

DeepFakes for Privacy: Investigating the Effectiveness of State-of-the-Art Privacy-Enhancing Face Obfuscation Methods

Mohamed Khamis
mohamed.khamis@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Habiba Farzand
habiba.farzand@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Marija Mumm
2232817m@student.gla.ac.uk
University of Glasgow
Glasgow, United Kingdom

Karola Marky
karola.marky@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

ABSTRACT

There are many contexts in which a person's face needs to be obfuscated for privacy, such as in social media posts. We present a user-centered analysis of the effectiveness of DeepFakes for obfuscation using synthetically generated faces, and compare it with state-of-the-art obfuscation methods: blurring, masking, pixelating, and replacement with avatars. For this, we conducted an online survey (N=110) and found that DeepFake obfuscation is a viable alternative to state-of-the-art obfuscation methods; it is as effective as masking and avatar obfuscation in concealing the identities of individuals in photos. At the same time, DeepFakes blend well with surroundings and are as aesthetically pleasing as blurring and pixelating. We discuss how DeepFake obfuscation can enhance privacy protection without negatively impacting the photo's aesthetics.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

DeepFakes, privacy, photos, computer vision, machine learning

ACM Reference Format:
Mohamed Khamis, Habiba Farzand, Marija Mumm, and Karola Marky. 2022. DeepFakes for Privacy: Investigating the Effectiveness of State-of-the-Art Privacy-Enhancing Face Obfuscation Methods. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI 2022), June 6–10, 2022, Frascati, Rome, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3531073.3531125

1 INTRODUCTION

Privacy issues can arise when individuals appear in photos they are not aware of. These issues are amplified by the increasing ubiquity of photo capturing devices in public environments, such as smartphone, surveillance and wearable cameras.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AVI '22, June 06–10, 2022, Frascati, Rome, Italy

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9719-3/22/06…$15.00
https://doi.org/10.1145/3531073.3531125

To protect privacy, a variety of privacy-enhancing photo obfuscation methods were proposed. Examples include face blurring [24] as done in Google Street view [7], pixelation [5, 22, 24], masking [22, 41], replacing users by avatars [31, 33] or cartoons [13]. In this paper, we specifically investigate DeepFakes for photo obfuscation. DeepFakes are a promising way to balance obfuscation effectiveness and photo aesthetics by replacing the faces of individuals by synthetic faces generated using a Generative Adversarial Network (GAN) [19]. Further, using DeepFakes the original information that humans were present in the scene is kept in contrast to inpainting, where individuals are completely removed and the missing part of the photo is filled in a way that is visually consistent with the background. Since DeepFakes are promising in terms of privacy but might also result in other implications, we investigate **1)** how effective DeepFakes are in concealing the identity of individuals, and **2)** the implications of using synthetic faces instead of more obvious obfuscation techniques. To this end, we conducted an online study (N=110) where participants guessed the identities of public figures whose faces were obfuscated using the aforementioned techniques, and rated their confidence in their guesses.

We found that our 110 participants were most successful in identifying public figures obfuscated by blurring (95.96%), followed by pixelating (85%), avatar (75%), masking (59.18%) and finally DeepFakes (29.03%). Overall, feedback from participants indicates that DeepFake obfuscation blends well with photos. However, it even blends too well that some were concerned about the ethical implications as it may mislead viewers. We conclude the paper by discussing the implications of privacy-aware DeepFake obfuscation and how to responsibly leverage this technology in an ethical way.

2 RELATED WORK

2.1 Obfuscating Individuals in Photos

Obfuscation is "the production of noise modeled on an existing signal in order to make a collection of data more ambiguous, confusing, harder to exploit, more difficult to act on, and therefore less valuable" [2]. This makes obfuscation promising for privacy protection in photos. At the same time, an important factor to consider when using privacy-aware obfuscations is their impact on the visual appeal of the photo i.e., photo aesthetics. For this reason, previous work in the HCI and security communities studied

obfuscation methods in terms of both: their effectiveness in concealing identities and their impact on perceived visual aesthetics of photos. Obfuscation methods that have been studied extensively include blurring [1, 16, 27, 28], pixelating [5, 23, 28, 39], masking [23, 28, 42], replacement with avatar [28, 33], cartooning [13, 33] and inpainting [8, 28, 31, 41].

Most related to our work is the work by Li et al [28] which evaluated the perception and effectiveness of multiple obfuscation methods with 271 participants. The study concluded that blurring and pixelating are ineffective whereas inpainting was found to be effective. In terms of user perceptions, their user study revealed that blurring, pixelating, inpainting and replacement with avatar were favored. An issue with inpainting however is that it removes information that individuals were present in the original photo. This may impact the users' memories in a negative and unethical way [6]. There are also some situations in which inpainting may not be possible. For example, if removing an individual results in a gap between two people in a group photo, the resulting photo may look unrealistic or fabricated. The study by Li et al. [28] also showed also showed that blurring, although not as effective in privacy protection, was among the preferred techniques, conforming with its extensive usage in research and practice [1, 16, 27].

Ilia et al. [16] proposed a system that blurs individuals' faces in the photo based on their privacy preferences. Their system was able to handle the problem of conflicting interest between multiple individuals associated with the photo by blurring the faces depending on the viewer of the photo. Li et al. [25] extended this work and proposed a solution that does not require individual's input every time the individual is tagged in a photo. However, controlling the content of the photo could be a better approach than controlling its recipient when considering the privacy of individuals, as it is not always possible to collect their sharing preferences. To this end, Hasan et al. [10] identified visual features which could help to distinguish between subjects and bystanders in the photo and used these features for training multiple machine learning models that distinguish them, achieving a mean accuracy of 93% when human raters had 100% agreement on whether they were shown a subject or a bystander. A natural follow-up step would be to obfuscate all of the bystanders in the photo, but this could result in that photo losing its visual appeal. In a follow up study by Hasan et al. [11], they found that minimizing the obfuscated areas in the photo results in more appealing photos. They compared various obfuscation techniques when applied on different scene elements like personal belongings and screen content, and investigated masking, blurring, pixelating, edge detection and silhouette. They found that *"stronger filters increase perceived privacy and decrease perceived information content, satisfaction, and aesthetics"* [11]. They also found that aesthetics and satisfaction depend on the size of the obfuscated area. The more obfuscated area, lesser the satisfaction and poor aesthetics of the photos. This implies that to maximize photo satisfaction and aesthetics, the obfuscated area should be minimized. Hasan et al. showed that applying beautification transformations such as abstract, cartoon and color did not significantly improve aesthetics of the photos containing obfuscated objects or people [12].

The aforementioned research influenced two key decisions in our work: First, prior research motivated us to compare DeepFake obfuscation to the following obfuscation techniques: blurring, pixelating, masking, and avatar replacement because these were the ones that were shown to be promising either in obfuscation effectiveness, user perception, or both [11, 26, 28]. Second, we learned from the work by Hasan et al. [11] that the obfuscated area should be minimized to maintain high aesthetics. Thus, we decided to obfuscate the user's face only and not their entire body.

2.2 DeepFake for Privacy Protection

Photo manipulation using DeepFakes could be categorized into four types: 1) Attribute Manipulation, 2) Expression Swap, 3) Identity Swap, and 4) Entire Face Synthesis [38].

Attribute Manipulation (aka face editing [38]) relies mostly on GANs to alter features like age, skin color and other so-called soft biometrics [9]. Expression Swap [38] replaces facial expressions of individuals thereby manipulating their face to an extent but not enough to conceal their identity. Identity Swap [38] replaces a face with another one. This technique benefits the film industry, but could also be misused for the creation of non-consensual pornography and fake news [38]. Kietzmann et al. [20] presented a number of examples using DeepFake's identity swap technique. While identity swap may hide the face, it creates ethical issues and is unlawful [35]. It is also not suitable for our purposes as it would violate the privacy of the individual whose face is used, and may also mislead the viewers. The final category is entire face synthesis, which is a technique that creates non-existent face photos. An example of this technique is the DeepPrivacy face anonymisation architecture developed by Hukkelås et al. [15]. DeepPrivacy uses a GAN to produce a photo with a synthetic face that matches the original pose and background. We chose this approach as the most suitable one for privacy-aware obfuscation because it hides the original face and uses a non-existent face instead.

3 IMPLEMENTING THE OBFUSCATIONS

We implemented DeepFake obfuscation using a state-of-the-art method that generates synthetic fake faces for anonymization [15]. The other obfuscations were applied using the OpenCV library [30]. For comparability with the previous work by Li et al. [28], we contacted the authors to use the same parameters they used in their implementations of blurring, masking, and pixelating obfuscations. We also used OpenCV to detect and label the faces.

DeepFakes: To implement DeepFake obfuscations, we used the DeepPrivacy framework by Hukkelås et al. [15]. In a nutshell, this algorithm uses a generative adversarial network (GAN), a type of neural network, to generate fake faces while incorporating "style transfer". Style transfer allows customizing a fake face by imposing some facial characteristics such as the skin and hair color of another person. In our use case, this means that the generated fake face would have the same hair and skin color of the individual but still look different. The DeepPrivacy framework also puts the background and the pose of the face into consideration to create more realistic fake faces. DeepPrivacy's anonymization was evaluated by running a state-of-the-art face detector on generated photos. However, a comparison of how well humans can identify the obfuscated person has not been done. The authors of DeepPrivacy made their



Figure 1: The figure shows examples of the obfuscating a photo with four individuals using blurring, pixelating, masking, DeepFake obfuscation and Avatar obfuscation. In our study, we obfuscated photos of public figures.

source code and pre-trained networks publicly available [14]. We then replaced the original face with the newly generated fake face.

Blurring: Li et al. [28] used a Gaussian blur with a radius of 4 pixels. To replicate these conditions, we used the *GaussianBlur()* method of the ImageFilter Module [29], setting the blurring radius to be directly proportional to the photo's width × height, with photos of size 770 × 552 pixels having a blurring radius of 4.

Pixelating: Pixelating was applied by first downscaling the face and then upscaling it again to its original size but in a pixelated form. The generated face then replaced the original one. Li et al. [28] downscaled the pixels of the face to 15 × 15 pixels and then upscaled them back to achieve pixelating. For this, we used the *resize()* method of the Pillow's Image Module [32], with the size parameter set to be 15 × 15 pixels for photos of size 770 × 552 pixels and larger. For smaller photos, we set the size parameter to be directly proportional to the product of the photo's width and height. Decreasing the size from 15 × 15 pixels for smaller photos was necessary to achieve the same level of pixelating as faces were much smaller, and keeping the same size value would have resulted in clearly visible faces. However, increasing the size to be directly proportional for larger photos was not viable, as it would make the obfuscation more detailed and hence less effective. So we capped the parameter to 15 × 15 pixels for photos sized 770 × 552 pixels or larger.

Masking: We applied a black rectangle on individual's face using OpenCV's *rectangle()* function shadowing prior work [28].

Avatar: We used emojis [21] instead of a human avatar as it is neutral to gender and skin color. The emoji was resized to the size of the located face and placed over the face area.

Examples of the obfuscations are illustrated in Figure 1.

4 METHODOLOGY

To investigate the effectiveness of the obfuscation techniques for privacy protection, we conducted an online survey with 110 participants recruited through mailing lists, social media and word of mouth (Males=55, Females=54, 1 preferred not to disclose). Participants ages ranged between 19 and 59 (M=27.6, SD=8.98). We presented obfuscated photos of well-known public figures to participants. We applied the following five obfuscation techniques on each photo: blurring, pixelating, masking, DeepFake and avatar (emoji). As public figures, we chose Morgan Freeman, Rihanna, Elon Musk, Boris Johnson and Keanu Reeves. All photos had a neutral background and the public figures in the photos were wearing a suit or a black jacket.

At the beginning of the study, participants provided their consent and their demographics. Next, they were asked to guess the identity of public figures presented in a counterbalanced order. Guessing could be done by typing either the name of the public figure or any information about that figure. For example, typing UK prime minister instead of Boris Johnson was considered a correct guess because we did not want to disadvantage participants who knew the public figure but not their name. After each guess, the participants were asked to rate the statement on a 5-point Likert scale. At the end, the participants were presented with the original non-obfuscated photos and asked whether they knew the public figure. This step was to filter out responses of participants that did not recognize the public figures because of not knowing them.

5 RESULTS

5.1 Guess Success Rate

The success rates for identifying the obfuscated individuals were highest for blurring (M=95.96%, SD=4.04%), followed by pixelating (M=85%, SD=13.31%), avatar (M=75%, SD=25.52%), masking (M=59.18%, SD=26.89%), then DeepFakes (M=29.03%, SD=23.85%). We analyzed the effectiveness with generalized estimating equations (GEE) to fit a repeated measures logistic regression. Since the dependent variable *success* is binary, we used a binary logistic model. We considered the obfuscation method as a within-subjects variable and predicting factor. Table 1 details the statistical analysis. All obfuscation techniques had a significant effect on the success rate, except for blurring (p=.063) and pixelating (p=.303).

5.2 Confidence in Guesses

We analyzed the confidence expressed by participants. A Friedman test showed that there is a statistically significant difference in perceived confidence in guesses depending on which obfuscation

	B	Std. Error	95% Wald Conf. Int. Lower	Upper	Wald χ^2	df	Sig.
Avatar	-0.994	0.3416	-1.663	-0.324	8.464	1	0.004
Blurring	0.916	0.4923	-0.049	1.881	3.461	1	0.063
DeepFake	-2.906	0.3767	-3.644	-2.168	59.516	1	< 0.001
Masking	-1.609	0.3721	-2.338	-0.879	18.690	1	< 0.001
Pixelating	-0.362	0.3516	-1.051	0.327	1.060	1	0.303

Table 1: Effectiveness results from the GEE.

is used $\chi^2(4) = 48.016, p < 0.001$. Post-hoc pairwise comparisons were done using Wilcoxon signed-rank tests. Bonferroni correction was applied to correct the p-value due to multiple comparisons resulting in a significance level set at p < 0.005. The highest confidence was recorded for the blurring technique (M=4.71, SD=0.15), followed by pixelating (M=4.35, SD=0.29), avatar (M=4.28, SD=0.23), masking (M=3.79, SD=0.40), DeepFakes (M=3.41, SD=0.62). Significant differences were found between blurring and each of: pixelating (Z = -3.967, p<0.001), masking (Z = -6.248, p < 0.001), DeepFakes (Z = -5.481, p < 0.001), and Avatar (Z = -4.815, p < 0.001). Significant differences were also found between masking and pixelating (Z = -3.846, p < 0.001), and between DeepFakes and pixelating (Z = -3.413, p < 0.005). The remaining pairs were not significantly different: Avatar vs Pixelating (p = 0.033), DeepFakes vs Masking (p = 0.703), Avatar vs Masking (p = 0.024), and Avatar vs DeepFakes (p = 0.082). The results suggest that participants were least confident in guessing DeepFakes and most confident in guessing blurred faces.

5.3 Limitations

There were four occasions when the participants guessed the public figure correctly, however, they answered negatively when shown the original photo and were asked if they knew the public figure. In these cases, the responses were not discarded, as correct full names of the public figures were provided by the participants, meaning that they did know the person in the photo. Further, we report on instances of public figures in front of neutral backgrounds. However, surroundings might provide cues that leak the obfuscated person's identity. Thus, our results do not consider environment cues.

6 DISCUSSION

We found that DeepFake obfuscation is significantly effective in protecting identities; participants were less successful in identifying public figures obfuscated using DeepFakes compared to other methods. Participants were also significantly less confident when guessing against DeepFake obfuscations compared to pixelating and blurring. The results on the effectiveness of blurring, pixelating, masking and avatar are in line with prior work; masking and avatar are more effective than pixelating and blurring, and the last two are largely ineffective in obfuscating familiar people [11, 12, 27, 28].

6.1 Declaring DeepFakes & Ethical Implications

While DeepFake obfuscation is a promising way to protect privacy, it may also be leveraged for unethical use such as impersonation [37]. In our implementation, we reduce possible negative uses of DeepFake obfuscation by using synthetically generated faces only rather than using other people's faces. Still, before DeepFake obfuscation becomes widely available to the public, we argue that there is a need to "declare" that photos have tampered with whenever any obfuscation method is used. One particularly promising way to achieve this is by using the JPEG Fake Media standard [18] to indicate in the metadata of the generated file that it includes DeepFakes. This metadata can then be scanned by systems to which the photos are uploaded so that the system can warn viewers that the said photos have been tampered with. Further, the distinction could also be made by preserving the digital fingerprint such as date and location. Other efforts in this direction include the Content Authenticity Initiative [17], which aims to preserve provenance and attribution data for digital content to counter misinformation. Blockchain has been suggested as a means to facilitate tracking the origin of photos of videos and changes made to them [40]. Furthermore, contextualized training and education were shown to assist in improving awareness and detection of DeepFakes [36].

6.2 Can DeepFake Obfuscation Impact the User's Memory?

A study by Elagroudy et al. [6] investigated the effect of privacy-aware obfuscations on user's memories. They found that ambiguous life-logs with obfuscations may distort memories. In their study, participants who experienced obfuscated photos of an event they attended remembered more details about the event, but were more likely to remember incorrect details. This was attributed to the retrieval induced forgetting phenomenon in which humans may recall incorrect details due to inaccuracies in the cues they are examining [3, 34]. If these preliminary findings are generalizable, then DeepFake obfuscation, like any other privacy-aware obfuscation method, may induce wrong memories. This has ethical implications and should be investigated. On a positive note, if DeepFake obfuscation can indeed impact memories, then it could also have potential use cases in treating people with Post-Traumatic Stress Disorder by altering their memories. However, this is also a controversial topic with ongoing debates about its ethical implications [4].

7 CONCLUSION AND FUTURE WORK

We presented a user-centered evaluation of the effectiveness of DeepFake obfuscation compared to state-of-the-art obfuscation methods using an online questionnaire study. Our results show that DeepFake obfuscation is promising. In terms of privacy protection, it is significantly effective in concealing identities; opposed to the other obfuscation methods, participants were not able to identify the public figures that are obfuscated using DeepFakes. We discussed the ethical implications of using DeepFakes for obfuscation and suggested ways to ensure transparency, such as the use of the JPEG Fake Media Standard.

ACKNOWLEDGMENTS

This work was supported an EPSRC New Investigator Award (grant number EP/V008870/1), and by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has also been funded by the UK EPSRC under grant number EP/S035362/1. This publication was partially supported by the Excellence Bursary Award by the University of Glasgow.

REFERENCES

[1] Andrew Besmer and Heather Lipford. 2009. Tagged Photos: Concerns, Perceptions, and Protections. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. ACM, New York, NY, USA, 4585–4590. https://doi.org/10.1145/1520340.1520704

[2] Finn Brunton and Helen Nissenbaum. 2015. *Obfuscation: A User's Guide for Privacy and Protest*. The MIT Press.

[3] Michael A Ciranni and Arthur P Shimamura. 1999. Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25, 6 (1999), 1403.

[4] Jacek Debiec. 2012. Memory reconsolidation processes and posttraumatic stress disorder: promises and challenges of translational research. *Biological psychiatry* 71, 4 (2012), 284–285.

[5] Jelle Demanet, Kristof Dhont, Lies Notebaert, Sven Pattyn, and André Vandierendonck. 2007. Pixelating familiar people in the media: Should masking be taken at face value? *Psychologica belgica* 47, 4 (2007), 261–276.

[6] Passant Elagroudy, Mohamed Khamis, Florian Mathis, Diana Irmscher, Andreas Bulling, and Albrecht Schmidt. 2019. Can Privacy-Aware Lifelogs Alter Our Memories?. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313052

[7] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent. 2009. Large-scale privacy protection in Google Street View. In *2009 IEEE 12th International Conference on Computer Vision*. 2373–2380.

[8] Amanna Ghanbari and Mohsen Soryani. 2011. Contour-based video inpainting. In *2011 7th Iranian Conference on Machine Vision and Image Processing*. IEEE, 1–5.

[9] Ester Gonzalez-Sosa, Julian Fierrez, Ruben Vera-Rodriguez, and Fernando Alonso-Fernandez. 2018. Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation. *IEEE Transactions on Information Forensics and Security* 13, 8 (2018), 2001–2014.

[10] Rakibul Hasan, David Crandall, Mario Fritz, and Apu Kapadia. 2020. Automatically detecting bystanders in photos to reduce privacy risks. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 318–335.

[11] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173621

[12] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2019. Can privacy be satisfying? On improving viewer satisfaction for privacy-enhanced photos using aesthetic transforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[13] Eman T. Hassan, Rakibul Hasan, Patrick Shaffer, David Crandall, and Apu Kapadia. 2017. Cartooning for Enhanced Privacy in Lifelogging and Streaming Videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1333–1342.

[14] Hakun Hukkelas. 2021. DeepPrivacy. https://github.com/hukkelas/DeepPrivacy Retrieved November 3, 2021.

[15] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. 2019. DeepPrivacy: A Generative Adversarial Network for Face Anonymization. In *Advances in Visual Computing*. Springer International Publishing, 565–578.

[16] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 781–792.

[17] JPEG. 2021. Content Authenticity Initiative. https://contentauthenticity.org/ Retrieved November 3, 2021.

[18] JPEG. 2021. JPEG Fake Media Standard. https://jpeg.org/items/20200803_fake_media.html Retrieved November 3, 2021.

[19] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[20] Jan Kietzmann, Linda W Lee, Ian P McCarthy, and Tim C Kietzmann. 2020. Deepfakes: Trick or treat? *Business Horizons* 63, 2 (2020), 135–146.

[21] kissclipart. 2021. Smiley Face Background. https://www.kissclipart.com/emojis-png-clipart-emoji-sticker-r1m975/ Retrieved November 3, 2021.

[22] Pavel Korshunov, Andrea Melle, Jean-Luc Dugelay, and Touradj Ebrahimi. 2013. Framework for objective evaluation of privacy filters. In *Applications of Digital Image Processing XXXVI*, Vol. 8856. International Society for Optics and Photonics, 88560T.

[23] Pavel Korshunov, Andrea Melle, Jean-Luc Dugelay, and Touradj Ebrahimi. 2013. Framework for objective evaluation of privacy filters. In *Applications of Digital Image Processing XXXVI*, Andrew G. Tescher (Ed.), Vol. 8856. International Society for Optics and Photonics, SPIE, 265 – 276. https://doi.org/10.1117/12.2027040

[24] Karen Lander, Vicki Bruce, and Harry Hill. 2001. Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 15, 1 (2001), 101–116.

[25] Fenghua Li, Zhe Sun, Ang Li, Ben Niu, Hui Li, and Guohong Cao. 2019. Hideme: Privacy-preserving photo sharing on social networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 154–162.

[26] Yifang Li, Wyatt Troutman, Bart P. Knijnenburg, and Kelly Caine. 2018. Human Perceptions of Sensitive Content in Photos. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[27] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Blur vs. Block: Investigating the Effectiveness of Privacy-Enhancing Obfuscation for Images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1343–1351.

[28] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.

[29] ImageFitler Module. 2021. ImageFilter Module. https://pillow.readthedocs.io/en/3.0.x/reference/ImageFilter.html Retrieved November 3, 2021.

[30] OpenCV. 2021. OpenCV. https://opencv.org/ Retrieved November 3, 2021.

[31] José Ramón Padilla-López, Alexandros Andre Chaaraoui, Feng Gu, and Francisco Flórez-Revuelta. 2015. Visual privacy by context: proposal and evaluation of a level-based visualisation scheme. *Sensors* 15, 6 (2015), 12959–12982.

[32] Pillow's. 2021. Pillow's Image Module. https://pillow.readthedocs.io/en/3.0.x/reference/Image.html Retrieved November 3, 2021.

[33] Chi-Hyoung Rhee and C Lee. 2013. Cartoon-like avatar generation using facial component matching. *Int. J. of Multimedia and Ubiquitous Engineering* 8, 4 (2013), 69–78.

[34] Daniel L Schacter. 1999. The seven sins of memory: insights from psychology and cognitive neuroscience. *American psychologist* 54, 3 (1999), 182.

[35] BBC Scotland. 2021. What are the laws against deepfake pornography across the UK? https://fb.watch/7HRvReEy1d/ Retrieved August 31, 2021.

[36] Rashid Tahir, Brishna Batool, Hira Jamshed, Mahnoor Jameel, Mubashir Anwar, Faizan Ahmed, Muhammad Adeel Zaffar, and Muhammad Fareed Zaffar. 2021. Seeing is Believing: Exploring Perceptual Differences in DeepFake Videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[37] Shahroz Tariq, Sowon Jeon, and Simon S Woo. 2021. Am I a Real or Fake Celebrity? Measuring Commercial Face Recognition Web APIs under Deepfake Impersonation Attack. *arXiv preprint arXiv:2103.00847* (2021).

[38] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. 2020. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion* 64 (2020), 131–148.

[39] Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann. [n.d.]. Filter Selection and Evaluation. ([n. d.]). http://www.mmi.ifi.lmu.de/pubdb/publications/pub/ZezschwitzTR_Privacy/ZezschwitzTR_Privacy.pdf

[40] Kaveh Waddell. 2018. The impending war over deepfakes. https://www.axios.com/the-impending-war-over-deepfakes-b3427757-2ed7-4fbc-9edb-45e461eb87ba.html Retrieved November 3, 2021.

[41] Xiaoyi Yu, Kenta Chinomi, Takashi Koshimizu, Naoko Nitta, Yoshimichi Ito, and Noboru Babaguchi. 2008. Privacy protecting visual processing for secure video surveillance. In *2008 15th IEEE International Conference on Image Processing*. IEEE, 1672–1675.

[42] Xiaoyi Yu, Kenta Chinomi, Takashi Koshimizu, Naoko Nitta, Yoshimichi Ito, and Noboru Babaguchi. 2008. Privacy protecting visual processing for secure video surveillance. In *2008 15th IEEE International Conference on Image Processing*. 1672–1675. https://doi.org/10.1109/ICIP.2008.4712094