol.

3.1 FFMF database extension

The FFMF database proposed in [9] consists of 10130 face images from 3 different image categories, original, uploaded, and beautified (with 3 different filters from Instagram). "Original" images did not undergo any processing, while "uploaded" images were only uploaded (no filter applied) on the SN, which will however perform some processing and "beautified" images are those on which the filters were applied. The original pictures present in the FFMF database were selected from the two publicly available datasets: CALFW [26] and VIP_attribute [27].

For this work, an extended version of the FFMF database is used and made available, obtained by 1) Filtering the original images presented in the FFMF dataset using five new beautification filters from two additional platforms, Snapchat and TikTok; 2) Beautifying 52 face videos from the COHFACE dataset [28]. COHFACE is composed of facial videos collected from 40 individuals and their physiological signals such as their HR.

The filters are chosen from the three most popular SNs[1]: Instagram, Snapchat, and TikTok. Besides being popular, the selected filters can be applied to already existing images or videos, an indispensable requirement for our study as the filtered images are created by modifying existing face images from biometric databases. In Table 2 the selected SN filters are presented as well as various types of filter modifications and different biometric features that can be altered by the FFMF. Trait modifications were assessed by visual inspection of pixel differences between original and filtered images. Examples of resulting images once beautification filters are applied are shown in Fig. 1. Unlike the other SNs, TikTok automatically adds a visible watermark image to all processed images/videos indicating the app logo and account identity (see Fig. 1). The watermark's position varies across all the TikTok FFMF database images, laying always on the corners of the images, thus, never occluding the central facial regions. By detecting and cropping all faces in the extended FFMF dataset (original, uploaded, and beautified via Instagram, Snapchat, and TikTok) using Viola-Jones algorithm, we also got rid of the TikTok watermark, which could have affected the assessment.

In Fig. 2 we present the pipeline for the FFMF dataset extension. The upload of content and application of filters on SNs has some requirements: filters have to be applied online, manually, and to one image or video at a time, with a maximum number of seconds (s) and a restriction of 30 frames per second (fps) for an uploaded video. Following all these prerequisites and to ease the process of manually

1. Since 2012, Facebook and Instagram belong to the same group thus sharing features such as filters.

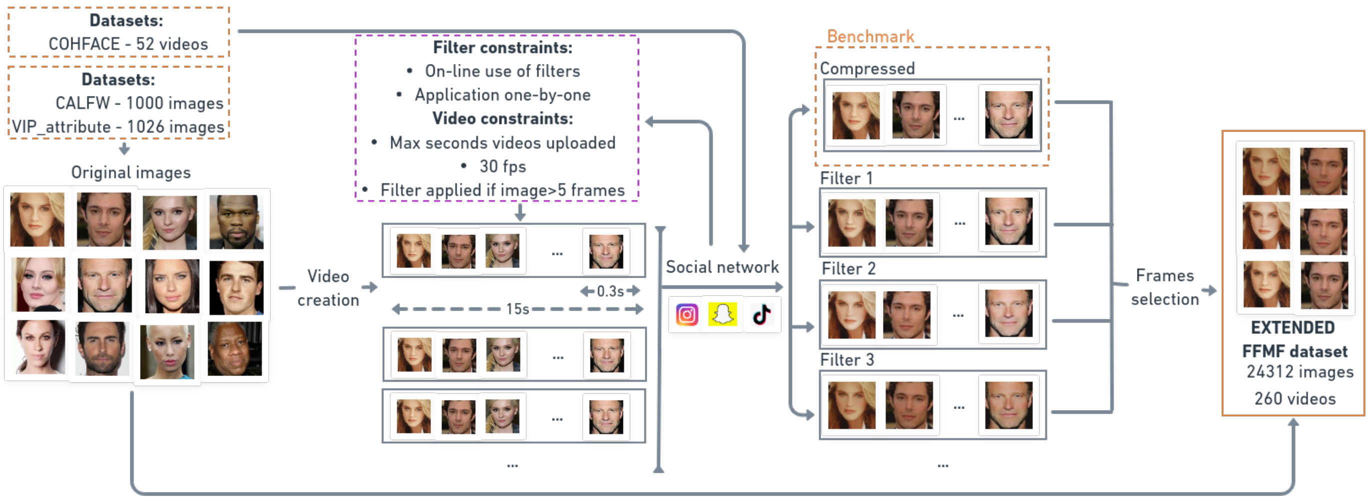


Fig. 2: Overview of the creation of new data for the FFMF extended database. The images from the source datasets are arranged in videos and passed through the different social media where all the original content is processed and filtered.

applying the filters, we created 15s videos with the images originally in CALFW and VIP_attribute, with each image displayed for 0,3 seconds (i.e. 10 frames). The videos are then passed through the different social networks to be filtered. Similarly, the videos from the COHFACE dataset are beautified with Instagram filters.

When content is uploaded to SNs, it goes through operations such as compression, resizing, and cropping, which have been proven to have a negative impact on facial processing tasks such as FR [29]. The original images were uploaded to and downloaded from the different social media platforms obtaining the "uploaded" category in the database. We consider "uploaded" as a baseline to isolate the impact that filters have on the different tasks studied, excluding all other factors listed above. By comparing the performance of the models on the original and uploaded sets, we ensure that any increase or decrease in performance displayed is solely due to the beautification filters.

The extended FFMF dataset is composed of 24312 images and 260 videos belonging to the three categories, original, uploaded, and beautified. When uploading them to an SN, the images were resized to 250x250. The videos retained their original width, length, and fps. According to their source database, the images and videos of the extended FFMF are referred to as 1) FFMF-VIP, 2) FFMF-CALFW and 3) FFMF-COHFACE. More information about the FFMF and how to obtain it can be found on the webpage[2].

3.2 Face Analysis Technologies

Face verification: Two state-of-the-art methods were found to obtain the highest performances on the FFMF and are thus presented in this article. ArcFace [30] or Additive Angular Margin Loss, is a face recognition solution based on a ResNet architecture. It explicitly optimizes the feature embedding to enforce higher similarity for intra-class samples and diversity for inter-class samples, which otherwise results in a performance gap for deep face recognition under large intra-class appearance variations. The second is MagFace [31], a revisited version of ArcFace that aims at addressing the problem of large intra-class variability of subjects' faces, which is stronger in the unconstrained acquisition case. MagFace encodes quality measures into the face representation through the magnitude-aware angular margin loss. By simultaneously enforcing cosine distance direction and magnitude, the learned face representation is more robust to the variability of faces in the wild.

Gender classification: Two popular open-source gender estimators *DeepFace*[3] and *cvlib*[4] are adopted. *DeepFace* provides most popular pre-trained models for face detection and FR along with its own models for gender classification. *cvlib* is a high-level open-source computer vision library for Python. It includes an AlexNet model trained for gender classification. Both *DeepFace* and *cvlib* return the labels "man", "woman", and associated probabilities, once a human face is passed to the gender models.
Regarding the use of the term "gender" in this article, since the annotations of the databases employed were mainly made manually by the researchers who collected the data, we presume that the gender was annotated based on the *assumed gender* of the persons portrayed. However, as the images in the databases are mostly images of celebrities (mainly actors, singers, and athletes), it is possible that a large majority of the annotations are in fact the *affirmed gender* of the people portrayed at the time of the data collection. We refer the reader to the databases' source papers for more information about the gender annotation.

Apparent age estimation: Deep EXpectation (DEX) [32] is a model for apparent age estimation based on Convolutional Neural Networks (CNNs) with VGG-16 architecture pre-trained on ImageNet. It extracts predictions from an ensemble of 20 age estimator networks from the subject's cropped face without explicitly using facial landmarks.

Weight estimation: Recent studies have proved that deep learning models allow for body weight estimation from a single facial image [33]. In 2018, Dantcheva *et al.* [27]

2. https://ffmf.eurecom.fr/

3. https://github.com/serengil/deepface
4. https://github.com/arunponnusamy/cvlib

proposed a ResNet architecture with 50 layers and a final regression layer to successfully estimate the height, weight, and BMI of a subject from a face image. For assessing the impact of FFMF on weight estimation, we selected likewise to [27] and [33] a ResNet50 structure.

Heart rate estimation: Remote photoplethysmography (rPPG) is a noninvasive technique of sensing the cardiovascular blood volume pulse through subtle color variations in the reflected light of human skin. Later studies have shown how a mobile phone camera has enough resolution to capture the subtle changes in skin color that lead to a successful HR estimation [34]. We assessed the effect of beautification filters on remote HR estimation through the network proposed in [34], a 3-dimensional CNN that directly estimates the HR from RGB face videos.

3.3 Evaluation Protocol

To assess the impact of digital beautification on the different biometric tasks, the following experimental protocol was designed: each face analysis model is evaluated (i) on the ”original” images from FFMF-extended; (ii) on the ”uploaded” images (no beautification filters); (iii) on the ”beautified” images. The results obtained on (i), (ii), and (iii), are then compared for the assessment of the impact of the different filters. For each biometric task described in Section 3.2, the same overall protocol is adopted and then metrics specific for the assessment of the single biometric task are applied. In particular, experiment (ii), is designed to set aside the impact of uploading images to SNs – which often undergo image compression, resizing, and cropping operations – from the actual impact of the beautification.

Metrics: In our experiments, we provide extensive reporting of the metrics and carefully analyze the results, following international standards and allowing comparison with previous work. To assess face verification, the metrics and terminology recommended by the ISO/IEC standard [35] are adopted: False match rate (FMR): proportion of zero-effort impostor attempt samples falsely declared to match the compared non-self template; False non-match rate (FNMR): proportion of genuine attempt samples falsely declared not to match the template of the same characteristic from the same user supplying the sample; Detection error trade-off (DET) curve: modified ROC curve which plots error rates on both axes (false positives on the x-axis and false negatives on the y-axis). In addition, we present accuracy and AUC, metrics often reported in works on FR.

Global accuracy and accuracy per class are presented in the gender classifier assessment. For age and weight, we provide the Mean and Standard Deviation (Mean and SD) of the difference between the computed age from the original and processed images in years and the difference between the real and predicted weight in kg. We also report the Mean Absolute Error (MAE) in years and kg respectively and the Pearson's correlation coefficient (ρ). Additionally, we include the Percentage of Acceptable Predictions (PAP) for the weight estimation network. This metric represents the percentage of predictions with an error smaller than 10% of the initial weight, indicating a reasonable error in medical applications. To assess the beautification filters' impact on HR estimation we employ similarly the SD in bpm of the heart rate error, the MAE in bpm and the ρ coefficient.

In Section 5.1, we propose the Structural Similarity Index (SSIM), an index commonly used for assessing the perceived changes between two compared images, as a metric to quantify the *aggressivity* of a filter. The SSIM index is based on the combination of 3 different terms: luminance $l(I, J)$, contrast $c(I, J)$, and pixel structure $s(I, J)$. Given two RGB images I and J:

$$SSIM(I, J) = [l(I, J)]^{\alpha}[c(I, J)]^{\beta}[s(I, J)]^{\gamma} \tag{1}$$

with μ_I, μ_J, σ_I, σ_J and σ_{IJ} the mean of I and J, the standard deviation of I and J, and the covariance of I and J, respectively.

4 Experimental Setup

Face verification: Our ArcFace implementation is based on the one provided by the InsightFace project[5]. We studied different setups during our experiments, such as different backbones (ResNet50 vs ResNet100). Our MagFace implementation with iResNet100 (improved residual network) architecture as backbone is based on the github project by Irving Meng[6]. All the models were pretrained on the MS1MV2 dataset [36].

A largely adopted state-of-the-art protocol for face verification (FV) is used [37]. Ten folds of images are created by randomly selecting images from the dataset. Each fold contains 300 matching pairs and 300 non-matching pairs, for a total of 6000 face comparisons (10×600). This random selection is uniform across databases of the ”original”, ”uploaded” and ”beautified” facial images. Face landmarks are computed for each face image using RetinaFace [38] face detector to take into account the possible displacement of facial features due to the application of the filters. In the event that the face landmark estimation via RetinaFace fails for one of the face images in the list of pairs, the corresponding comparison is excluded from the test.

Age: The datasets CALFW and VIP_attribute are not annotated by age therefore we will assess the impact of beautification filters on apparent age. In [26], the authors of the CALFW dataset ranked the images of each subject from younger to older they estimated the age of each image using DEX. In our experiments with DEX, the implementation provided in GitHub by *siriusdemon* is used[7].

Weight: Experiments are performed on the FFMF data originally from the VIP_attribute dataset. Following the same protocol as in [33], 800 original images (400 female and 400 male) are selected for the training. The results in Section 5.5 are computed using the remaining 226 (113 female and 133 male) identities. The ResNet50 was trained during 10 epochs and the final regression layer during 10 more epochs. Adam optimizer is used, with a learning rate 0.01. The adopted loss function is Huber loss with $\delta = 1$.

HR: We trained the 3D-CNN for 10 epochs, on 96 videos from 24 subjects of the COHFACE dataset. Adam optimizer is selected with a learning rate set to 0.001 and a categorical cross-entropy loss function. In Section 5.6 the experiments are carried out in the testing set of COHFACE, composed of 260 videos from 12 subjects not present at the training.

5. https://github.com/deepinsight/insightface
6. https://github.com/IrvingMeng/MagFace
7. https://github.com/siriusdemon/pytorch-DEX

TABLE 3: SSIM coefficient between the original and processed images of the FFMF extended database. Results are apportioned by gender: female (F) and male (M).

SN	Img. Processing	CALFW (F)	CALFW (M)	VIP (F)	VIP (M)
Ig	Uploaded	97	97	98	98
	Thinner_face	92	92	91	92
	Relax! You Pretty!	80	82	81	81
	Glam Grain	77	78	77	77
Sc	Uploaded	96	96	96	96
	Fresh vibes	82	82	81	82
	Fresh light	91	91	91	92
	Mellow glow	89	89	89	89
Tk	Uploaded	97	97	98	98
	Belle	94	94	93	93
	Spring glow	94	94	93	94

5 RESULTS AND DISCUSSIONS

In this section, we employ similarity measures to quantify the difference between natural and digitally beautified images. To assess the impact of digital beautification on the different technologies presented in 3.2, the following experiments were designed: the technology is evaluated (i) on the "original" images from FFMF-extended; (ii) on the "uploaded" images (no beautification filters); (iii) on the "beautified" images. The performances are presented in tables where the highest values are indicated in **bold** text and the lowest values are <u>underlined</u>. Best performances are highlighted in green and worst in red.

5.1 Filter Aggressivity Assessment

In [9], the property *aggressivity* of a filter is presented for the first time. A filter is said to be more *aggressive* if it modifies a larger number of facial features than others. In this paper, to further quantify the *aggressivity* of a filter, we propose the use of SSIM. We present the SSIM index between the original and processed ("uploaded" and "beautified") images on the extended FFMF dataset in Table 3. For the three SNs, the images already present a degradation when uploaded to the platforms without any beautification filter application, with Snapchat images having a slightly higher penalization. We can observe that the TikTok filters (Belle and Spring glow) and the Instagram filter Thinner_face have a higher similarity with the original image when compared to other filters. This is consistent with the information presented in Table 2, where we can see how those filters modify a smaller number of biometric features. The difference is greater for the Snapchat filters Fresh light and Mellow glow, which modify a bigger number of facial traits. Nevertheless, Fresh vibes strongly penalizes the SSIM score indicating that although a small number of features is modified, their variations are significant. Finally, a strong impact of Instagram filters (Relax! You Pretty and Glam Grain) on facial traits is observed (lowest SSIM score). No significant SSIM discrepancies in terms of gender are observed, meaning that filters equally impact the SSIM of the female and male.

5.2 Impact on face verification

Table 4 reports the performance in terms of accuracy and AUC. Mellow glow, Glam Grain, "uploaded" by Sc, and "original" all achieve the lowest accuracy or AUC values. Interestingly, for ArcFace w/ ResNet100, the lowest performance in terms of accuracy is that of the original images, which means that somehow the application of the filters has improved the accuracy of face recognition. In line with this statement, if we look at the *positive* impact, and thus the higher accuracy and AUC values, it is the Belle filter that achieves the best results overall, followed by Fresh light, "uploaded" by SC, and "uploaded" by Ig. This corroborates what is reported in Ueda *et al.* [10] about light make-up, namely slight facial modifications aimed at beautification can even improve recognition performance.

The results presented in Table 5 and illustrated in Fig.3 report the performances of face verification in terms of FMR and FNMR. They should be read as the rate of authorized users wrongly rejected (FNMR) at fixed rates of impostors (i.e. unauthorized users) wrongly accepted (FMR). It can be observed that at FMR1000, the difference between the filters is wider. Meaning that, for a system with higher security, i.e. that allows only $1/1000$ impostor to access the system, the impact of filters is more significant. Regarding ArcFace, the results, both for negative and positive impact, converge most clearly on two filters. Regarding the negative impact, and thus the highest error rates, Glam Grain, the most aggressive filter in our selection, achieves the overall highest error rates, with Thinner_Face, Relax! You Pretty!, and "uploaded" by Tiktok achieved the highest error rates for some of the tests too. The largest performance drop is obtained at FMR1000 with Glam Grain for ResNet50 and Thinner_Face for ResNet100, where about 10% more authorized persons would be rejected when using the Glam Grain/Thinner_Face filter compared to original. As for the positive impact, the filter Belle, which is the least aggressive, clearly improves face verification performance for all tests. This means that images processed with this filter are less likely to be falsely rejected in line with Ueda *et al.* [10] and Table 4. As for MagFace, the largest drops in performance are achieved by Mellow glow, and Spring glow. While the largest improvements are obtained by Relax! You Pretty!, Fresh light, and "uploaded" by Tk. This indicates that the impact of compression and beautification filters depends on the architecture and the loss function used by the FR model.

5.3 Impact on gender classification

In Table 6, the assessment of the two selected open-source gender classifiers is presented. Initially, when "original" images are passed to both estimators, we remark some disparities in their performance for male and female subjects. We can observe a significant disadvantage for the female subjects, indicating a potential bias towards this group.

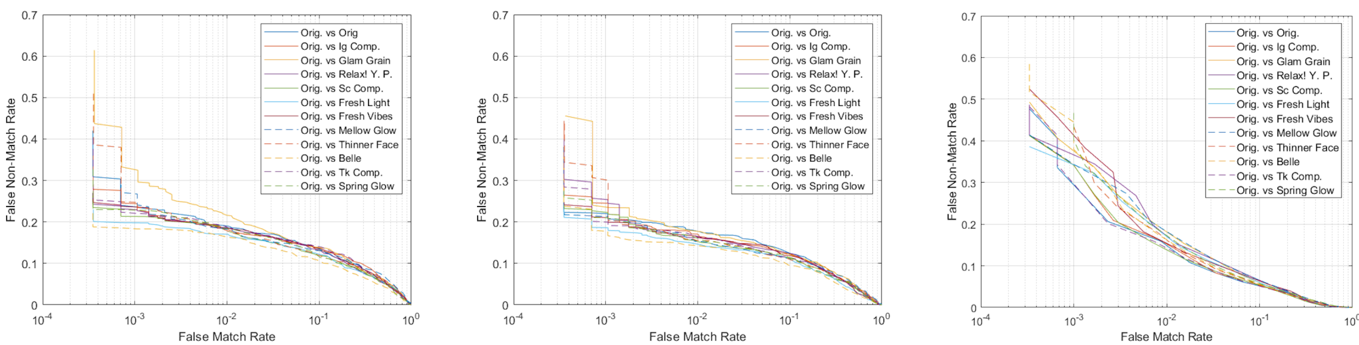
Regarding the "uploaded" images we can see that *cvlib* is globally more robust than *DeepFace*, with less fluctuation in the results. When the filters are applied, the use of digital beautification enhances the ability of gender estimators to correctly classify female subjects. This is more explicit for *DeepFace*. We can see that for both datasets the male accuracy remains stable when filters are used. Exceptionally, for Relax! You Pretty! the accuracy for the male class drops. However, when filters are applied, the accuracy of the female group increases for almost all filters therefore the global accuracy augments accordingly. This trend is more evident for the filters with a lower SSIM as shown in Table 3, such as the Snapchat filters and the Instagram filters Relax!

TABLE 4: Assessment of the impact of beautification filters on *face verification*. Accuracy, corresponding standard deviation (±) and the Area Under the ROC Curve (AUC) are reported. The *higher* the value, the better.

		FFMF-CALFW ArcFace w/ ResNet 50		FFMF-CALFW ArcFace w/ ResNet 100		FFMF-CALFW MagFace w/ iResNet 100	
Experiment		**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**
Original		0.9351	0.9030±0.0041	0.9444	0.9074±0.0059	0.9810	0.9417±0.0076
Ig	Uploaded	0.9308	0.9047±0.0086	0.9437	0.9123±0.0071	0.9805	**0.9418±0.0063**
	Thinner_Face	0.9381	0.9090±0.0053	0.9481	0.9157±0.0056	0.9802	0.9373±0.0070
	Relax! You Pretty!	0.9379	0.9030±0.0087	0.9464	0.9145±0.0089	0.9769	0.9245±0.0099
	Glam Grain	0.9322	0.8963±0.0054	0.9424	0.9096±0.0073	0.9773	0.9295±0.0095
Sc	Uploaded	0.9398	0.9085±0.0048	0.9410	0.9203±0.0056	**0.9818**	0.9398±0.0060
	Fresh light	0.9401	0.9118±0.0092	0.9543	**0.9231±0.0083**	0.9783	0.9300±0.0090
	Fresh vibes	0.9351	0.9046±0.0113	0.9488	0.9139±0.0105	0.9768	0.9315±0.0078
	Mellow glow	0.9300	0.9103±0.0100	0.9441	0.9195±0.0087	0.9753	0.9238±0.0087
Tk	Uploaded	0.9340	0.9045±0.0063	0.9478	0.9149±0.0064	0.9814	0.9398±0.0081
	Belle	**0.9493**	**0.9160±0.0074**	**0.9584**	0.9216±0.0079	0.9793	0.9360±0.0084
	Spring glow	0.9418	0.9038±0.0093	0.9498	0.9159±0.0087	0.9771	0.9317±0.0091

TABLE 5: Assessment of the impact of beautification filters on *face verification*. Performances are reported in terms of FNMR (%) at fixed values of FMR. The *lower* the value, the better.

		FFMF-CALFW ArcFace w/ ResNet 50			FFMF-CALFW ArcFace w/ ResNet 100			FFMF-CALFW MagFace w/ iResNet 100		
Experiment		**FMR10**	**FMR100**	**FMR1000**	**FMR10**	**FMR100**	**FMR1000**	**FMR10**	**FMR100**	**FMR1000**
Original		13.35	18.95	23.61	12.38	17.63	21.98	4.57	15.90	40.60
Ig	Uploaded	13.11	18.33	24.4	12.2	17.07	24.5	4.70	16.07	34.20
	Thinner_Face	11.84	18.17	24.7	10.89	15.57	**30.05**	4.67	18.00	43.40
	Relax! You Pretty!	12.97	18.68	22.95	**12.48**	16.19	25.37	5.87	15.47	34.47
	Glam Grain	13.88	**21.58**	**32.47**	12.34	**17.77**	23.48	5.40	20.17	40.67
Sc	Uploaded	11.98	17.92	23.07	11.14	15.2	22.27	4.57	15.63	34.13
	Fresh light	12.17	17.03	19.83	10.84	14.51	18.6	5.20	19.73	32.23
	Fresh vibes	13.59	18.45	24.05	11.87	16.21	21.01	5.50	18.13	45.70
	Mellow glow	13.08	18.24	26.73	11.16	15.43	21.29	**6.00**	16.63	34.37
Tk	Uploaded	**14.16**	18.78	22.04	11.35	16.43	19.94	4.43	15.87	41.33
	Belle	10.71	16.47	18.32	9.53	14.3	17.72	4.93	17.60	44.50
	Spring glow	11.4	18.54	22.98	11.51	16.19	20.98	5.23	**20.70**	**46.67**



(a) ArcFace w/ ResNet50 (b) ArcFace w/ ResNet100 (c) MagFace w/ iResNet100

Fig. 3: (Color online) *Face Verification*: Detection error tradeoff (DET) curves.

You Pretty and Glam grain. Concerning *cvlib*, we remark fewer gender classification differences than *DeepFace* when non-filtered images were provided, nevertheless the observed impact of filters on this estimator is greater. Although the impact is minor for most of the filters, when Glam grain is applied to FFMF-CALFW and FFMF-VIP or when Mellow Glow or Belle are applied to FFMF-CALFW, not only the accuracy of the female subjects increase significantly but also the male class suffers a notable drop. When Spring glow is used, both classes suffer from a decrease in performance. In general, we observe how the most popular open-source gender estimators are not balanced by the classes they aim to predict and, similarly, the use of digital beautification via SN affects the female and male classes in different ways. While males are more difficult to recognize when beautification is applied, women benefit from this pre-processing.

In Table 2, we can see that filters enlarge the lips or shrink the nose size, modify the skin color and/or add some virtual makeup by including ARF, making their effect similar to the use of makeup, acting as foundation or lipstick or imitating some makeup techniques such as face contouring. Researchers have associated the use of makeup with an increase of apparent femininity [39], by studying the impact of cosmetics on automated gender estimation, proving that it can negatively impact the employed neural networks, which then fail to classify men and women [11]. An analogous behavior is observed in our experiments where the use of filters increases the image femininity leading to an increase in female accuracy and a decrease in male recognition.

5.4 Impact on age estimation

Table 7 presents the results of the apparent age estimation experiments in the FFMF-VIP and FFMF-CALFW. When the images are "uploaded" to SNs, we observe small changes in

TABLE 6: Assessment in terms of classification accuracy of the impact of beautification filters on *gender classification*. The *higher* the value, the better.

		FFMF-CALFW						FFMF-VIP					
		DeepFace			cvlib			DeepFace			cvlib		
Experiment		All	Female	Male	All	Female	Male	All	Female	Male	All	Female	Male
	Original	0.93	0.87	**0.99**	**0.81**	0.81	0.81	0.92	0.85	0.99	**0.96**	**0.98**	0.95
Ig	Uploaded	0.93	0.88	**0.99**	**0.81**	0.81	0.81	0.86	0.75	0.98	**0.96**	0.97	**0.96**
	Thinner_Face	0.92	0.85	**0.99**	**0.81**	0.80	0.81	0.87	0.76	0.98	**0.96**	0.97	0.95
	Relax! You Pretty!	0.94	0.90	**0.99**	0.80	0.82	0.78	0.87	0.79	0.96	**0.96**	0.97	0.94
	Glam Grain	0.95	0.91	**0.99**	0.80	**0.83**	0.77	0.94	0.88	0.99	**0.96**	**0.98**	0.94
Sc	Uploaded	0.93	0.88	**0.99**	0.80	0.79	0.82	0.94	0.89	0.99	**0.96**	0.97	0.95
	Fresh light	**0.96**	**0.93**	**0.99**	0.79	0.81	0.78	0.96	0.92	**1**	0.95	**0.98**	0.93
	Fresh vibes	0.95	0.91	**0.99**	0.79	0.81	0.78	0.96	0.94	0.99	0.94	0.95	0.94
	Mellow glow	0.95	0.92	**0.99**	0.79	**0.83**	0.76	**0.97**	**0.95**	0.99	0.94	0.99	0.90
Tk	Uploaded	0.93	0.88	**0.99**	**0.81**	0.80	0.83	0.95	0.90	0.99	**0.96**	0.96	0.95
	Belle	0.95	0.91	0.98	0.80	0.75	**0.85**	0.93	0.89	0.98	**0.96**	0.96	0.95
	Spring glow	**0.96**	**0.93**	**0.99**	0.74	0.74	0.74	0.95	0.90	0.99	0.94	0.97	0.92

TABLE 7: Assessment of the impact of beautification filters on *age estimation*. The *lower* the Mean, SD and MAE, the better. The *higher* the ρ, the better.

		FFMF-CALFW						FFMF-VIP					
		Female			Male			Female			Male		
Experiment		SD	MAE	ρ	SD	MAE	ρ	SD	MAE	ρ	SD	MAE	ρ
Ig	Uploaded	0.784	0.571	**0.997**	0.708	0.583	**0.998**	1.002	0.804	**0.992**	0.936	0.859	**0.995**
	Thinner_Face	1.149	0.879	0.994	0.961	0.840	0.996	1.152	0.875	0.990	1.234	1.118	0.991
	Relax! You Pretty!	1.996	1.764	0.982	1.809	1.770	0.989	3.601	3.516	0.906	**3.701**	4.893	0.922
	Glam Grain	2.252	2.188	0.979	1.847	2.023	0.988	3.579	3.195	0.918	3.411	4.554	0.932
Sc	Uploaded	1.159	0.887	0.994	0.975	0.778	0.996	1.73	1.882	0.979	1.507	1.609	0.987
	Fresh light	2.114	2.252	0.981	2.047	2.538	0.894	3.182	3.467	0.931	2.740	3.930	0.957
	Fresh vibes	2.682	3.177	0.969	2.455	3.350	0.978	**3.762**	**4.097**	0.904	3.150	4.518	0.943
	Mellow glow	3.008	2.984	0.962	2.756	3.360	0.976	3.645	3.179	0.909	2.984	3.663	0.949
Tk	Uploaded	1.254	0.900	0.993	1.004	0.755	0.996	1.750	1.693	0.979	1.520	1.481	0.987
	Belle	2.357	2.781	0.978	2.050	2.796	0.984	3.596	4.002	0.914	3.027	4.681	0.948
	Spring glow	**4.642**	**5.769**	0.915	**2.823**	**4.601**	0.971	3.679	3.755	0.923	3.449	**5.622**	0.932

the estimated age reported. When beautification filters are applied the correlation coefficient remains high, never going below 0.91. Even so, diverse effects can be observed. The study on the FFMF-VIP ”beautified” dataset reveals that the application of almost all the filters modify the apparent age estimation with similar strength. The computed MAE across all groups of images in the FFMF-VIP group ranges between 4.0 and 3.1 years for the female subjects and between 5.6 and 3.6 years for the male subjects. Nevertheless, the Instagram filter Thinner_face does not fall in this interval having close to no impact on the predicted age. The computed MAE intervals are both wider and higher for the male subjects indicating that the use of filters implies a higher disturbance when predicting the age of males. Moreover, we can observe an increase in the MAE and Mean for the male subjects with respect to the MAE and Mean of the female subjects when Spring glow, Relax! You pretty!, and Glam grain are applied.

On the FFMF-CALFW set, there are no big discrepancies between the estimated age from the ”uploaded” and ”beautified” images for the female and male subjects. A special case stands for the filter Spring glow delivering a MAE of 5.7 and 4.6 years for female and male respectively.

The presented discrepancies in terms of gender can be explained by some unwilling network bias induced at the training stage. Similar to the discussion in Section 5.3, the use of filters can be compared to the digital addition of makeup. In our experiments, we observed that the DEX estimator is more robust to disturbances induced by FFMF and ARF for female subjects than for the male category. Makeup variability is likely not present in the training set for male, making this class more sensitive to filter disturbances.

5.5 Impact on weight estimation

TABLE 8: Assessment of the impact of SN filters on *weight estimation* on the FFMF-VIP. The *lower* the Me, SD and MAE, the better. The *higher* the ρ and PAP, the better.

Experiment		Me	SD	MAE	ρ	PAP
	Original	2,79	22,82	8,52	**0,68**	**68 %**
Ig	Uploaded	2,55	22,94	8,65	0,67	54%
	Thinner_Face	2,42	**23**	8,69	0,67	54%
	Relax! You Pretty!	3,92	22,78	8,89	0,65	53%
	Glam Grain	2,86	22,42	9,16	0,64	47%
Sc	Uploaded	2,62	22,86	8,82	0,66	52%
	Fresh light	5,50	22,18	9,76	0,63	49%
	Fresh vibes	4,33	22,14	**9,85**	0,61	49%
	Mellow glow	4,83	22,27	9,83	0,61	44%
Tk	Uploaded	2,37	22,57	8,98	0,65	50%
	Belle	3,56	22,39	9,21	0,63	48%
	Spring glow	**5,52**	22,03	9,81	0,63	48%

In Table 8 the results of the weight network are presented. When predicting the weight from the ”uploaded” images we observe that TikTok has the strongest impact on the weight model, increasing the network’s MAE by almost 0.5kg. The effect of Snapchat and Instagram filters is similar, increasing slightly the SD and MAE while dropping the PAP around 15%. Comparing ”uploaded” with ”beautified” images, we can observe stability in terms of SD, an increase in MAE ranging from 0.4 kg to 1.3 kg, a decrease of ρ up to 0.5 points and a decrease of PAP of a maximum of 8%. Glam Grain, Fresh vibes, Fresh light, Mellow glow, and Spring glow are the filters with a bigger impact on the MAE, making it drop of more than $0,5$ kg with respect to their corresponding ”uploaded” images. As shown in Table 2, those filters modify biometric features such as nose and face contour, which are important for predicting the weight from face images being facial contour the most crucial region for this task [33]. The Me of Thinner_face, the lowest among all categories of images, suggests that the weight estimator assigns lower weights for this type of data, consistent with the fact that the main effect of this filter is to shrink the face. The biggest drop in PAP with respect to the "uploaded" images is caused by Glam Grain and Mellow glow, filters that, as reported in Table 2, apply the highest amount of modifications to face images and, consequently, present the lowest SSIM on the FFMF-VIP as shown in Table 3.

5.6 Impact on heart rate estimation

TABLE 9: Assessment of the impact of SN filters on *HR estimation* on FFMF-COHFACE. The *lower* the SD and MAE, the better. The *higher* the ρ, the better.

Experiment	SD	MAE	ρ
Original	7.23	5.5	**0.62**
Uploaded	8.06	6.73	0.53
Thinner_face	8.67	6.31	0.51
Relax! You Pretty!	**10.98**	**9.1**	0.18
Hawaii_grain	10.12	8.32	0.38

In Table 9, the evaluation of the HR estimator on the videos of the FFMF dataset is presented. Since Instagram is the SN that includes a wider range and a higher usage of filters, this study focuses on filters from this platform.

The results displayed in Table 9 confirm our hypothesis, the application of beautification filters severely penalizes the HR estimation from face videos with higher SD and MAE values and lower correlation coefficient ρ between the real and the predicted HR when a filter is applied. More specifically, for the filters that apply skin smoothing, as presented in Table 2 Relax! You Pretty! and Hawaii_grain, the high SD values indicate a low model precision. Parallel to this metric, a value of $\rho < 0.40$ and $\rho < 0.20$ indicate low correlation and uncorrelation for the filters Relax! You Pretty! and Hawaii_grain, respectively. For Thinner_face, we can see a smaller increase of error than for the other filters studied which can be explained by the fact that this filter does not apply any skin smoothing as shown in Table 2.

In this experiments, we proved that beautification filters prevent HR from being successfully detected on face videos since most filters apply a skin smoothing technique erasing information usually encoded in skin pixels. In this case, filters reveal themselves as an inexpensive way to protect those attributes and, consequently, the users' privacy.

6 Conclusions

In this paper, we analyze the current trend of digital face beautification via SN filters and its impact on the biometric information recovered from faces. In addition to extending the FFMF database with 14182 new face images "uploaded" and "beautified" with 5 new filters from 2 new social networks (Snapchat and TikTok), we also included 240 new videos, obtained by compressing and beautifying face videos from the COHFACE database through Instagram filters. We propose to *quantify* the effect of each filter by using SSIM to define their *aggresivity*. The impact of each selected filter in the different studied tasks is summarized in Table 10. Filters that reported a low SSIM value, such as Glam Grain and Relax! You Pretty!, have consistently demonstrated a significant drop in the performance of the models in all experiments. Belle, the filter that shows a higher SSIM value, has been proven to act as light makeup thus increasing the user's feminity leading to an improvement in gender classification for females and an increase in face recognition performances, supporting the findings of Ueda *et al.* [10]. However, filters undoubtedly act as a disturbance factor for AI-face facial biometric models. To conclude, we list possible steps to take in order to mitigate the filter effect on facial processing tasks. We advise including filtered images in the training step of biometric systems such as FR or gender estimation, since digital beautification is now a common variation in the real world. The impact on microsignals extraction, such as HR, which can be erased from filtered videos, cannot be compensated in the training step. Therefore, we propose two approaches; 1) the design of beautification filters that do not degrade hidden microsignals, allowing the machine to detect them, while beautifying the video for the human eye, 2) to restore the concealed microsignal after filtering the video.

TABLE 10: Impact of each filter on the analyzed tasks taking the compressed images as a baseline. Filters are ranked from lower to higher aggressivity according to SSIM values.

	Filter name	FR	Gender	Age	Weight	HR
A						
G	Belle	+	++	++	+	-
G	Spring glow	+	++	+++	++	-
R	Thinner_Face	++	+	+	+	+
E	Fresh Lights	+	++	++	++	-
S	Mellow glow	++	+	++	+++	-
S	Fresh vibes	++	+	++	+++	-
I	Relax! Your Pretty!	++	++	++	++	+++
V	Glam Grain*	+++	+++	++	+++	-
E	Hawaii Grain*	-	-	-	-	+++

+: Low impact, ++: Medium impact, +++: High impact.
*Glam Grain and Hawaii Grain have similar aggresivity.

References

[1] A. Dantcheva, C. Velardo, A. D'Angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739–777, Jan. 2011.

[2] A. Nait-Ali, "Hidden biometrics: Towards using biosignals and biomedical images for security applications," in *International Workshop on Systems, Signal Processing and their Applications, WOSSPA*. IEEE, 2011, pp. 352–356.

[3] N. Mirabet-Herranz, "Social media filters: Beautification for humans but a critical issue for ai," *Science Talks*, vol. 9, 2024.

[4] C. Lavrence and C. Cambre, ""do i look like my selfie?": Filters and the digital-forensic gaze," *Social Media+ Society*, vol. 6, no. 4, p. 2056305120955182, 2020.

[5] A. Javornik, B. Marder, J. B. Barhorst, G. McLean, Y. Rogers, P. Marshall, and L. Warlop, "'what lies behind the filter?'uncovering the motivations for using augmented reality (ar) face filters on social media and their effect on well-being," *Computers in Human Behavior*, vol. 128, p. 107126, 2022.

[6] J. Yang, J. Fardouly, Y. Wang, and W. Shi, "Selfie-viewing and facial dissatisfaction among emerging adults: A moderated mediation model of appearance comparisons and self-objectification," *International Journal of Environmental Research and Public Health*, vol. 17, no. 2, 2020.

[7] P. Riccio and N. Oliver, "Racial bias in the beautyverse: Evaluation of augmented-reality beauty filters," in *European Conference on Computer Vision*. Springer, 2022, pp. 714–721.

[8] P. Hedman, V. Skepetzis, K. Hernandez-Diaz, J. Bigun, and F. Alonso-Fernandez, "On the effect of selfie beautification filters on face detection and recognition," *Pattern Recognition Letters*, vol. 163, pp. 104–111, 2022.

[9] N. Mirabet-Herranz, C. Galdi, and J.-L. Dugelay, "Impact of digital face beautification in biometrics," in *EUVIP 2022, 10th European Workshop on Visual Information Processing*, 2022.

[10] S. Ueda and T. Koyama, "Influence of make-up on facial recognition," *Perception*, vol. 39, no. 2, pp. 260–264, 2010.

[11] C. Chen, A. Dantcheva, and A. Ross, "Impact of facial cosmetics on automatic gender and age estimation algorithms," in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 2. IEEE, 2014, pp. 182–190.

[12] C. Rathgeb, P. Drozdowski, D. Fischer, and C. Busch, "Vulnerability assessment and detection of makeup presentation attacks," in *2020 8th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2020, pp. 1–6.

[13] C. Rathgeb, D. Dogan, F. Stockhardt, M. De Marsico, and C. Busch, "Plastic surgery: An obstacle for deep face recognition?" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 806–807.

[14] S. Suri, A. Sankaran, M. Vatsa, and R. Singh, "On matching faces with alterations due to plastic surgery and disguise," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2018, pp. 1–7.

[15] R. Singh, M. Vatsa, H. S. Bhatt, S. Bharadwaj, A. Noore, and S. S. Nooreyezdan, "Plastic surgery: A new dimension to face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 441–448, 2010.

[16] C. Rathgeb, A. Dantcheva, and C. Busch, "Impact and detection of facial beautification in face recognition: An overview," *IEEE Access*, vol. 7, pp. 152 667–152 678, 2019.

[17] A. Bharati, M. Vatsa, R. Singh, K. W. Bowyer, and X. Tong, "Demography-based facial retouching detection using subclass supervised sparse autoencoder," in *2017 IEEE international joint conference on biometrics (IJCB)*. IEEE, 2017, pp. 474–482.

[18] H. Chen, W. Li, X. Gao, and B. Xiao, "Aep-gan: Aesthetic enhanced perception generative adversarial network for asian facial beauty synthesis," *Applied Intelligence*, pp. 1–28, 2023.

[19] P. Riccio, B. Psomas, F. Galati, F. Escolano, T. Hofmann, and N. M. Oliver, "Openfilter: A framework to democratize research access to social media ar filters," in *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022.

[20] E. Derman, C. Galdi, and J.-L. Dugelay, "Integrating facial makeup detection into multimodal biometric user verification system," in *2017 5th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2017, pp. 1–6.

[21] V. Albiero, K. Zhang, M. C. King, and K. W. Bowyer, "Gendered differences in face recognition accuracy explained by hairstyles, makeup, and facial morphology," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 127–137, 2021.

[22] N. Kose, N. Erdogmus, and J.-L. Dugelay, "Block based face recognition approach robust to nose alterations," in *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2012, pp. 121–126.

[23] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, "Digital face beautification," in *ACM Siggraph 2006 Sketches*, 2006, pp. 169–es.

[24] C. Botezatu, M. Ibsen, C. Rathgeb, and C. Busch, "Fun selfie filters in face recognition: Impact assessment and removal," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 5, no. 1, 2022.

[25] A. Libourel, S. Husseini, N. Mirabet-Herranz, and J.-L. Dugelay, "A case study on how beautification filters can fool deepfake detectors," in *IWBF 2024, 12th IEEE International Workshop on Biometrics and Forensics*, 2024.

[26] T. Zheng, W. Deng, and J. Hu, "Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments," *arXiv preprint arXiv:1708.08197*, 2017.

[27] A. Dantcheva, F. Bremond, and P. Bilinski, "Show me your face and i will tell you your height, weight and body mass index," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018.

[28] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," *preprint arXiv:1709.00962*, 2017.

[29] N. Bousnina, J. Ascenso, P. L. Correia, and F. Pereira, "Impact of conventional and ai-based image coding on ai-based face recognition performance," in *2022 10th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2022, pp. 1–6.

[30] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019.

[31] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," 2021.

[32] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015.

[33] N. Mirabet-Herranz, K. Mallat, and J.-L. Dugelay, "New insights on weight estimation from face images," in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 2023, pp. 1–6.

[34] ——, "Deep learning for remote heart rate estimation: A reproducible and optimal state-of-the-art framework," in *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges*.

[35] "Information technology — Biometric performance testing and reporting — Part 1: Principles and framework," International Organization for Standardization/International Electrotechnical Commission, Standard, May 2021.

[36] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European conference on computer vision*. Springer, 2016, pp. 87–102.

[37] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[38] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *CVPR*, 2020.

[39] M. Carter, "Facials: the aesthetics of cosmetics and makeup," vol. 8, 1998.

Nelida Mirabet-Herranz received her Bachelor's degree in Mathematics from the Universidad de Valencia in 2018 and her Master's degree in Data Science from EURECOM in 2020. She obtained her Ph.D. in Computer Vision and Biometrics from Sorbonne University in 2024 and is currently a postdoctoral researcher at EURECOM, Sophia Antipolis.

Chiara Galdi received her joint supervision PhD from the University of Salerno, Italy, and from Telecom ParisTech, France, in 2016. She is currently Assistant Professor at EURECOM, Sophia Antipolis, France. Her current work focuses on the domain of biometrics, image forensics, and security and fairness of AI-based systems.

Jean-Luc Dugelay obtained his PhD in Information Technology from the University of Rennes in 1992. He is currently a Full Professor at EURECOM, Sophia Antipolis. His current work focuses on Image Engineering & Security. He has co-authored 300+ publications. He is the founding Editor-in-Chief of the EURASIP Journal on image and Video Processing (SPRINGEROPEN).