Impact of Digital Face Beautification in Biometrics

Nelida Mirabet-Herranz, Chiara Galdi, Jean-Luc Dugelay
dept. of Digital Security
EURECOM
Sophia Antipolis, France
{mirabet, galdi, dugelay}@eurecom.fr

***Abstract*—Face retouching is a widespread procedure available across a huge spectrum of modern applications. Among them, social media offer different filters to beautify face pictures by performing operations such as skin smoothing, addition of virtual makeup, as well as deforming certain facial features, for instance by widening the eyes or making the nose thinner. In this work, the effect of different facial feature modification filters (FFMF) on face recognition (FR), gender classifiers and a weight estimator are studied. To this end, popular FFMF are applied to face images of the publicly available CALFW and VIP_attribute databases. Such filters distort or modify biometric features, affecting the ability of automatic FR systems to recognize individuals. The results show that the application of FFMF to face images penalizes the accuracy of FR systems and affects the estimation of other facial traits such as gender and weight.**

***Index Terms*—Digital beautification, Social media filters, Face recognition, Gender classification, Weight estimation**

I. Introduction

Beautification is the alteration of the appearance of a subject to make it more visually attractive. This can be done by modifying the perceived or real shape and texture of different features. Facial features can be manipulated in the physical domain (plastic surgery and makeup presence) or in the digital domain (beautification filters).

Selfies, photographs typically taken with a smartphone that a subject takes of oneself are becoming more popular since technology has grown and adapted to the subject needs. Although Yang *et al.* [1] stated that selfies can be a positive way to allow people to express themselves, they also pointed out in their study the adverse effects selfies could bring to users and whether they could be related to facial dissatisfaction due to appearance comparisons. Since it has been proved that filtered faces are some of the most heavily engaged photos on social media [2], the use of filters on images has increased. Nowadays social media platforms offer many different filters to modify and beautify images by using editing tools. These modifications include the application of alterations to different face features such as skin, eyes, and nose, e.g., by enlarging the lips or smoothing the skin. Beautification has been already pointed out as a compromising factor for Face Recognition (FR) systems since the use of plastic surgery, cosmetics or retouching might be profound hence substantially altering facial traits [3], [4]. More particularly, the use of social media filters on a daily basis poses a new challenge. Filters are not necessarily used with the aim of compromising FR systems but might as well affect the capability of detecting the face and recognizing individuals [5]. Since social media evidence can be used to create event timelines, reveal conspiracy intents, and establish connections between people [6] the impact of different type of filters in FR needs to be assessed in order to avoid major security breaches.

Little attention has been given to the impact that such facial filters might have on other tasks. Despite the fact that soft biometrics are usually estimated from other biometric traits [7], a direct estimation from the subject's face is also possible. Scenarios where a face image has been altered with a filter causing the enlargement, shrinking, and sharpening of different facial traits are potentially able to alter the behaviour of gender and weight estimators too. However, to the best of the authors' knowledge, the impact of social media filters on the estimation of other soft biometric traits has not been investigated yet.

Motivated by the above, in this work, we explore the impact of social media filters on different face analysis tasks. The analysis of influencing factors, including face beautification filters, relevant for the final decision of a FR system, is an essential step to understand and improve the underlying processes involved. We select as our object of study filters that modify different facial traits but are not necessarily noticeable by the human eye, which are hereafter referred to as "Facial Features Modification Filters" (FFMF), since an image of this kind could be taken as original, e.g. passport ID picture, when instead it has been modified. The main contributions of this work are the following: 1) A new database, called FFMF dataset, is created, consisting on 5 types of images (original, compressed, and filtered in three different manners) from the CALFW and VIP_attribute publicly available databases. 2) We conduct the first study, to our knowledge, that assess the impact of different FFMF on FR. ArcFace [8], commonly used as a baseline for FR, is used as reference FR pipeline. 3) To the best of our knowledge, the first study on the impact of social media filters on the estimation of two soft biometric traits, gender and weight, from face images is proposed.

The rest of this paper is organized as follows. In Section II a review of the state-of-the-art works that assess the problem of digital beautification is presented. Section III details the FR, gender classification, and weight estimation models used, as well as the FFMF dataset. The results of the models for each type of image present in the database are summarized in Section IV. Finally, conclusions and future research directions are reported in Section V.

II. RELATED WORKS

Beautification is the process of making visual alterations to the perceived shape and texture of a human face in order to increase the beauty of the subject. These modifications can therefore compromise the use of FR systems in security applications since they distort or modify biometric features. Different types of facial beautification include the use of real-time social media filters, plastic surgery, and makeup.

Makeup: Facial cosmetics have the ability to substantially alter the facial appearance. The use of makeup can visually modify the proportions of different face traits such as eyes and cheekbones. The use of makeup has been proved as an effective attack for FR [9], [10]. In this type of attack, a person might apply a high amount of makeup in order to imitate the facial appearance of a target user with the purpose of impersonation. Related to this issue, in [10] the vulnerability of different open-source FR systems such as ArcFace [8] is assessed with regards to different makeup presentation attacks. A makeup attack detection scheme is proposed comparing face depth data with face depth reconstructions obtained from RGB images of potential makeup presentation attacks.

Plastic surgery: Facial plastic surgery can be divided into two classes; Reconstructive Plastic Surgery, which rectifies the face features anomalies e.g., cleft lip, palate birthmarks and Cosmetic Plastic Surgery that enhances the visual appearance of the facial structures and characteristics. When an individual gets plastic surgery, the facial characteristics are naturally transformed to an extent that alters the facial appearance. Consequently, these individuals may become unknown to the already existing FR systems, including their reference templates. These surgery changes pose a challenge to automatic FR technology [3]. Although facial plastic surgery is usually employed for cosmetic and scars treatment to improve the person's appearance, it might also be used by criminals to 'manipulate' their facial identity with the intent to deceive FR systems [11], [12].

Social media filters: Face retouching is a widespread application available across a huge spectrum of modern smartphone applications. A social media filter is a photo effect that is applied to images, including face images, before publishing them. These effects can range from simple RGB to black-and-white image transformation to different facial shape modifications. Facial social media filters can be divided into four classes:

- Colour Adjustment Filters (CAF): Change in real-time the captured image colours. Also modify the image contrast or illumination.
- Smoothing Filters (SF): Smooth and blurry face skin, as if a layer of foundation was applied.
- Facial Features Modification Filters: Enlarge, shrink, and sharpen different lines of the face. The most common ones create a thinner nose, a more defined jawline, or increase the lips' size. FFMF are not necessarily noticeable by the human eye since they alter the facial features without adding unnatural or extraneous elements or textures to the face. These types of re typically used in combination with SF.
- Augmented Reality Filters (ARF): An Augmented Reality (AR) filter is defined as a mask-like filter that incorporates virtual elements to face images in real time such as long eyelashes, makeup, puppy ears, or flower crowns, sometimes occluding parts of the face.
- Immersive AR Face Filter (IARFF): Place the users' face into a virtual 3D scene in real time.

Among the filters presented above just the SF and FFMF are filters that modify the subject's face for beautification in a nearly unnoticeable way for the human eye. CAF affect the whole image and are frequently applied in landscapes pictures while ARF and IARFF play an amusement role.

The first to present a categorisation of beautification and the challenges caused by facial beautification were Rathgeb *et al.* in their survey [3], where they give for the first time attention to facial retouching in the digital domain. Two unpublished works address the impact of filters in FR tasks. In [5] Hedman *et al.* assess for the first time the impact of CAF and ARF on FR and propose a method to inverse the applied manipulation. A counter-filter is developed by implementing a modified version of the U-NET segmentation network. More recently, Botezatu *et al.* investigate the impact of "fun" selfie filters, filters that fall in the category of ARF, and they evaluate their impact on different face detector and recognition models [13]. To mitigate the effect of the filters, the authors also propose a GAN-based filter removal algorithm consisting of a segmentation module, a perceptual network, and a generation module.

In our work, we aim to assess for the first time the impact on different systems of beautification filters, i.e. filters that affect the shape of different facial traits and initially are not noticeable for the human eye, more specifically SF and FFMF.

III. EXPERIMENTAL SETUP

In this Section, we present the FFMF dataset as well as the models used to evaluate the impact of the filtered images on FR, gender classification, and weight estimation.

A. Datasets

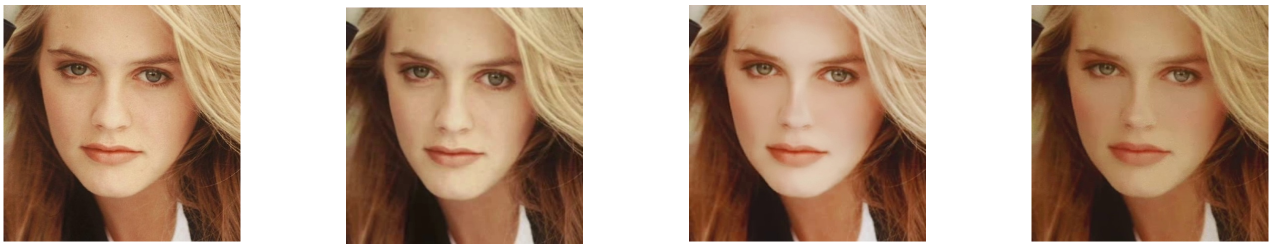
The FFMF dataset is composed by filtered images initially presented in the two publicly avilable dataset CALFW [1] and VIP_attribute [2].

CALFW: The Cross-Age LFW (CALFW) [14] is an improved version of the LFW face dataset [15], where more face pairs with age gaps were included to add age variation and intra-class variance while keeping the same identities as in the LFW dataset. It is a public benchmark for face verification with face photographs designed for studying the problem of unconstrained FR. The CALFW dataset contains 4,025 individuals with 2, 3 or 4 images for each person.

[1]http://whdeng.cn/CALFW/?reload=truedownload
[2]http://www.antitza.com/VIP_attribute-dataset.htmldownload

TABLE I
SUMMARY OF SOME OF THE MOST POPULAR INSTAGRAM FILTERS. THE TABLE INCLUDES THE NUMBER OF VIDEOS FOR WHICH THE FILTER WAS APPLIED, THE NUMBER OF VIEWS OF THE MOST POPULAR VIDEO AND THE FACIAL TRAITS THAT THE FILTER MODIFIES.

Name	Number of videos	Views	Color adjustment	Skin smoothing	Face contour	Eyes	Nose	Lips
Thinner_face	20	16K	x		x	x		
California Dreamin'	113K	1,4M	x	x			x	x
Relax! You Pretty!	236K	2,4M	x	x			x	x
Hawaii grain	127K	300K	x	x		x	x	x
Glam grain	170K	1,4M	x	x		x	x	x
Baby boo	57K	900K	x	x		x	x	x



Original image. Filter: Thinner_face Filter: Relax! You Pretty! Filter: Glam grain

Fig. 1. Example of images on the FFMF dataset. The original one is shown on the left and the remain 3 are the ones modified by the Instagram FFMF filters.

VIP_attribute: The VIP_attribute [16] is a dataset composed of facial images annotated for gender, height, weight, and Body Mass Index (BMI), which has been used to prove that facial images contain discriminatory information pertaining to those traits. The database consists of mainly frontal face images of celebrities collected from the WWW. It contains one image of each of the 1026 (513 female and 513 male) subjects enrolled in it. Since the VIP_attribute dataset is distributed in an already tightly cropped version, the authors of [16] provided us with the original dataset. This was necessary for the filtering step, if the face is cropped too tight the Instagram filter might not detect it thus making impossible the pplication.

Beautified datasets: Our FFMF dataset is composed by 10130 images belonging to three categories, original, beautified, and compressed:

- Original: The 2026 original images were obtained by taking all the 1026 images presented in VIP_attribute and a subset composed by 1000 images from the CALFW. Later studies have shown how some commercial face gender classification systems perform better on male and on light faces [17] promoting the creation of balanced dataset in order to mitigate bias [18]. The VIP_attribute dataset is already gender balanced, but in order to keep the FFMF dataset proportional, we selected from CALFW 125 female and 125 male subjects with 4 images per identity obtaining 500 images per gender. The list of selected subjects will be provided upon request.
- Compressed: When uploading/downloading images on social media, the images typically go through image compression and eventually other operations such as resizing and cropping [19], which may as well impact the FR and soft biometric evaluation. The objective of this work is to assess whether the FFMF have an impact on FR and soft biometric models. We thus need to exclude other image processing steps, such as image compression, from the evaluation. For this reason, the original images were uploaded and downloaded to/from Instagram obtaining 2026 compressed pictures that are used as a baseline to measure the impact of the beautification filters.
- Beautified: In order to apply the filters to the selected original images, we summarize in Table I some of the most popular filters on Instagram. The filters presented in the Table could also be applied to already existing images or videos, making it easier for us to apply them to the existing images of CALFW and VIP_attribute datasets. Table I reports the name of the filter applied, the number of videos (also known as *Reels*) that were modified using the selected filter, and the popularity of the most viewed modified video. It also presents different image and facial factors marking with a cross which ones are altered by the filter. We selected the two most popular filters for our experiments, that is *Relax! You Pretty!* and *Glam Grain*, as well as another filter *Thinner_face* that, although not one of the most popular is interesting for the purposes of our study since it modifies a different feature than the others: the face contour. The beautified part of the dataset is then created by applying the three selected Instagram filters to the original images resulting in 6078 filtered images. An example of the resulting images is shown in Figure 1.

B. Face recognition network

In order to assess the impact of FMMF on FR, ArcFace [8] is selected, a FR solution that reports high verification performance on two most widely used FR benchmarks LFW [15] and YTF [20]. ArcFace uses ResNet as backbone architecture and proposes a novel Additive Angular Margin Loss to obtain highly discriminative features for FR.

In our experiments with ArcFace, the implementation provided by the InsightFace project is used[3], ResNet100 is chosen as backbone. The pre-trained network, available on the InsightFace website, was trained on the MS1MV2 [21] dataset.

C. Gender classifiers

For evaluating the effect on gender classification of FFMF, we selected two among the most popular open source models for gender estimation. *DeepFace* [22] is a FR and facial attribute analysis library for Python. While the open-sourced DeepFace's facial recognition module works with existing state-of-the-art models, it includes its own models for facial attribute analysis, more specifically for gender. When the facial attribute analysis is performed four labels belonging to age, gender, emotion and race are returned. *cvlib* [23] is a high level open source computer vision library for Python. It focuses on the task of face and object detection and it includes a model for gender classification, AlexNet network trained on Adience dataset [24]. Once a face is detected, it is passed on to the function to recognize gender returning the labels (man, woman) and associated probabilities.

D. Weight estimatior

Dantcheva *et al.* [16] conducted a study of height, weight, and BMI estimation from a single subject's facial image by implementing a ResNet architecture with 50 layers. They report promising result demonstrating that weight can be estimated from a single facial shot. For assessing the impact of FFMF on weight, we selected likewise to [16] a ResNet50 structure. The model was trained on 800 images of the VIP_attribute dataset during 10 epochs and the final regression layer during 10 epochs more. Adam optimizer was selected with learning rate 0.01. The loss function was Huber loss with $\delta = 1$.

IV. Experimental results

In this Section, the impact of digital beautification with FFMF is presented, both on the verification of individuals' identity and on the extraction of soft biometric traits, in particular gender and weight.

A. Impact on face recognition

In order to assess the impact of digital beautification of FR, the following experiments were designed: the face verification performance is evaluated (i) on the original images from CALFW database; (ii) on the compressed images to assess the impact of compression only (no beautification filters); (iii) on the three selected beautification filters. The objective is to observe possible differences in the verification performance according to the different image processing.

ArcFace is used for assessing face verification performance for the different experiments. The reported performance metrics are the following:

- False match rate (FMR);
- False non-match rate (FNMR);
- Detection error trade-off (DET) curve;
- For the sake of comparison with previous works, classification accuracy and Area Under the ROC Curve are also reported.

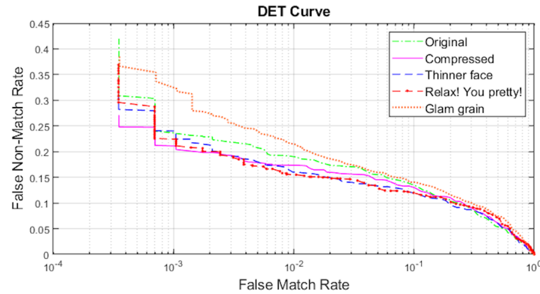


Fig. 2. Detection error tradeoff (DET) curve of ArcFace [8].

Prior to applying ArcFace, face landmarks are computed for each face image using RetinaFace [25] face detector. The face landmarks are recomputed for each variation: original, compressed, and the three filters. This is done to take into account the possible displacement of facial features due to the application of the filters.

As for the face verification protocol, the same protocol as proposed in [15] is adopted. Ten folds of images are created by randomly selecting images from the FFMF dataset. Each fold contains 300 matching pairs and 300 non-matching pairs, for a total of 6000 face comparisons (10×600). However, in case the face landmark estimation failed for a face image, the corresponding comparison is excluded from the test. The same set of pairs is used for all evaluations. ResNet100 backbone and batch size of 64 is used for the evaluation.

Table II reports the performance in terms of accuracy and AUC. As can be observed, the filter Glam Grain is the one presenting a stronger negative impact on the FR performance. Thinner_face have a hardly noticeable impact, while Relax! You Pretty on the contrary, shows to slightly improve the verification performance.

The results presented in Table III, and Fig. 2 report the perfomance of face verification in terms of FMR and FNMR. They should be read as the rate of people wrongly rejected (FNMR) at fixed rates of impostors wrongly accepted (FMR). The lower the FNMR, the better. It can be observed that at $FMR = 0.1\%$, the difference between the filters are wider. Meaning that, for example, for a system with higher security, i.e. that allows only $1/1000$ impostor to access the system, about 10% more authorised persons would be rejected when using the Glam Grain filter compared to "original".

B. Impact on soft biometrics

Gender classification. In table IV we present the accuracy (in %) of the two open source gender classifiers: *DeepFace* (deepface) and *cvlib* (cvlib). In addition to provide the overall results of each network, we also present the performance of the network for the female (deepface_f and cvlib_f) and male

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

TABLE II
PERFORMANCE ASSESSMENT OF ARCFACE [8] FOR FACE VERIFICATION ON 5000 SUBJECTS OF THE FFMF DATASET. WE PRESENT THE ACCURACY AND CORRESPONDING STANDARD DEVIATION ($\pm$), AND THE AREA UNDER THE ROC CURVE (AUC).

(%)	**Original**	**Compressed**	**Thinner_face**	**Relax! You Pretty!**	**Glam grain**
Accuracy	90.74 ± 0.59	90.89 ± 0.88	91.37 ± 0.75	91.62 ± 0.74	89.81 ± 0.96
AUC	94.44	93.36	93.53	93.26	92.32

TABLE III
ARCFACE [8] FACE VERIFICATION PERFORMANCE VALUES COMPARISON. FNMR AT FIXED VALUES OF FMR.

FMR (%)	**Original**	**Compressed**	**Thinner_face**	**Relax! You Pretty!**	**Glam grain**
10	12.38	12.98	12.47	11.98	14.05
1	17.63	17.35	17.30	15.50	21.38
0.1	21.98	21.11	25.53	22.42	32.23

(deepface_m and cvlib_m) subjects. The results are computed using 5130 (2565 female and 2565 male) face images from the FFMF dataset. We use the images annotated by gender provided in its initial form by the VIP_attribute database.

As initially thought, for the compressed images both models suffer from a decrease of performance, with *cvlib* showing to be more robust. We can also observe a potential bias for both networks since the female images are more heavily penalized due to compression. These compressed images will be our baseline for further comparisons. When the filters are applied, an interesting behaviour is reflected on the gender accuracy. We can observe how the general performance of the model tends to improve the more *aggressive* the filter is, which is specially noticeable for *DeepFace* whose performance, initially stated as 86,84%, increases up to 93,56%. We consider a filter as more *aggressive* if it modifies a larger number of face features as presented in Table I. If we look in more details the accuracy of the networks per gender, we can remark how the increase of performance is due to a rise of accuracy in both models for the female subjects in opposition to the male accuracy that decreases the more *aggressive* is the filter. If we look at the images presented in Figure 1 we can note how the effect of some filters is similar to the use of some makeup, such as foundation or lipstick. It has been proved that makeup is associated with women femininity [26] and some works have already addressed how facial cosmetics can be used to confound automated gender estimation schemes [27]. This explains how the use of these filters increases the number of images classified as women thus leading to an increase on accuracy for this category.

Weight estimation. In Table V the results of the weight network are presented. No pretrained model of the network presented in [16] is available, therefore we select as training set 800 original images (400 female and 400 male) belonging initially to the VIP_attribute dataset. The results in Table V are computed using the remain 226 (113 female and 133 male) identities annotated by weight provided in its initial form by the VIP_attribute database as well as the compressed and filtered versions of them, making a total of 1130 face images for testing. Several metrics are used to assess the FFMF impact on the weight algorithm, the Pearson's correlation coefficient (ρ) and the Mean Absolute weight Error (MAE) in kilograms (kg), which are the most common unit of measurement in weight researches. In addition, we present the mean and standard deviation (Me and STDe) in Kg of the weight error and the Percentage of Acceptable Predictions (PAP) introduced by [28]. The PAP represents the percentage of the prediction whose error is smaller than 10% of the initial weight, i.e. a reasonable error in medical applications.

The MAE, the Pearson's correlation coefficient between the prediction and the original weights and the PAP in Table V confirm how the presence of filters confuse the weight network leading to higher MAE and lower ρ and PAP specially for Glam Grain, the most *aggressive* filter. The Me of the filter Thinner_face, the lowest among all categories of images, suggests that the weight estimator assigns lower weights for this type of data, which is consistent with the fact that the main effect of this filter is to shrink the face.

V. CONCLUSION

Deep convolutional neural networks have achieved high accuracy in different tasks such as FR and gender classification surpassing in some cases human-level performance. In this work, we address the new trend of digital face beautification achieved through the application of social media filters and more specifically its impact on the information recovered from filtered faces through deep learning structures. We created the new FFMF dataset, composed by 10130 images including original, Instagram compressed, and filtered images with three different types of FFMF that alter different facial traits such as face contour and eyes. We tested ArcFace [8] on the FFMF dataset showing that the network, even if robust, it is not unaffected by the use of FFMF. A similar behaviour was observed for the weight estimation model for which the accuracy decreases when more *aggressive* filters are applied. Finally, for gender classification, it is observed that despite more *aggressive* filters penalize more heavily the accuracy, the femininity score of the images is instead increased leading to a higher and lower accuracy for female and male subjects, respectively.

TABLE IV
PERFORMANCE ASSESSMENT OF TWO POPULAR OPEN-SOURCE GENDER CLASSIFIERS ON 5160 SUBJECTS THE FFMF DATASET. THE RESULTS REPORTED INCLUDE OVERALL PERFORMANCE, AND ACCURACY PER GENDER.

(%)	deepface	deepface_f	deepface_m	cvlib	cvlib_f	cvlib_m
Original	91,81	84,40	99,22	96,68	98,05	95,32
Compressed	86,84	75,63	98,05	96,58	97,46	95,71
Thinner_face	87,13	76,02	98,24	96,68	97,66	95,71
RelaxYouPretty	87,62	79,14	96,10	96,00	97,07	94,93
Glam grain	93,56	88,10	99,02	96,10	98,05	94,15

TABLE V
PERFORMANCE ASSESSMENT OF THE WEIGHT ESTIMATOR ON 1130 SUBJECTS OF THE FFMF DATASET.

	Original	Compressed	Thinner_face	Relax! You Pretty!	Glam grain
MAE	8,52	8,65	8,69	8,89	9,16
Me	2,79	2,55	2,42	3,92	2,86
STDe	22,82	22,94	23	22,78	22,42
ρ	0,68	0,67	0,67	0,65	0,64
PAP	68%	54%	54%	53%	47%

ACKNOWLEDGMENT

This work has been partially supported by the European CHIST-ERA program via the French National Research Agency (ANR) within the XAIface project (grant agreement CHIST-ERA-19-XAI-011).

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] A. Ross, S. Banerjee, C. Chen, A. Chowdhury, V. Mirjalili, R. Sharma, T. Swearingen, and S. Yadav, "Some research problems in biometrics: The future beckons," in *2019 International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–8.

[5] 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," *arXiv preprint arXiv:2110.08934*, 2021 (Unpublished).

[6] A. Powell and C. Haynes, "Social media data in digital forensics investigations," in *Digital Forensic Education*. Springer, 2020.

[7] V. Cantoni, M. Porta, C. Galdi, M. Nappi, and H. Wechsler, "Gender and age categorization using gaze analysis," in *2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems*. IEEE, 2014, pp. 574–579.

[8] 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.

[9] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?" in *2012 IEEE Fifth international conference on biometrics: theory, applications and systems (BTAS)*. IEEE, 2012, pp. 391–398.

[10] 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.

[11] 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.

[12] 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.

[13] C. Botezatu, M. Ibsen, C. Rathgeb, and C. Busch, "Fun selfie filters in face recognition: Impact assessment and removal," *arXiv preprint arXiv:2202.06022*, 2022 (Unpublished).

[14] 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.

[15] 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.

[16] 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.

[17] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*. PMLR, 2018, pp. 77–91.

[18] K. Kärkkäinen and J. Joo, "Fairface: Face attribute dataset for balanced race, gender, and age," *arXiv preprint arXiv:1908.04913*, 2019.

[19] A. Berthet, F. Tescari, C. Galdi, and J.-L. Dugelay, "Two-stream convolutional neural network for image source social network identification," in *2021 International Conference on Cyberworlds (CW)*, 2021.

[20] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *CVPR 2011*. IEEE, 2011, pp. 529–534.

[21] 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.

[22] S. I. Serengil, "deepface," 2020. [Online]. Available: https://github.com/serengil/deepface

[23] A. Ponnusamy, "cvlib - high level computer vision library for python," https://github.com/arunponnusamy/cvlib, 2018.

[24] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on information forensics and security*, vol. 9, no. 12, pp. 2170–2179, 2014.

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

[26] M. Carter, "Facials: the aesthetics of cosmetics and makeup," *Literature & Aesthetics*, vol. 8, 1998.

[27] 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.

[28] C. Velardo and J.-L. Dugelay, "Weight estimation from visual body appearance," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2010, pp. 1–6.