A. Brömme, C. Busch, A. Dantcheva, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2019,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 1

Style Your Face Morph and Improve Your Face Morphing Attack Detector

Clemens Seibold[1], Anna Hilsmann[2], Peter Eisert[3]

Abstract: A morphed face image is a synthetically created image that looks so similar to the faces of two subjects that both can use it for verification against a biometric verification system. It can be easily created by aligning and blending face images of the two subjects. In this paper, we propose a style transfer based method that improves the quality of morphed face images. It counters the image degeneration during the creation of morphed face images caused by blending. We analyze different state of the art face morphing attack detection systems regarding their performance against our improved morphed face images and other methods that improve the image quality. All detection systems perform significantly worse, when first confronted with our improved morphed face images. Most of them can be enhanced by adding our quality improved morphs to the training data, which further improves the robustness against other means of quality improvement.

Keywords: biometric spoofing, face morphing detection, image quality improvement

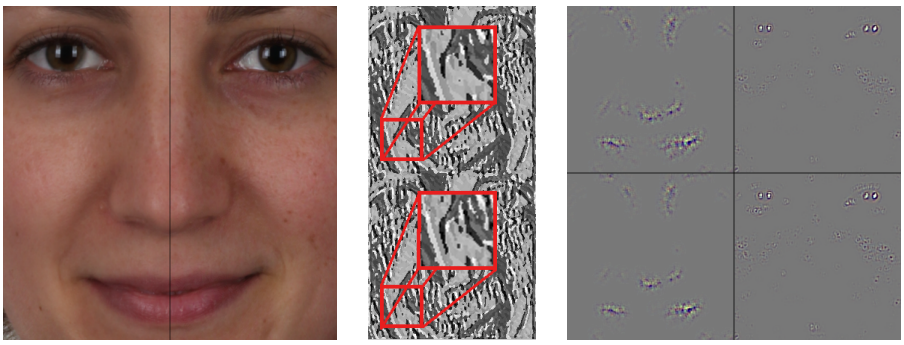
1 Introduction

Ferarra et al. [FFM14] showed that a synthetic face image that looks similar to two different subjects and contains biometric characteristics of both can easily be created with freely-available tools. Both subjects can use this image to verify their identity against a biometric verification system. This attack is called face morphing attack. It can be performed by a non-rigid alignment and blending of face images of two subjects. The simplicity of this attack and the fact that an applicant for an official document like a passport or other ID-cards can provide his/her own printed image in most European countries make face morphing attacks a real and dangerous threat to the integrity for biometric verification systems to border control. Therefore, the detection of such attacks is essential for the reliability of biometric verification systems.

In [Sc19] and [MW18], an overview on the state of the art of morphing attack detection (MAD) methods is provided. One important step in the process of creating a morphed face image (morph) is the blending of two aligned images. During this process high frequency details, like wrinkles, scars or pore structures are smoothed or get lost and the resulting image appears dull. Hence, these characteristics can provide evidence for image manipulations and MAD systems are likely to use these characteristics for detection.

In this paper, we present a style transfer based method that improves the image quality of morphs and counters image degenerating effects caused by the creation process of morphs.

[1] Fraunhofer HHI, Berlin, Germany, clemens.seibold@hhi.fraunhofer.de
[2] Fraunhofer HHI, Berlin, Germany, anna.hilsmann@hhi.fraunhofer.de
[3] Fraunhofer HHI and Humboldt University Berlin, Berlin, Germany, peter.eisert@hhi.fraunhofer.de



(a) (b) (c)

Fig. 1: (a) simple morph (left half) and improved morph (right half). (b) Binarized Statistical Image Features for a simple morph (top) and an improved morph (bottom). (c) visualizes two feature maps [ZF14] in a later layer with a relative strong difference between simple and improved morph of a DNN that was trained for MAD (top: simple morph, bottom: improved morph). The differences are hardly recognizable in the image as well as in the feature spaces.

We show that being prepared against improved morphs is essential and considering a improvement method can also help to be robust against others. Our improved morphs are more often not detected by MAD systems that were trained without knowledge of these improved morphs, although they only differ slightly compared to simple morphs in image and in feature space, see `Fig.1`. However, most systems can be improved by adding style transfer based improved morphs to the data during training. In addition to robustness against our style transfer based improved morphs, we show that training on these improved morphs can also increase the robustness against other kinds of image quality improvement methods like histogram equalization or sharpening filters.

The key contributions of this paper are:

- We propose a Deep Neural Network (DNN) based method that improves the quality of morphs by countering the image degeneration caused by additive blending.
- We evaluate the vulnerability of several state of the art MAD systems against our improvement method.
- We show that considering our improved morphs during training can improve the robustness of detection systems also against other image quality improvement methods.

In contrast to [Hi17], [SSV18] and [FFM19] who show that image degeneration operations like adding noise, double-scaling or a print-scan process can worsen a MAD system's performance significantly and need to be considered, we improve the image quality by adapting neural style transfer and study the effects of such improvement on MAD systems.

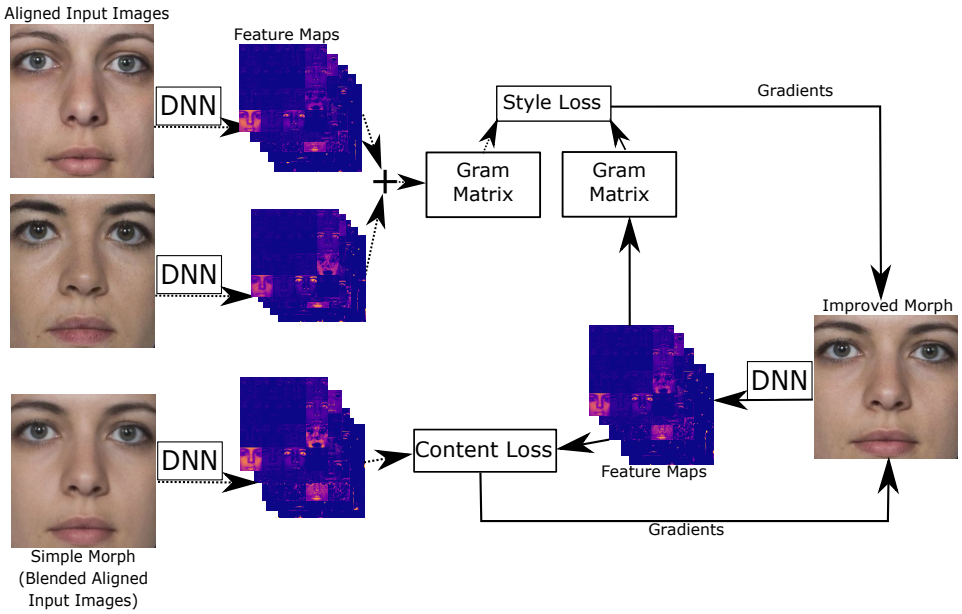


Fig. 2: Style transfer for the improvement of morphs.

In the next section, we provide theoretical background on style transfer and our adaptation to improve the quality of morphs. We give a short description of the analyzed MAD systems, our data sources and pre-processing in Section 3. Section 4 contains our research questions and our experimental results.

2 Style Transfer for Morph Enhancement

Gatys et al. [GEB15] proposed a method that changes the style of an image (content image) to the style of another image (style image) while preserving the original content, e.g. they transformed still images to have the characteristics of famous paintings. The concept of this approach is to transform both input images into a content and a style space and to find an image that is close to the content image in content space and close to the style image in style space. Both spaces are defined by feature maps of a neural network. In contrast to Gatys et al., we do not want to transfer an image to a target style, but want the style of our image being similar to the styles of two different face images. For this purpose, we define our target style as average style of both input face images that are used for the creation of a morph, see `Fig.2`.

Mathematical Model for Style Transfer The content of an image is defined as vectorized feature maps $F_j^l \in \mathbb{R}^{M_l}$ of selected layers in the neural networks and the style of an image as Gram matrices $G^l \in \mathbb{R}^{N_l \times N_l}$ of F^l, with N_l being the number of features maps

in layer l and M_l the number of pixels of a feature map in layer l. The element $G^l_{i,j}$ of the Gram matrix is the inner product of the vectorized feature map i and j in layer l: $G^l_{i,j} = F^l_i (F^l_j)^\mathsf{T}$. In order to transform the style of an image, we look for an image I that minimizes the loss function

$$L(I) = \sum_l v_l C(I)_l + \sum_l w_l S(I)_l,$$

where C_l and S_l are the content and style loss (the weighted square difference of the content/style of I and the target content/style) of layer l and v_l and w_l are weights for the content and style loss.

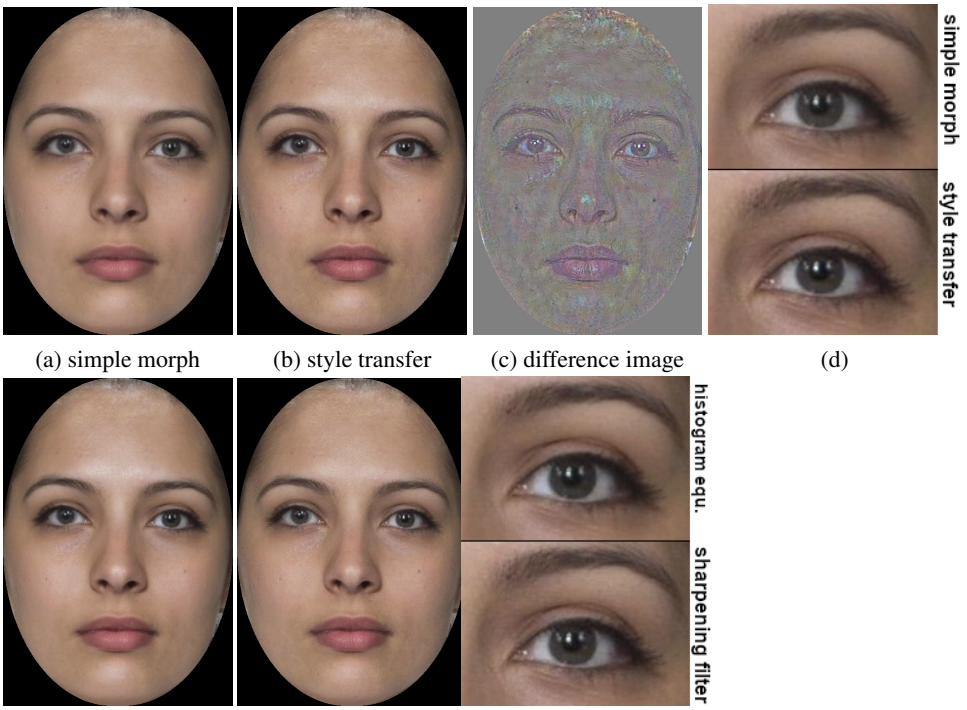
We use a VGG-19 network trained for object recognition to calculate the feature maps and minimize $L(I)$ using the gradient based Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm with box constraints (L-BFGS-B) [Zh97] to create a new morph with the style characteristics of the original image. Style transfer can handle images of any size, since only the convolutional layers of the neural network are used and no fixed size for the feature maps are needed.

Improving Blended Faces Using Style Transfer Before extracting the style from our genuine input images, we align and crop the inner part of the face, see `Fig.3a`. The style is extracted as described above for both images and finally averaged to get a target style, see `Fig.2`. The content is provided by cropping the same region of the aligned and blended face image. We initialize the L-BFGS-B algorithm with the simple blended image and use the layers 'conv1_2', 'conv2_2', 'conv3_4', 'conv4_4', 'conv5_4' for the content representation and 'conv1_1', 'conv2_1', 'conv3_1', 'conv4_1', 'conv5_1' for the style representation.

`Fig.3` shows examples for different means of image quality improvement of morphs. `Fig.3b` shows the style transfer based improved version of the simple morph shown in `Fig.3a`. The differences between the improved and simple version are visualized in `Fig.3c`. Fine structures on the skin and features like moles are enhanced and edges that are smoothed due to the blending, e.g. in and around the eyes, are reinforced to look sharp again, see `Fig.3d`. `Fig.3e` and `Fig.3f` show the results of other means of post-processing that aim to recover the sharpness of the original images. `Fig.3e` shows the result of sharpening the image using unsharp masking and `Fig.3e` the image after transforming the intensity of the simple morph so that its histogram approximately matches the histogram of one of the input images used for this morph.

3 Experimental Setup

Morphed Face Image Generation and Improvement For the generation of our morphs, we use the all-automatic face morphing pipeline in [Se17]. The factor for the geometry warping and the morphing factor for the blending of the images was set to 0.5 for the generation of the morphed images. We apply the improvement of the morphs on the *simple blended* face images before the seamless region cloning is applied (see [Se17]).



(a) simple morph (b) style transfer (c) difference image (d)

(e) histogram equalization (f) sharpening filter (g)

Fig. 3: Examples of different means for the image quality improvement of morphs: (c) shows difference image between (a) and (b) with enhanced contrast, (d) enlarged part of (a) and (b), (g) enlarged part of (e) and (f)

Face Image Collection and Pre-processing We collected face images from different publicly available datasets[4] and from our internal face databases. In total, we use about 2,000 face images in our experiments, after removing images with bad quality or violating requirements for passports, e.g. without frontal view. We split these genuine face images into a training dataset including 70% of all images, test dataset (20%) and validation dataset (10%) to avoid overfitting of our detectors based on neural networks. When selecting pairs for the generation of morphs, we ensure that both subjects have the same gender, are from the same database and are used equally frequent.

We process all images to have a standardized region and size for the detectors as in [RRB16, Ra17, Se18]. We rotate each image such that the eyes are on a horizontal line and crop the smallest bounding box that includes mouth and eyebrows. Finally, this region is scaled to 224x224 pixels.

[4] BU-4DFE, Chicago Face Database, FERET, LondonFace Database, PUT, scFace, Utrecht, CyberExtruder Ultimate Face Matching Data Set

Face Morphing Detectors In our study, we analyze five different MAD systems. In the following, a short overview on these detectors is provided.

The image degeneration based MAD system presented in [Kr17] relies on the number of edge describing features that are detected in an image and the change of the amount of these features after compression. The classification is performed using a pruned C4.5 decision tree.

One detector presented in [RRB16] is based on Local Binary Patterns (LBP). The histogram of the 59 uniform local binary pattern is calculated to extract features and a support-vector machine is used for classification.

A second detector proposed in [RRB16] is based on Binarized Statistical Image Features (BSIF) with a filter size of 11×11 and a bit length of 12. The histogram of the resulting image with 12-bit depth is calculated and a support-vector machine is employed to obtain a classifier.

Two MAD systems based on deep neural networks [Se18] use the VGG19-architecture and start the training with on object classification pretrained DNNs. One network is directly trained on genuine images vs. morphed faces (DNN naive), while the other network (DNN MC) is first pretrained on partial morphs.

4 Results

Experimental Goals We define three goals to study the quality of our style transfer based morph improvement approach and its effects on MAD systems.
(G1.1): We analyze whether the MAD systems can abstract from simple morphs and also detect the improved morphs. We train the five described detection systems on genuine images and simple (not improved) morphs and test them on genuine images, simple morphs and different kinds of improved morphs.
(G1.2): We study if the MAD systems can be adapted to robustly detect our style transfer based improved morphs by replacing half of the morphs during the training by these improved morphs.
(G2): We analyse the biometric quality of the improved morphs in terms of realistic morph acceptance rate (*rMAR*) [Hi17] and morph acceptance rate (*MAR*) using a biometric verification system[5]. We set the threshold for this system such that we have a FAR of 0.1% (according to the vendor specification) as recommended by FRONTEX for automated border control [FR15] and a even more strict rate of 0.01%.

Evaluation Metrics We consider face morphing attack detection as kind of presentation attack detection (PAD) and use the PAD metrics *bona fide presentation classification error rate* (BPCER) and *attack presentation classification error rate* (APCER), which are defined in ISO/IEC 30107-3 [In17]. `Tab.1a` and `Tab.2a` show the performance of the

[5] Verilook 10.0/MegaMatcher 10.0 Faces Identification Thechnology Algorithm Demo

studied MAD systems in BPCER and APCER, separated by the kind of improvement that is applied to the morphs. In addition to the not post-processed morphs (simple) and our presented improvement method (improved), we use a sharpening filter (sharp), which uses the unsharp masking technique, histogram equalization (HEQU), and histgram equalization after our syle transfer based method (imp.+HEQU) as post-processing steps to improve the quality of our morphs. `Tab.1b` and `Tab.2b` show the BPCER at different fixed APCER for the MAD systems that allow an adjustment of the error rates by changing the decision threshold.

		APCER(%)				
Detector	BPCER(%)	simple	improved	sharp	HEQU	imp. + HEQU
Features [Kr17]	32.6	17.3	54.6	43.6	49.7	74.7
LBP [RRB16]	25.4	21.1	60.3	58.5	35.1	65.7
BSIF [RRB16]	13.3	17.3	54.9	39.4	24.7	63.1
DNN naive [Se18]	1.5	1.0	30.7	3.1	32.5	72.6
DNN complex MC [Se18]	1.5	0.5	27.1	2.6	29.1	62.9

a) BPCER and APCER at default threshold of the MAD systems

		BPCER(%) at fixed APCER				
Detector	APCER(%)	simple	improved	sharp	HEQU	imp. + HEQU
LBP [RRB16]	10.0	40.5	82.3	69.2	55.4	83.9
	5.0	54.4	90.0	82.8	67.7	90.3
	1.0	79.5	95.6	95.9	91.3	97.2
BSIF [RRB16]	10.0	24.1	72.6	46.9	33.9	75.4
	5.0	35.9	78.5	67.2	46.7	85.1
	1.0	78.5	88.7	89.2	74.4	93.3
DNN naive [Se18]	10.0	0.5	15.4	0.5	29.7	74.9
	5.0	0.5	26.9	1.3	45.4	83.9
	1.0	1.5	46.7	7.7	81.0	95.4
DNN complex MC [Se18]	10.0	0.3	7.2	0.3	26.9	61.8
	5.0	0.3	16.4	0.8	51.3	84.4
	1.0	1.0	38.5	2.3	93.8	98.5

b) BPCER at different fixed APCER

Tab. 1: G1.1 Performance of different MAD systems trained on simple morphs and genuine images

All detectors that are trained on genuine images and simple morphs only, performed worse in detecting style transfer based improved morphs and even worse on style transfer based improved and histogram equalized morphs. The APCER increases up to more than 62% for all detectors. The Detection Error Tradeoff (DET) curves in `Fig.4` show that this is not only a matter of threshold of the classifier, but for any given APCER the BPCER is always much worse for style transfer and histogram equalized improved morphs (dashed red line) than for simple morphs (dashed green line). The MAD systems based on DNNs show the worst absolute and relative loss of performance, while all other methods also perform extremely poor on style transfer based and histogram equalized improved face morphs.

Including the style transfer based improved morphs in our training data, increases the detection rate of all kinds of improved morphs for all detectors. The detection rate of the style transfer based improved and histogram equalized morphs increases also for nearly all detectors. The detectors adapted differently to this new kind of morph, while the detection systems that are most vulnerable to improvement methods (the DNNs) can best adapt. The DET curves for the detection of simple morphs (solid green line) of the MAD systems that were trained with also on style transfer based improved morphs are slightly worse for all but the BSIF, but for other means of attacks (solid red for style transfer based and histogram equalized morphs and solid black for all mentioned attacks) far better.

`Tab.3` shows that our improvement method slightly worsens the biometric quality of the morphed faces, but we still have very high realistic morphing acceptance rates (rMAR)[Hi17].

		APCER(%)				
Detector	BPCER(%)	simple	improved	sharp	HEQU	imp. + HEQU
Features [Kr17]	33.8	17.3	43.6	30.7	50.0	72.2
LBP [RRB16]	32.6	25.8	38.4	50.5	33.8	43.0
BSIF [RRB16]	17.4	10.6	31.7	34.8	19.6	38.7
DNN naive [Se18]	1.5	2.8	7.2	4.4	15.7	29.9
DNN complex MC [Se18]	1.8	1.8	3.4	2.1	9.3	6.4

a) BPCER and APCER at default threshold of the MAD systems

		BPCER(%) at fixed APCER				
Detector	APCER(%)	simple	improved	sharp	HEQU	imp. + HEQU
LBP [RRB16]	10.0	49.2	62.8	74.4	59.0	70.5
	5.0	56.2	75.4	81.5	77.7	81.0
	1.0	84.9	87.4	89.2	89.2	89.2
BSIF [RRB16]	10.0	19.2	44.9	47.2	28.7	55.6
	5.0	32.1	63.3	64.1	49.7	70.8
	1.0	67.4	82.6	79.0	70.0	86.2
DNN naive [Se18]	10.0	0.5	1.0	1.0	6.9	11.3
	5.0	1.0	4.4	1.3	11.8	26.4
	1.0	7.7	11.0	10.3	50.0	60.8
DNN complex MC [Se18]	10.0	0.0	0.0	0.0	1.3	0.8
	5.0	0.8	0.8	0.8	4.4	3.3
	1.0	1.8	3.6	3.8	21.0	28.5

b) BPCER at different fixed APCER

Tab. 2: G1.2 Performance of different MAD systems trained on improved morphs, simple morphs and genuine images

5 Summary and Discussion

We introduced a method that improves the quality of morphs to be a step ahead of the attacker and analyze a broader range of attacks. The performance of different MAD systems drop significantly when they are first confronted with our style transfer based improved

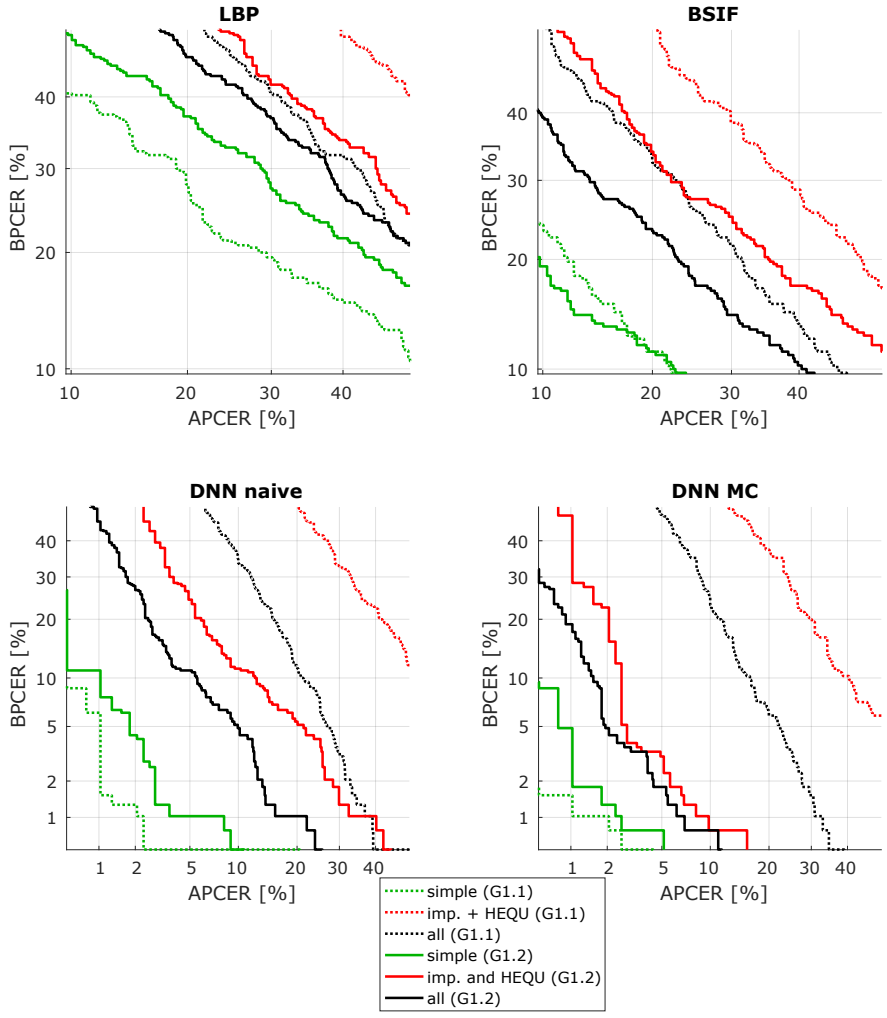


Fig. 4: DET curve of the different detectors. Since the Feature-based detector uses a classification tree, no DET curve was calculated.

Morph type	rMAR1000	rMAR10000	MAR1000	MAR10000
Simple Morphs	96.0%	90.5%	98.0%	95.3%
Improved Morphs	93.2%	87.0%	96.6%	93.4%
Sharp	95.5%	90.6%	97.8%	95.3%
HEQU	95.5%	90.5%	97.7%	95.2%
Imp.+HEQU	92.9%	86.5%	96.4%	93.1%

Tab. 3: G2 Biometric evaluation of improved morphs in comparison with simple morphs

morphs. After including our improved morphs in the training data, most of the MAD systems get significantly better in detecting them. They are also able to generalize and the detection rate for other means of post-processing that improve the image quality also increases. We achieve the best detection rates for all means of quality improved morphs by the MAD systems based on deep neural networks.

We studied the effects of image quality improvement on MAD systems that are based on handcrafted and learned features and showed that they are sensitive to subtle changes of the image. More sophisticated methods, which are based on physical models like reflection analysis [SHE18] or on biometric comparisons [FFM18], might be more robust against our improvements, since the reflections on the face and the biometric differences are only changed slightly.

Our method for style transfer based improvement of morphs is easy to implement using a deep learning framework and does not need special or expensive resources. Hence, it is a realistic scenario that an attacker would try to improve morphs using style transfer and thus it should be considered in the evaluation of MAD systems.

Acknowledgment

The work in this paper has been funded in part by the German Federal Ministry of Education and Research (BMBF) through the Research Program ANANAS under Contract No. FKZ: 16KIS0511.

References

[FFM14] Ferrara, M.; Franco, A.; Maltoni, D.: The magic passport. In: IEEE International Joint Conference on Biometrics. 2014.

[FFM18] Ferrara, M.; Franco, A.; Maltoni, D.: Face Demorphing. IEEE Trans. Information Forensics and Security, 2018.

[FFM19] Ferrara, M.; Franco, A.; Maltoni, D.: Face morphing detection in the presence of printing/scanning and heterogeneous image sources. ArXiv e-prints, January 2019.

[FR15] FRONTEX: Best Practice Technical Guidelines for Automated Border Control (ABC) Systems. 2015.

[GEB15] Gatys, L. A.; Ecker, A. S.; Bethge, M.: A Neural Algorithm of Artistic Style. CoRR, abs/1508.06576, 2015.

[Hi17] Hildebrandt, M.; Neubert, T.; Makrushin, A.; Dittmann, J.: Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps. In: IWBF 2017. April 2017.

[In17] International Organization for Standardization: , ISO/IEC 30107-3:2017 Information technology – Biometric presentation attack detection – Part 3: Testing and reporting, 2017.

[Kr17] Kraetzer, C.; Makrushin, A.; Neubert, T.; Hildebrandt, M.; Dittmann, J.: Modeling Attacks on Photo-ID Documents and Applying Media Forensics for the Detection of Facial Morphing. In: IHMMSec '17. 2017.

[MW18] Makrushin, A.; Wolf, A.: An Overview of Recent Advances in Assessing and Mitigating the Face Morphing Attack. In: 26th European Signal Processing Conference, EUSIPCO 2018, Roma, Italy, September 3-7, 2018. 2018.

[Ra17] Raghavendra, R.; Raja, K. B.; Venkatesh, S.; Busch, C.: Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images. In: CVPRW. 2017.

[RRB16] Raghavendra, R.; Raja, K. B.; Busch, C.: Detecting morphed face images. In: BTAS. 2016.

[Sc19] Scherhag, U.; Rathgeb, C.; Merkle, J.; Breithaupt, R.; Busch, C.: Face Recognition Systems Under Morphing Attacks: A Survey. IEEE Access, 7, 2019.

[Se17] Seibold, C.; Samek, W.; Hilsmann, A.; Eisert, P.: Detection of Face Morphing Attacks by Deep Learning. In: IWDW 2017, Magdeburg, Germany. 2017.

[Se18] Seibold, C.; Samek, W.; Hilsmann, A.; Eisert, P.: Accurate and Robust Neural Networks for Security Related Applications Exampled by Face Morphing Attacks. ArXiv e-prints, June 2018.

[SHE18] Seibold, C.; Hilsmann, A.; Eisert, P.: Reflection Analysis for Face Morphing Attack Detection. In: 26th European Signal Processing Conference (EUSIPCO). 2018.

[SSV18] Spreeuwers, L.; Schils, M.; Veldhuis, R.: Towards Robust Evaluation of Face Morphing Detection. In: 26th European Signal Processing Conference (EUSIPCO). 2018.

[ZF14] Zeiler, M. D.; Fergus, R.: Visualizing and Understanding Convolutional Networks. In: Computer Vision – ECCV 2014. 2014.

[Zh97] Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-scale Bound-constrained Optimization. ACM Trans. Math. Softw., 1997.