A. Brömme, C. Busch, A. Dantcheva, K. Raja, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2020 1

Simulation of Print-Scan Transformations for Face Images based on Conditional Adversarial Networks

Aleksandar Mitkovski[1], Johannes Merkle[2], Christian Rathgeb[2,3], Benjamin Tams[2], Kevin Bernardo[3], Nathania E. Haryanto[3], Christoph Busch[3]

Abstract: In many countries, printing and scanning of face images is frequently performed as part of the issuance process of electronic travel documents, e.g., ePassports. Image alterations induced by such print-scan transformations may negatively effect the performance of various biometric subsystems, in particular image manipulation detection. Consequently, according training data is needed in order to achieve robustness towards said transformations. However, manual printing and scanning is time-consuming and costly.

In this work, we propose a simulation of print-scan transformations for face images based on a Conditional Generative Adversarial Network (cGAN). To this end, subsets of two public face databases are manually printed and scanned using different printer-scanner combinations. A cGAN is then trained to perform an image-to-image translation which simulates the corresponding print-scan transformations. The goodness of simulation is evaluated with respect to image quality, biometric sample quality and performance, as well as human assessment.

Keywords: Biometrics, face, print-scan transformation, simulation, generative adversarial network.

1 Introduction

Face recognition technologies are frequently utilized for the verification of electronic travel documents, *e.g.*, in automated border crossings. In various countries, the issuance process of electronic travel documents requires applicants to provide face images in digital or analogue form (printed). This has already been identified as security gap since an applicant could manipulate his face image before submitting it to the issuance authority. Possible facial image alterations range from simple retouching [RDB19] to morphing [Sc19], where the latter type of manipulation causes a serious security risk, as shown by Ferrara *et al.* [FFM14]. It was found that human observers achieve only low accuracy in detecting such face image manipulations [RKB17, Rö19]. This necessitates the integration of automated procedures with the aim of reliably detecting face image manipulations.

Recently, different benchmarks [Rö19, Ng20, Ra20] have been conducted to compare the performance of manipulation detection schemes proposed in the scientific literature. The majority of state-of-the-art detection systems relies on machine learning techniques, *e.g.*, deep learning, which usually require a huge amount of training data. In order to ensure

[1] Hochschule Fulda, Fulda, Germany
[2] secunet Security Networks AG, Essen, Germany, {johannes.merkle,christian.rathgeb}@secunet.com
[3] da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Darmstadt, Germany, christoph.busch@h-da.de

high detection performance and generalizability, the used training data should resemble variations present in real-world scenarios. In particular, detection algorithms which are used to analyze face images stored in electronic travel documents are required to exhibit robustness towards print-scan transformations. However, a manual creation of a reasonable amount of printed and scanned face images is time-consuming and costly. Ferrara *et al.* [FFM19] demonstrated that the simulation of print-scan transformations during training can significantly improve the performance of face morphing attack detection on real printed and scanned face images. For this purpose, they employed the scheme of Lin and Chang *et al.* [LC99] which makes use of some mathematical models. Other similar schemes aiming at the simulation of print-scan transformations have been suggested in different application contexts, *e.g.*, [So05, Ei11].

In contrast to published works, we present a GAN-based approach for simulating print-scan transformations. A GAN is a machine learning model in which two neural networks compete with each other to become more accurate in their predictions and to be able to analyze, capture, and resemble the variations within a dataset. A subclass are cGANs which have been found to be suitabe for image-to-image translations [Go14, Is17, Zh17]. Specifically, the term *style transfer* is used to describe the operation of recomposing one image in the style of another (group of) image(s).

In this work, we train a cGAN to perform an image-to-image translation which resembles a print-scan transformation. We obtain face images from two public face database which are printed and scanned employing two printer-scanner combinations. These images are then used together with their original counterparts to train a cGAN to perform an image-to-image print-scan transformation. For each printer-scanner combination, 20 models trained with a different number of epochs are applied to simulate print-scan transformations. Finally, resulting test images are assessed in a comprehensive manner considering many relevant factors such as image quality, biometric sample quality, and human recognition. The obtained results confirm that the proposed cGAN-based approach is capable of realistically simulating print-scan transformations.

This paper is organized as follows: Sect. 2 summarizes the face databases used. The proposed architecture and training of the cGAN is described in detail in Sect. 3. The assessment of the simulated print-scan transformation is presented in Sect. 4. Finally, conclusions are drawn in Sect. 5.

2 Databases

From two public databases, *i.e.*, the FERET [Ph98] and FRGCv2 [Ph05], face images have been manually selected which meet the face image quality standards for electronic travel documents, as specified by the International Civil Aviation Organization (ICAO) [In15]. These images exhibit, among other requirements, full-frontal pose, uniform illumination, good focus, a neutral face expression with open eyes and no visible teeth, and neutral background. For these images, we adjusted the alignment of the face by suitable scaling, rotation and padding/cropping to ensure that the ICAO requirements with respect to the

eyes' positions are met. Overall, 529 and 984 face images have been selected from the FERET and FRGCv2 databases, respectively. Subsequently, all face images are printed and scanned using two printer-scanner combinations listed in Table 1. The resulting image sets are referred to as IS1 and IS2. That is, both image sets contain all 1,513 face images selected from the FERET and FRGCv2 databases, but have been printed and scanned using different devices. Example face images of the original database, IS1, and IS2 are shown in Fig. 1.

Tab. 1: Printer-scanner combinations used for database creation.

Image set	Printer	Scanner	Properties
IS1	Fujifilm Frontier 5700R Minlab	Epson DS-50000	300 dpi, 24-bit RGB, print on matte photo paper
IS2	Developed by professional photo studio	Canon Imagerunner Advance 4535i	600 dpi, 24-bit RGB, print on glossy paper

(a) original (b) IS1 (c) IS2

Fig. 1: Example images of used databases. Best viewed in electronic format (zoomed in).

The printing and scanning process was conducted in a semi-automatic manner. For this purpose, a software tool was implemented which enables the arrangement of images in the size of passport face images according to ICAO [In15] on A4 paper sheets, 20 images per page. Additionally, markers were included in the top left and the right bottom corners of each face image to facilitate a subsequent segmentation. Further, the filenames of images where included as QR-codes. The resulting sheets were then printed on photo paper and scanned, *cf.* Table 1. Finally, the face images were automatically extracted from the scanned sheets and the corresponding filenames were assigned. Fig. 2 shows examples of face images after printing and scanning.

3 Proposed approach

A GAN architecture [Go14] consists of a generator model G for outputting synthetic images according to a given distribution, and a discriminator model D that classifies images as real (from the dataset) or fake (generated). The discriminator model is updated directly,

Fig. 2: Example images of after printing and scanning with segmentation makers and filenames in form of QR-codes. Best viewed in electronic format (zoomed in).

whereas the generator model is updated via the discriminator model. As such, the two models are trained simultaneously in an adversarial process where the generator seeks to fool the discriminator and the discriminator seeks to better identify the generated images.

In a cGAN, G generates the images not just from internal noise z (as in a tradional GAN) but also based on an input image x; the conditional distribution of the output image $G(x,z)$ given the input x is supposed to resemble that of real image translations (x,y). The discriminator is provided both with a source image x and a target image, and must determine whether the target is a real image y or an output $G(x,z)$ of the generator. An example is the well-known Pix2Pix framework of Isola *et al.* [Is17]. The generator is trained via adversarial loss,

$$\mathscr{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x,G(x,z))], \tag{1}$$

which encourages the generator to generate plausible images in the target domain. The generator is also updated via L1 loss measured between the generated image and the expected output image,

$$\mathscr{L}_{L1}(G) = \mathbb{E}_{x,y,z}[||y - G(x,z)||_1]. \tag{2}$$

This additional loss helps the generator model to create translations nearer to the ground-truth, resulting in,

$$G^* = \arg\min_G \max_D \mathscr{L}_{cGAN}(G,D) + \lambda \mathscr{L}_{L1}(G). \tag{3}$$

The Pix2Pix cGAN has been demonstrated on a range of image-to-image translation tasks such as converting maps to satellite photographs, black and white photographs to color, and sketches of products to product photographs [Is17].

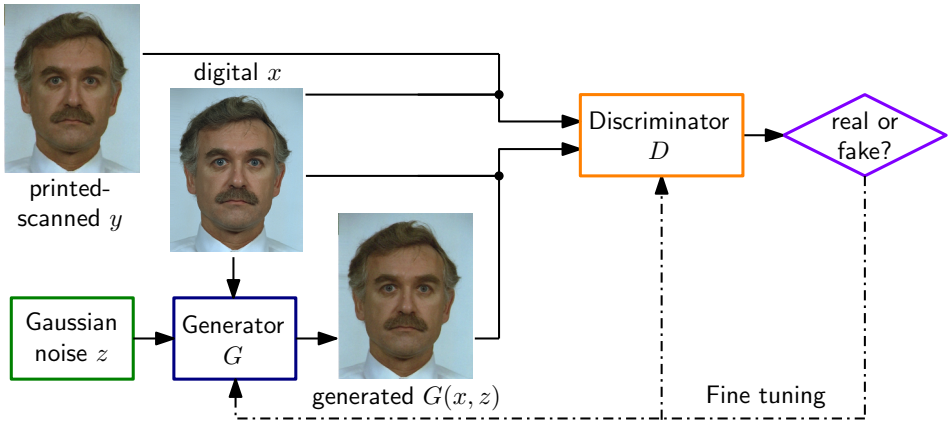


Fig. 3: Concept of the proposed cGAN-based print-scan transformation.

Our approach builds on the Pix2Pix framework [Is17] and is depicted in Fig. 3. Alternatively, a Cycle-Consistent Adversarial Network (CycleGAN) [Zh17] could be employed which allows for the training of an unpaired image-to-image translation. However, since a sufficient number of images is available in digital and printed and scanned form, the Pix2Pix framework turns out to be suitable.

As suggested in [Is17], an image classifier named PatchGAN is used instead of a traditional discriminator. While a traditional discriminator maps the complete image to a single scalar which expresses the probability whether the image is real or generated, the PatchGAN splits the image into small local patches. The L1 norm in the loss function already encourage the generator to correctly represent the coarse structures of the target image. Therefore, the discriminator must only assess if the generator's outputs resemble the fine structures in a realistic way. To this end, the discriminator only requires small patches of the image as input. By restricting the input of the discriminator to small images patches, the size of the discriminator, and number of its parameters can be greatly reduced. In our case, the coarse structures, for which correct representation is encouraged by the L1 norm, are the features of the subject depicted in the input image which are still visible in the print-scan transformed image, and the fine structures investigated by the discriminator are the artefacts induced by the print-scan transformation. For computing the loss function, the images are divided in $N \times N$ patches, and after passing the corresponding pairs of patches to the discriminator the resulting outputs are averaged to estimate whether the image is real or generated.

We used a PatchGAN with fixed kernel size of 4×4, a fixed stride of 2×2, and 70×70 patches as input. During training, horizontal mirroring was applied for data augmentation. This resulted in approximately 3,000 face images per image set. The training on each image set has been performed with 2,400 randomly chosen face image pairs. Training has been conducted separately for each image set with up to 200 epochs each and with a batch size of 1, as suggested in [Is17]. Each epoch uses the entire training set. One

data batch passes the neural network 480,000 times. For GANs, it is often difficult to find the time when training should be stopped. Therefore, we save and evaluate the model after every 10 epochs, resulting in 20 models for each image set. Subsequently, print-scan transformations have been performed on 100 randomly chosen face images of the remaining ones. Note that the number of test images was restricted by the database size as well as time constraints. Examples of simulated print-scan transformations are depicted in Fig. 4.

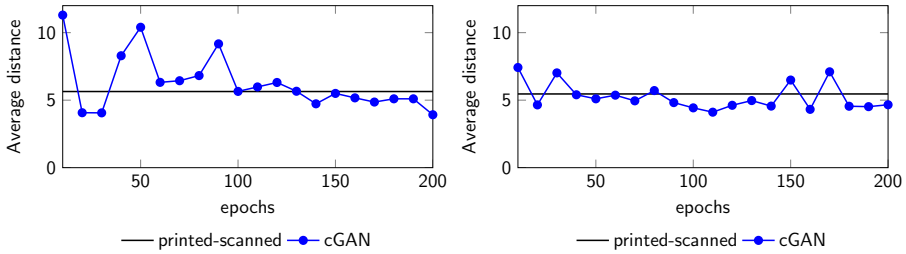
(a) digital (b) printed-scanned (c) generated

Fig. 4: Example images of IS1 (top row) and IS2 (bottom row). Models trained with 180 epochs were used to generate images in the rightmost column. Best viewed in electronic format (zoomed in).

4 Assessment

Firstly, the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [MMB12] is used to determine an appropriate number of epochs used during training. BRISQUE calculates a no-reference image quality score which is sensitive to various distortions, *e.g.* blur. Three scores, s_d, s_p, and s_g, are estimated for a digital face image and its printed-scanned as well as generated counterpart, respectively. For such triples of face images we then estimate the distances of scores of digital images to printed-scanned, *i.e.*, $|s_d - s_p|$, and to generated ones, *i.e.*, $|s_d - s_g|$. Eventually, the average distance μ is reported. Thereby, we

measure whether the generated images resemble the effects of a real print-scan transformation. Note that BRISQUE scores of s_d are generally smaller than those of s_p or s_g. Further, note that since the generator models a probabilistic function that describes the image modifications induced by the print-scan transformation, its output cannot perfectly match the target images. The results are plotted in Fig. 5. In partcular on IS1, higher variations can be observed for training with less 100 epochs. For the subsequent experiments (including human assessment) we only consider cGANs which have been trained with 180 or 190 epochs as these configurations yield good results on both image sets.



(a) IS1 (b) IS2

Fig. 5: Average distance of BRISQUE scores across training epochs on both image sets.

In the second experiment, the effects of printing and scanning on face recognition are analyzed, in particular sample quality assessment and performance. To this end, the FaceQNet algorithm [He19] and the ArcFace system [De19] are used for quality estimation and comparison, respectively. The impact on sample quality is computed using the scores of FaceQNet and ArcFace in the same manner as for BRISQUE scores before. In the case of recognition, unconstrained probe face images have been additionally chosen from the FERET and FRGCv2 face databases. Lastly, the Root-Mean-Square Error (RMSE) between the digital reference image and the printed-scanned as well as generated image is estimated for each triple of face images. To do so, image pairs are firstly aligned using the Scale-Invariant Feature Transform (SIFT)-based Fiji tool [Lo99, Sc12] and the RMSE is estimated for each color channel. As final score, the average RMSE over all color channels is calculated. All assessment algorithms process the cropped face regions since these facial image parts are most relevant. It is important to note that the considered assessment algorithms produce scores in different ranges. Obtained results including average distances μ and the standard deviations σ for the real printed-scanned images and the cGAN-based approach trained with different numbers of epochs (ep.) are summarized in Table 2.

Tab. 2: Obtained results for different quality assessment algorithms on both image sets.

Algorithm	**FaceQNet** ($\times 10^2$)				**ArcFace** ($\times 10^2$)				**BRISQUE**				**RMSE**			
Image set	**IS1**		**IS2**		**IS1**		**IS2**		**IS1**		**IS2**		**IS1**		**IS2**	
Score	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Printed-scanned	1.01	0.796	0.892	0.634	1.96	1.52	0.524	0.421	5.64	4.1	5.46	3.73	12.4	2.55	21.0	2.75
cGAN (180 ep.)	0.995	0.746	0.946	0.671	2.68	2.12	0.471	0.348	5.1	3.92	4.55	3.57	12.9	2.95	22.4	2.47
cGAN (190 ep.)	1.12	0.78	1.02	0.654	2.87	1.95	0.512	0.382	5.48	3.92	4.52	3.4	12.9	2.97	22.0	2.75

It can be seen that for the considered assessment algorithms, the generated images yield effects similar to those of the real printed-scanned images. With only a very few outliers,

e.g., effects on ArcFace scores in IS1, this is true for both used printer-scanner combinations. Note that the proposed approach also resemble image set specific variations which can be observed from ArcFace and RMSE scores.

In addition to the above evaluation, a human assessment has been conducted in a second experiment. For this purpose, 35 experts form secunet Security Networks AG were asked to rate images with one to three stars depending on how closely these resemble real printed and scanned face images. After an explanatory introduction to the experiment, they were presented with 20 triples of face images, *i.e.* a real printed-scanned image, a cGAN-based generated image, and an image to which a style transfer based on DeepArt[3] has been applied, see Fig. 6. The latter images were created by transforming the digital images providing their printed-scanned counterpart image as desired style image to the web-based DeepArt style transfer application. Triples of face images were presented to the participants in a randomized order. Obtained results in terms of average rating are shown in Fig. 7.

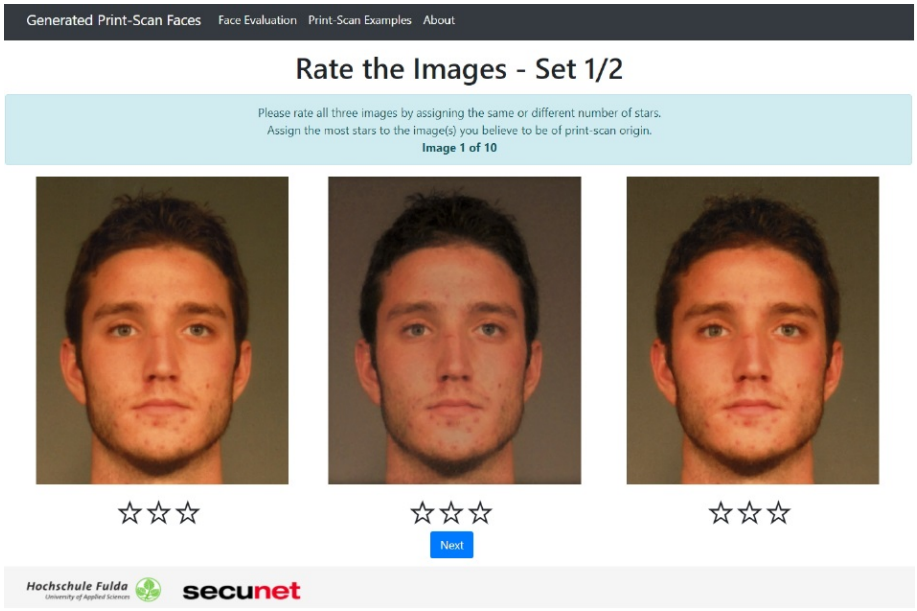


Fig. 6: Screenshot of the rating application in the human assessment experiment.

In can be observed that the proposed cGAN-based approach outperforms the DeepArt style transfer in terms of visual perception. The ratings obtained by the cGAN-based simulation of print-scan transformations are close to those given for real printed-scanned face images.

5 Conclusions

The aim of this work was to generate cGAN-based images that are virtually indistinguishable from real printed and scanned images. In a comprehensive assessment it has been

[3] DeepArt: `https://deepart.io/`

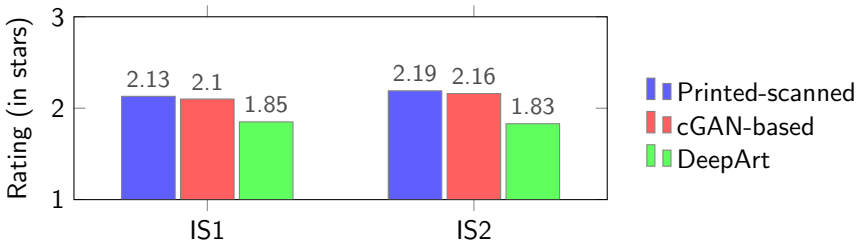


Fig. 7: Obtained results for the human assessment experiment.

shown that the presented approach is capable of simulating the effects of real printed and scanned face images for different printer-scanner combinations. That is, our cGAN-based simulation of print-scan transformations can be used to automatically generate training data as input for face image manipulation detection systems which is subject to future work.

Acknowledgements

This research work has been partly funded by the Federal Office of Information Security (BSI) through the FACETRUST Project, the German Federal Ministry of Education and Research and the Hessen StateMinistry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[De19] Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Conf. on Computer Vision and Pattern Recognition (CVPR). 2019.

[Ei11] Eid, A. H.; Ahmed, M. N.; Cooper, B. E.; Rippetoe, E. E.: Characterization of Electrophotographic Print Artifacts: Banding, Jitter, and Ghosting. IEEE Trans. on Image Processing, 2011.

[FFM14] Ferrara, M.; Franco, A.; Maltoni, D.: The magic passport. In: Int'l Joint Conf. on Biometrics (IJCB). 2014.

[FFM19] Ferrara, M.; Franco, A.; Maltoni, D.: Face morphing detection in the presence of printing/scanning and heterogeneous image sources. CoRR, abs/1901.08811, 2019.

[Go14] Goodfellow, I.; Mirza, J. Pouget-Abadie M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y.: Generative Adversarial Nets. In: Advances in Neural Information Processing Systems (NIPS). 2014.

[He19] Hernandez-Ortega, J.; Galbally, J.; Fiérrez, J.; Haraksim, R.; Beslay, L.: FaceQnet: Quality Assessment for Face Recognition based on Deep Learning. In: Int'l Conf. on Biometrics (ICB). 2019.

[In15] International Civil Aviation Organization: ICAO Doc 9303, Machine Readable Travel Documents – Part 9: Deployment of Biometric Identification and Electronic Storage of Data in MRTDs (7th edition). Technical report, ICAO, 2015.

[Is17] Isola, P.; Zhu, J.; Zhou, T.; Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks. In: Conf. on Computer Vision and Pattern Recognition (CVPR). 2017.

[LC99] Lin, C.; Chang, S.: Distortion Modeling and Invariant Extraction for Digital Image Print-and-Scan Process. In: Int'l Symposium on Multimedia Information Processing (ISMIP). 1999.

[Lo99] Lowe, David G.: Object Recognition from Local Scale-Invariant Features. In: Int'l Conf. on Computer Vision (ICCV). 1999.

[MMB12] Mittal, A.; Moorthy, A. K.; Bovik, A. C.: No-Reference Image Quality Assessment in the Spatial Domain. IEEE Trans. on Image Processing, 2012.

[Ng20] Ngan, M.; Grother, P.; Hanaoka, K.; Kuo, J.: Face Recognition Vendor Test (FRVT) Part 4: MORPH Performance of Automated Face Morph Detection. Technical Report NISTIR 8292, National Institute of Technology (NIST), 2020.

[Ph98] Phillips, P. J.; Wechsler, H.; Huang, J.; Rauss, P. J.: The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing, 1998.

[Ph05] Phillips, P. J.; Flynn, P. J.; Scruggs, T.; Bowyer, K. W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W.: Overview of the Face Recognition Grand Challenge. In: Conf. on Computer Vision and Pattern Recognition (CVPR). 2005.

[Rö19] Rössler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Niessner, M.: FaceForensics++: Learning to Detect Manipulated Facial Images. In: Int'l Conf. of Computer Vision (ICCV). 2019.

[Ra20] Raja, K.; Ferrara, M.; Franco, A. et al.: , Morphing Attack Detection – Database, Evaluation Platform and Benchmarking, 2020. arXiv 2006.06458.

[RDB19] Rathgeb, C.; Dantcheva, A.; Busch, C.: Impact and Detection of Facial Beautification in Face Recognition: An Overview. IEEE Access, 2019.

[RKB17] Robertson, D. J.; Kramer, R. S. S.; Burton, A. M.: Fraudulent ID using face morphs: Experiments on human and automatic recognition. PLOS ONE, 2017.

[Sc12] Schindelin, J.; Arganda-Carreras, I.; Frise, E. et al.: Fiji: an open-source platform for biological-image analysis. Nat Meth, 2012.

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

[So05] Solanki, K.; Madhow, U.; Manjunath, B. S.; Chandrasekaran, S.: Modeling the print-scan process for resilient data hiding. In: Security, Steganography, and Watermarking of Multimedia Contents VII. 2005.

[Zh17] Zhu, J.; Park, T.; Isola, P.; Efros, A. A.: Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: Int'l Conf. on Computer Vision (ICCV). 2017.