A. Brömme, C. Busch, N. Damer, A. Dantcheva, M. Gomez-Barrero, K. Raja,
C. Rathgeb, A. Sequeira and A. Uhl (Eds.): BIOSIG 2021,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021

Ad-hoc and Deep-Learning based Face Anonymization Experiments

Jan Heimes[1]

Abstract: In today's digitized world, everyone leaves information on the Internet that can be used to identify the specific person. In this paper, we give an overview of common techniques for anonymizing facial images. The implementations consist of different methods to anonymize faces, as well as execute them with different strengths. The experiment results show that the higher the level of anonymization, the harder it is to recognize the person and the more encrypted the private information of our biometric data we leave in the world.

Keywords: biometrics, privacy, face recognition, anonymisation, computer vision, GAN, Neural Networks.

1 Introduction

We're seeing an increase in privacy concerns, especially these days, when there are big debates about Facebook's, Google's, and Microsoft's data collection. This is mainly due to the growing availability of information and the linking of our personal data with digital applications. We are online for much of our lives in highly digitized jobs. Above all, we engage in social media and perform work-related tasks there. This has gained strength with the Corona pandemic, as the home office has become the norm and digital conferencing via video platforms has become comSmonplace. Data relating to our own biometric characteristics has been increasingly injected into the Internet. Unfortunately, this information can be misused to identify or impersonate individuals. The biometric feature we most strongly associate with a person's privacy is their face.

This issue is being forced by increasing data protection regulations, such as the *General Data Protection Regulation GDPR* [Gr18] and increasingly by data breaches [WMS16]. The EU *GDPR* requires that companies obtain consent from individuals before privacy-sensitive data can be used or stored by a third-party company.

The goal of this project is to survey available face s

[1] Technical University of Denmark (DTU), Richard Petersens Plads, 324 DK-2800 Kgs. Lyngby Denmark, s202260@student.dtu.dk

2 Related Work

Existing studies focused purely on maintaining security. Therefore, the main goal was predominantly to remove as much sensitive and personal information as possible [Gr06]. This followed that anonymized image could not be detected. However, the non-private sensitive features are also lost. The biggest extreme of such a technique is the complete blackout method Fig. 1, where the whole face is simply covered with a black filter. No information can be extracted. However, this is not the perfect solution either.

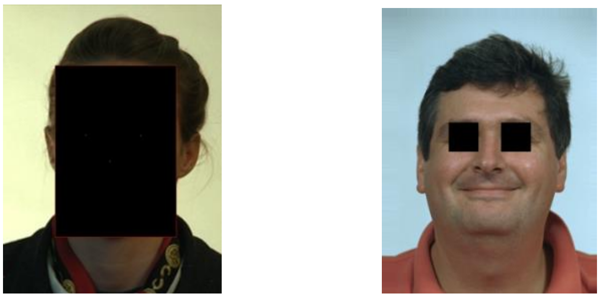


Fig. 1 Black out method for face or eyes

The intention to encrypt only the complete data has changed in recent studies, such as preserving privacy-insensitive features of the original image [Du14]. Nowadays, a technique that provides a trade-off between preserving the security and the utility of the image is sought [Mu13].

2.1 Ad-hoc Methods

Methods without considering the utility are named *Ad-hoc* techniques.
The following are some which are used later in this paper:

1. *blur*: smoothing filter with large strength [Gr06]
2. *pixelization*: similar to blur without smoothing, reducing resolution [Gr06]
3. *blackout*: a black rectangle which hides the face [NSM05]
4. *mask*: a black rectangle overlayed over a specific face part [NSM05]
5. *noise*: values of random pixels are changed and create “noise”

In terms of security preservation, blackout and masks techniques are the simplest methods. They also have the highest level of anonymization. Both can reduce recognizability by 100% by simply blacking out / erasing the face. [NSM05]. *Noise* that obscures more than 50% of the face can also reduce the probability of correct identification [NSM05].

To remove privacy sensitive information on Google Street View faces are blurred. [Go21]. However, techniques such as pixilation and blurring have been shown to be unreliable despite current use. [NSM05] [Gr06]. In both naive and reverse detection, neither of these anonymization techniques yields good results.

2.2 Neural Networks for face anonymization

Modern techniques for facial anonymization are often based on deep learning models. They focus on so-called *Generative Adversial Networks* (GANs). GAN methods can anonymize faces, but they provide significantly better utility [WY18] [Du14]. GAN-based neural network (NN) is even able to anonymize faces on the video recordings [RJLR18]. Even though it anonymizes faces, it retains the action of the person. Therefore, it can recognize actions of that person [RJLR18]. Moreover, GAN neural networks can remove identification features of the face while preserving the race, gender, and age of the person [Du14] [WY18].

One method used in this report is *DeepPrivacy*, whose architecture is based on the GAN mentioned above. Hukkelås et al [HML21] have developed an architecture for automatic anonymization of faces Fig. 2. This anonymization technique is used later in this paper. The pipeline first removes the face, then detects where the most important parts of the face are located. Finally, new faces are augmented, looking as realistic as possible.

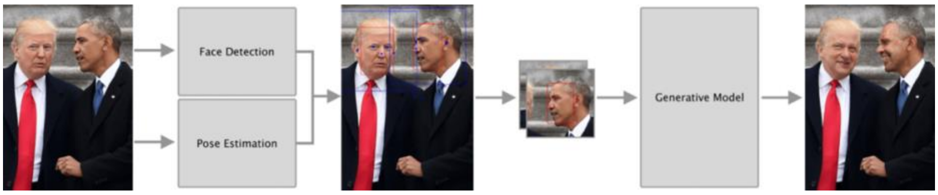


Fig. 2 DeepPrivacy pipeline: cropped, de-faced, and fed to the generator. [HML21]

To quantify the success of the GAN, *Hukkelås* et al. measures whether the resulting augmented faces are detectable. The detection is carried out by *DSFD* shown in Fig. 3 (Dual Shot Face Detector) a state-of-the-art face detector proposed by Wang et al [Li18]. Hukkelås states that 99.3% percent of augmented faces are still detectable [HML21].

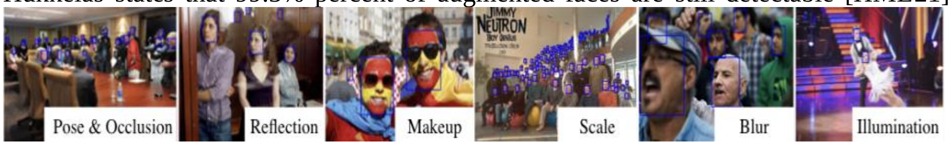


Fig. 3 DSFD for, blur, illumination, pose, occlusion, reflection, and makeup. [Li18]

3 Experiments

The following techniques are evaluated: (1) *pixelization*, (2) *blur*, (3) *noise*, (4) *DeepPrivacy*. The focus lays on the creation of privacy through anonymisation of each method without considering the utility. results.

3.1 Database

The experiments rely on a dataset provided by *Hochschule Darmstadt.*
The database is named FERET [Ph98]. To fulfil validation considerations and avoid unrelated errors, which would mislead outcomes a big subset of 500 images is used.

3.2 Face Recognition Software

One of the most widespread Face Recognition packages is *face recognition* [Ge18]. It is a publicly available written in Python. It has been used. Also, the lightweight face recognition *DeepFace* has been used with the Model *Facebook DeepFace*. However, *DeepFace* has a huge computation time for a bigger dataset. Hence the due to validation aspects of using a reasonable size of dataset the *face recognition* package is considered in this report.

3.3 Face Recognition Software

To evaluate the performance of the anonymisation following cases are considered:

1. *Naive Recognition*: Matching of original images to the anonymised images [NSM05]. In this setup gallery consist of original images, the anonymised images are matched against the gallery.
2. *Reverse Recognition:* Matching of anonymised images to the original images [NSM05]. In this setup gallery consist of the anonymised images, the original images
3. *Parrot Recognition:* Matching of anonymised images to anonymised images [NSM05]

In this report the focus lays on *Naive Recognition*, while the *Reverse Recognition* is shortly touched. The *Parrot Recognition* has not been taken into this report.

3.4 Figures

Pixelation, *blur* and *noise* are based on a similar code base. Their code was executed in the terminal. Where before the code is executed, it is decided which recognition technique from *HaarCascade* should be used (either recognition of the eyes or the whole face or recognition of the face with special focus on recognition of glasses). Glasses can cause problems because the lenses reflect the light, which causes difficulties for the face recognition program. The anonymization techniques were performed for the entire frontal face and additionally only for the eyes with emphasis on glasses to recognize faces more accurately.

For all ad hoc methods, only pixels within this processed box are used and reinserted into the image. For each face image, the script generates 500 images with kernel sizes of different strengths ***k***. The larger the value of k, the stronger the abstraction from the original image.

1. For *blur* a face with ***k*** = 1 indicates 1x1 kernel. This means the abstracted image is identical to the original image.

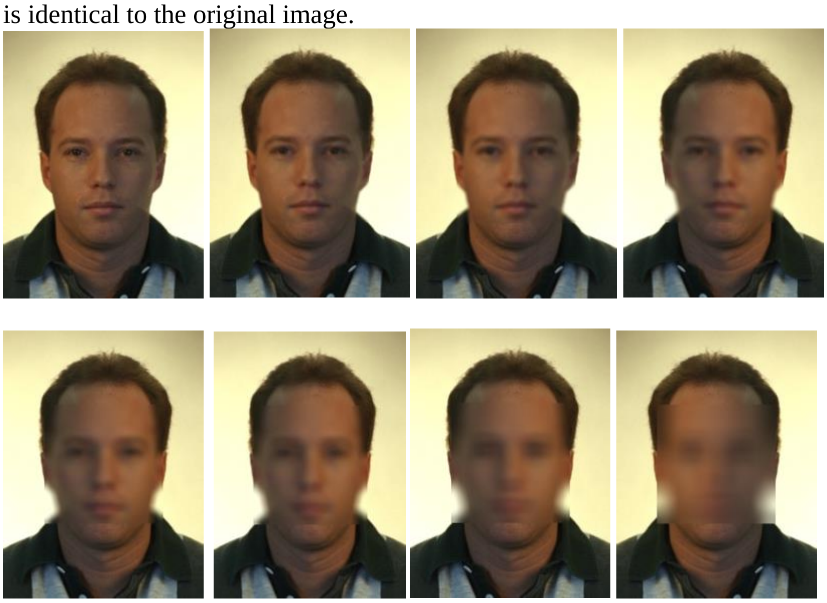


Fig. 4 Anonymization method blur with kernel size k = [5, 10, 15, 20, 30, 35, 50, 90]

2. For the *pixelization*, the region of the face is down sampled to be represented in *k x k* pixels. Different strengths of ***k*** are used.

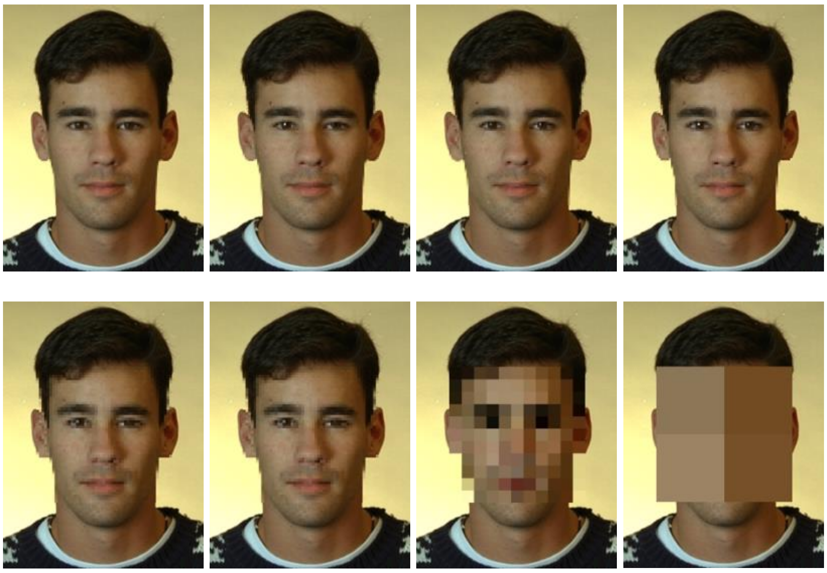


Fig. 5 Anonymization method pixelate with kernel size k = [10, 15, 20, 30, 35, 50, 90, 99]

3. , ***k*** % of the pixels are assigned a random value (red, green, blue)

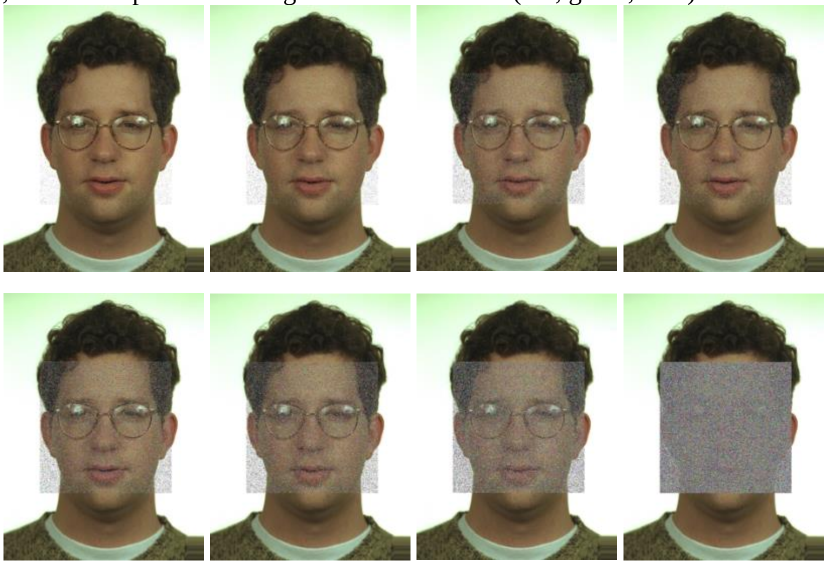


Fig. 6 Anonymization method noise with kernel size k = [5, 10, 15, 20, 30, 35, 50, 90]

3.5 DeepPrivacy Implementation

To test how *DeepPrivacy* can make faces unrecognizable, the following experiment compares combinations of images with the same person. This section describes the method used to evaluate anonymization.

The experiment's pipeline uses the same subset of the FERET database [Ph98] as the ad hoc algorithms. All these images are then anonymized using *DeepPrivacy*.
Subsequently, both sets of images; the anonymized set and the original are iterated in pairs to form and compare every possible image pair.

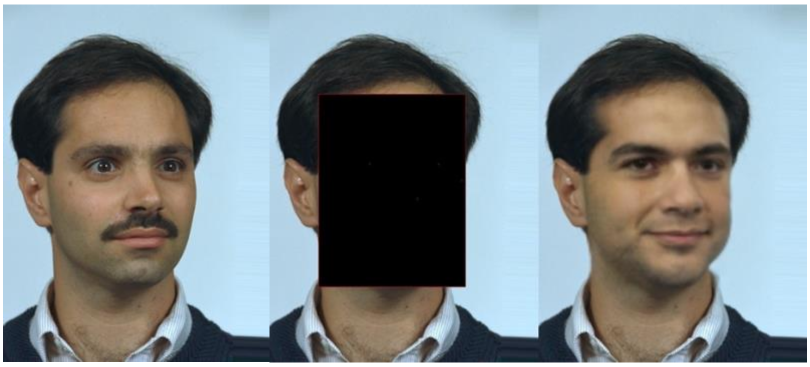


Fig. 7 Illustration of DeepPrivacy Anonymization technique

4 Evaluation naïve and reverse recognition

Facial anonymization can significantly reduce the probability of recognition while maintaining a certain degree of realism, as seen in the image above Fig. 7.

After anonymization, the resulting images are generally almost as distant from the image as an image that is a completely different person Fig. 8. However, a completely different person is still more anonymized than the same person anonymized using the *DeepPrivacy* method. The generative network in GAN generates faces not to misdirect them, but to fill the missing gap using the contextual environment.

Hence, we correctly classify approximately 72% of mated attempts as mated (true accept), and approximately 78% of non-mated attempts as non-mated (true rejects) Fig. 9. However, some errors are present. Approximately 22% non-mated attempts are incorrectly classified as mated (false accepts, type I error), and approximately 28% of mated attempts as non-mated (false rejects, type II error).

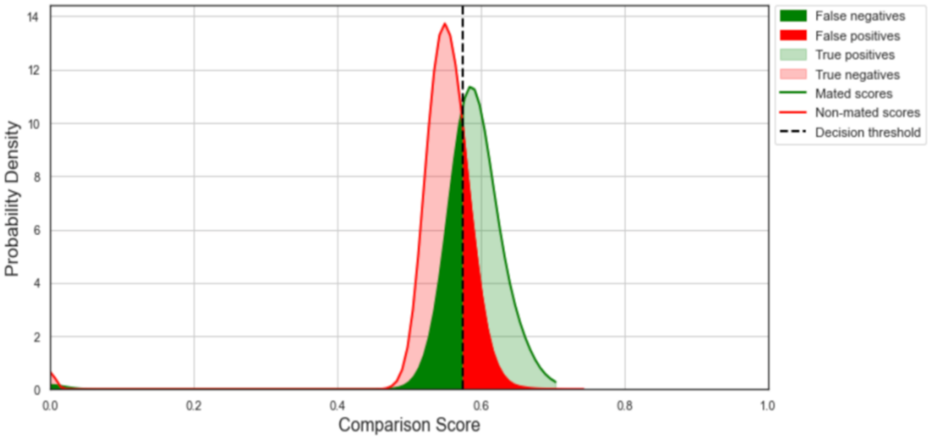
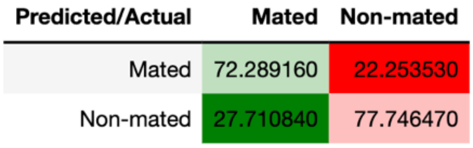


Fig. 8 Illustration of images anonymized with DeepPrivacy; threshold = 0.575



Predicted/Actual	Mated	Non-mated
Mated	72.289160	22.253530
Non-mated	27.710840	77.746470

Fig. 9 Confusion matrix for *DeepPrivacy* Method

In the figure shown below Fig. 10 is the DET-curve for the *DeepPrivacy* anonymization. We can realize the equal error rate ERR of around 22%. With false match rate *FMR in %* on the abscise and false non-match rate *FNMR in %* on the ordinate we observe that *FMR* equals around 0.1% when *FNMR* equals ~100% and *FNMR* ~ 1% when FMR equals ~100%.

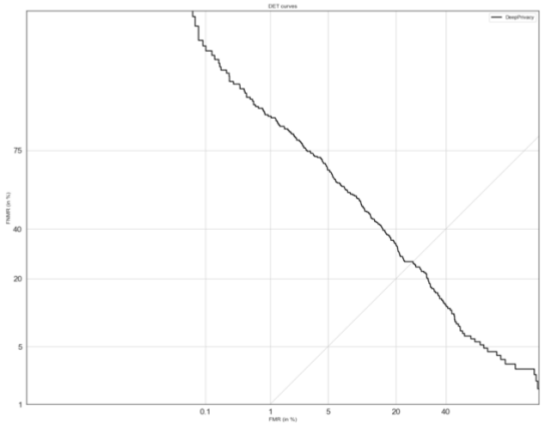


Fig. 10 DET-curve for DeepPrivacy anonymization method

In particular, the focus is on the False Matching Rate and False Non-Match Rate. In addition, the equal error rate plots are examined against the methods strength of ***k*** of each method. To consider the usefulness of an image, the Failure-to-Acquire Rate is considered (the face recognition software is not able to detect a face in the image).

For naive recognition, the statistics are generated for each strength ***k***. A similarity match result is called a mated result if it is the result of comparing two samples of the same biometric feature of a user. It is called non-mated or impostor score if it is the comparison of two biometric samples coming from different users.

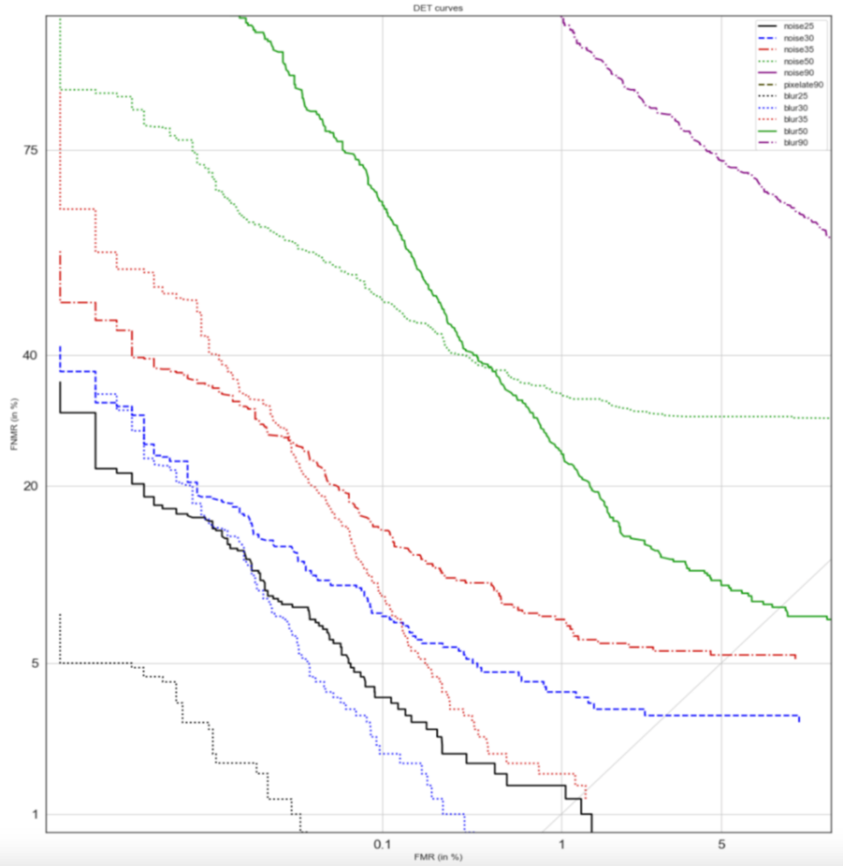


Fig. 11 DET curves for different anonymization methods with different k

We can see the whole spectrum of the trade-off between false positive occurrences (FMR) and false negative occurrences (FNMR). Also drawn (faint grey) is the so-called equal-error-rate (EER) line which merely goes through points where the FMR and

FNMR are equal. In this graph it is just slightly recognizable in the right lower corner. We benchmark the biometric performance of three different systems in this graph, namely *blur, noise*, and *pixelate*.

We can thus conclude that across the whole spectrum, system *blur* 30 performs the best while *noise* 30 also performs well. They have the lowest error rates compared to the other systems. The biometric performance (accuracy) is equal-error-rate (EER). It can be computed by interpolating on FMR and FNMR values or read from the figure directly.

It simply denotes the point at which the two error rates are equal, i.e. FMR == FNMR. That is the point, where the grey diagonal line starting at origin intersects with the DET curves for the respective systems. Hence, we can observe that the EER for *noise* 30 would be around 3%, whereas for *blur* 50 it would be already 10% and for *blur* 90 it would be already 40%.

We can plot the anonymization methods individually while comparing the DET curves for different k of the same anonymization method. We can realize that for the anonymization method *noise* and a ***k*** = 90 a DET curve does not exist anymore.

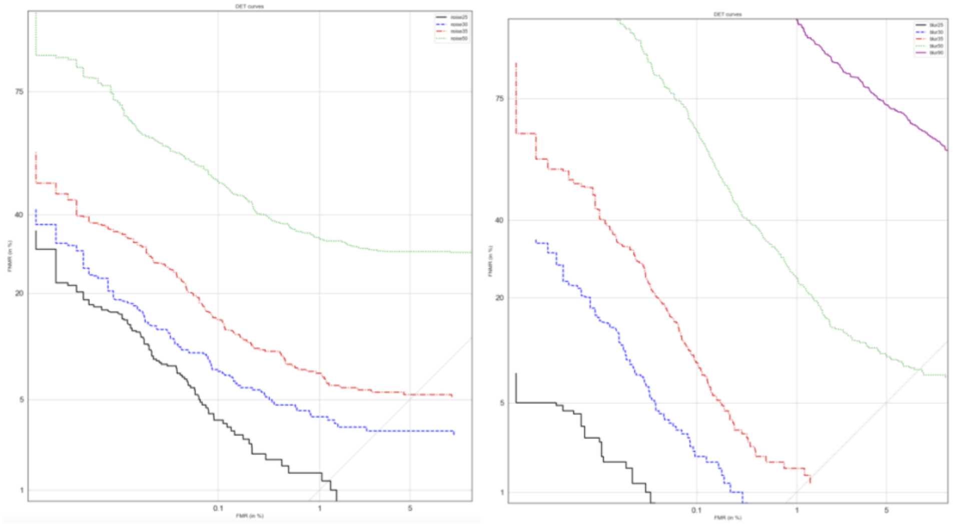


Fig. 12 DET curves of anonymization method noise [left] and blur [right], naïve recognition with k = [20, 25, 30 50] and k = [25, 30, 35, 50, 90] respectively

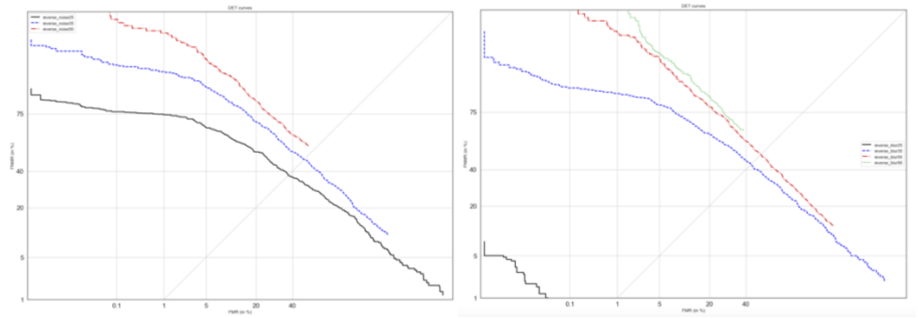


Fig. 13 DET curves of anonymization method noise [left] and blur [right], reverse recognition with k = [25, 35 50]and k = [25, 35 50, 90] respectively

We are able observe that the ERR for *noise* and *blur* in terms of the reverse recognition method are around 40%. Hence the face recognition performs worse recognizing the original faces from the anonymized faces, then the naïve recognition.

5 Outlook

Regarding privacy preservation of *ad-hoc methods*; reverser recognition results in lower rates of recognition than naïve. However, if we also consider the importance of image utility the *ad-hoc methods* do not provide a useful result. *DeepPrivacy* anonymizes the faces with keeping a high-level utility.

In total we were able observe, that the higher the degree of anonymization the more encrypted the biometric information. The scope of implementation and testing was limited due to technical constraints. This could be solved if the tests were produced in a serverless EC2 instance of AWS, which has greater capability than the relatively simple laptop used to run the experiments. As for future iterations of this experiment, the utility should take a bigger role apart from the encryption of biometric data.

References

[DU14] Du, Liang; Yi, Meng; Blasch, Erik; Ling, Haibin: GARP-face: Balancing privacy protection and utility preservation in face de-identification. In: IEEE International Joint Conference on Biometrics. IEEE, pp. 1–8, 2014.

[Ge18] Geitgey, Adam: Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning, Nov 2018.

[Go21] Google: Google street view, 2021

[Gr06] Gross, Ralph; Sweeney, Latanya; De la Torre, Fernando; Baker, Simon: Model-based face de-identification. In: 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). IEEE, pp. 161–161, 2006.

[Gr18] Gruschka, Nils; Mavroeidis, Vasileios; Vishi, Kamer; Jensen, Meiko:, Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR, 2018.

[HML21] Hukkel as, H akon; Mester, Rudolf; Lindseth, Frank: , DeepPrivacy: A Generative Adversarial Network for Face Anonymization, 2021.

[Li18] Li, Jian; Wang, Yabiao; Wang, Changan; Tai, Ying; Qian, Jianjun; Yang, Jian; Wang, Chengjie; Li, Ji-Lin; Huang, Feiyue: DSFD: Dual Shot Face Detector. CoRR, abs/1810.10220, 2018.

[Mu13] Muraki, Tomoya; Oishi, Shintaro; Ichino, Masatsugu; Echizen, Isao; Yoshiura, Hiroshi: Anonymizing face images by using similarity-based metric. In: 2013 International Conference on Availability, Reliability and Security. IEEE, pp. 517–524, 2013.

[NSM05] Newton, Elaine M; Sweeney, Latanya; Malin, Bradley: Preserving privacy by de-identifying face images. IEEE transactions on Knowledge and Data Engineering, 17(2):232–243, 2005.

[Ph98] Phillips, P Jonathon; Wechsler, Harry; Huang, Jeffery; Rauss, Patrick J: The FERET database and evaluation procedure for face-recognition algorithms. Image and vision computing, 16(5):295–306, 1998

[WMS16] Wheatley, Spencer; Maillart, Thomas; Sornette, Didier: The extreme risk of personal data breaches and the erosion of privacy. The European Physical Journal B, 89(1), Jan 2016.

[WY18] Wu, Yifan; Yang, Fan; Ling, Haibin: Privacy-Protective-GAN for Face De-identification. arXiv preprint arXiv:1806.08906, 2018.