On Improving Face Generation for Privacy Preservation

Werner Bailer and Martin Winter
DIGITAL – Institute for Information and Communication Technologies
JOANNEUM RESEARCH
Graz, Austria
{firstname.lastname}@joanneum.at

***Abstract*—Replacing faces in image and video content with generated ones (e.g., using generative adversarial networks, GANs) has gained attention recently, as it enables resolving privacy issues in image and video data used for visualization purposes or training data in multimedia analysis and retrieval systems. Privacy issues should be addressed when visual content enters the system, as identifying and removing content later (which may be necessary due to the shifts in legislation and users' increased awareness) is a tedious and costly task. This paper proposes two improvements of face generation: First, we propose the use of portrait segmentation on the training data of the GAN, in order to generate images that are not only cropped to the face region, which may cause artifacts during insertion. Second, we add a face detection term to the loss function, in order to better guide the training process. The results show that these modifications enable creating uncropped face images achieving the same or better performance than for closely cropped images. We use the detectability of the generated faces as an evaluation metric, discuss the limitations of such a metric and propose enhancements for better comparability. We also demonstrate that the aim of anonymization is achieved by running face recognition on the modified images from the LFW data set.**

***Index Terms*—face generation, generative adversarial networks, evaluation, face detection, privacy**

I. Introduction

Privacy is an increasing concern in multimedia applications, both triggered by increasing awareness of users and legislation such as the General Data Protection Regulation (GDPR) [1] in the European Union. Naturally, images and videos showing identifiable persons commonly appear in training data, query examples, visualization of search results, etc. There are many cases where it is just an inevitable side-effect that persons appear in the content, even if identifiable persons in the content are not relevant in the respective applications. Example application domains include traffic and navigation, construction or tourism, where the objects of interest are depicted in public space, and (identifiable) persons may also be visible. For visualization purposes, to retrain machine learning tools (or migrate to future technology) and to enable traceability of the results of multimedia analytics systems, it is useful to store the visual content and not discard it after its use for training. Thus privacy issues should be addressed more fundamentally.

978-1-7281-4673-7/19/$31.00 ©2019 IEEE

The use of face swapping to replace faces (or even entire persons) with artificially generated ones (e.g., using generative adversarial networks, GANs) has thus become increasingly proposed recently [2], [3]. This is motivated by the fact that a method for privacy preservation should meet three goals: (i) prevent identification by inverting the method, (ii) avoid insertion of repeated patterns that could bias machine learning methods and (iii) provide visually pleasant results, in particular for content intended to be used for visualization purposes. While acceptable results are achieved with state of the art methods, two shortcomings can be identified. First, existing GAN-based face generation methods use closely cropped face regions as training data and thus produce also that kind of results. During face swapping, this often produces artifacts, in particular near the chin or hairline (see Figure 2a). Removing this spatial constraint of the GAN negatively impacts the quality of the resulting images, caused by the large variability of backgrounds (see Figure 4a). Second, while many of the resulting faces are of good quality, some are still deteriorated. The fact that this is not avoided means that the discriminator of the GAN does not sufficiently penalize their generation.

The contributions of this paper address these two issues (i.e., avoiding artifacts due to close cropping, improving detectability of faces) by performing head region segmentation as part of the generation workflow, and thus enabling the generation of uncropped faces, and by integrating a face detection loss to better guide the training of the GAN. In addition, we discuss the evaluation of GANs for face generation. Section II discusses related work, and Section III describes the proposed approach. We present evaluation results based on assessing the accuracy of face detection and the anonymization capabilities using face recognition and discuss issues with evaluation metrics in Section IV. Section V concludes the paper.

II. Related Work

There is a range of methods that can be applied to support privacy protection of visual content. A recent survey that provides a good overview can be found in [4]. The authors of [5] apply the concept of k-anonymity to face recognition. Using a training set, they average eigenface descriptors of the faces, so that discriminability of faces is reduced to groups of at least k individuals. One drawback of this approach is that it may require updates across many images, if new faces are added which are outliers wrt. the previous set of faces, and could thus break the k-anonymity. A similar approach of de-identification for content with a closed set of faces is described in [6], and performing the k-same test in real-time. In videos, this approach also ensures consistent de-identification of all appearances of one person. In [7], an approach for face swapping for privacy preservation is proposed. The tool contains a library of previously collected faces, and performs replacement using face detection and registration of facial landmarks. A more recent approach applying face swapping for privacy protection is described in [8], also using faces from a predefined database.

Several recent works employ GANs for face generation. In [9], an approach for face swapping under unconstrained conditions using a fully convolutional neural network for segmentation is proposed. A region-separative GAN (RSGAN) is used in [3] for face swapping based on latent face and hair representations. [10] propose an addition to GAN called conditional GAN (cGAN) and apply it to face generation. The approach allows conditioning the training data on certain distributions. The authors demonstrate that conditioning generation on the 36 face attributes in the LFW dataset [11] improves the quality of generated images. They also show that a subset of attributes guides the generated faces accordingly. The faces are closely cropped. cGANs are used in [12] and MTCNN [13] is used in the evaluation process to measure the fraction of correctly generated faces. A recent work [2] proposes to replace both the face and the appearance of the full body by generating clothing given a human segmentation in an image. The face generation is a separate process, using DCGAN [14]. Another face swapping approach using DCGAN for the generation process is [15]. Both work with closely cropped faces in the generation process. [16] propose an extension of GAN called Bayesian GAN, and evaluate the method among others for face generation, also cropping faces closely. [17] aim at generating faces that preserve attributes such as age or race, also using cropped images. The proposed privacy preserving GAN (PP-GAN) includes a face verificator, to ensure that the original and replaced faces are sufficiently different, and a regulator that enforces structural similarity of the original and replaced face. MTCNN [13] is used in the evaluation to determine the rate of generated faces.

For segmentation, [18] propose an automatic portrait segmentation method based on the FCN-8s framework [19]. The method leverages position and shape information by registering a canonical portrait image template with the target image using the results of facial landmark detection. They also propose a data set for this task. Head segmentation[1] has been also proposed based on the scene segmentation models PSPNet [20], U-Net [21] and TiramisuNet [22].

To the best of our knowledge, there are no face generation approaches that use segmentation in order to better handle the generation of larger face regions. While face detectors are sometimes used in the evaluation, there seems to be no approach integrating a detector to constrain the generation process.

[1]https://github.com/xubiuit/head-segmentation

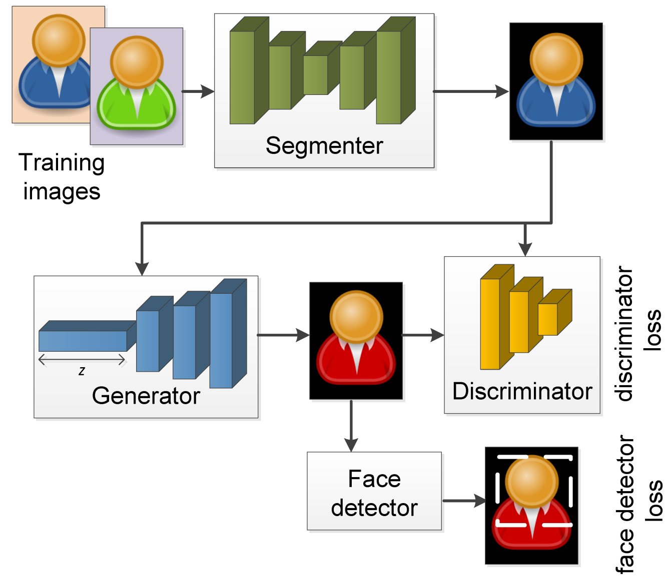

Fig. 1. The structure of the proposed approach: adding a segmenter to process the training data, and a face detector to generate an additional component of the loss function.

III. Proposed Approach

We use DCGAN [14] as the basis of our approach, to which we add support for segmentation and integrate face detection into the loss function. An overview of the approach is shown in Figure 1. We use a Tensorflow implementation of DCGAN[2], which makes more frequent updates to the generator, in order to slow the convergence of the discriminator. We keep the training setup of this implementation, i.e. using a batch size of $n_b = 64$ and the dimension z of the distribution used for sampling, as well as the hyperparameters used for training (25 epochs, learning rate 10^{-4}). The actual face swapping is performed as described in [15].

A. Segmented face images

In order to avoid close cropping of face images we use larger portrait images, but eliminate the diverse background that might negatively impact the output quality of the GAN. We apply fully automatic segmentation of the portrait region to the training images fed in the GAN. For segmentation we use a Tensorflow implementation[3] of the portrait segmentation method proposed in [18]. The results are in many cases acceptable, however, we found that the segmentation mask has in some cases holes. In order to handle this issue, a morphological closing operation (using a square kernel with a size of 20% of the image height) is applied to the resulting segmentation mask. After this postprocessing step of the mask,

[2]https://github.com/carpedm20/DCGAN-tensorflow
[3]https://github.com/Corea/automatic-portrait-tf

it is used to set the background region of the portrait image to black.

B. Face detection loss

The idea is to guide the training of the GAN by adding a face detector that complements the discriminator. The discriminator learns the differences between training and generated images, and scores them in the loss. However, there is no distinction between differences that impact the "faceness" of the generated image, and those that do not. We thus generate a set of n_b faces in every iteration, and apply a face detector to it. The face detector is expected to detect n_b faces, and we use the number of detected faces n_{det} as the face detection loss $l_f = 1 - \min\left(\frac{n_{det}}{n_b}, 1\right)$. The min function is a safeguard to ensure that the loss does not fall below 0 in case of false positives. However, as the input images consist only of images showing single faces the occurrence of false positives is extremely rare. By adding the detector and the resulting loss to the network, our approach is in terms of structure most similar to [17], which also adds components outputting additional loss terms.

We use the Viola-Jones face detector [23] for determining the loss based on face detection. The choice to use a rather basic face detector has been made for two reasons: First, a simple detector will be more selective and also discard only partially correct and harder to recognize faces. As we only apply it to generated face images, there is only little impact of false positives early in the training stage. Second, we ensure that we use a different method in the generation process than for the evaluation (cf. Section IV).

As applying the face detector requires generating samples and applying a fixed procedure that is not modified during training, we decided to add the loss as a kind of regularization term to the loss of the discriminator in each iteration. Thus no gradient needs to be calculated specifically for the detection loss and it penalizes generated images that contain less detectable faces. In the current implementation and the reported experiments we do not apply a weighting factor to the face detection loss.

An alternative approach would be to implement the face detector as a set of operations in a branch of the CNN, and determine the face detection loss from the output of this branch. In this case the equation for l_f can be modified to a leaky version (similar to e.g. leaky ReLU), i.e.,

$$ll_f = \begin{cases} 1 - \alpha n_{det} & \text{if } n_{det} < 0 \\ \alpha(n_{det} - n_b) & \text{if } n_{det} > n_b \\ 1 - \frac{n_{det}}{n_b} & \text{otherwise,} \end{cases}$$

with α being a small constant.

IV. Evaluation

In our experiments, the DCGAN is trained on the CelebFaces Attributes (CelebA) dataset [24], which contains around 200K images of more than 10K individuals. We train for 25 epochs with 3,165 iterations each. For each of the experiments, we use a set of 6,400 images (training or generated, respectively).

a) Evaluation metric: The evaluation of GANs is an active research question, and no generally accepted best practice exists. There are several recent works comparing evaluation metrics and pointing out different shortcomings [25]–[27]. Many of the generic method mostly focus on assessing differences in the distribution of the natural and generated samples. Another recent work [28] proposes two metrics based on image classification, to assess the diversity and the precision of the generated images. In the case of faces, the classifier measuring precision, trained on original images and tested on generated images ("GAN-test" in the terminology of [28]) can be realized by a pretrained classifier that discriminates faces and non-faces. In this work, we are interested in measuring differences in precision. We follow the approach used in [17] and [12] to use Multi-task CNNs (MTCNN) [13] for this purpose, and use the ratio of detected faces as a metric for the quality of the generated faces

In addition, we verify how well the anonymization of faces works. We perform an experiment on the Labeled Faces in the Wild (LFW) [11] data set, using the face recognition approach proposed in [29]. We compare the number of true/false positives/negatives between the original data set and a modified data set, where all faces have been replaced.

b) Results: Figure 2 compares the result of inserting a face into an image using the closely cropped generated image, and an image generated using the proposed approach, applying segmentation and face detection instead of cropping. It is visible that the artifacts caused by close cropping can be avoided. For objective evaluation, we measure the face detection accuracy for images generated using the proposed method. In addition, we provide results for three baselines: training (i.e., natural) face images, generated closely cropped faces as they are commonly used, and the generated uncropped portrait images including the background. The results are shown in Figure 3, at different decision thresholds of the face detector. Examples of generated faces are shown in Figure 4.

The proposed method using uncropped images achieves a similar detection accuracy than the closely cropped faces. However, at higher selectivity of the face detection probability (threshold ≥ 0.99), the proposed method outperforms closely cropped images. This means that with stricter discrimination of what is considered a face or not a face, the artifacts cause the detection to fail, while the proposed method still provides detectable faces.

Looking just at the baselines, one interesting result is that the generated images have mostly higher detection accuracy than natural images, which could be an indication that the GAN overfits. Despite the fact that the uncropped images with background are clearly less visually pleasing, the detection accuracy is higher than for the closely cropped ones. Visual inspection also shows that the diversity of the uncropped images with background is lower.

c) Experiment on anonymization quality: In order to assess how well the proposed approach reaches the aim of

(a) (b)

Fig. 2. Examples of inserted face (a) using cropped generated image and (b) using uncropped and segmented generated image.

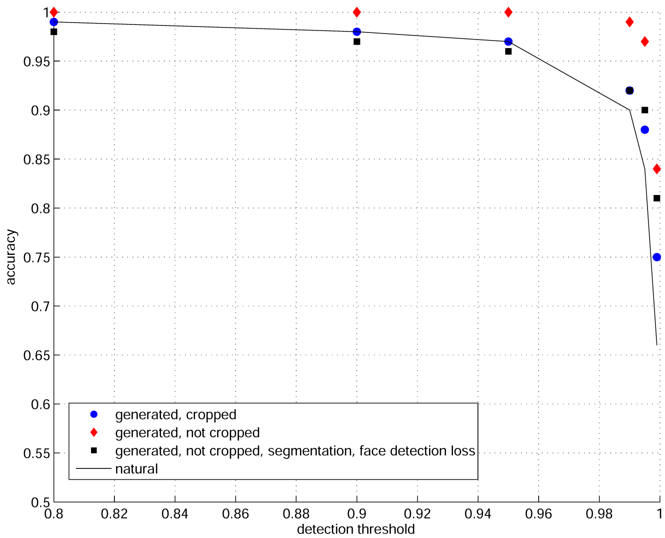

Fig. 3. The accuracy of the proposed approach (squares) at different thresholds of the face detection probabilities of the face detector, compared to the baselines with (circles) and without (diamonds) cropping. The accuracy for images from the training set is shown as solid line.

anonymizing faces in visual content, we use LFW as an established face recognition data set. We train the face recognition approach as proposed in [29] on the LFW training data. In brief, this approach uses MTCNNs [13] for face detection, extracts deep features using FaceNet and uses online random forests as a classifier. We run the trained algorithm on the original LFW test data set, and on a modified version, in which all faces have been swapped using the proposed method.

Table I summarizes the results. For the unknown faces in the collection, the true and false positive rates are very similar between the original and swapped faces. As expected, the false negative rate reaches nearly 1.0 when swapping is applied. The 0.44% of recognized known faces (i.e., true and false positives of faces in the collection) contain two types of errors: The first error type (0.30%) is matching a face of another person than the one which has been replaced. This type of false positive occurs also on the original data for 0.10% of the faces, i.e., this error increases by 0.20%. However, at the same time, the false positives for unknown faces (i.e., those not in the collection) decreases from 0.96% to 0.54%. This means that overall, the

	original	**swapped**
true positives for all known faces	79.39%	0.14%
false positives for all known faces	0.10%	0.30%
false negatives for all known faces	20.51%	99.56%
true negatives for unknown faces	99.04%	99.46%
false positives for unknown faces	0.96%	0.54%

TABLE I
RESULTS OF RUNNING FACE RECOGNITION ON THE LFW DATASET, USING THE ORIGINAL TEST DATA, AND AFTER APPLYING FACE SWAPPING TO THE TEST DATA.

number of false matches does not increase as a result of applying face swapping. The second error (0.14%) consists of correct matches of faces in cases where face swapping has not been applied because the detection of the face in the source image failed. Closer investigation has revealed that many of these cases are caused by JPEG compression artifacts.

Another class of error concerns cases where face detection in the swapped image failed, and thus recognition could not be performed. This is the case for 679 out of the 11,647 images in the data set. In 22 cases, this problem occurs to due to re-encoding of the image as JPEG, while in the other cases the geometric transformations to fit the generated face into the target image cause the face detector to fail. These cases require further investigation, and might be addressed by selecting generated faces that are more similar in pose to the target rather than random ones.

d) Ablation study: We also report results for the cases where only one of the proposed components is included, i.e., applying only segmentation and using the face detection loss for uncropped images with background. These results are shown in Figure 5. Using only segmentation, but not integrating the face detection loss, results in consistently lower accuracy across all thresholds of the face detector, and the difference increases at higher thresholds. Using only the face detection loss on the uncropped, unsegmented image also results in lower detection accuracy than the proposed method.

e) Discussion: There are three main insights concerning the evaluation. First, in our results the use of the detection threshold has major impact on the relative ranking of the methods. More generally, this means that the choice of the parameters of a classifier used for evaluation will influence results in a way that is neither linear nor easy to predict. To address this issue, we propose a metric of *accuracy at natural accuracy* (inspired by precision at rank measures in information retrieval, but lacking a natural notion of rank), which is defined as the accuracy reported by the classifier at the parameterization where it reports a certain accuracy level on natural data. For example, in our results, the threshold 0.990 corresponds to an accuracy on the natural data of 0.90. We would thus report the corresponding results as *acc@0.90*.

Second, in particular the results on the uncropped images show that neither the face detector nor the discriminator capture all the aspects of what defines a realistic face in the presence of background. For some of the experiments, the accuracy of face detection and the perceived quality of the faces diverge. A future extension could be the use of an

(a) (b) (c) (d)

Fig. 4. Examples of generated faces without cropping: (a) baseline, (b) with face detection loss, (c) with segmentation and (d) with segmentation and face detection loss.

automatic approach for facial beauty prediction (e.g., [30]–[32]), in addition to the face detector.

Third, the results show that the aim of anonymization on a realistic data set is achieved. However, there are cases where face detection failed, often due to compression artifacts. This highlights the vulnerabilities of CNNs to small perturbations in the data, which is an issue that needs more attention in future work.

V. Conclusion

In this paper, we have proposed two improvements for the face generation using GANs, in particular targeting preservation of privacy of images and videos in multimedia analysis and retrieval applications. First, we used portrait segmentation on the training data of the GAN, in order to generate images that are not only cropped to the face region, which may cause artifacts during insertion. Second, we integrated a face detection term to the loss function, in order to better guide the training process and improve the quality of generated faces. We used the detectability of the generated faces as an evaluation metric for comparing the proposed approach, and assess the impact of the two modifications. The results show that these modifications enable creating uncropped face images achieving the same or better performance than for closely cropped images. One insight from the evaluation is that the accuracy measured from face detection should be expressed with reference to the accuracy of the same face detector on natural images. The other insight is that the detection performance alone does not always correlate well with the perceived quality of the generated faces, thus the use of facial beauty prediction methods should be considered.

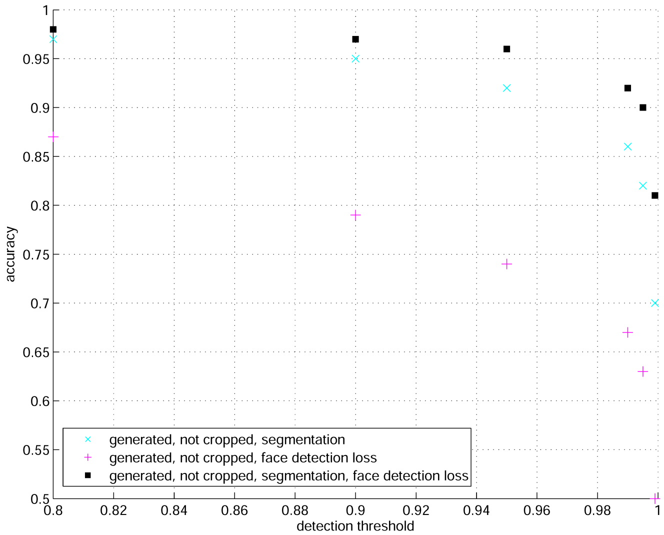

Fig. 5. Ablation study. The proposed approach (squares) is compared at different thresholds of the face detection probability with a version only using the segmentation (×) and a version only using the face detector (+).

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 732461 MARCONI ("Multimedia and Augmented Radio Creation: Online, iNteractive, Individual").

REFERENCES

[1] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L119, pp. 1–88, May 2016. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[2] K. Brkic, I. Sikiric, T. Hrkac, and Z. Kalafatic, "I know that person: Generative full body and face de-identification of people in images," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on. IEEE*, 2017, pp. 1319–1328.

[3] R. Natsume, T. Yatagawa, and S. Morishima, "Rsgan: Face swapping and editing using face and hair representation in latent spaces," *arXiv preprint arXiv:1804.03447*, 2018.

[4] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta, "Visual privacy protection methods: A survey," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4177–4195, 2015.

[5] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[6] L. Meng, Z. Sun, and O. T. Collado, "Efficient approach to de-identifying faces in videos," *IET Signal Processing*, vol. 11, no. 9, pp. 1039–1045, 2017.

[7] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: automatically replacing faces in photographs," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3. ACM, 2008, p. 39.

[8] S. Mahajan, L.-J. Chen, and T.-C. Tsai, "Swapitup: A face swap application for privacy protection," in *Advanced Information Networking and Applications (AINA), 2017 IEEE 31st International Conference on*, 2017, pp. 46–50.

[9] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*. IEEE, 2018, pp. 98–105.

[10] J. Gauthier, "Conditional generative adversarial nets for convolutional face generation," *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester*, vol. 2014, no. 5, p. 2, 2014.

[11] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled faces in the wild: A survey," in *Advances in face detection and facial image analysis*. Springer, 2016, pp. 189–248.

[12] X. Cao, S. R. Dulloor, and M. C. Prasetio, "Face generation with conditional generative adversarial networks," http://cs229.stanford.edu/proj2017/final-reports/5238859.pdf, Standord University, Tech. Rep., 2017.

[13] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[14] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *CoRR*, vol. abs/1511.06434, 2015.

[15] W. Bailer, "Face swapping for solving collateral privacy issues in multimedia analytics," in *International Conference on Multimedia Modeling*. Springer, 2019, pp. 169–177.

[16] Y. Saatchi and A. G. Wilson, "Bayesian gan," in *Advances in neural information processing systems*, 2017, pp. 3622–3631.

[17] Y. Wu, F. Yang, and H. Ling, "Privacy-protective-gan for face de-identification," *arXiv preprint arXiv:1806.08906*, 2018.

[18] X. Shen, A. Hertzmann, J. Jia, S. Paris, B. Price, E. Shechtman, and I. Sachs, "Automatic portrait segmentation for image stylization," in *Computer Graphics Forum*, vol. 35, no. 2. Wiley Online Library, 2016, pp. 93–102.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[20] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2881–2890.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[22] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 1175–1183.

[23] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[24] 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.

[25] Q. Xu, G. Huang, Y. Yuan, C. Guo, Y. Sun, F. Wu, and K. Weinberger, "An empirical study on evaluation metrics of generative adversarial networks," *arXiv preprint arXiv:1806.07755*, 2018.

[26] A. Borji, "Pros and cons of gan evaluation measures," *arXiv preprint arXiv:1802.03446*, 2018.

[27] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are gans created equal? a large-scale study," in *Advances in neural information processing systems*, 2018, pp. 698–707.

[28] K. Shmelkov, C. Schmid, and K. Alahari, "How good is my gan?" in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 213–229.

[29] M. Winter and W. Bailer, "Incremental training for face recognition," in *MultiMedia Modeling - 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8-11, 2019, Proceedings, Part I*, 2019, pp. 289–299. [Online]. Available: https://doi.org/10.1007/978-3-030-05710-7_24

[30] J. Gan, L. Li, Y. Zhai, and Y. Liu, "Deep self-taught learning for facial beauty prediction," *Neurocomputing*, vol. 144, pp. 295–303, 2014.

[31] J. Xu, L. Jin, L. Liang, Z. Feng, D. Xie, and H. Mao, "Facial attractiveness prediction using psychologically inspired convolutional neural network (pi-cnn)," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1657–1661.

[32] F. Chen, X. Xiao, and D. Zhang, "Data-driven facial beauty analysis: prediction, retrieval and manipulation," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 205–216, 2018.