

Does Melania Trump Have a Body Double from the Perspective of Automatic Face Verification?

Khawla Mallat¹^a, Fabiola Becerra-Riera², Annette Morales-González²^b, Heydi Méndez-Vázquez²
and Jean-Luc Dugelay¹^c

¹Digital Security Department, EURECOM Campus Sophia Tech, 450 route des Chappes F-06410 Sophia Antipolis, France

²Advanced Technologies Application Center (CENATAV) 7A #21406 Siboney, Playa, P. C. 12200, Havana, Cuba
{mallat, dugelay}@eurecom.fr; {fbecerra, amorales, hmendez}@cenatav.co.cu

Keywords: Information Verification, Automatic Face Verification, Conspiracy Theory, Image Quality.

Abstract: With the growing number of users getting updated about current events through social media, the spread of misinformation is increasing and thus endorsing conspiracy belief. During the last presidential election campaign in the USA, the conspiracy theory claiming the existence of a body double that stands in for the former first lady Melania Trump had made international news headlines. Fighting the spread of misinformation is crucial as it is threatening the society by manipulating the public opinion. In this paper, we explore whether automatic face verification can help in verifying widespread misinformation on social media, dealing particularly with the conspiracy theory related to Melania Trump replacement. We employed four different state-of-the-art descriptors for face recognition to verify the integrity of the claim of the studied conspiracy theory. In addition, we assessed the impact of different image quality metrics on the variation of the face verification scores. Two sets of image quality metrics were considered: acquisition-related metrics and subject-related metrics.

1 INTRODUCTION


This is not Melania! In October 2017, the hypothesis of the existence of a body double of Melania TRUMP appeared for the first time on social networks. Then regularly, what is now referred to as ‘*The Melania Trump replacement conspiracy theory*’¹ by mass media, reappears in the front of the stage, at least once a year, with a peak during the last presidential election campaign. Face is used by all of us as the main human trait to recognize each other in daily life. This is why, even if some other criteria have been used to demonstrate that Melania is not Melania, like her attitude and behavior or her body height, most arguments advanced by people are related to her facial appearance.


Algorithms based on artificial intelligence, image/video processing and computer vision are designed and used to fight against fake news that include doctored images (Mahfoudi et al., 2019) and/or AI-based generated videos (i.e. deepfakes) (Dolhan-


sky et al., 2020). The issue addressed in this article is different in the sense that the open question is not about the integrity of the picture but about the physical presence of the real Melania or of a body double in the scene recorded by the camera. So Biometrics (and more particularly Automatic Face Recognition), more than image forensic tools, is the appropriate technology to address the problem.

There are two possibilities in case of the non-existence of a body double of Melania: either people are complotists in the sense that they are aware that Melania has no double body but claim the contrary or, people are of good faith and consider by visual inspection that there are two different people on photographs: Melania and a body double based on some possible visual differences among faces. Visual inspection by human is subjective and today less accurate than a digital and objective inspection by machines for many tasks related to face analysis (Phillips and O’Toole, 2014). This is why automatic face verification may significantly help in the process for forming a robust opinion concerning this supposed conspiracy theory.

In this work, we use automatic face verification to conduct an analysis of a group of images of Melania

^a <https://orcid.org/0000-0003-2116-8219>

^b <https://orcid.org/0000-0003-2716-3144>

^c <https://orcid.org/0000-0003-3151-4330>

¹<https://w.wiki/4MVP>

Trump, in order to try to determine if she has a body double or not.

2 AUTOMATIC FACE VERIFICATION

In this section, we present primarily the image samples of Melania Trump on which our study has been based. Then, we introduce the state-of-the-art descriptors of face recognition employed in this study followed by the experimental protocol and results that will help answering the question ‘*Does Melania Trump has a body double?*’.

2.1 Face Samples

The presented study was driven by the interest of the French metropole television M6 in the conspiracy theory case related to Melania Trump². The TV channel provided us with a set of images of Melania Trump collected from social media, mainly Twitter, collected by a team of journalists working on the topic. These images were labeled by the social media users as the ‘*real*’ Melania Trump or as the ‘*body double*’. The total number of the image samples of Melania Trump is 30. Considering the labels provided by the journalists, 16 images were labeled as the ‘*real*’ Melania Trump and 14 as the ‘*body double*’. Faces were detected using RetinaFace (Deng et al., 2020) and only faces of Melania Trump were selected and then aligned. Figure 1 shows the preprocessed face images. At first glance, one could see that there is an important discrepancy between the two sets of images. First, the set of the alleged ‘*body double*’ contains 6 out of 14 images in which Melania is wearing large dark glasses which covers up the ocular area, while it is the case for only 1 out of 16 images for the ‘*real*’ Melania. In addition, one could see that the set of images of the ‘*real*’ Melania contains images where Melania is captured with a smiling face, open eyes and proper illumination conditions, while the images contained in the set of the alleged ‘*body double*’ of Melania seem to be taken from video recordings, where Melania appears with blinking eyes, head down and low illumination conditions.

2.2 Face Recognition Models

In this study, we employ four state-of-the-art face recognition descriptors that achieved more than 99.3% of accuracy when evaluated on LFW

dataset (Huang et al., 2007). For each face descriptor, feature embeddings are extracted from Melania’s face images and compared against each other using cosine similarity to perform the face verification task.

LightCNN (Wu et al., 2018). is an implementation of CNN for face recognition designed to have fewer trainable parameters than a vanilla CNN and to handle noisy labels. This network introduces a new concept of max-out activation in each convolutional layer for feature filter selection. We use the pretrained network with 29-layers on LFW dataset (Huang et al., 2007) to obtain embeddings of 256-dimension from face images.

MobileSqueezeNet (mob, 2019). is a lightweight CNN released with Intel’s OpenVino framework and model zoo, for face recognition in re-identification scenarios. It is based on MobileNet V2 (Howard et al., 2017) backbone, which consists of 3x3 inverted residual blocks with squeeze-excitation attention modules. After the backbone, the network applies global depthwise pooling and then uses 1x1 convolution to create the final 256-dimensional embedding vector. We used the pretrained model `face-reidentification-retail-0095`.

SphereFace (Liu et al., 2017). refers to angularly distributed feature embeddings designed for face recognition. In contrast to employing Euclidean margin to the CNN’s learned features, SphereFace proposes to use an angular softmax (named A-Softmax) to learn discriminative face features with a novel geometric interpretation. A-Softmax loss can be interpreted as constraining learned features to be discriminative on a hypersphere manifold, which intrinsically matches the prior that face images lie on a manifold. We use the 512-dimensional SphereFace embeddings extracted using the pretrained model on CASIA-WebFace dataset (Yi et al., 2014).

ArcFace (Deng et al., 2019). is Additive Angular Margin loss function, a new loss function for face recognition designed to improve the discriminative power of the learnt feature embeddings. The proposed loss function aims to optimise the geodesic distance margin by considering the correspondence between the angle and arc in the normalised hypersphere. We used a model pretrained on CASIA-WebFace dataset (Yi et al., 2014) to extract embeddings of 512-dimension from face images.

²<https://youtu.be/9HJKKYeJdTk?t=1239>

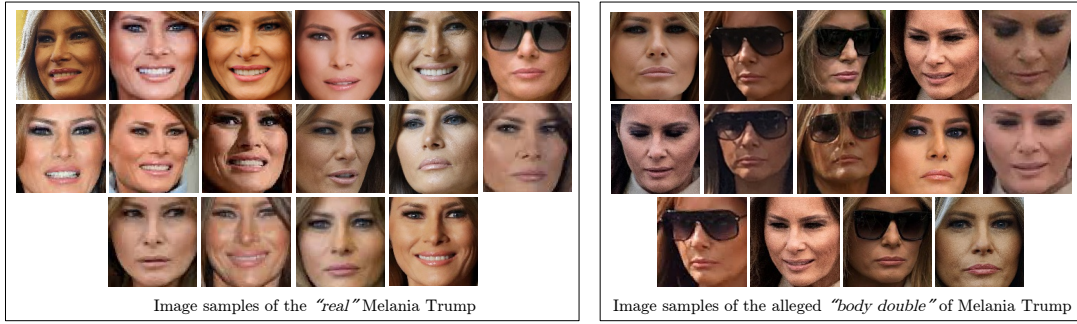
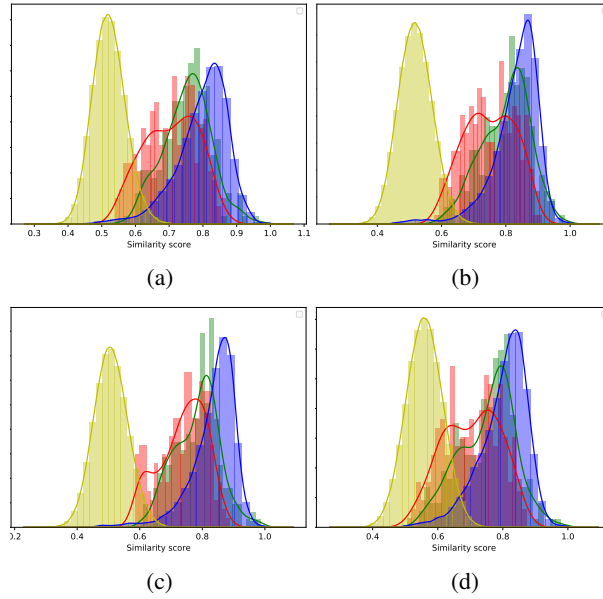


Figure 1: Preprocessed face images of the ‘real’ Melania and the alleged ‘body double’.



■ Genuines (Melania) ■ Imposters (Melania) ■ Genuines (Celebrities) ■ Imposters (Celebrities)

Figure 2: Similarity score distribution of ‘real’ against ‘body double’ Melania Trump (a) Arcface (b) LightCNN (c) MobileSqueezeNet (d) SphereFace.

2.3 Experimental Protocol and Results

Assuming we have 2 identities for Melania Trump, the first consists of the ‘real’ Melania and the second consists of the alleged ‘body double’, we performed a face verification experiment using an All-vs-All comparison scheme. For each identity, we performed the face verification experiment by considering a different reference face image at a time, each reference face image was compared against all the remaining image samples of the 2 identities, resulting in 224 genuine trials and 422 imposter trials. *Genuine* pairs belong to the same identity, while *Imposters* pairs correspond to the comparison of one identity vs. the other, meaning the ‘real’ Melania against the alleged ‘body double’. We illustrate in Figure 2 the score distributions of Genuines in green and of Imposters in red using the four selected face descriptors. We also introduced the

score distributions of imposters and genuines deduced from a dataset, collected from the web by the authors, of female celebrities in their 50s similar in profile to Melania Trump. This dataset contains 50 identities, with 30 images per identity.

It can be seen that the genuine (in green) and the imposter (in red) score distributions overlap substantially, although, there is a slight difference between the two score distributions. This difference might be due to the discrepancy in image samples between the set of ‘real’ Melania and the alleged ‘body double’ as depicted in section 2.1. However, we observe that the genuine and the imposter similarity score distributions deduced from the celebrities dataset are almost completely disjoint, as the dataset contains different identities. In section 3, we evaluate the impact of different factors that might affect the automatic face

Table 1: Equal error rate (EER) and area under the curve (AUC) of the different face verification systems.

	ArcFace		LightCNN		MobileSqueezeNet		SphereFace	
	EER (%)	AUC (%)	EER (%)	AUC (%)	EER (%)	AUC (%)	EER (%)	AUC (%)
Melania	39.3	66.2	37	68	38.8	66.2	39.3	65.5
Celebrities	0.27	99.2	0.22	99.2	0.20	99.4	0.43	98.7

verification score as an attempt to justify the slight difference between the green and the red score distribution. In addition, one could see that the distributions of genuine and imposter scores deduced from the Melania data generally fall into the genuine score distribution deduced from the celebrities dataset, with a minor overlap with the imposters score distribution.

Table 1 reports the equal error rates (EER) and the area under the curve (AUC) deduced from the face verification experiment performed on Melania image samples and on the celebrities dataset. We note that for distinct identities, as it is the case for the celebrities dataset, 0.02% of EER and 0.99% of AUC are reported in average. However, when computed for the Melania face verification experiment, the EER increased to 0.38% and the AUC decreased to 0.66% in average, implying that the 2 considered identities (the ‘real’ Melania and the alleged ‘body double’) are actually the same. The results reported by the different face recognition models, in figure 2 and table 1 are highly comparable. The selection of 4 different state-of-the-art face recognition models was motivated by the need to validate our conclusions.

3 FACTORS AFFECTING AUTOMATIC FACE VERIFICATION

In this section, we suppose that Melania Trump does not have a body double according to the significant overlap between genuine and imposter score distributions deduced from Melania image set, illustrated in section 2.3. However, in this section, we attempt to justify the difference between these two distributions highlighted hereabove. There are several factors that could affect the utility of a face image for a specific task. In this section, since we are particularly interested in face verification, we explored the impact of key quality metrics (Méndez-Vázquez et al., 2012), (Mallat et al., 2019) which were considered according to the particular conditions of our case of study and that are subject and environmental/sensor-related.

To highlight the impact of a given quality metric on the face verification performance, we sorted all the 30 image samples of Melania Trump according to

their corresponding quality measure in an ascending order. Feature embeddings were extracted from all the images. Then, the affinity matrix of the cosine similarity scores is computed between feature embeddings of Melania Trump’s images in an All-vs-All comparison scheme. Hereafter, only results obtained using ArcFace feature embeddings will be reported as all the four face descriptors depict the same behaviour.

3.1 Acquisition-related Image Quality Metrics

From simply looking at the image samples of Melania Trump, one could see a prominent difference between the ‘real’ and the alleged ‘body double’ sets mainly in the image acquisition quality.

Brightness. (Zohra et al., 2017) Depending on the acquisition conditions and the camera devices, images could often be affected by variations in terms of illumination and color leveling (the degree of lightness or darkness of a color) which result in brightness variations. Brightness is the perception caused by the luminance of an object, a face in our case. Major fluctuations of these measures could drastically change the appearance of subjects causing different skin color and shaded or extremely brighten zones that could even partially occlude discriminative face parts. Since brightness does not visually appear to be homogeneous across the Melania Trump image set, we decided to analyse how much it really affects the performance of face verification.

Face Luminance. (Méndez-Vázquez et al., 2012) Face luminance deals with the distribution of light across the face and it is measured in specific face regions that are prominent for the face verification process. The face image quality based on luminance, was measured by averaging the normalized mean luminance values for a set of triangles, representing key face regions, that are more likely to be affected by changes in illumination. Since illumination does not appear to be homogeneously distributed across the face, an assessment of the variation of face luminance on the face verification score is conducted.

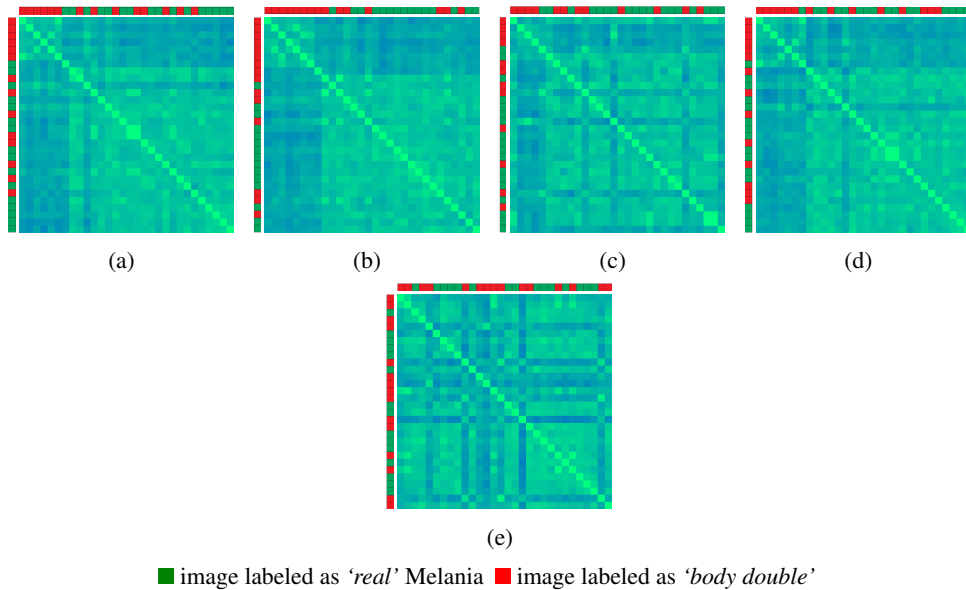


Figure 3: Affinity matrix of cosine similarity between ArcFace embeddings of Melania Trump’s images, sorted according to: (a) Brightness (b) Face luminance (c) Contrast (d) Exposure (e) Sharpness. Blue cells reflect a low similarity score (closer to 0) whereas green cells correspond to high similarity score (close to 1).

Exposure. (Shirvaikar, 2004) is the amount of light that reaches the acquisition sensor and it can be controlled using aperture and shutter settings. When the face image is correctly exposed, the image provides rich details in both dark shadow regions as well as bright regions. The Melania Trump set of images contain mainly underexposed images in which the dark shadow regions lose fine details and appear as pure black regions because of receiving too little light.

Contrast. (Gao et al., 2007) Low contrast images are commonly due to poor illumination conditions during the image acquisition, wrong setting of the aperture and shutter speed. Low contrast face images contain underexposed and overexposed regions that result in loss of geometrical and color information of facial attributes. Image contrast is defined as the separation between the darkest and brightest areas of the image, and it can be computed as the difference in color intensities.

Sharpness. (Méndez-Vázquez et al., 2012) In blurred images, the face regions (eyes, nose, mouth, etc.) are usually not sufficiently clear or well delimited due to the loss of visual information by confusing and mixing shapes, textures and colors. To decide to which extent more or less defined/sharp edges are influencing the face verification process, we applied an unsharp masking process to sharpen the original face image and average the resulting pixel intensities to obtain a sharpness quality score.

3.2 Subject-related Image Quality Metrics

The second main difference between the set of the ‘real’ and the alleged ‘body double’ is related to the physical appearance of Melania herself. One could see that the set of the ‘real’ Melania shows images of her with the head straight and a glowy face, whilst for the set of the alleged ‘body double’, Melania appears either wearing sunglasses or with the eyes closed with a varying head pose. Accordingly, 3 quality metrics that are subject-related were selected to evaluate their impact on face verification similarity score.

Occlusion by Sunglasses. (Méndez-Vázquez et al., 2012) Since several samples of the alleged ‘body double’ set show Melania wearing sunglasses, our focus was drawn on the occlusion of the periorbital region, which has been shown to be a very discriminative region in face recognition, and its impact on face verification performance. Therefore, to measure the quality of a face image based on the use of sunglasses, we first used a CNN classifier as sunglasses detector, and then we employed the output probability of the class representing “non sunglasses” as final quality score: face images with higher ‘non sunglasses’ probability are more likely to be good quality images and vice versa.

Femininity. (Mallat et al., 2017) A main factor that might have affected the face verification performance

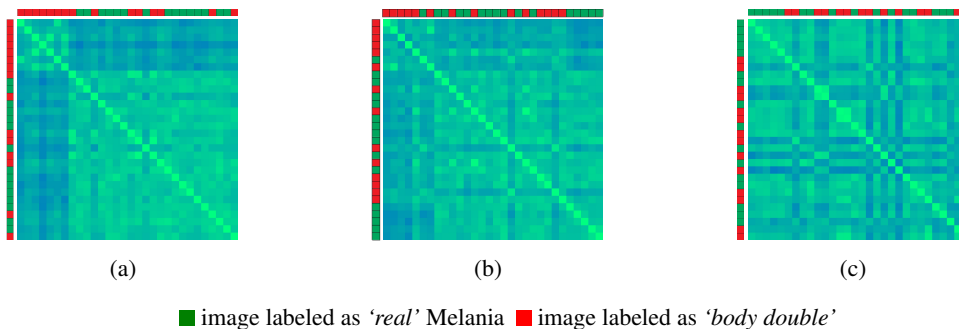


Figure 4: Affinity matrix of cosine similarity between ArcFace embeddings of Melania Trump’s images, sorted according to: (a) Occlusion by sunglasses (b) Femininity (c) Head pose. Blue cells reflect a low similarity score (closer to 0) whereas green cells correspond to high similarity score (close to 1).

can clearly observed in the Melania Trump images is her physical appearance. In the set of the ‘real’ Melania, she appears with radiant face and make up on, whilst in the alleged ‘body double’ set, Melania seems to be wearing too little make up or no make up at all, with a wearied face. Consequently, we propose to quantify the femininity appearance of Melania Trump on all the image samples and evaluate their impact on face verification performance. For this purpose, a gender estimation model (Antipov et al., 2016) was employed to deliver a score representing the probability of being a male or a female. We consider the femininity score as the probability of being a female. Mallat et al. (Mallat et al., 2017) proved that the more makeup is applied (up to a certain level), the higher the femininity score is.

Head Pose. (Méndez-Vázquez et al., 2012) The complexity of dealing with multiple head rotations under uncontrolled conditions has long been a bottleneck in face recognition. Different head poses usually imply the occlusion or deformation of some face regions, which can lead to the mismatch of two samples of the same person just because they were captured from different angles. It is worth noting that general Melania postures are basically frontal and just a few samples present some yaw or pitch variation, in particular for the alleged “body double” set.

3.3 Discussion

Figure 3 reports the affinity matrices of the cosine similarity scores that highlight the impact of image quality metrics that are related to the image acquisition: brightness, face luminance, sharpness, exposure and contrast. Blue cells reflect a low similarity score (closer to 0) whereas green cells correspond to high similarity score (close to 1).

In Figures 3a, 3b, 3c and 3d, one could see that images with low quality yield low similarity scores

and they correspond to image samples from the alleged ‘body double’, whilst high quality images deliver high similarity score and they are mostly associated to image samples of ‘real’ Melania. For the exposure quality metric, the lines of the affinity matrix, in Figure 3c, that depict low similarity scores for images with higher exposure correspond to specific image samples in which Melania Trump appears either with sunglasses or with the head down. Figure 3e shows a small correlation between sharper images and high similarity scores. However, we can deduce from the matrix that there are different and more relevant factors that are influencing the face verification performance. The cross-shaped blue area that depict low similarity scores from Figure 3e corresponds to the comparison of four images of average sharpness but considerably affected by illumination and using sunglasses; this is also the case of the two isolated sharp images producing the blue lines that converge in the bottom right corner of the matrix.

Figure 4 portrays the affinity matrices of the cosine similarity scores by sorting the image samples in an ascending order according to occlusion by sunglasses, femininity and head pose quality metrics, in order to highlight their impact on face verification performance.

In Figures 4a and 4b, one could perceive that the images with low quality scores yield low similarity scores and correspond to the image samples of the ‘body double’, whereas the images with the highest quality coincide with samples of the ‘real’ Melania and they deliver relatively higher similarity scores. The linear behaviour observed in Figure 4a is a confirmation that sunglasses have high impact on the variation of face verification scores when comparing the set of ‘real’ Melania with the alleged ‘body double’. The two crossed shaped blue areas, in Figure 4b, depicting low similarity scores for higher femininity values correspond to two samples of Melania wearing sunglasses. However, Figure 4c shows the pose vari-

ation is barely influencing the face verification performance: comparisons between good and bad quality samples output higher or lower similarity scores alike. This behaviour is an indicator that the used face descriptors are robust enough to the slight head pose variation across Melania's image samples.

Among the selected quality metrics, one could affirm that brightness, face luminance and occlusion by sunglasses are the most relevant quality metrics affecting the face verification performance, and to a lesser degree with exposure, contrast and femininity. Brightness, face luminance, exposure and contrast being all related to the illumination conditions during the image acquisition depict the same behaviour of the affinity matrix that could be seen in figures 3a, 3b, 3c and 3d. The slight shift between the genuine and imposter score distributions deduced from Melania Trump image samples, observed in subsection 2.3, is certainly due to the fact that the set of the 'real' Melania contains overall images of better quality compared to the set of the alleged 'body double', and obviously not because of the existence of a Melania Trump 'body double'.

4 CONCLUSION

This paper proposes a first attempt to explore whether automatic face verification can help in fighting the spread of conspiracy theories and if it could be used as an information verification tool. A further extensive study should be performed on a wider set of identities and their associated body doubles. In this paper, different state-of-the-art descriptors of face recognition were used to compare the 'real' Melania image set to the images of the alleged 'body double'. According to the proposed study, the answer to the question 'Does Melania Trump have a body double?' is definitely a *No*. In the studied case, automatic face verification invalidates the polemical conspiracy theory claiming the existence of a replacement of Melania Trump during the presidency term of Donald Trump. This disclaimer could only be reliable if we presume that the provided face image samples were not priorly manipulated. Recently, a new type of image manipulation intended to trick the deep learning based models into delivering an erroneous output. This type of image manipulation, called adversarial attacks, is performed by adding an adversarial perturbation that can be imperceptible to the human eye (Bisogni et al., 2021). It could be imagined, for instance, that a replacement of Melania Trump exists indeed and that the associated image samples were manipulated in a way that the replacement will be recognized as Melania Trump.

REFERENCES

- (2019). Mobilesqueezenet: face reidentification retail 0095. t.ly/MqxX.
- Antipov, G., Berrani, S.-A., and Dugelay, J.-L. (2016). Minimalistic cnn-based ensemble model for gender prediction from face images. *Pattern recognition letters*.
- Bisogni, C., Cascone, L., Dugelay, J.-L., and Pero, C. (2021). Adversarial attacks through architectures and spectra in face recognition. *Pattern Recognition Letters*.
- Deng, J., Guo, J., Ververas, E., Kotsia, I., and Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). 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.
- Dolhansky, B., Howes, R., Pflaum, B., Baram, N., and Ferrer, C. C. (2020). The deepfake detection challenge (dfdc) dataset. *arXiv preprint arXiv:2006.07397*.
- Gao, X., Li, S. Z., Liu, R., and Zhang, P. (2007). Standardization of face image sample quality. In *International Conference on Biometrics*, pages 242–251. Springer.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.
- Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. (2017). Sphreface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220.
- Mahfoudi, G., Tajini, B., Retraint, F., Morain-Nicolier, F., Dugelay, J. L., and Marc, P. (2019). Defacto: Image and face manipulation dataset. In *27th European Signal Processing Conference (EUSIPCO)*.
- Mallat, K., Damer, N., Boutros, F., and Dugelay, J.-L. (2019). Robust face authentication based on dynamic quality-weighted comparison of visible and thermal-to-visible images to visible enrollments. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–8.
- Mallat, K., Galdi, C., and Dugelay, J.-L. (2017). Evaluation of the facial makeup impact on femininity appearance based on automatic prediction. COMET 2017, 2nd Cosmetic Measurement & Testing Symposium, Cergy-Pontoise, France.
- Méndez-Vázquez, H., Chang, L., Rizo-Rodríguez, D., and Morales-González, A. (2012). Evaluación de la calidad de las imágenes de rostros utilizadas para la iden-

- tificación de las personas. *Computación y Sistemas*, 16(2):147–165.
- Phillips, P. J. and O’Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, 32(1):74–85.
- Shirvaikar, M. V. (2004). An optimal measure for camera focus and exposure. In *Thirty-Sixth Southeastern Symposium on System Theory, 2004. Proceedings of the*, pages 472–475. IEEE.
- Wu, X., He, R., Sun, Z., and Tan, T. (2018). A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*.
- Yi, D., Lei, Z., Liao, S., and Li, S. Z. (2014). Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.
- Zohra, F. T., Gavrilov, A. D., Duran, O. Z., and Gavrilova, M. (2017). A linear regression model for estimating facial image quality. In *2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pages 130–138. IEEE.