Realistic Dreams: Cascaded Enhancement of GAN-generated Images with an Example in Face Morphing Attacks

Naser Damer[12], Fadi Boutros[12], Alexandra Moseguí Saladié[1], Florian Kirchbuchner[12], Arjan Kuijper[12]
[1]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[2]Mathematical and Applied Visual Computing, TU Darmstadt, Darmstadt, Germany
Email: naser.damer@igd.fraunhofer.de

Abstract

The quality of images produced by generative adversarial networks (GAN) is commonly a trade-off between the model size, its training data needs, and the generation resolution. This trad-off is clear when applying GANs to issues like generating face morphing attacks, where the latent vector used by the generator is manipulated. In this paper, we propose an image enhancement solution designed to increase the quality and resolution of GAN-generated images. The solution is designed to require limited training data and be extendable to higher resolutions. We successfully apply our solution on GAN-based face morphing attacks. Beside the face recognition vulnerability and attack detectability analysis, we prove that the images enhanced by our solution are of higher visual and quantitative quality in comparison to unprocessed attacks and attack images enhanced by state-of-the-art super-resolution approaches.

1. Introduction

Generative Adversarial Networks (GAN) provided a series of interesting solutions for supervised image synthesis. Some of these solutions focused on improving the training stability, without improving the generated images quality. Other solutions, like the Cycle-Consistent Adversarial Networks, focused on producing images of higher quality and resolutions, but with very high network complexity and training needs. These conditions might not be possible for some applications, and even if available, these solutions are still limited to certain resolution and quality issues, generally appearing as generation artifacts on the images. One of the recent biometric publications based on GANs focused on foreseeing unknown face morphing attacks generated by GANs [6]. Although theoretically interesting, and such as many GAN-based solutions, these images suffered from unrealisticly low resolution and apparent artifacts. These limitations motivated our work to build an image enhancement solution specifically designed for GAN-generated images.

In this work, we propose an image enhancement approach to transfer low-resolution and artifact-effected GAN-generated images into a more realistic form, in terms of quality and resolution. We build our solution on the basis of multi-cascaded refinement networks to have low training data requirements and be extendable to higher resolutions. To demonstrate the abilities of our solution, we applied it on images that were manipulated in the latent space of the GAN, GAN-based face morphing attack images. These attacks were recently foreseen, however, with limitations in resolution and quality. Applying our solution on these images proved to produce images that are visually and quantitatively more similar to the bona fide images in comparison to two recently published super-resolution approaches, as well as different types of unprocessed morphing attack images. The resulting enhanced morphing attacks were also tested for detectability and vulnerability of face recognition systems. The vulnerability evaluation proved that the image enhancement approach maintain the strength of the original attacks and thus do not manipulate the identity information.

2. Related work

Due to the latest achievement in deep learning techniques, the generative models benefited from the deep architecture and achieved very promise results. One of the most dominant generative models is the GAN [14]. Within face biometrics, different solutions used GAN architectures. For example, Riggan *et al*. [30] proposed a convolution neural network that aims to map edge-based features, locally and globally, from the polarimetric to the gray visible domain. Zhang *et al*. [37] recently proposed Synthesizing colored faces from thermal images using Conditional GAN. Given that GANs are data hungry, they had to use data augmentation to train the model, resulting in generated images with high level of visible artifacts.

Different models can also be used for face image generation. For examples, deep convolutional Generative Adversarial network (DCGAN)[27] and Boundary Equilibrium

978-1-7281-1522-1/19/$31.00 ©2019 IEEE

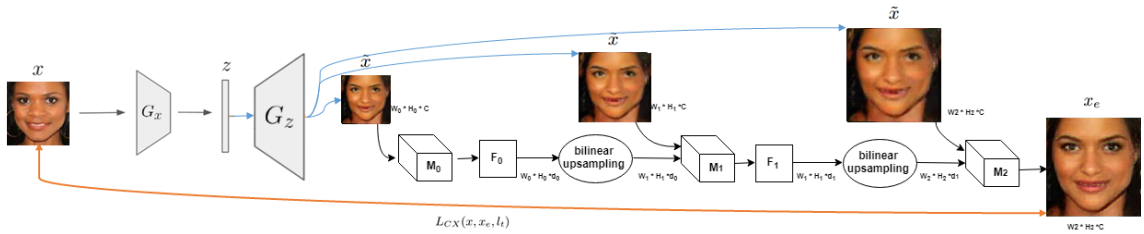

Figure 1: The proposed GAN-generated image enhancement solution during training. The training images x are decoded then re-generated by the GAN network (here, MorGAN network [6]). These generated images $\bar{x}$ are re-sized to the input sizes of the cascaded modules to produce an enhanced image x_e. The cascaded network is trained to minimize the CL between its output x_e and the original image x.

Generative Adversarial Networks (BEGAN) [3]. DCGAN introduced the Convolution Neural Network (CNN) into the discriminator and the generator. BEGAN introduced equilibrium factor that controls the model training by balancing the discriminator and generator. These GAN models significantly improved the training stability, but they did not improve the generated images quality. Where both models trained on (64×64 px) image size. However, some GAN approaches such as Cycle-Consistent Adversarial Networks (CycleGAN) [38] and Image-to-Image Translation with Conditional Adversarial Nets (Pix2Pix) [18] were able to achieve a higher resolutions images, but it ends with adding more complexity to the model. CycleGAN consists of four neural networks (two generators and tow discriminators). Training such a big model is computationally costly and requires large databases that are unavailable for some applications. Even if this is achieved, these models still have resolution and quality (due to generation artifacts) limitations that might fail below their application requirements. This work aims at solving this issue by proposing a specifically trained post-generation image enhancement solution. Image enhancement solutions have been a major direction of research in the computer vision community, with recent works achieving impressive results such as the MemNet [34] and the Information Distillation Network (IDN) [15]. However, unlike this work, these methods did not address the issue of enhancing automatically generated images, but rather super-resolutionizing natural images.

The possibility of creating a morphed face image attack out of two images of two subjects was introduced by Ferrara et al. [11]. Different solutions were later developed to detect face morphing attacks [28, 5]. However, most of these solutions were based on morphing attacks created by interpolating facial landmarks detected in the morphed face images, manually or automatically, i.e. facial landmark-based attacks (LMA). Face morphing attacks and their detection is a field were the use of GAN to create novel morphing attacks have been recently proposed to foresee future attacks and enhance detection of unknown attacks [6]. These attacks proved to be hard to detect when previously unknown and were the focus on further studies targeting variations in the morphing process [7], multi-detector fusion [8], and detection based on PRNU [9]. However, these attacks lacked a realistic resolution due to the limitation of the used GAN architecture. These morphing attack images will be the basis for testing our proposed image enhancement solution.

3. Methodology

This section presents the main contribution of this work, the cascaded image generated image enhancement. In the following, we motivate and present our enhancement methodology. We also discuss the GAN based morphing face image generation that will be used as the base for our experiments.

3.1. Cascaded image enhancement (CIE)

The solution we propose to enhance GAN-generated images was motivated by the requirements and challenges associated with the problem. The GAN generated images often look blurry at some parts and might contain other artifacts, they are also of lower resolution than expected for many applications. Therefore, our goal is to suppress the generation artifacts and increase the image resolution. Our solution also aims at having low training data size requirements and be easily scalable to higher resolutions by design. These requirements motivated our design of an image enhancement solution. To be able to generate high resolution images while having relatively small training data, we based our solution on the cascaded refinement network (CRN) [4]. The CRN considers multi-scale information and is based on training a relatively small number of parameters, leading to lower need of training data. The CRN is a convolutional neural network that consists of inter-connected refinement modules which was recently proposed for segmentation tasks. Each module consists of only three layers, input, intermediate, and output layer. The first module considers the lowest resolution space (4×4 in our case). This resolution is duplicated in the successor modules until the last module, matching the target image resolution. The second to last module expects two inputs, the output representation of the previous module and the input image at the specific resolution of the module. The target image (enhanced) resolution in our experiments was set to 128×128 pixels, to be comparable to the original resolution of the

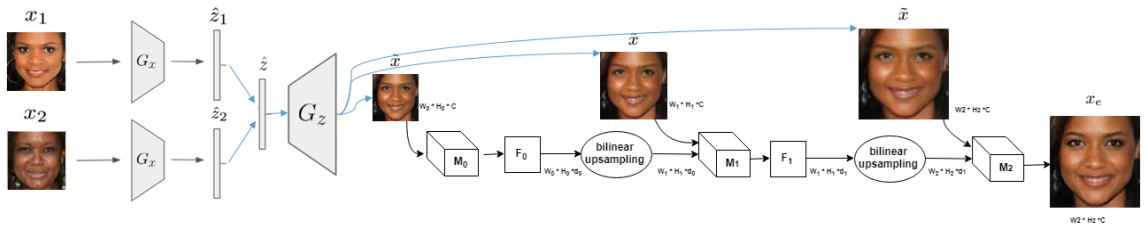

Figure 2: The morphing of two images in the latent vector space as described in [6], followed by the deployment of our trained CIE solution to transform the generated image $\bar{x}$ (64×64 pixels) into the enhanced image x_e (128×128 pixels).

face area of our basic pre-generation database.

The CRN training is controlled by a loss function. The loss function should be invariant to exact scale, alignment, and exact shape of the object in the image, a face in our case. This requires a loss function that neglects outliers on the pixel level (in comparison to pixel-level loss [18, 36]). The Gramm loss [13] can satisfy the other requirements, however, it does not constrain the content of the generated image as it describes it globally, which is problematic for image generation. To achieve the required properties, we chose the contextual loss function (CL) [24].

In our case, the source image is a GAN-generated image $\bar{x}$, the enhanced image $x_e = E(\bar{x})$ is the image generated by our CIE solution. During the training of our CIE model, the target image is the image x encoded by the GAN into $z = G_x(x)$, then regenerated by the generator into $\bar{x} = G_z(z)$, Where G_x and G_z are the GAN encoder and generator, respectively. The CL function is calculated between the target x and the enhanced images x_e. The target-enhanced loss maintains the properties of the target image in the enhanced image, e.g. target image style and content. The loss was calculated between image embeddings extracted by a pre-trained VGG-19 [33] network trained on the ImageNet database [10]. The total loss is calculated as given in [24] and formulated as:

$$L_{CX}(x, x_e, l_t) = -log(CX(\Phi^{l_t}(x_e), \Phi^{l_t}(x))), \qquad (1)$$

CX is the rotation and scale invariant contextual similarity [24]. Φ is a perceptual network, VGG19 in our work. $\Phi^{l_t}(y)$ is the embeddings vectors extracted from the image y at layer l_t of the perceptual network respectively. Here, l_t is the `conv3_2` and `conv4_2` layers, as motivated in [24]. The overall solution during training is illustrated in Figure 1, where the encoder and generator are pre-trained MorGAN network [6], all the details on the used GAN are given in [6]. In Figure 1, training image x are encoded and regenerated to $\bar{x}$ by the MorGAN network. The loss tries to optimize the CRN to produce an image x_e out of $\bar{x}$, enhanced in a way that would make it have similar properties to x. During deployment of our CIE, the input can be any image produced by the MorGAN network. In our implementation, x is 128×128 pixels (down-sampled for the MorGAN encoder to 64×64), $\bar{x}$ is 64×64 pixels, and x_e is 128×128 pixels. Our CRN contains six modules starting from 4×4 pixels input, up to 128×128 pixels input, Figure 1 contains only three module to ease the illustration.

3.2. Creating the attacks

MorGAN attacks are created using the approach and network trained in [6]. There, the face morphing is going to be performed in three steps given two face images x_1 and x_2. First, the two images are encoded into the latent space $\bar{z}_1$ and $\bar{z}_2$. Then, the resulting latent vectors are linearly interpolated with a factor $\beta = 0.5$ into a morphed latent vector $\bar{z}$. Finally, the interpolated latent vector is decoded into the image space to form the MorGAN morphed image $\bar{x}$. Figure 2 shows this MorGAN morphing process, followed by the deployment of our CIE approach to transform the generated image $\bar{x}$ (64×64 pixels) into the enhanced image x_e (128×128 pixels). The structure and training of the MorGAN encoder and generator are discussed in details in [6].

4. Experimental setup

4.1. Basic database

The solution and evaluation are built on the the CelebA [21]. CelebA is composed of 202,599 face images of 10,177 identities and 40 attribute binary vectors. The size of the images is 178×218 pixels, with the face region of interest is 120×120 on average. To cover the frontal image condition in the International Civil Aviation Organisation (ICAO) travel document requirements [16], all non-frontal images are filtered out by detecting the central coordinate of the eyes and the upper coordinate of the nose. The two distances between each of the two eyes and the nose landmarks are calculated, and if the ratio of the difference between these distances to any of them was more than 0.05, the image is neglected. Further filtering was performed based on the provided attributes, images labeled as blurry, with glasses, or with hat, was also filtered out.

As a starting point, 500 key images of 500 identities were manually chosen, split evenly between males and females. These 500 images were chosen to have neutral impression, good illumination quality, and no occlusion. Each of these images is matched twice, with two different images of different identities. The selection of these two was made so they are the most similar identities to the key image. The

similarity was measured by the Euclidean distance between the OpenFace representations [2]. Each of the 500 key images was paired twice. This resulted in 1000 morphing pairs and 1500 bona fide references. For each of the 1500 bona fide identities, a second bona fide image was chosen to be a bona fide probe.

Each of the 1000 morphing attacks pairs is used to create an attack. These attacks are created by the LMA and using the MorGAN approach [6]. LMA is performed by detecting 68 landmarks on the face as proposed in [19]. The mean face points for each image are calculated and each image is then warped to sit on these coordinates after performing the Delaunay Triangulation [20]. Only the facial area is morphed and stitched into one of the original morphed images [22]. MorGAN attacks are created for the same image pairs using the approach and network trained in [6]. The databases created here are referred to as MorGAN and LMA databases, in correspondence to the morphing approach. The MorGAN data was of the size 64×64 pixels, while the LMA was down-samples from the original size (120×120 pixels) to 64×64 for fair comparison. A version of the LMA data was not down-sampled and brought directly to a common size of 128×128 pixels, this version is referred to as full size LMA (FSLMA). The database is split into a disjoint (identity and image) and equal train and test sets. These sets are based on a random split of the initial 500 key reference images and therefore, each includes 750 bona fide references, 750 bona fide probes, and 500 attack images from each of both attack types.

Out of the filtered CelebA [21] database, 2500 images of 2500 identities were randomly selected to train the CIE. These identities did not overlap with the identities used to create the morphing attacks. This enhancement training data will be noted as ETD.

4.2. Image enhancement

Our goal is to enhance the quality of the GAN-generated images. The example we are using is the face morphing attacks created by the MorGAN network. Here, we create an enhanced version of the MorGAN database using our CIE solution and three baseline solutions. The CIE is trained always on the ETD images as described in Section 3.1. We trained CIE solution on a system with system i5-6500 CPU 16 GB RAM and NVIDIA GTX 1050 Ti 4 GB on-board. The training was run for 40 epochs, patch size of one, and 1e-4 learning rate. The training takes slightly under 4 hours. The resulting enhanced MorGAN is noted as EMorGAN and it has the same structure as the MorGAN data. The same probes and reference images associated with the MorGAN data are also used here, however, they are up-sampled from their original size (120×120 pixels) to 128×128 pixels to match the EMorGAN data. These are also used along with the FSLMA data. While the bona fide data associated with the MorGAN and LMA is down-sampled to 64×64 pixels to match the MorGAN and LMA attacks resolution.

As a baseline for our CIE approach, we utilize the Information Distillation Network (IDN) [15] and two versions of the MemNet approach [34], the MemNet-M10R0 and the MemNet-M6R6. The DIN approach is a deep but compact convolutional network that directly reconstruct a high resolution image from an original low resolution image [15]. The DIN model consists of a feature extraction block, stacked information distillation blocks, and reconstruction block. The MemNet is a super-resolution approach [34] based on a very deep persistent memory network that contains a recursive unit and a gate unit to mine persistent memory through an adaptive learning process. The recursive unit learns multi-level representations under different receptive fields. Two versions with different depths of the MemNet approach are used, the MemNet-M10R0 and the MemNet-M6R6. We used the pre-trained networks provided by the authors of both solutions [34, 15]. The three approaches are applied on the MorGAN database to double its resolution to 128×128 pixels. The resulting databases are structured identically to the MorGAN and EmorGAN databases, and are noted by the name of the enhancement approach, IDN, MemNet-M10R0 and the MemNet-M6R6.

In the next section, we present and discuss a visual qualitative samples of the bona fide images, the LMA attacks, and the MorGAN attacks with our proposed enhancement and baseline enhancement solutions. To have a quantitative comparison of the enhanced image quality, we report a number of image quality measures suggested in [35]. Namely, as defined in the relative references, the sharpness [12], blur [25], exposure [32], global contrast factor (GCF) [23], contrast [39, 12], and brightness [39]. In general, the goal is that the generated image would have similar image quality to the original images, here, the bona fide images.

4.3. Vulnerability and detectability

We investigate the **vulnerability** of face recognition algorithms to the produced MorGAN attacks before and after enhancement (EMorGAN), along with the LMA and FSLMA attacks. This vulnerability of two pre-trained face recognition systems are tested, the OpenFace as described in [2] and the VGG-face as described in [26]. Given an aligned and cropped face image, this pre-trained network produces a highly discriminant representation (feature vector) of 128 elements. On the other hand, VGG-Face [26] is based on the VGG-Very-Deep-16 CNN. The feature representation is extracted from the last max pooling layer which gives an output of 7x7x512. Face representations, whether from VGG and OpenFace, are compared by calculating the Euclidean distance between two representation vectors.

This vulnerability is discussed by showing the comparison score distributions of imposter, genuine, and morphed

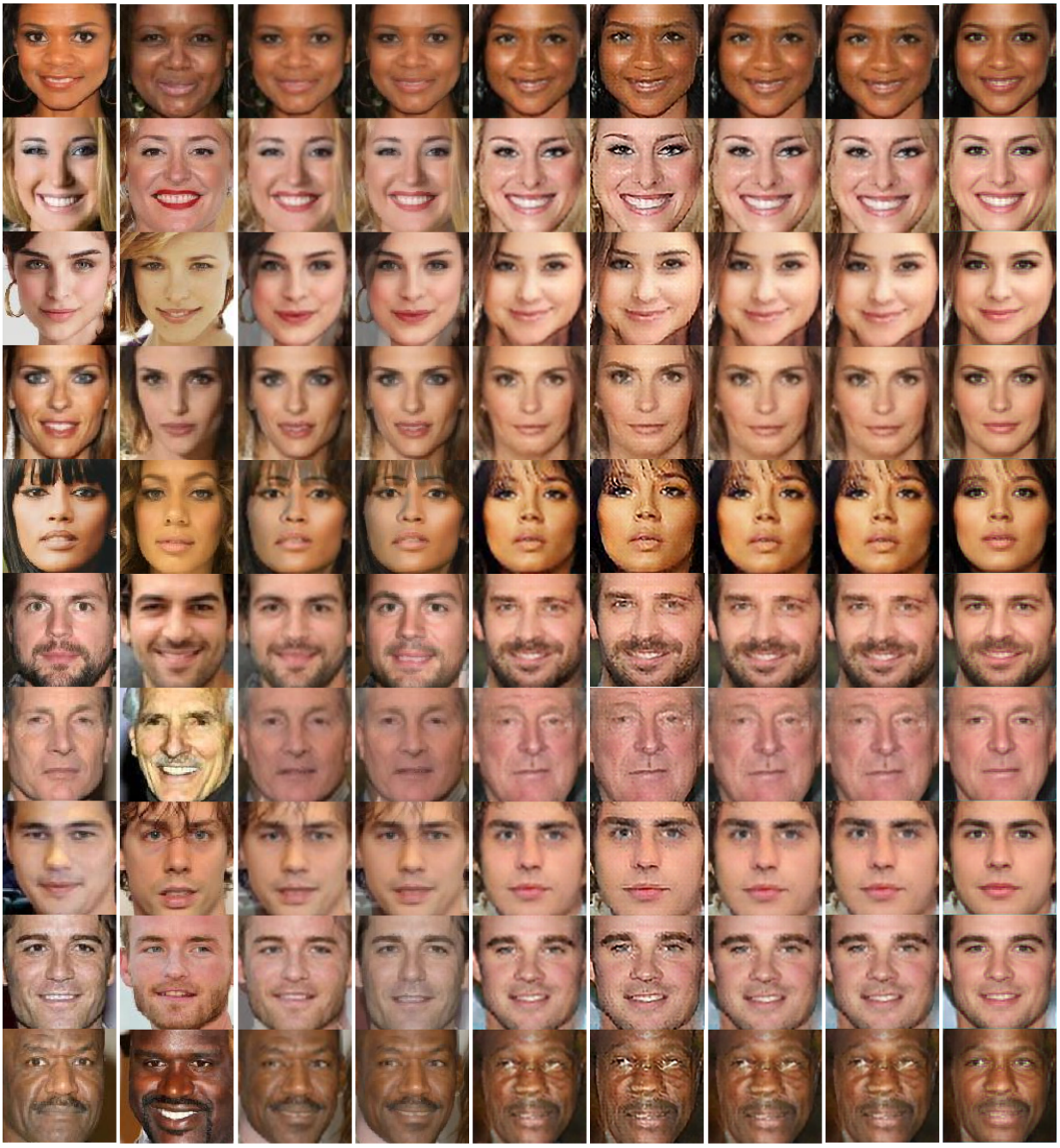

Figure 3: Sample images showing from the first to last column, left to right, identities images involved in creating the morphing attack (ID1 and ID2). These are followed by the morphed images LMA, FSLMA, MorGAN. Then the enhanced MorGAN by the IDN [15], MemNet-M6R6 [34], MemNet-M10R10 [34], and our EMorGAN. Five randomly selected male and female samples are illustrated. The reduction in the artifacts is clear in the EMorGAN images in comparison to the MorGAN and the baseline enhancement approaches.

face attack comparison to each of the probe original identities contained in the morph. To measure the attacks ability to simultaneously match both original identities, we plot the comparison scores between the morphing attacks and their two original identities images in the probe set. These comparisons scores are shown with respect to the threshold at the equal error rate (EER) operational point to get a relative measure of the attack success. The comparisons are made of images from reference and probe sets. The genuine and imposter comparisons are results of the 1500 bona fide ref-

	Sharpness	Blur	Exposure	GCF	Contrast	Brightness
Bona fide	0.248	0.318	0.165	7.414	0.461	0.627
LMA	0.193	0.197	0.143	6.347	0.471	**0.614**
FSLMA	0.197	0.231	0.144	6.391	0.472	**0.614**
MorGAN [6]	**0.226**	**0.260**	**0.164**	**7.185**	**0.458**	0.652
IDN [15]	0.303	0.491	**0.163**	7.121	0.453	0.668
MemNet-M10R10 [34]	0.220	0.221	0.160	6.996	0.454	0.664
MemNet-M6R6 [34]	0.220	0.228	0.160	6.996	0.454	0.664
EMorGAN (ours)	**0.235**	**0.268**	0.174	**7.746**	**0.464**	**0.645**

Table 1: Image quality measures different raw and enhanced images. Each of the measures is given as a mean value over the full database. The quality similarity between the EMorGAN and the Bona fide images is clear in comparison to the baseline super-resolution solutions. The two most similar (to bona fide) databases in each quality metric are in bold.

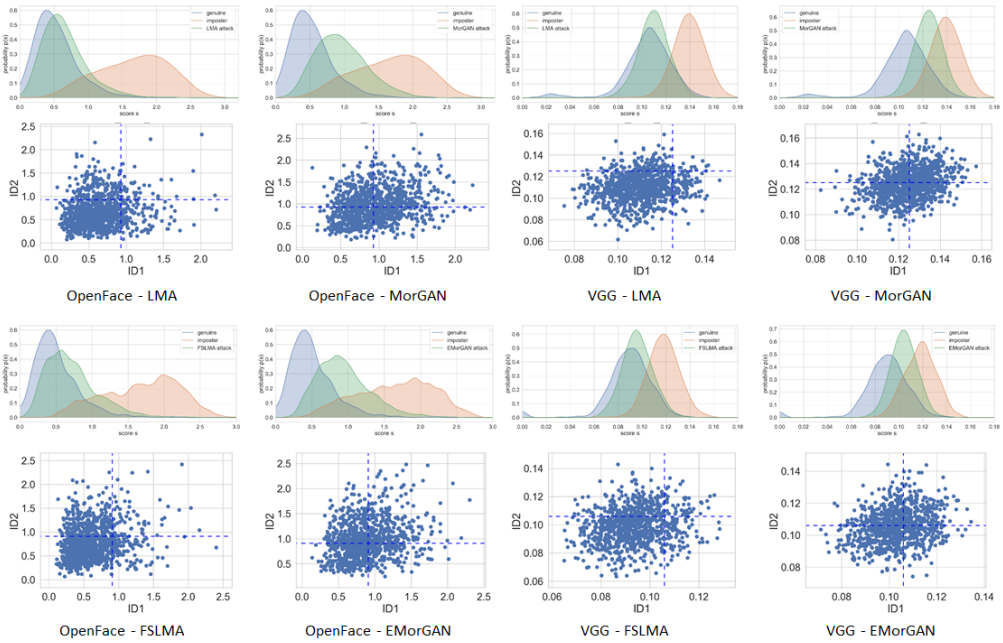

Figure 4: Vulnerability of two face recognition approaches (OpenFace & VGG) to attacks of the LMA, FSLMA, MorGAN, and EMorGAN databases. Top: the comparison score (dissimilarity) distributions of genuine (blue), imposter (red), and attack (green) comparisons. Bottom: morphing attacks comparison scores in comparison to the dotted line representing the threshold at EER.

erences cross-compared (NxN) with the 1500 probes. The morphing attacks score distribution is based on comparing the 1000 morphed images, each with their corresponding two identities in the probe set.

Our attack **detectability** evaluation aims at enabling a wider range of conceptual evaluation and more diverse coverage by considering image feature extraction methods of two different natures. One is the hand crafted classical image descriptors, the Local Binary Pattern Histogram (LBPH) [1]. The second is based on transferable deep-CNN features. Both types of features were previously utilized for the detection of face morphing attacks of similar nature to LMA [28][29]. The LBPH features are extracted from the cropped face image. A histogram is calculated for each block of an 8x8 grid of blocks in the face image. These histograms are concatenated to produce the final feature vector describing the image. Each LBP is extracted within a radius of one pixel and eight neighbor pixels. The transferable deep-CNN features are extracted using the well performing, and relatively small OpenFace NN4.SMALL2 model [2]. The extracted feature vector from an image, whether from CNN or LBPH, is classified by a support vector machine (SVM) classifier, to be originated from a morphed or a bona fide image. The approach based on direct single image feature vector classification is referred to as FV. The SVM utilizes a Linear kernel given the database size. The SVM hyperparameters are found using Bayesian optimization. The SVM classifier produces a decision score that represent the confidence degree of the input image being a morphed one rather than a bona fide one. The training and evaluation were performed on the identity-disjoint training and testing sets subsequently.

The performance of the morphed face detection is presented as a trade-off between two error rates, the Attack

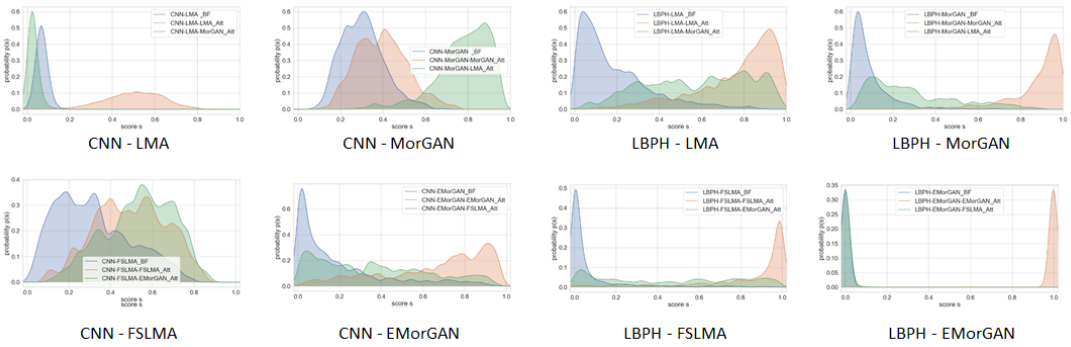

Figure 5: Detectability of the various morphing attack databases using CNN-based and LBPH features, for the two cases where the detector is trained on LMA, FSLMA, MorGAN, or EMorGAN attacks. Bona fide in blue, known attacks in red, and unknown attacks in green.

Presentation Classification Error Rate (APCER) and Bona Fide Presentation Classification Error Rate (BPCER) as defined by the ISO/IEC 30107-3 [17] and advised by recent works [31]. Here, the APCER is the proportion of morphed face presentations incorrectly classified as bona fide presentations. The BPCER is the proportion of bona fide presentations incorrectly classified as morphed face attacks. The detection decision thresholds producing fixed APCER rates on known attacks are calculated. These threshold represent possible decision thresholds chosen for system deployment and they depend on the detection performance of known attacks. BPCER values achieved at fixed APCER rates are reported to enable direct comparison between different solutions at a wide range of operation points, these BPCER rates only depend on the bona fide images, and thus are the same for known and unknown attacks. APCER rates of unknown attacks are reported on the thresholds previously assigned for the fixed known attacks APCER rates, to measure the detectability of unknown attacks. Lower values of BPCER and APCER indicates higher detection performance.

5. Results

To discuss **image enhancement** results, Figure 3 shows five male and five female random samples of morphing attacks and their pre-morphing image pairs. The figure shows the LMA, FSLMA, and MorGAN attacks. Then it shows, in the next columns, the MorGAN attacks after going through the baseline super-resolution solutions (DIN [15], MemNet-M6R6 [34], and MemNet-M10R10 [34]) and our proposed CIE solution (EMorGAN). The LMA and FSLMA shows clear blurring artifacts due to the image blending. The MorGAN images show typical artifacts of GAN-generated images, especially in the fine details of the face. The IDN solution generally increases the sharpness of the image, however, at some places it also increases the sharpness of the generation artifacts. Images produced by the MemNet have suppressed artifacts, however, they seem less sharp. One can clearly see that the EMorGAN images produced by our proposed CIE solution are more realistic in comparison to the baseline method and the different raw attacks. In a visual comparison, one can notice that the EMorGAN images are the most similar to the unprocessed bona fide images in the first two columns.

APCER	**1%**	**10%**	**20%**	**30%**
CNN-LMA-LMA	0%	0%	0%	0%
CNN-MorGAN-MorGAN	91.3%	68%	54.1%	39.5%
CNN-FSLMA-FSLMA	90.1%	54.1%	38.3%	31.2%
CNN-EMorGAN-EMorGAN	66.4%	28.2%	17.0%	10.8%
LBPH-LMA-LMA	36.1%	8%	3.2%	1.6%
LBPH-MorGAN-MorGAN	1.2%	0%	0%	0%
LBPH-FSLMA-FSLMA	35.6%	4.4%	2.0%	0.5%
LBPH-EMorGAN-EMorGAN	0%	0%	0%	0%

Table 2: BPCER rates achieved at fixed APCER rates of known attacks.

APCER	**1%**	**10%**	**20%**	**30%**
CNN-LMA-MorGAN	100%	100%	100%	100%
CNN-MorGAN-LMA	0%	0%	0.2%	0.2%
CNN-FSLMA-EMorGAN	0.0%	6.6%	14.9%	20.7%
CNN-EMorGAN-FSLMA	9.9%	40.8%	57.3%	73.2%
LBPH-LMA-MorGAN	7%	36.4%	51.4%	62.8%
LBPH-MorGAN-LMA	74.8%	100%	100%	100%
LBPH-FSLMA-EMorGAN	8.2%	51.9%	706%	84.1%
LBPH-EMorGAN-FSLMA	100%	100%	100%	100%

Table 3: APCER values of unknown attacks achieved at the detection decision thresholds that produced certain fixed APCER of known attacks.

To quantify the visual quality of the results, we present a comparison between the bona fide images, the different raw attack, and the enhanced MorGAN attacks using the three baselines and our CIE solution. A better enhanced image here should have similar quality to the bona fide ones. The comparisons are listed in Table 1. The table shows that in five out of six metrics, our proposed CIE approach produced images that are more similar to the bona fide images than the other enhancement baseline solutions. This proves the validity of our visual observations in Figure 3. The EMorGAN images were also constantly more similar

in these quality metrics to the bona fide images than the LMA and FSLMA attacks. This similarity makes these attacks, in principle, harder to detect for detection algorithms based on image quality observations. This leads to foreseeing more advanced attacks and thus developing solutions that are ready for novel attacks.

Besides evaluating the attack strength, comparing the face recognition **vulnerability** of the MorGAN and EMorGAN attacks aims at proving that our CIE solution maintains the identity of the input image. Figure 4 presents information about the vulnerability of the VGG-Face and OpenFace face recognition solutions to the LMA, MorGAN, FSLMA, and EMorGAN attacks. The distributions in the figure 4 show the comparison scores distributions produced by the genuine, imposter, and attack comparisons. As expected, the distributions of genuine and imposter comparisons are quite separated in both face recognition solutions, which indicates correct verification decision keeping in mind the uncontrolled captures in the database. When it comes to attacks, a successful attack imitates a genuine scenario, which means that it should produce comparison scores distribution similar to the genuine one. Knowing that, and under the different experiment scenarios, it is noticeable that the LMA and FSLMA attacks produces scores that fall almost completely in the genuine distribution range, which indicates strong attacks. On the other hand, MorGAN and EMorGAN attacks produce scores that are relatively less similar to the genuine one, however, still relatively separate from the imposter (i.e. still contains successful attacks). Knowing that both attacks can be successful, although with different degrees of success, we have to make sure that this works to attack both identities involved in the attack with the same degree. To demonstrate this property, we plot the comparison score between the attacks and the first involved identity vs. the one with the second identity in the scatter plots of Figure 4. The dotted lines in these plots represent the threshold value that achieves EER. This helps putting the plotted scores in perspective knowing that lower scores are stronger attacks. Ideally, if the morphing is performing as expected, most of the comparison scores will occur in the diagonal line. This is the case for all attacks. One can notice that the face recognition vulnerability to both MorGAN and EMorGAN is similar. This prove the identity preservation within our CIE solution without significantly manipulating the identity information.

Table 2 presents the **detectability** of LMA, FSLMA, MorGAN, and EMorGAN attacks, given that they are known, i.e. the detector is trained on the same type of attacks. Each experiment setting is noted by the feature used (CNN/LBPH), the data used for training, and the data used for testing (e.g. Feature-training-testing). CNN-based features performed very good in detecting LMA attacks but failed on the higher resolution FSLMA. It performed poorly in detecting MorGAN and EMorGAN attacks, however, more so on the MorGAN. On the contrary, LBPH features performed fairly in detecting MorGAN and EMorGANattacks and less so for LMA and FSLMA attacks. To measure the generalization ability of these detectors on unknown attacks, Table 3 presents the APCER values produced by unknown attacks, given that the system was configured (decision threshold) on fixed APCER values of the known attacks. For fair comparison, only the attacks of same resolution are paired. It is noticeable that the MorGAN attacks were very hard to detect by systems trained on the LMA attacks. Detecting the EMorGAN attacks with a system trained on FSLMA failed as well, however to a lower degree. This was the case for both CNN-based and LBPH features, although CNN-based features performs worse again with MorGAN attacks. The same scenario can be seen when an LBPH-based solution is trained on MorGAN or EMorGAN and faced by LMA or FSLMA attacks. On the contrary, detecting LMA/FSLMA attacks by a MorGAN/EMorGAN-trained CNN-based detector performed quite well. Figure 5 presents a deeper look into the discussed results by visualizing the detection score distribution for the different detection experiment settings.

The comparison of detectability and vulnerability between MorGAN and the enhanced EMorGAN attacks must be made while keeping in mind that the MorGAN attacks are of lower resolution and contains some artifacts, while the EMorGAN attacks are more realistic for manual inspection as illustrated earlier.

6. Conclusion

GAN-generated images are often faced by limitations in quality and resolution, especially if a purpose-specific training data is limited. This work presents an image enhancement approach that successfully increases the resolution and suppresses the generation artifacts in such images. This solution is based on the multi-scale cascaded refinement network, making it scalable to higher resolutions and simple to train. The presented method is applied on the recently proposed GAN-generated face morphing attacks. The evaluation proves the visual and quantitative quality enhancement in the images in comparison to different raw morphing attack images and images enhanced by state-of-the-art super-resolution solutions. This is accompanied with vulnerability and detectability analysis, proving that the enhancement approach did not eliminate the vulnerability of face recognition systems to the attacks after enhancement.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the National Research Center for Applied Cybersecurity CRISP.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Czech Republic, May, 2004. Proceedings, Part I*, volume 3021 of *LNCS*, pages 469–481. Springer, 2004.

[2] B. Amos, B. Ludwiczuk, and M. Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[3] D. Berthelot, T. Schumm, and L. Metz. Began: boundary equilibrium generative adversarial networks. *European Conference of Computer Vision Workshops (ECCVW)*, 2018.

[4] Q. Chen and V. Koltun. Photographic image synthesis with cascaded refinement networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 1520–1529. IEEE Computer Society, 2017.

[5] N. Damer, V. Boller, Y. Wainakh, F. Boutros, P. Terhörst, A. Braun, and A. Kuijper. Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts. In *Pattern Recognition - 40th German Conference, GCPR 2018, Stuttgart, Germany, October 10-12, 2018, Proceedings*, LNCS. Springer, 2018.

[6] N. Damer, A. M. Saladie, A. Braun, and A. Kuijper. MorGAN: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In *9th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2018, Los Angeles, California, USA, October 22-25, 2018*. IEEE, 2018.

[7] N. Damer, A. M. Saladie, S. Zienert, Y. Wainakh, P. Terhörst, F. Kirchbuchner, and A. Kuijper. To detect or not to detect: The right faces to morph. In *International Conference on Biometrics, ICB 2019, 4-7 June, 2019, Crete, Greece*. IEEE, 2019.

[8] N. Damer, S. Zienert, Y. Wainakh, A. M. Saladie, F. Kirchbuchner, and A. Kuijper. A multi-detector solution towards an accurate and generalized detection of face morphing attacks. In *22nd International Conference on Information Fusion, FUSION 2019, Ottawa, Canada , July 2-5, 2019*. IEEE, 2019.

[9] L. Debiasi, N. Damer, C. Rathgeb, U. Scherhag, A. M. Saladi, A. Uhl, C. Busch, and F. Kirchbuchner. On the detection of gan-based face morphs using established morph detectors. In *Image Analysis and Processing - ICIAP 2019 - 20th International Conference, Trento, 9-13 September, 2019, Proceedings*, Lecture Notes in Computer Science. Springer, 2019.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[11] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics, Clearwater, IJCB 2014, FL, USA, September 29 - October 2, 2014*, pages 1–7. IEEE, 2014.

[12] X. Gao, S. Z. Li, R. Liu, and P. Zhang. Standardization of face image sample quality. In *International Conference on Biometrics*, pages 242–251. Springer, 2007.

[13] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2414–2423. IEEE Computer Society, 2016.

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014.

[15] Z. Hui, X. Wang, and X. Gao. Fast and accurate single image super-resolution via information distillation network. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 723–731. IEEE Computer Society, 2018.

[16] International Civil Aviation Organisation (ICAO). ICAO Draft Technical Report: Portrait quality (reference facial images for MRTD). technical report, Version 0.9, 2017.

[17] International Organization for Standardization. ISO/IEC DIS 30107-3:2016: Information Technology Biometric presentation attack detection P. 3: Testing and reporting, 2017.

[18] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5967–5976. IEEE Computer Society, 2017.

[19] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874, 2014.

[20] D.-T. Lee and B. J. Schachter. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242, 1980.

[21] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[22] S. Mallick. Face morph using opencv c++ / python. https://www.learnopencv.com/face-morph-using-opencv-cpp-python/, 2016.

[23] K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgathofer. Global contrast factor-a new approach to image contrast. *Computational Aesthetics*, 2005:159–168, 2005.

[24] R. Mechrez, I. Talmi, and L. Zelnik-Manor. The contextual loss for image transformation with non-aligned data. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings*, volume 11218 of *Lecture Notes in Computer Science*, pages 800–815. Springer, 2018.

[25] N. D. Narvekar and L. J. Karam. A no-reference image blur metric based on the cumulative probability of blur detection (cpbd). *IEEE Transactions on Image Processing*, 20(9):2678–2683, 2011.

[26] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In X. Xie, M. W. Jones, and G. K. L. Tam, editors, *Proceedings of the British Machine Vision Conference 2015, BMVC 2015, Swansea, UK, September 7-10, 2015*, pages 41.1–41.12. BMVA Press, 2015.

[27] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[28] R. Ramachandra, K. B. Raja, and C. Busch. Detecting morphed face images. In *8th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2016, NY, USA, September, 2016*, pages 1–7. IEEE, 2016.

[29] R. Ramachandra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-cnn features for detecting digital and print-scanned morphed face images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, USA, July 21-26, 2017*, pages 1822–1830. IEEE Computer Society, 2017.

[30] B. S. Riggan, N. J. Short, and S. Hu. Thermal to visible synthesis of face images using multiple regions. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 30–38, March 2018.

[31] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. J. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Ramachandra, and C. Busch. Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *International Conference of the Biometrics Special Interest Group, BIOSIG 2017, Darmstadt, Germany, September 20-22, 2017*. GI / IEEE, 2017.

[32] M. V. Shirvaikar. An optimal measure for camera focus and exposure. In *System Theory, 2004. Proceedings of the Thirty-Sixth Southeastern Symposium on*, pages 472–475. IEEE, 2004.

[33] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[34] Y. Tai, J. Yang, X. Liu, and C. Xu. Memnet: A persistent memory network for image restoration. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 4549–4557. IEEE Computer Society, 2017.

[35] P. Wasnik, K. B. Raja, R. Ramachandra, and C. Busch. Assessing face image quality for smartphone based face recognition system. In *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*, pages 1–6. IEEE, 2017.

[36] L. Xu, J. S. J. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 1790–1798, 2014.

[37] T. Zhang, A. Wiliem, S. Yang, and B. Lovell. TV-GAN: Generative adversarial network based thermal to visible face recognition. In *2018 International Conference on Biometrics (ICB)*, pages 174–181, Feb 2018.

[38] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*, 2017.

[39] F. T. Zohra, A. D. Gavrilov, O. Z. Duran, and M. Gavrilova. A linear regression model for estimating facial image quality. In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2017 IEEE 16th International Conference on*, pages 130–138. IEEE, 2017.