Simulating Print/Scan Textures for Morphing Attack Detection

Juan Tapia and Christoph Busch
Hochschule Darmstadt
da/sec-Biometrics and Internet Security Research Group
Darmstadt, Germany
juan.tapia-farias, christoph.busch@h-da.de

Haoyu Zhang, Raghavendra Ramachandra, Kiran Raja
Norwegian University of Science and Technology (NTNU)
Norwegian Biometrics Laboratory (NBL)
Norway
haoyu.zhang, raghavendra.ramachandra, kiran.raja@ntnu.no

***Abstract*—Morphing Attack Detection is a relevant topic aiming to detect attempts of unauthorised individuals who want access to a "valid" identity. One of the main scenarios is printing morphed images and submitting the respective print in a passport application process. In order to improve the detection capabilities and spot such morphing attacks, it will be necessary to have a more realistic data set representing the passport application scenario with the diversity of devices and the resulting printed scanned or compressed images. Creating training data representing the diversity of attacks is a very demanding task because the training material is developed manually or semi-automatically. This paper proposes a transfer-style pixel-wise network for a general-purpose method to automatically create digital print/scan face images and use such images in the training of a Morphing Attack Detection (MAD) method. Our proposal can reach an Equal Error Rate (EER) of 5.13% with Random Forest and 3.17% using MobileNetV2 on the FRGCv2 database between manual print/scan and synthetics print/scan with 600 dpi. This method opens a new insight into developing attack datasets easily and time efficiently.**

I. Introduction

Synthetic face images have been utilised in many fields, as realistic images can be created with Generative Adversarial Networks (GAN). Most state-of-the-art approaches are based on convolutional neural networks, transfer learning and GAN to transfer the domain properties from one image to another domain. These methods need pairing images of the same objects to perform a pixel-wise transfer style. On the other hand, unpaired images can be used to transfer the style of one object image to another image unrelated to or from a different object. In this context, the pairs pixel-wise transfer method can generate synthetic images applied to morphing faces to create and transfer the style from printed/scanned images. Large numbers of such images will support the training of Morphing Attack Detection (MAD) systems using these methods, which is a relevant approach to improve training database diversity, which was previously consisting of only digital domain images.

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

It is a relevant topic to detect attempts of unauthorised individuals who want to use a "valid" identity document issued for another individual. In recent years several such cases have been reported, most of them related to illegal border crossings [1]. Morphing can be understood as a technique to seamlessly combine two or more look-alike facial images from a subject and an accomplice who, for example, could apply for a valid passport by exploiting the accomplice's identity. A morphing attack takes place in the enrolment process stage. The threat of morphing attacks is known for border crossing or identification control scenarios. It can be broadly divided into two types: (1) Single Image Morphing Attack Detection (S-MAD) techniques and Differential Morphing Attack Detection (D-MAD) methods [2], [3], [4].

Many countries issue electronic Machine-Readable Travel documents (eMRTD) passports in real-life scenarios based on the applicant's printed face photo. Some countries offer online portals for passport renewal, where citizens can upload their own face photo [5]. However, in most countries, the passport applicant supplies a printed image to the government authority issuing the identity document. It is subsequently scanned and embedded in the identity document, both on the data page and in the chip of the eMRTD. Accepting a printed face image excludes supervision of the face capture process by design; the applicant supplies a facial image, and its provenance cannot be conclusively verified. Then, to detect a morphed image, which was printed and scanned, it will be necessary to train S-MAD and D-MAD systems on a large number of such samples.

In order to improve the previous limitations of a low number of morphed images, we developed an image-to-image pairing method based on the Pix2Pix algorithm [6] to generate simulated print/scan images in order to complement the existing database that contains mainly digital versions of MA with several morphing tools. It is essential to highlight that manually creating print/scan images is very time demanding.

We describe the scenario as follows: First, a set of bona fide and morphed images are selected. Second, a printed version of digital images is created using high-quality glossy paper. Afterwards, printed images are captured with a desktop scanner. Later, these images are manually checked for quality, occlusion or other artefacts. It is essential to highlight that this process must be repeated image by image for each new

scanning device, which complies with the technical regulation defining the passport application process.

In summary, the contributions of this work are as follows.

- A pixel-wise transfer style method was proposed to create synthetic digital print/scan 600 dpi versions from digital images.
- Two backbones, UNet256 and ResNet50, were trained from scratch and evaluated to get high-quality face images based on Frechet Inception Distance scores.
- A Leave-One-Out protocol was applied to the FRGCv2 database based on Random Forest and MobileNetV2 with four morphing tools to show the utility of our proposed method in terms of high MAD accuracy.
- The image results from the pixel-wise process reach a very similar average EER to the manual process, according to our results.

The rest of the article is organised as follows: Section II summarises the related work on MAD. The database description is explained in Section III. The metrics are explained in Section V. The experiment and results are presented in Section VI. We conclude the article in Section VII.

II. Related Work

Most of the approaches for image-to-image translation that are recently reported in the literature use transfer techniques based on Deep Learning (DL), where Generative Adversarial Networks(GAN) are applied to the input image in order to translate the content to the target domain [7].

Zhu et al. [8] proposed an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. The goal is to learn a mapping $G : X \rightarrow Y$ such that the distribution of images from $G(X)$ is indistinguishable from the distribution Y using an adversarial loss. Qualitative results are presented on several tasks where paired training data does not exist, including collection style transfer, season transfer, and photo enhancement.

Gatys et al. [9] proposed a new parametric texture model to tackle this problem. Instead of describing textures based on a model for the early visual system, they use a convolutional neural network – a functional model for the entire ventral stream – as the foundation for this texture model. The feature information is extracted by 16 convolutional and five pooling layers.

Johnson et al. [10] proposed utilising perceptual loss functions to train feed-forward networks for real-time texture transfer tasks. Li and Wand [11] combined the Markov Random Fields model with deep neural networks, which was later extended to semantic style transfer.

Karras et al. [12] developed the StyleGAN2 network. This GAN is an extension of the progressive growing GAN that is an approach for training generator models capable of synthesising huge high-quality images via the incremental expansion of discriminator and generator models from small to large images during the training process.

Regarding MAD, Raghavendra et al. [13] explored the transfer learning approach using the Deep Convolutional Neural Networks (D-CNN) to detect both digital and print-scanned versions of morphed face images. This work explored a feature-level fusion of the first fully connected layer from pre-trained VGG19, and AlexNet networks.

Debiasi et al. [14] and Scherhag et al. [15] proposed to exploit the image noise patterns by Photo Response Non-Uniformity (PRNU) analysis, where the unique PRNU-pattern is extracted and analysed.

Mitkovski et al. [7] proposed a method based on a conditional generative adversarial network to generate print and scan images. The goodness of simulation is evaluated with respect to image quality, biometric sample quality and performance, as well as human assessment.

Ferrara et al. [16] proposed an approach based on Deep Neural Networks for morphing attack detection. In particular, the generation of simulated printed-scanned images and other data augmentation strategies; and pre-training on large face recognition datasets. The author used the Progressive Morphing Database (PMDB) [17] for network training. This dataset contains 6.000 morphed images automatically generated starting from 280 subjects selected from a different dataset.

Damer et al. [18] proposed a pixel-wise morphing attack detection (PW-MAD) approach where they train a network to classify each pixel of the image rather than only having one label for the whole image. They also evaluated unknown re-digitising attacks. Additionally, they created a new face morphing attack dataset with digital and re-digitised samples, namely the LMA-DRD dataset. However, this dataset presents limited printed and scanned images with only 276 bona fide attacks.

III. Databases

We employ the FRGCv2 database [19] for our experiments in this work. The selection of this database was motivated because we developed the morphed images in two sets: No post-processing and the print/scan at 600 dpi version, which means printing and scanning images one by one was done with a resolution of 600 dpi. The original images have a size of 360×480 pixels. All the images are aligned by the connecting line between the eye centres using the Dlib library.

The database was processed two times for different stages. First, to create synthetic images in a pixel-wise algorithm. We developed pairs of images side by side to represent the original digital images and the manually created printed/scanned version in 300 dpi. Both images must have the same subject, size, and image aligned. The side-by-size images have a final size of 720×480 pixels.

As a second process, the FRGCv2 database and the output of the synthetic version generated by the pixel-wise method were used to train a classifier based on Random Forest (RF) with 300 trees [4] and a MobileNetV2 network to detect bona fide and morph images. With this setup, it is possible to evaluate the influence of our pixel-wise transfer style method. Table I shows the number of images.

The following algorithms were used to create the morphed images:

- FaceFusion is a proprietary morphing algorithm developed for IOS app [1]. This algorithm creates high-quality morph images without visible artefacts.
- FaceMorpher is an open-source algorithm to create morph images [2]. This algorithm introduces some artefacts in the background.
- OpenCV-Morph, this algorithm is based on the OpenCV implementation [3]. The images contain visible artefacts in the background and some areas of the face.
- Face UBO-Morpher [20]. The University of Bologna developed this algorithm. The resulting images are of high quality without artefacts in the background.

TABLE I
MORPHING SOFTWARE TOOLS AND THE NUMBER OF IMAGES CREATED BY EACH TOOL WITH IMAGES FROM FRGCV2.

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

IV. METHOD

Our method is based on a transfer style network called pixel-to-pixel (Pix2Pix), as illustrated in Figure 1. It takes two face images as the input, bona fide digital and bona fide handcrafted print/scan versions from the same subject and delivers as outputs the original images translated to the new domain (print/scan) as is shown in Figure 2, this is depicted as the two images passing through two identical copies of the CNN block, but in practice, only one copy is stored in memory, and it takes the two images, producing the two embeddings.

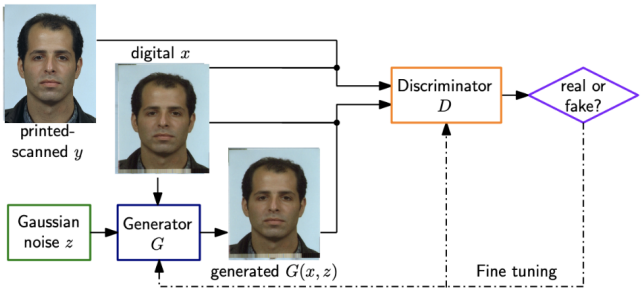

Fig. 1. Illustration of our pixel-wise print/scan simulation network [7].

V. EVALUATION METRICS

A. Sample Quality with Frechet Inception Distance

One of the difficulties with GAN algorithms, and in particular when applied to face images, is how to assess the quality of the resulting (synthesised) images. Currently, a suite of qualitative and quantitative metrics has been proposed to assess the performance of a GAN model based on the quality and diversity of the generated synthetic images [21], [22]. In this work, we use the Frechet Inception Distance (FID) [22]. The FID metrics allow us to compare different GAN models with the print/scan image results. The FID score measures the objective quality of the new print/scan synthetic images.

Frechet Inception Distance (FID) compares the similarity between two groups of images, A and B. First, to compute the FID, all images from set A and set B have to be processed by an Inception v3 network, pre-trained on ImageNet [23]. Then, the 2,048 feature vector of the Inception-v3-pool3-layer is stored for each image. Finally, the distributions of A and B in the feature space are compared using Equation 1, where μ_A and μ_B are the mean values of the distributions A and B, respectively, and Σ_A and Σ_B are the covariances of the two distributions.

$$FID = \|\mu_A - \mu_B\|^2 + Tr\left(\Sigma_A + \Sigma_B - 2(\Sigma_A \cdot \Sigma_B)^{1/2}\right) \quad (1)$$

B. Morphing Attack Detection Accuracy

The ISO/IEC 30107-3 standard[4] presents methodologies for the evaluation of the detection performance of MAD algorithms for biometric systems. The APCER metric measures the proportion of attack presentations—for each different PAI—incorrectly classified as bona fide presentation. This metric is calculated for each PAI, where the worst-case scenario is considered. Equation 2 details how to compute the APCER metric, in which the value of N_{PAIS} corresponds to the number of attack presentation images, where RES_i for the ith image is 1 if the algorithm classifies it as an attack presentation (morphed image), or 0 if it is classified as a bona fide presentation.

$$APCER_{PAIS} = 1 - \left(\frac{1}{N_{PAIS}}\right) \sum_{i=1}^{N_{PAIS}} RES_i \quad (2)$$

Additionally, the BPCER metric measures the proportion of bona fide presentations mistakenly classified as morphing attack presentations. The BPCER metric is formulated according to equation 3, where N_{BF} corresponds to the number of bona fide presentation images, and RES_i takes identical values of those of the APCER metric.

$$BPCER = \frac{\sum_{i=1}^{N_{BF}} RES_i}{N_{BF}} \quad (3)$$

These metrics effectively measure to what degree the algorithm confuses morphed images with bona fide images and vice versa. The APCER, and BPCER, metrics are dependent on a decision threshold. A Detection Error Trade-off (DET) curve is also reported for all the experiments. In the DET curve, the D-EER value represents the trade-off when the APCER is equal to the BPCER. Values in this curve are presented as percentages.

[1] www.wearemoment.com/FaceFusion/
[2] https://github.com/alyssaq/face_morpher
[3] www.learnopencv.com/face-morph-using-opencv-cpp-python
[4] https://www.iso.org/standard/79520.html

VI. Experiments and Results

A. Image Generation

Portrait pictures from FRGC were used to generate new print/scan images. For image generation based on pix2pix, three different convolutions neural networks were explored based on UNet256 and ResNet50. For ResNet, we are fine-tuning the network in order to improve the results based on layers: block6 and block9. The batch size was set up to 16 and 200 epochs.

Fig. 2. Example of the images generated by pix2pix.

Table II shows the FID scores for three proposed backgrounds. The first columns show different handcrafted methods and morphing tools used. The second column shows the FID value (a lower value is better) between the manual print/scan images and each morphing tool. It is essential to highlight that this value is the goal to reach for our pix2pix GANs for automatic print/scan version. Columns three up to five shows the FID score reached for UNet256, ResNet-6blocks, and ResNet-9blocks. The best results, which means lower differences with column 2, were obtained by UNet (column 3).

TABLE II
Summary result for pix2pix generation based on FID score.

Source		FID (↓)		
FRGC Handcrafted	Handcrafted Bona fide PS600 Baseline	PIX2PIX UNet256	PIX2PIX ResNet50 Block6	PIX2PIX ResNet50 Block9
Bona fide PS600	0	**16.985**	52.623	77.83
FaceMorpher PS600	63.71	**79.29**	100.589	117.75
FaceFusion PS600	114.975	**127.874**	186.16	166.075
OpenCV PS600	28.023	**39.824**	75.119	107.088
UBO PS600	18.559	**20.066**	58.114	95.847

B. Morphing Attack Detection

A Leave-One-Out (LOO) protocol was applied to train the morphing attack detection system, which means in the first round, UBO-Morpher was used to compute the morphing images used for the train, and testing was carried out with FaceMorpher, OpenCV-Morpher and FaceFusion. OpenCV Morpher was used for training and testing in the second round, with morphed images created with FaceFusion, FaceMorpher, UBO-Morpher, and so on. All datasets allow subject-disjoint results to be computed; no subject has an image in both the training and the testing subset.

TABLE III
Summary results and difference between handcrafted and synthetics images for RF.

Morphs Tool-Out UBO-Morpher	PS600 LBP-VER EER (%)	PS600+Syn LBP-VER EER (%)	PS600+Syn LBP-VER EER (%)	PS600 LBP-HOR EER (%)	PS600 LBP-HOR EER (%)	PS600+Syn LBP-HOR EER (%)
FaceFusion	13.49	24.86	5.64	12.91	24.42	5.31
FaceMorpher	13.09	17.96	5.91	12.39	17.20	6.01
OpenCV Morpher	11.51	20.38	4.46	10.87	20.15	4.09
Average	12.69	21.06	**5.33**	12.15	13.38	**5.13**

In order to train our S-MAD classifier, the FRGCv2 database was partitioned to have 70% training and 30% testing data. Different kinds of features were extracted from faces based on Uniform Local Binary Patterns (uLBP) for all experiments. The histogram of the Uniform Local Binary Patterns (uLBP) was used for texture. For the uLBP all radii values were explored from uLBP81 to uLBP88. The image's vertical (uLBP_VERT) and horizontal (uLBP_HOR) concatenation divided into eight patches was also explored.

After feature extraction, we fused that information at the feature level by concatenating the feature vectors from different sources into a single feature vector that becomes the input to the classifier. All features were extracted after applying our proposed transfer texture method.

The S-MAD system was trained based on a Random Forest classifier, which considered each feature extraction method described above as input. Three experiments were developed:

Experiment 1: We evaluated the LOO protocol applied to the handcrafted morphing set with 600 dpi for this experiment (PS600-based morphing set).

Experiment 2: This test evaluated the LOO protocol applied to print and scan images generated for our pixel-wise proposal to simulate 600 dpi (PS600-Synt-morphing set). Figure 3 show DET curve results for Experiments 1 and 2.

Figure 3, show the DET curves for UBO-Morpher in a LOO protocol for RF. In parenthesis is shown the EER for each morphing tool. The four DET curves show the MAD results between PS600 handcrafted and PS600 handcrafted plus synthetics pixel-wise generated. FaceMorpher was identified as the most challenging morphing tool.

Experiment 3: We evaluated the best results based on the LOO protocol but applying a MobileNetV2 convolutional neural network. This network was finetuned in order to obtain the best results considering the low number of images. The Learning rate applied was $1e-5$, 100 epochs, Adam optimiser, alpha rate of 1.4. and data-augmentation.

Table III and Table IV shows the EER for each feature applied to PS600 Handcrafted and PS600 Handcrafted plus synthetics on the FRGCv2 database and the fourth morphing tool for RF and MobileNetV2, respectively. Overall, we can obtain an EER between the real PS600 and Synthetics PS600 of 5.13% for RF and 3.17% for MobileNetV2.

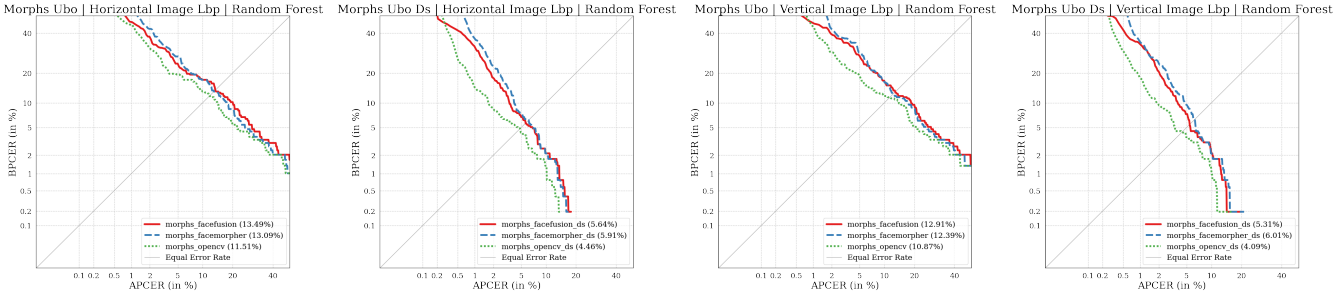

Fig. 3. DET Curves comparing handcrafted manual PS600 images versus PS600 handcrafted+synthetics (DS) images using a Random Forest Classifier and UBO-Morpher as LOO. Left to right: Horizontal LBP concatenation handcrafted PS600, PS600 DS. Vertical LBP concatenation handcrafted PS600, PS600 DS.

TABLE IV
SUMMARY RESULTS AND THE DIFFERENCE BETWEEN HANDCRAFTED AND SYNTHETICS IMAGES FOR MOBILENETV2.

Morphs Tool-Out UBO-Morpher	PS600 LBP-VER EER (%)	PS600+Syn LBP-VER EER (%)	PS600+Syn MobileNetV2 EER (%)
FaceFusion	13.49	5.31	1.33
FaceMorpher	13.09	6.01	5.39
OpenCV Morpher	11.51	4.09	2.81
Average	12.69	5.13	**3.17**

VII. CONCLUSIONS

This work shows that it is feasible to create print/scan from digital databases in order to improve the MAD and increase the number of images and scenarios available from training more robust classifiers and developing generalisation capabilities. The pixel-wise must be improved even more in order to reach lower FID scores and different image resolutions; thus, we can get more realistic images and reduce the EER.

REFERENCES

[1] M. Torkar, "Morphing cases in slovenia," in *Proc. Intl. Face Performance Conf. (IFPC)*, 2022.

[2] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," *IEEEAccess*, 2019.

[3] S. Venkatesh, R. Ramachandra, K. Raja, L. Spreeuwers, R. Veldhuis, and C. Busch, "Morphed face detection based on deep color residual noise," in *Ninth Intl. Conf. on Image Processing Theory, Tools and Applications (IPTA)*, 2019, pp. 1–6.

[4] J. E. Tapia and C. Busch, "Single morphing attack detection using feature selection and visualization based on mutual information," *IEEE Access*, vol. 9, pp. 167 628–167 641, 2021.

[5] R. Raghavendra, K. B. Raja, and C. Busch, "Detecting morphed face images," in *IEEE 8th Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–7.

[6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Conf. on Computer Vision and Pattern Recognition (CVPR), 2017 IEEE*, 2017.

[7] A. Mitkovski, J. Merkle, C. Rathgeb, B. Tams, K. Bernardo, N. E. Haryanto, and C. Busch, "Simulation of print-scan transformations for face images based on conditional adversarial networks," in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2020, pp. 1–5.

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2017, pp. 2242–2251.

[9] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015.

[10] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer Intl. Publishing, 2016, pp. 694–711.

[11] C. Li and M. Wand, "Combining markov random fields and convolutional neural networks for image synthesis," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2479–2486, 2016.

[12] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *CoRR*, vol. abs/2006.06676, 2020.

[13] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, "Transferable deep-cnn features for detecting digital and print-scanned morphed face images," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1822–1830.

[14] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch, "PRNU-based detection of morphed face images," in *6th Intl. Workshop on Biometrics and Forensics*, 2018, pp. 1–6.

[15] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl, "Detection of face morphing attacks based on PRNU analysis," *Trans. on Biometrics, Behavior, and Identity Science (TBIOM)*, 2019.

[16] M. Ferrara, A. Franco, and D. Maltoni, "Face morphing detection in the presence of printing/scanning and heterogeneous image sources," *IET Biometrics*, vol. 10, no. 3, pp. 290–303, 2021.

[17] ——, "Face demorphing," *IEEE Trans. on Information Forensics and Security*, vol. 13, no. 4, pp. 1008–1017, 2018.

[18] N. Damer, N. Spiller, M. Fang, F. Boutros, F. Kirchbuchner, and A. Kuijper, "Pw-mad: Pixel-wise supervision for generalized face morphing attack detection," in *Advances in Visual Computing*, G. Bebis, V. Athitsos, T. Yan, M. Lau, F. Li, C. Shi, X. Yuan, C. Mousas, and G. Bruder, Eds. Cham: Springer Intl. Publishing, 2021, pp. 291–304.

[19] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 947–954 vol. 1.

[20] K. Raja, M. Ferrara, A. Franco, L. Spreeuwers, I. Batskos *et al.*, "Morphing attack detection - database, evaluation platform and benchmarking," *IEEE Trans. on Information Forensics and Security*, November 2020.

[21] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *CoRR*, vol. abs/1606.03498, 2016.

[22] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conf. on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.