Beauty Score Fusion for Morphing Attack Detection

Juan E. Tapia
juan.tapia-farias@h-da.de
Mathias Ibsen
mathias.ibsen@h-da.de
Christoph Busch
christoph.busch@h-da.de

da/sec-Biometrics
and Internet Security Research Group
Hochschule Darmstadt
Germany, DE.

Abstract

Face attractiveness is a subjective topic studied in order to determine when one person's face or image is "beautiful" based primarily on symmetrical and shaped facial traits. This topic has recently received attention because it may help determine if one face image was digitally modified to simulate different identities. In this work, we explored the capability to be applied to morphing images. Thus, an automatic face beauty score approach was trained using deep learning techniques based on human votes. This beauty score was applied to morph images created from different numbers of subjects, and a classifier was trained with latent vectors derived from two feature extractors and sent to a differential classifier focused on a score fusion scheme. Our finding raised the question of the impact of beautifying images and their effects on evading morph detection.

1 Introduction

Nowadays, many digital tools are available that have the purpose of modifying and altering facial images, such as digital makeup using apps, retouching and morphing images, to change the physical face appearance with the intent to remove imperfections or to reach a look similar to another subject or celebrity [13, 25, 26]. Beyond that, further changes can be made to face images in the digital domain, e.g. enlarging of the eyes. On the other hand, one subject and its accomplishments can create morph images to use the same picture in a passport application. The use of these tools to improve the face appearance is mostly called "beautification" [11, 25, 26, 33].

Moreover, in many countries, the applicant for electronic travel documents can provide the photo used for the identity document himself in printed form. That is, various types of image editing, including facial beautification, can be performed prior to the submission of the facial image to the administration and hence can negatively affect the performance of a facial recognition system, e.g., for automated border control.

The main hypothesis for this work is that a "less beautiful" face is correlated with bona fide images, which present more imperfection on the face skin, than the morph images, which present smoother faces. In that context, "less beautiful" means more imperfection on faces, and "most beautiful" means smothering faces, such as morph images.

Face images are manipulated to make their shapes closer to the average faces and are perceived as more attractive. The influences of symmetry and averageness are often confounded in studies based on full-face views of faces. Attackers or criminals have exploited these relationships and obtained access to border control applications with valid passports using retouching, averaging techniques and morphed face images [33, 35].

The Morphing Attack Detection (MAD) aims to detect unauthorised individuals who present an identity document with a morphed image to gain access to a "valid" identity, for instance, in border control processes. Morphing can be understood as a technique to combine *two or more* look-alike facial images from one subject and an accomplice who could apply for a valid passport. As both contributing subjects can be verified against the morphed face image, the identity of the accomplice can be exploited. Morphing takes place in the enrolment process stage. Automatic detection of morphing attacks can be broadly divided into two types: (1) Single Image Morphing Attack Detection (S-MAD) techniques (no-reference MAD) and Differential Morphing Attack Detection (D-MAD) methods [30]. The S-MAD scenario is more challenging as the decision needs to be taken on a single image without a trusted image available for the same subject [5, 32, 34]. This approach is based on D-MAD according to real-life applications.

A morphing attack's success depends on the decision of human observers, especially a passport identification expert. The real-life application for a border police expert who compares the passport reference image of the traveller (digital extracted from the embedded chip) with the facial appearance of the traveller [29, 30] is too hard because of the improvements of the morphing tools and because of the difficulty for the human expert to localise facial areas, in which morphing artefacts are present [22].

To validate our hypothesis, we analyse the creation of morph images with many subjects (from 2 to 16) and estimate their automatic beauty score based on human votes. Further, we estimated differences using feature vectors from bona fide and morph faces obtained from deep face representations and their classification scores, which are fused with beauty scores to distinguish between bona fide and morphed face images.

The main contributions of this work are as follows:

- The automatic beauty scoring method was explored in order to show the feasibility of using beauty information to distinguish bona fide from morphing face images.
- Two new beauty networks were trained based on ICAO requirements for face images to explore how deep learning can also be used to automatically estimate the beauty score based on human votes.
- A challenge differential approach based on two different face recognition feature extractions was explored in order to train a classifier and measure the performance of D-MAD.
- A fusion score approach was proposed in order to measure the incorporation of the beauty score as complementary information for D-MAD.

This paper is organised as follows: The related work is described in Section 2. The database and metrics are in Section 3 and 4. The methods are described in Section 5. The experiments and results in Section 6, and the conclusion is presented in Section 7.

2 Related work

The state-of-the-art for facial beauty assessment has been studied considering different approaches using psychological points of view [17, 35], machine learning with face landmark detection, geometrical features [18, 27], and feature extraction based on texture, shape, or frequency [4, 19]. Deep learning techniques based on transfer learning approaches with several Convolutional Neural Networks (CNN) such as VGG, ResNet and Others [13, 31, 33]. Moreover, Generative Adversarial Networks (GAN) have been used in order to generate synthetic images with different levels of beauty [8, 11].

Valentine et al. [35] explored the perception of beauty from a psychological point of view and analysed the symmetrical and average features from frontal and profile (lateral) views. Both full-face and profile views were perceived as less attractive than the average shape. It is concluded that averageness is independent of any effect of symmetry on the perceived attractiveness of female faces.

Lian et al. [18] provided the SCUT-FBP5500 database created by votes from 60 volunteers with about 5,500 face images. They proposed different approaches for beauty scoring based on handcrafted feature extraction and CNN, such as AlexNet and ResNet. The result shows the superiority of CNN-based approaches over handcrafted and landmark-based.

Diamant et al. [8] propose generating and beautifying facial images using conditional beauty level, creating a variant of Progressive GAN. To verify the validity of this idea, they trained a predictive model on the SCUT-FBP5500 dataset and tested it on 200 random images from CelebAHQ [20]. 60 VGG models were trained, one per human rater, with the weights initialised from VGG trained on ImageNet [6].

Zhitong et al. [12] propose to generate synthetic face beauty images based on GANs using three modules: adaptive instances normalisation, cycle consistency and identity preservation beauty transformation. A Face Beauty Predictor (FBP) is introduced to ensure the output face's beauty score equals the target. FBP is a pre-trained CNN that can predict the beauty score of a given face, and the parameters of FBP are fixed during the whole training process. They also used the SCUT-FBP5500 as the face beauty dataset.

Inspired by the previous works in [18], we proposed to create a "BeautyNet" to extrapolate the knowledge of 60 human voters used in the SCUT-FBP5500 database to be adapted to FERET and FRGCv2 images based on ICAO constraints and to create an Automatic Facial Beauty score applied to bona fide and morph images. This new score was integrated into the MAD processing pipeline as additional information.

3 Database

In this work, four different databases were used: SCUT-FBP5500 [18], FERET [23], FRGCv2 databases [24] and MultiMorph [9].

The SCUT-FBP5500 database was used to train the "BeautyNet" and to estimate the beauty score. This dataset has a total of 5,500 images, and it is one of the most used in the literature. All the images are labelled with beauty scores ranging from [1, 5] provided by a total of 60 volunteers aged from 18-27 (average 21.6), where the beauty score 1 means "low beauty" and 5 means "high beauty" [18] as is shown in Figure 1. According to Liang et al., [18], 330,000 human beauty scores were evaluated using a web-based GUI system to obtain the facial beauty scores for each image.

The FERET dataset is a subset of the Colour FERET Database, generated in the context of the Facial Recognition Technology program technically handled by the National Institute

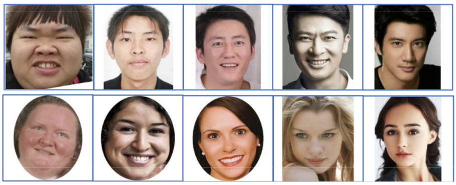

Figure 1: Example of SCUT-FBP5500 face images manually labelled. Left: "Less beauty score" (index 1). Right: "Most beauty score" (index 5).

Figure 2: Example of MultiMorph images. From left to right: Morph images using $K = 2$, 4, 8, 16. Where K is the number of face images.

of Standards and Technology (NIST). It contains 569 bona fide images.

The FRGCv2 dataset used in this work is a constrained subset of the second version of the Face Recognition Grand Challenge dataset. It contains 979 bona fide face images.

Using different morphing tools, the FERET and FRGCv2 databases were used to create the morph images based on the protocol described in [30].

All the images were captured in a controlled scenario and included variations in pose and illumination. FRGCv2 presents images that are more compliant with the passport portrait photo requirements. The images contain illumination variation, different sharpness and changes in the background. The original images have a size of 720×960 pixels. For this paper, the faces were detected, and images were resized and reduced to 224×224 pixels. These images still fulfil the resolution requirement of the intra-eye distance of 90 pixels defined by ICAO specification 9303 for Machine Readable Passports [15].

The MultiMorph dataset (MM) contains Morph faces from multiple K contributors and was created from 923 manually labelled FERET and FRGCv2 face database subjects [9]. This dataset *was used to evaluate and analyse the morph average hypothesis*. We replaced the bona fide with FERET and FRGC images that were not used in the morphing process. The quality assurance was manually performed for the face images of these subjects. Then, the morphing algorithm Face Morpher [3] was used. Table 1 shows a summary of the databases and a number of images used in this work.

Table 1: Number of images used for FERET, FRGCv2, MM and SCUT dataset.

Datasets	Nº Subjects	Bona fide	Morphs
FRGCv2	979	979	964
FERET	569	569	569
MultiMorph (MM)	3,692	923	10,123
SCUT-FBP5500	5,500	5,500	N/A

3.1 Morphing Tools

The following algorithms were used to create morph images from pairs of parent images: FaceFusion [2], FaceMorpher, OpenCV-Morph, [1] and UBO-Morpher [10]. These morphing tools represent different qualities of morphed images in terms of background noise,

Table 2: Morphing tool software and the number of images created by each method. The number of images is per dataset (FRGCv2/FERET).

Database	Nº Subjects	Bona fide	Morphs
FaceFusion	533/529	984/529	964/529
FaceMorpher	533/529	984/529	964/529
OpenCV-Morph	533/529	984/529	964/529
UBO-Morpher	533/529	984 /529	964/529

Figure 3: Example of Morphing tool images. Left to right: Subject 1, FaceMorpher, FaceFusion, OpenCV, UBO-Morpher and Subject 2.

artefacts, and others. Table 2 shows a summary of the images per each morphing tool. Figure 3 presents an example of each morphing tool.

4 Metrics

4.1 Morphing Attack Classification Error Rate

The detection performance of biometric MAD algorithms is standardised by ISO/IEC 20059 [16]. The most relevant metrics for this study are the Morphing Attack Classification Error Rate (MACER) and the Bona fide Presentation Classification Error Rate (BPCER).

The MACER metric measures the proportion of morphing attacks incorrectly classified as bona fide presentations in a specific scenario. The BPCER measures the proportion of bona fide presentations incorrectly classified as morphing attacks. The computation method is detailed in Equation 1, where the value of N corresponds to the number of morphing presentation images, Res_i is 1 if the ith image is classified as morphed, or 0 if it was classified as a bona fide presentation.

$$MACER = \frac{\sum_{i=1}^{N_M} 1 - Res_i}{N_M}, \qquad BPCER = \frac{\sum_{i=1}^{N_{BF}} Res_i}{N_{BF}} \tag{1}$$

On the other hand, the BPCER metric measures the proportion of bona fide presentations wrongly classified as a morph. BPCER can be computed using Equation 1, where N_{BF} is the amount of bona fide presentation images, and Res_i takes the same values described in the MACER metric. Together, the two metrics determine the performance of the system, and they are subject to a specific operation point.

Finally, BPCER_AP and the Equal Error Rate (EER) are used to analyse the system performance on a specific operating point. The latter is the operating point where MACER and BPCER are equal. This operating point corresponds to the intersection with the diagonal line in a Detection Error Trade-off (DET) curve, which is also reported for all the experiments.

5 Proposed Method

A new framework was proposed based on the state-of-the-art for facial attractiveness assessment and differential morphing attack detection, as shown in Figure 4. This new framework includes four different stages: Beauty scoring, Feature extraction, Classifier, and Score level fusion.

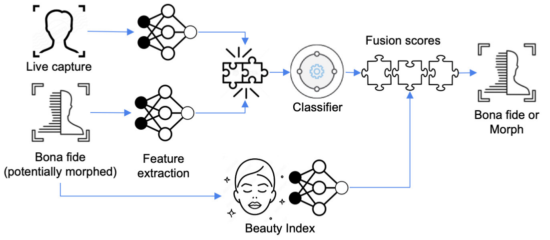

Figure 4: Beauty Morphing Attack Detection Framework.

5.1 Beauty Scoring

One of the challenges of this work is to develop a network named "BeautyNet" that allows us to create labels representing the beauty scores for any other morph dataset. For this task, we re-train a new predictor based on VGG-16 and VGG-19 [31]. Contrary to the SCUT-FBP5500 database, which is less in the wild faces, our images to be assessed are ICAO compliant. These networks were selected due to their simplicity compared to the state-of-the-art CNNs.

This architecture is attractive in comparison to other architectures by reducing the size of the CNN filters to 3×3 and increasing the depth of the network up to 19 layers. In their paper [31], the authors show that the greater depth of the network positively influences the classification results and that the VGGNet model is capable of generalising for a wide range of tasks and databases.

In order to analyse the influence of the new beauty scores, a comparison between bona fide and morphed images was created with different numbers of subjects. Figure 5 shows a Probability Distribution (PD) of the beauty score used to demonstrate and evaluate our hypothesis regarding that morph images are considered more beautiful while more subjects are used to create the morph. Blue curves represent the beauty score for bona fide images, and the Pink curve represents the beauty score for $K = 2$, 4, 8 and 16, respectively. The figures show that the PD is displaced to the right when the K value is increased.

5.2 Feature extraction

Deep Face Representations (DFR) are extracted from the reference and probe image using two different algorithms: ArcFace [7], and MagFace [21].

5.2.1 ArcFace

The ArcFace is based on the ResNet-50 and ResNet-100 architecture and uses Additive Angular Margin Loss to obtain highly discriminative features for face recognition [14].

ArcFace was shown to achieve state-of-the-art recognition performance on various challenging datasets. The pre-trained deep face recognition network is used as a feature extractor, i.e. the deep representations extracted by the neural network (on the lowest layer). Feature vectors comprising 512 floats are extracted from the suspected image and from the trusted

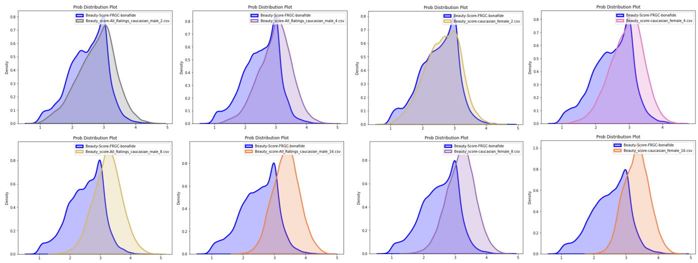

Figure 5: Left to Right: PD of the FRGC database (Bona fides) versus MALE/FEMALE face images probability distribution from MultiMorph for $K = 2$, $K = 4$, $K = 8$ and $K = 16$.

live capture face images. It is expected that alterations induced by morphed images will also be reflected in extracted deep-face features. Due to the high generalisation capabilities of deep face recognition systems with respect to textural changes of the skin face, such changes might be more pronounced in the case of morph and smothered faces induced through the beautification process. A similar procedure was applied to MagFace as feature extractors.

5.2.2 MagFace

MagFace tackles problems with previous angular margin-based losses, which are quality-agnostic and hence can lead to unstable within-class distribution as high-quality face images can stay at the decision boundary, whereas lower-quality images can be at the class centre. This is especially a problem in unconstrained in-the-wild recognition. To solve this, MagFace encodes quality information into the feature representation by considering the feature magnitude as a quality indicator.

5.3 Classifier

A Support Vector Machine (SVM) was used for training and classification, and the subtraction of embedding features was estimated from pairs of feature vectors extracted from the suspected (morphed) image and from the trusted live captured images. Specifically, an element-wise subtraction of feature vectors is performed [30]. In the training stage, difference vectors were extracted for each feature extractor, and SVMs with Radial Basis Function (RBF) kernels were trained to distinguish between bona fide and morphed beauty face images. Data normalisation was applied as the feature elements of extracted feature vectors are expected to have different ranges.

5.4 Score Level Fusion

The trained SVM generates a normalised MAD score in the range [0, 1]. Subsequently, a score level fusion is performed by testing the weight and sum score operation. The sum rule is used to obtain a fused score based on the final decision between the SVM score and the beauty score. The beautify score for each image is computed and normalised also in the range [0, 1] in the standalone process using the BeautyNet described in section 5.1 for FERET and FRGCv2 databases.

6 Experiment and Results

This section describes and explores a D-MAD approach to classify bona fide and morphing images based on the premise that morph faces are more "beautiful". In this case, our limitation is that the FERET and FRGC morphing sets were created using only the two subjects. For this task, we use three different experiments: Exp1.) D-MAD using only beauty scores. Exp2.) D-MAD using SVM scores from embedding. Exp3.) D-MAD using the fusion of beauty and MAD scores.

OpenCV and FaceMorpher images from FERET were used to train and estimate the SVM classification score. The FaceFusion and UBO-Morpher from FRGC images were used for testing. This selection was applied based on the quality of the output morphing tool, and it is considered a challenging scenario, as is shown in Figure 3.

6.1 Experiment 1: Beauty score

The SCUT-FBP5500 database was divided into 60% training, 20% validation and 20% tests. According to the network's requirements, the image size was reduced to $224 \times 224 \times 3$. Data augmentation is based on flip images, random rotation by 10 degrees, and centring crop. Our trained model is used for beauty prediction by using the L2-norm distance loss. Each raw RGB image is resized as 256×256, and then a 227×227 random crop of raw images is obtained to feed into VGG-16 or VGG-19. We set the batch size as 16, epochs=50, and weight decay coefficient as $5e-4$.

6.2 Experiment 2: Deep Face Representation (DFR)

In order to compare our results with the state-of-the-art [18, 28], the differential approach proposed by [30] was re-implemented. For DFR, a feature extractor based on ResNet50/100 for the ArcFace and MagFace was evaluated.

6.3 Experiment 3: Score Fusion

A fusion approach was explored to combine the SVM training scores from differential embeddings and the beauty scores. The beauty score was computed in a stand-alone operation. An alpha factor (α) was used to determine the contribution of the SVM and beauty scores. A grid search was used to determine the alpha value (α) as follows in the equation:

$$x = DFR_{Score} * \alpha_1 + Beauty_{Score} * (1 - \alpha_1) \tag{2}$$

Table 3, shows the EER results for the FRGCv2, UBO morphing tool as a test set. It is essential to highlight that the DFR system was trained with the FERET database. The table compares two feature extractors as backbones, ResNet50 and ResNet100, for two face recognition systems: ArcFace and MagFace, with α equal to 0.8.

From the results, we can observe that the beauty score alone performs the lower results in classifying bona fide and morph. This result is supported because any features extracted from the face recognition system were used for prediction.

Our implementation of the retrained MAD approach achieved the best results. However, when we fused the information with the beauty score, the results could not be considered a general improvement for MAD because they were based on morphed images created from only two subjects.

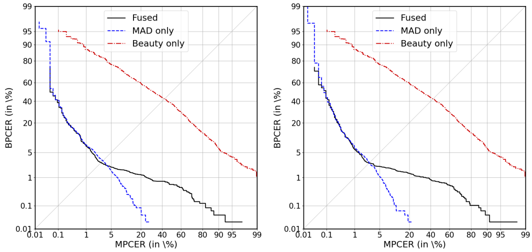

Figure 6: DET Comparison using UBO database as a test-set. Left to right: ArcFace and MagFace with alpha = 0.8 and ResNet100 as a feature extractor.

Figure 6 shows the DET curve for experiments 1, 2 and 3, considering a contribution of 80% for the SVM score (MAD) and 20% for the beauty score and ResNet100 as a feature extractor for ArcFace and MagFace respectively. According to our findings, after analysing **a)** the beauty information alone, **b)** MAD only and **c)** Fused score information, we can conclude that the fusion of beauty score helps only in some cases to improve the D-MAD results. These results are strongly correlated with the number of subjects used to create the morph. The results are sometimes competitive but also depend on the feature extraction process and face recognition. However, if more subjects or contributors are used to generate the morph images, the beauty score will be valuable to outperform the results according to Figures 5. The best results for the MAD approach were reached with MagFace, ResNet100 and α equal to 0.8 with an EER of 3.82%. The MAD approach presents better performance and consistency in most scenarios with ResNet50/100 as a backbone. The fused score approach of the MAD and beauty scores allows us to reach similar results and improve marginally in some cases, according to EER, as shown in Figure 6 and Table 3. The best results for our approach were reached with ArcFace, ResNet100 and α equal to 0.8 with an EER of 3.5%.

Table 3: EER results for D-MAD considering feature extraction from face recognition with ResNet100 with $\alpha = 0.8$. and UBO morphed as a test set.

Metric	ArcFace - EER (%)		MagFace - EER (%)	
Beauty	41.5	41.5	41.5	41.5
D-MAD	7.42	**3.9**	4.63	3.82
Fused - (MAD+Beauty **Ours**)	10.13	**3.5**	4.82	**3.81**

7 Conclusion

This work analysed whether morphed images created from FERET and FRGCv2 exhibit complement properties extracted from the beauty score. Using state-of-the-art morphing tools, an empirical evaluation of different morphed images created from $K = 2, 4, 8$ and 16 contributors showed that morph faces generally achieve a higher beauty score than random bona fide.

It is important to remark that the morphed images on the test set were created using only two subjects ($K = 2$) because this is the most realistic scenario. Therefore, we can not exploit the best condition with the $K = 8$ or 16 subjects. This condition is also extended to a state-of-the-art database because morphed images were also created based on morph images with two subjects.

Acknowledgements: This work is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 883356 and by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] Opencv. https://https://learnopencv.com/tag/face-morphing/. Last accessed: 2023-01-26.

[2] FaceFusion. www.wearemoment.com/FaceFusion, . Last accessed: 2023-01-26.

[3] FaceMorpher. https://github.com/yaopang/FaceMorpher/tree/master/facemorpher, . Last accessed: 2023-01-26.

[4] P. Aarabi, D. Hughes, K. Mohajer, and M. Emami. The automatic measurement of facial beauty. In *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, volume 4, pages 2644–2647 vol.4, 2001. doi: 10.1109/ICSMC.2001.972963.

[5] Laurine Dargaud, Mathias Ibsen, Juan Tapia, and Christoph Busch. A principal component analysis-based approach for single morphing attack detection. In *Proc. of the IEEE/CVF Winter Conf. on Appl.of Computer Vision (WACV) Workshops*, pages 683–692, January 2023.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[7] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[8] Nir Diamant, Dean Zadok, Chaim Baskin, Eli Schwartz, and Alex M. Bronstein. Beholder-gan: Generation and beautification of facial images with conditioning on their beauty level. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 739–743, 2019. doi: 10.1109/ICIP.2019.8803807.

[9] André Dörsch, Christian Rathgeb, Mathias Ibsen, and Christoph Busch. Are average faces master faces? In *2022 Intl. Workshop on Biometrics and Forensics (IWBF)*, pages 1–6, 2022. doi: 10.1109/IWBF55382.2022.9794554.

[10] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face demorphing. *IEEE Transactions on Information Forensics and Security*, 13(4):1008–1017, 2018. doi: 10.1109/TIFS.2017.2777340.

[11] Wenming Han, Fangmei Chen, and Fuming Sun. Facial beauty study based on 3d geometric features. In *2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML)*, pages 155–161, 2021. doi: 10.1109/PRML52754.2021.9520726.

[12] Zhitong Huang and Ching Yee Suen. Identity-preserved face beauty transformation with conditional generative adversarial networks. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 7273–7280, 2021. doi: 10.1109/ICPR48806.2021.9413167.

[13] M. Ibsen, C. Rathgeb, D. Fischer, P. Drozdowski, and C. Busch. Digital face manipulation in biometric systems. In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, Advances in Computer Vision and Pattern Recognition, pages 27–43. Springer Verlag, 2022.

[14] Mathias Ibsen, Lázaro J. González-Soler, Christian Rathgeb, and Christoph Busch. Tetraloss: Improving the robustness of face recognition against morphing attacks, 2024.

[15] International Civil Aviation Organization. Machine readable passports – part 9 – deployment of biometric identification and electronic storage of data in eMRTDs. http://www.icao.int/publications/Documents/9303_p9_cons_en.pdf, 2021. Last accessed: 2021-11-23.

[16] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC CD 20059. Information Technology – Methodologies to evaluate the resistance of biometric recognition systems to morphing attacks*. International Organization for Standardization, 2023.

[17] Judith H. Langlois and Lori A. Roggman. Attractive faces are only average. *Psychological Science*, 1(2):115–121, 1990. doi: 10.1111/j.1467-9280.1990.tb00079.x.

[18] Lingyu Liang, Luojun Lin, Lianwen Jin, Duorui Xie, and Mengru Li. Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. *ICPR*, 2018.

[19] Huali Liao and Jiexun Yang. Design and realization of real-time face beauty system based on android. In *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pages 253–261, 2022. doi: 10.1109/AEECA55500.2022.9919042.

[20] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[21] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. MagFace: A universal representation for face recognition and quality assessment. 2021.

[22] Robert Nichols, Christian Rathgeb, Pawel Drozdowski, and Christoph Busch. Psychophysical evaluation of human performance in detecting digital face image manipulations. *IEEE Access*, 10:31359–31376, 2022. doi: 10.1109/ACCESS.2022.3160596.

[23] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[24] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, et al. Overview of the Face Recognition Grand Challenge. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 947–954, 2005.

[25] C. Rathgeb, C.-I. Satnoianu, N. E. Haryanto, K. Bernardo, and C. Busch. Differential detection of facial retouching: A multi-biometric approach. *IEEE Access*, 8:106373–106385, 2020. doi: 10.1109/ACCESS.2020.3000254.

[26] C. Rathgeb, P. Drozdowski, and C. Busch. Detection of makeup presentation attacks based on deep face representations. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3443–3450, 2021. doi: 10.1109/ICPR48806.2021.9413347.

[27] Miriam Redi, Nikhil Rasiwasia, Gaurav Aggarwal, and Alejandro Jaimes. The beauty of capturing faces: Rating the quality of digital portraits. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–8, 2015. doi: 10.1109/FG.2015.7163086.

[28] U. Scherhag, C. Rathgeb, and C. Busch. Performance variation of morphed face image detection algorithms across different datasets. In *6th Intl. Workshop on Biometrics and Forensics*, pages 1–6, 2018.

[29] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch. Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7:23012–23026, 2019.

[30] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, and Christoph Busch. Deep face representations for differential morphing attack detection. *IEEE Transactions on Information Forensics and Security*, 15:3625–3639, 2020. doi: 10.1109/TIFS.2020.2994750.

[31] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. URL https://api.semanticscholar.org/CorpusID:14124313.

[32] J. E. Tapia and C. Busch. Single morphing attack detection using feature selection and visualization based on mutual information. *IEEE Access*, 9:167628–167641, 2021.

[33] Juan Tapia and Christoph Busch. Alphanet: Single morphing attack detection using multiple contributors. In *2023 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, 2023. doi: 10.1109/WIFS58808.2023.10374584.

[34] Juan E. Tapia and Christoph Busch. Face feature visualisation of single morphing attack detection. In *2023 11th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6, 2023. doi: 10.1109/IWBF57495.2023.10157534.

[35] T. Valentine, S. Darling, and M. Donnelly. Why are average faces attractive? the effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin*, Review 11:482–487, 2004.