PRACTICAL IMAGE OBFUSCATION WITH PROVABLE PRIVACY

Liyue Fan

University at Albany, State University of New York
liyuefan@albany.edu

ABSTRACT

An increasing amount of image data is being generated nowadays, thanks to the popularity of surveillance cameras and camera-equipped personal devices. While such image data can be shared widely to enable research studies, it often contains sensitive information, such as individual identities, location indications, etc. Therefore, the image data must be sanitized before sharing with untrusted parties. Current image privacy-enhancing solutions do not offer provable privacy guarantees, or sacrifice utility to achieve the standard ϵ-differential privacy. In this study, we propose a novel image obfuscation solution based on metric privacy, a rigorous privacy notion generalized from differential privacy. The key advantage of our solution is that our privacy model allows for higher utility by providing indistinguishability based on image visual similarity, compared to the current method with standard differential privacy. Empirical evaluation with real-world datasets demonstrates that our method provides high utility while providing provable privacy guarantees.

Index Terms— Image Obfuscation, Provable Privacy

1. INTRODUCTION

The amount of image data captured nowadays is rapidly increasing, thanks to the popularity of surveillance cameras and camera-equipped personal devices. Such image data can be widely shared to enable research studies, such as for energy efficiency optimization [1] and social relation recognition [2]. However, image data may contain various types of sensitive information, such as identity, location, health, belief, etc. To protect individual privacy, the image data must be sanitized before sharing with un-trusted parties.

Existing image privacy-enhancing solutions obfuscate regions-of-interest (ROIs) in an image, such as faces and texts, using standard methods such as pixelization and blurring. However, due to rapid development of machine learning, standard obfuscation methods are no longer effective in privacy protection. Studies have shown that machine learning models, especially convnet-based models, are highly adaptable to obfuscated data. For instance, McPherson et al. [3] showed that up to 96% of obfuscated faces can be re-identified. More sophisticated obfuscation methods have been proposed to enhance privacy and utility. For instance, GANs (generative adversarial nets) have been adopted for image obfuscation, e.g., by *inpainting* the head region as in Sun et al. [4], and by modifying the identity while preserving action detection as in Ren et al. [5]. Such approaches may heavily rely on training data, and yet do not provide *provable* privacy guarantees.

Differential privacy [6] has become the state-of-the-art paradigm for sanitizing statistical databases. It guarantees that an adversary is not able to distinguish between a pair of *neighboring* databases, differing in at most one record, by observing the output of the private algorithm. However, very little has been done toward publishing image data with such rigorous privacy notions. We recently developed an image pixelization method with ϵ-differential privacy [7]. However, the utility of the pixelized image is quite low, due to the perturbation required to achieve standard differential privacy.

The **goal** of this study is to develop an image obfuscation solution that provides provable privacy guarantees, without compromising utility. The challenge is three-fold: (1) we need to identify a rigorous privacy model that allows for a balanced trade-off between privacy and utility; (2) it is not straight-forward to define indistinguishable secret pairs in the image domain; (3) the adversary's background knowledge about the secret image needs to be characterized accordingly. The specific contributions of this paper are:

- We propose an image obfuscation solution that achieves metric privacy [8], a generalized notion based on differential privacy [6]. Specifically, our solution guarantees distance-based indistinguishability, providing better utility compared to standard differential privacy.
- We adopt Singular Value Decomposition (SVD), which is known to preserve perceptual similarity between images. As a result, our solution guarantees indistinguishability among visually similar images, e.g., those with the same singular matrices but slightly different singular values, providing strong privacy protection even in worst-case scenarios.
- We design a randomized sampling mechanism and prove that it satisfies metric privacy. In a non-straightforward effort, our work extends previous results in the 2-dimensional geo-space in [9] to an arbitrary k-dimensional metric space.
- We empirically evaluate our obfuscation method with widely used image datasets and study the privacy and util-

978-1-5386-9552-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICME.2019.00140

ity performance. We show our method achieves higher utility than differentially private pixelization, while providing provable privacy guarantees.

2. RELATED WORKS

Image Privacy Methods. Standard obfuscation has been used to sanitize sensitive ROIs when sharing data with untrusted parties. Such techniques include blacking, pixelization (or mosaicing) and blurring. However, recent studies [10, 3] have shown that the standard obfuscation methods are ineffective, due to the adaptability of convnet-based models. In particular, McPherson et al. [3] showed that obfuscated faces can be re-identified up to 96%; Oh et al. [10] showed that even with black fill-in faces, body and scene features can be utilized to re-identify 70% of the people. Recently, GANs (generative adversarial nets) have been adopted for visual data obfuscation, e.g., by inpainting the head region as in Sun et al. [4], and by modifying the identity while preserving action detection as in Ren et al. [5]. Such approaches may heavily rely on training data, and yet do not provide *provable* privacy guarantees. Moreover, the inpainted image may still breach privacy, e.g., up to 51.7% identity recognition reported in [4].

Differential Privacy.While differential privacy [6] provides rigorous privacy guarantees for individual records in the database, it is challenging to apply the standard differential privacy notion to non-aggregated data. Although it has been considered in learning deep models [11], only one study enables image data publication with differential privacy. We proposed in [7] an ϵ-differentially private method for image pixelization. The notion of neighboring databases was adapted to protect the presence/absence of any person or object captured by m pixels in the input image. However, the approach pixelizes the entire image and the quality of the obfuscated image is quite low. Moreover, the proposed privacy model may be overly strong by guaranteeing indistinguishability between image pairs where m pixels can change arbitrarily, thus inflicting high utility loss. The current study proposes a relaxed privacy model for image data to preserve utility and develops an efficient mechanism to achieve it.

3. PRELIMINARIES

An image I is considered as a matrix of pixels $\{I(i,j)\}$, where i and j index rows and columns respectively. In this paper, we focus on greyscale images; however, our method can be extended to multiple channels.

3.1. Perceptual Image Transformation

Rather than perform privacy perturbation directly on pixels or pixel-level aggregation as in [7], our approach quantitatively models the perceived information from an image and then applies rigorous privacy protection. The rationale is that many image transformations that inflict pixel value changes may not significantly affect the human perception of the image content, for instance, after JPEG image compression [12] or adding a small constant to every pixel. The challenge is thus to effectively model what can be perceived in an image, despite the aforementioned transformations.

In this paper, we consider Singular Value Decomposition (SVD) to capture the perceptual information in input images, as perceptual image hashing methods [13] based on SVD were shown to robustly hash visually similar images, such as after compression, rotation, and cropping. Such methods [13] employed SVD to extract most of the geometric structure and characteristics of the image data. The intuition of SVD is that any real or complex matrix A can be decomposed into a product of three matrices, i.e., $A = U\Sigma V^T$, where U and V are left and right singular vector matrices, Σ is a non-negative diagonal matrix, consisting of the singular values. Intuitively, the singular vectors in U and V, capture the geometric *features* in an image, while the singular values in Σ can be interpreted as the *magnitude* of each feature.

3.2. Metric Privacy

While the standard *differential privacy* [6] is a rigorous privacy notion, it is only applicable to publishing aggregate statistics. The problem studied in this paper wishes to publish image content, rather than aggregate statistics about the image data. Therefore, it requires a more general privacy notion for data that belongs to an arbitrary domain of *secrets*. In a recent paper, the authors of [8] extended the principle of differential privacy and proposed a generalized notion, i.e., metric privacy. Essentially, it defines a *distance* metric between secrets and guarantees a level of indistinguishability proportional of the distance. Specifically, given an arbitrary set of secrets $\mathcal{X}$ with a metric $d_{\mathcal{X}}$:

Definition 1. [8] A mechanism $K : \mathcal{X} \to \mathcal{P}(\mathcal{Z})$ satisfies $d_{\mathcal{X}}$-privacy, if and only if $\forall x, x' \in \mathcal{X} : d_{\mathcal{P}}(K(x), K(x')) \leq d_{\mathcal{X}}(x, x')$, or equivalently:

$$K(x)(Z) \leq e^{d_{\mathcal{X}}(x,x')} K(x')(Z) \quad \forall Z \in \mathcal{F}_{\mathcal{Z}} \tag{1}$$

where $\mathcal{Z}$ is a set of query outcomes, $\mathcal{F}_{\mathcal{Z}}$ is a σ-algebra over $\mathcal{Z}$, and $\mathcal{P}(\mathcal{Z})$ is the set of probability measures over $\mathcal{Z}$.

Metric privacy guarantees that the output of a mechanism should be roughly the same, i.e., bounded by the distance $d_{\mathcal{X}}(x, x')$, between two inputs x and x'. For an adversary who observes the output space, it is challenging to infer the exact input, thus the privacy of the input is protected. With this generalized definition, we can define a private mechanism $K()()$ on any domain $\mathcal{X}$ and $\mathcal{Z}$. The metric $d_{\mathcal{X}}$ can be derived by scaling a standard metric d by a factor ϵ, i.e., $d_{\mathcal{X}} = \epsilon \cdot d$ as suggested in [8]. As a result, the guarantee of

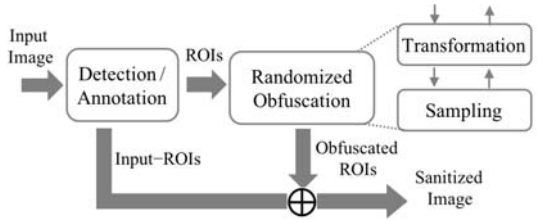


Fig. 1: Privacy-Preserving Image Sharing Framework

metric privacy also relies on ϵ: lower ϵ indicates higher indistinguishability, hence stronger privacy. When $d_{\mathcal{X}} = \epsilon \cdot d_H$ (where d_H is Hamming distance), the authors [8] have shown that metric privacy is equivalent to ϵ-differential privacy.

4. SOLUTION

We depict in Figure 1 the proposed solution for privacy-preserving image data publication. An image often contains one or more regions-of-interest (ROIs), such as faces, objects, text, etc., whose privacy needs protection. Such ROIs can be either detected automatically [14, 15] or annotated by data owners. Note that randomized obfuscation will be only applied to the ROIs, rather than to the whole image as in [7]. The obfuscated ROIs are then used in the output image for sharing. The obfuscation step involves two major components: transformation and sampling. A sensitive ROI will first be *transformed* to a feature vector, and the vector will go through the *sampling* step to achieve privacy guarantees; the sampled vector will be processed with *inverse* transform, resulting in the obfuscated ROI image.

Attack Model. We propose a strong attack model in order to provide privacy guarantees in worst-case scenarios. We assume an adversary who may have approximate knowledge about an input ROI. Specifically, the adversary knows the set of images that are visually similar to a given ROI (including the ROI image itself), e.g., with same singular matrices but somewhat different singular values. The adversarial goal is to infer the exact input image by observing the obfuscated image. We further assume the adversary is capable of performing the transformation. Therefore, given the obfuscated image, the adversary can transform it and produce the sampled vector, e.g., privacy-enhanced singular values, with which the adversary can try to infer the original singular values.

Privacy Assurance. Given an ROI, the goal of randomized obfuscation is to guarantee metric privacy in the transformed domain, which in return provides plausible deniability for the input ROI. Specifically, by observing the output of the sampling step (which can be achieved by the adversary's transformation of the obfuscated image), the adversary cannot distinguish between similar input vectors, e.g., similar sets of singular values. Hence the privacy of the input ROI is guaranteed. Thanks to Definition 1, our proposed solution provides rigorous privacy guarantees, despite an informed attacker.

4.1. Transformation

The transformation technique maps an input image (from here on, we refer to "image of an ROI" as "image" for brevity) to a feature vector. We denote the transformation as $\mathcal{F} : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^k$, which maps an input image I to a k-dimensional real vector $\boldsymbol{x_0}$. We denote Singular Value Decomposition (SVD) with $\mathcal{F}_{SVD,k}$, which yields the largest k singular values of an input image while setting the rest singular values to 0. Our choice of the transformation method is supported by the following considerations. Firstly, the transformation must be invertible: a vector generated by the sampling component can be transformed back to the image domain to obtain the obfuscated ROI image. We denote the inverse transform as $\mathcal{F}^{-1} : \mathbb{R}^k \rightarrow \mathbb{R}^{M \times N}$. For Singular Value Decomposition, we can derive the obfuscated image by multiplying a k-dimensional vector with the singular vector matrices. Secondly, the result of the transformation must capture high-level geometric information in the input image, offering ease to control the level of approximation in the transformed space. By increasing k, the reconstructed image $\mathcal{F}^{-1} \circ \mathcal{F}_{SVD,k}(I)$ provides a better approximation to the input image I. Thirdly, SVD has been shown to preserve image perceptual similarity [13]. Images that appear similar to the human eye would also exhibit high similarity in the transformed domain (e.g., under Euclidean distance [1]).

4.2. Sampling

Given the input feature vector $\boldsymbol{x_0}$, the private mechanism K performs random sampling in the space $\mathbb{R}^k$ according to certain probability distributions which can provide plausible deniability for $\boldsymbol{x_0}$. The following theorem states the privacy guarantee and the properties of the sampling distribution. Proof of the theorem is omitted for brevity.

Theorem 1. In a k-dimensional space, a mechanism $K()()$ that samples $\boldsymbol{x}$ given $\boldsymbol{x_0}$ according to the following probability density function satisfies $\epsilon \cdot d_k$-privacy

$$D_{\epsilon,k}(\boldsymbol{x_0})(\boldsymbol{x}) = C_{\epsilon,k} e^{-\epsilon \cdot d_k(\boldsymbol{x_0}, \boldsymbol{x})} \tag{2}$$

where d_k represents k-dimensional Euclidean distance and

$$C_{\epsilon,k} = \frac{1}{2}(\frac{\epsilon}{\sqrt{\pi}})^k \frac{(\frac{k}{2}-1)!}{(k-1)!} \tag{3}$$

assuming k is even without loss of generality.

Sampling according to Equation 2 can be achieved as follows. We first convert the Cartesian coordinates of $\boldsymbol{x}$ to the hyper-spherical coordinate system with $\boldsymbol{x_0}$ at the origin, resulting in 1 radial coordinate and $k-1$ angular coordinates.

[1] We are aware of other loss functions. Given each dimension in the feature space is independent, e.g., singular vectors, L2-norm is a classic choice.

Two steps are taken next: (1) sampling the radial coordinate according to its marginal distribution; (2) uniformly sampling a point on the unit $(k-1)$-sphere. Multiplying the results of (1) and (2) gives the output, privacy-enhanced vector, which will be used to generate the obfuscated ROI. Note that in (1), numerical root finding methods, such as the Newton-Raphson and secant method, must be used as there is no analytical solution for sampling the radial coordinate.

5. EXPERIMENTS

5.1. Settings

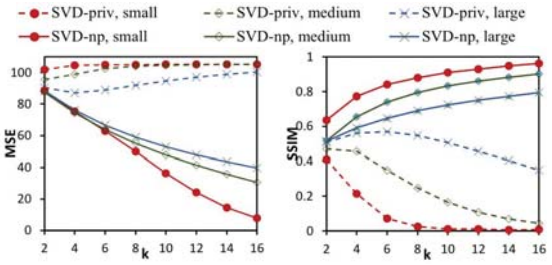
Datasets. Our method can be applied to sanitize faces, objects, and texts in general. We decide to focus on face images in this evaluation in order to illustrate the sensitivity of image data and to be fairly comparable with previous studies. In particular, we adopted the People In Photo Albums (PIPA) dataset [16]. The dataset contains head annotations (i.e., bounding rectangles), and has been widely used for learning person recognizers [10, 17] and image inpainting [4]. We adopted the PIPA test set and partitioned the annotated heads based on the size of the bounding rectangle, in order to study the impact of our private method on images of different sizes: the *small* partition contains 299 heads, the size of each between 256 and 900 pixels; the *medium* partition contains 4620 heads, the size of each between 901 and 10000 pixels; and the *large* partition contains 7785 heads, the size of each greater than 10000 pixels.

Metrics. To quantitatively measure the utility of images obfuscated by our solutions, we employed the commonly used Mean Square Error (MSE) defined between the original ROI and the obfuscated ROI. In addition, we adopted a widely used perceptual quality measure SSIM [18] , which considers perceived difference in structural information, in addition to luminance and contrast. In the following, we reported both measures between the obfuscated images and their sources. We also evaluated the computational efficiency: although omitted, results showed that our method incurs little overhead for privacy protection .

Setup. Our solution is prototyped in Python, running on 2.3 GHz i5 Intel Core with 16 GB memory. We used OpenCV and Numpy for Singular Value Decomposition, and Scipy for Newton's method for root finding. The parameters were set to default values, i.e., $\epsilon = 0.5$ and $k = 4$, unless specified otherwise. The average result among all images was reported.

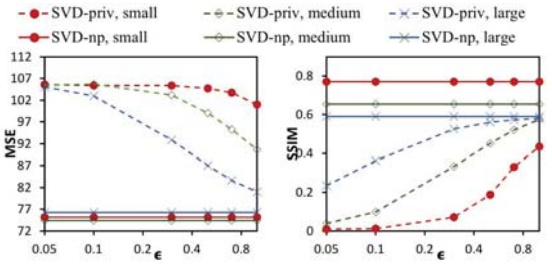
5.2. Results

Dimensionality k. We first present the impact of k, i.e., the dimensionality of the feature space, on the utility of the obfuscated image data. We evaluated our method, denoted by `SVD-priv`, against the non-private baseline `SVD-np`, which produces the reconstructed image using the k truncated singular values without the private sampling. We reported MSE and SSIM for each partition in Figure 2(a) and Figure 2(b). As k increases, the non-private baseline (solid lines) shows lower MSE, as a result of better approximation by the Singular Value Decomposition. As for our private method (dashed lines), we observe higher MSE errors with larger k values, as higher sampling errors are introduced by the private mechanism due to the increased dimensionality. Among different image partitions, the *large* partition yields highest MSE when using the non-private baseline (solid blue line with $\times$), as a large-size image requires a higher k for better approximation. When applying our method `SVD-priv`, the *large* partition shows the least impact under the privacy requirement (dashed blue line with $\times$), as the error introduced from sampling the k-dimensional vector is dispersed to a larger number of pixels. Moreover, there is an "elbow" point around $k = 4$ for `SVD-priv`, *large*, which captures the trade-off between the approximation error and the privacy error. For the *small* and *medium* partitions, the "elbow" point has not been observed. We believe that in those cases the privacy error dominates the approximation error for all k values, due to the relative small size of each image in those partitions. The SSIM results in Figure 2(b) are consistent with those measured in MSE. The best SSIM achieved by our private method is 0.590 for the *large* partition. This result is **comparable** to the recent head inpainting approach [4] (see Table 1, column "mask-SSIM" in [4]), with resulting SSIM from 0.186 to 0.679. In comparison, our method provides provable privacy guarantees without compromising utility.



(a) Mean Squared Error (MSE) (b) Structure Similarity (SSIM)

Fig. 2: Varying k



(a) Mean Squared Error (MSE) (b) Structure Similarity (SSIM)

Fig. 3: Varying ϵ

(a) *s*-1 (b) *s*-2 (c) *m*-1 (d) *m*-2 (e) *l*-1 (f) *l*-2

Fig. 4: Example images from PIPA partitions. Row 1 - original images ; Row 2 - images in Row 1 obfuscated by `SVD-priv` with $k = 4$ and $\epsilon = 0.5$.

Privacy ϵ. Here we present the privacy utility trade-off by varying the parameter ϵ in the following range $[0.05, 0.1, 0.3, 0.5, 0.7, 1]$. Recall that ϵ can be used to tune the level of indistinguishability as in Section 3.2. Lower ϵ indicates higher indistinguishability, hence stronger privacy, and vice versa. We plotted the utility of our method for each data partition in Figure 3. The non-private baseline is included for reference, whose performance should not be affected by the value of ϵ. As can be seen in Figure 3(a), when increasing the privacy parameter, i.e., from more private to less private, the MSE of our method starts to approach that of the non-private baseline. The relaxation of privacy benefits the *large* partition the most (dashed blue line with $\times$), as the sampling error is dispersed to a larger number of pixels. Figure 3(b) shows similar trends for the SSIM measure. We notice that when $\epsilon = 1$, `SVD-priv` yields around 60% SSIM for both *medium* and *large* partitions, demonstrating high utility. The performance gap between our method and the non-private baseline is the most significant for the *small* partition, indicating the highest privacy impact on utility for smaller images.

Qualitative Utility. In this section, we showcase the utility of our private method by presenting several original and obfuscated images in every partition. The example images are shown in Figure 4. Note that all images have been resized to fit and the difference in resolution can be observed when zoom in. It can be seen that our method yields high utility for the *medium* and *large* images, as discovered before. Higher amount of noise can be observed in the *small* images and we will study denoising methods in future work. Overall, we conclude that our method can be tuned, i.e., with larger ϵ values, to provide high utility.

Re-identification Attacks. Although our solution adopts a generalized privacy model that is different from standard differential privacy, we demonstrate that our solution can effectively mitigate practical re-identification attacks. We simulated the CNN-based attacks with the *AT&T* [19] faces database, outlined in [3]. One CNN model was trained for every obfuscation method, to explore how well the models can adapt to obfuscation. Subsequently, the model was used to attack the obfuscated testing instances via identity classification. Higher accuracy leads to a higher re-identification rate, indicating weaker protection offered by the corresponding obfuscation method. We provided in Table 1 the results of our method `SVD-priv` with a range of ϵ values, along with plain pixelization for each 16×16-pixel grid cell. With provable privacy and randomized mechanisms, our method results in lower re-identification risks, compared to standard pixelization. Furthermore, imposing stronger metric privacy further reduces the risk. For instance, `SVD-priv` with $\epsilon = 0.1$ yields the lowest risk.

	Pixelization	`SVD-priv`		
	16x16	$\epsilon = 0.1$	0.3	0.5
Accuracy	96.25	17.50	61.25	82.50

Table 1: Accuracy (in %) of Re-Identification Attacks

Fig. 5: Example images and their corresponding obfuscation. Row 1 - original *AT&T* images ; Row 2 - images in Row 1 obfuscated by `Pix-DP` [7] for each 16×16-pixel grid cell with $\epsilon = 0.3$; Row 3 - images in Row 1 obfuscated by our method `SVD-priv` with $k = 4$ and $\epsilon = 0.3$.

Comparison to Differentially Private Pixelization. We present several example images in the *AT&T* dataset and their corresponding obfuscation in Figure 5, in order to compare the utility of our method to the differentially private approach [7], denoted by `Pix-DP`. As can be seen, images in Row 3 exhibit more resemblance to the originals, compared to those in Row 2, despite the same ϵ value. This indicates that the standard indistinguishability achieved by `Pix-DP` may be overly strong in privacy protection. We conclude that our method yields higher utility than `Pix-DP`, while providing provable privacy guarantees.

6. CONCLUSION AND DISCUSSION

We have presented an image obfuscation solution that satisfies metric privacy. Our method protects the input image against an informed adversary, i.e., one who has knowledge about visually similar images to the input, and provides indistinguishability according to the similarity. As a result, our

method achieves a balanced trade-off between privacy and utility. Empirical evaluations with real-world image datasets demonstrated that our solution can mitigate re-identification attacks and achieves higher utility than differentially private pixelization, while providing provable privacy guarantees.

For future work, we will broadly evaluate the applicability of the proposed solution on different datasets, including objects and texts. We expect better utility results as those ROIs tend to be less complex than faces. We will also implement various denoising techniques to further improve the utility of the sanitized images, and develop the extension to color images. Last but not least, we will review other similarity preserving, invertible transforms in image processing literature, such as Discrete Cosine Transform, for further instantiation of our proposed solution.

ACKNOWLEDGEMENTS

The author would like to thank the anonymous reviewers for their valuable feedback. In addition, this material is based upon work supported in part by the National Science Foundation under Grant No. CNS-1755884. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

7. REFERENCES

[1] Varick L Erickson, Yiqing Lin, Ankur Kamthe, Rohini Brahme, Amit Surana, Alberto E Cerpa, Michael D Sohn, and Satish Narayanan, "Energy efficient building environment control strategies using real-time occupancy measurements," in *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM, 2009, pp. 19–24.

[2] Qianru Sun, Bernt Schiele, and Mario Fritz, "A domain based approach to social relation recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[3] Richard McPherson, Reza Shokri, and Vitaly Shmatikov, "Defeating image obfuscation with deep learning," *CoRR*, vol. abs/1609.00408, 2016.

[4] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz, "Natural and effective obfuscation by head inpainting," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[5] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo, "Learning to anonymize faces for privacy preserving action detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 620–636.

[6] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, pp. 265–284, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[7] Liyue Fan, "Image pixelization with differential privacy," in *Data and Applications Security and Privacy XXXII*, Florian Kerschbaum and Stefano Paraboschi, Eds., Cham, 2018, pp. 148–162, Springer International Publishing.

[8] Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi, "Broadening the scope of differential privacy using metrics," in *Privacy Enhancing Technologies*, Emiliano De Cristofaro and Matthew Wright, Eds., Berlin, Heidelberg, 2013, pp. 82–102, Springer Berlin Heidelberg.

[9] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, New York, NY, USA, 2013, CCS '13, pp. 901–914, ACM.

[10] Seong Joon Oh, Rodrigo Benenson, Mario Fritz, and Bernt Schiele, "Faceless person recognition: Privacy implications in social media," in *Computer Vision – ECCV 2016*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, Eds., Cham, 2016, pp. 19–35, Springer International Publishing.

[11] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, 2016, CCS '16, pp. 308–318, ACM.

[12] W.B. Pennebaker and J.L. Mitchell, *JPEG: Still Image Data Compression Standard*, Chapman & Hall digital multimedia standards series. Springer US, 1992.

[13] S. S. Kozat, R. Venkatesan, and M. K. Mihcak, "Robust perceptual image hashing via matrix invariants," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, Oct 2004, vol. 5, pp. 3443–3446 Vol. 5.

[14] Pierre Sermanet, Koray Kavukcuoglu, Soumith Chintala, and Yann LeCun, "Pedestrian detection with unsupervised multi-stage feature learning," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3626–3633.

[15] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5325–5334.

[16] Ning Zhang, Manohar Paluri, Yaniv Taigman, Rob Fergus, and Lubomir Bourdev, "Beyond frontal faces: Improving person recognition using multiple cues," 2015.

[17] Iacopo Masi, Stephen Rawls, Gerard Medioni, and Prem Natarajan, "Pose-aware face recognition in the wild," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[18] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[19] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, Dec 1994, pp. 138–142.

Practical Image Obfuscation with Provable Privacy

Liyue Fan, Department of Computer Science, University of North Carolina at Charlotte

Bibliographic reference:

L. Fan, "Practical Image Obfuscation with Provable Privacy," *2019 IEEE International Conference on Multimedia and Expo (ICME)*, Shanghai, China, 2019, pp. 784-789.
doi: 10.1109/ICME.2019.00140

Abstract:

An increasing amount of image data is being generated nowadays, thanks to the popularity of surveillance cameras and camera-equipped personal devices. While such image data can be shared widely to enable research studies, it often contains sensitive information, such as individual identities, location indications, etc. Therefore, the image data must be sanitized before sharing with untrusted parties. Current image privacy-enhancing solutions do not offer provable privacy guarantees, or sacrifice utility to achieve the standard ε-differential privacy. In this study, we propose a novel image obfuscation solution based on metric privacy, a rigorous privacy notion generalized from differential privacy. The key advantage of our solution is that our privacy model allows for higher utility by providing indistinguishability based on image visual similarity, compared to the current method with standard differential privacy. Empirical evaluation with real-world datasets demonstrates that our method provides high utility while providing provable privacy guarantees.

URL:

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8784836&isnumber=8784700

Draft: Image Obfuscation with Quantifiable Privacy

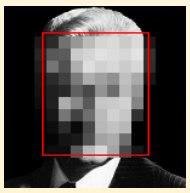
Liyue Fan Assistant Professor @ UNC Charlotte

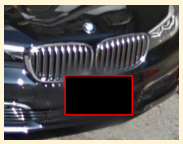
UNC CHARLOTTE
College of Computing and Informatics

Introduction

Image obfuscation is widely used to protect private content in photos, such as Google street view [1] and journalism [2]. Some opular obfuscation techniques:

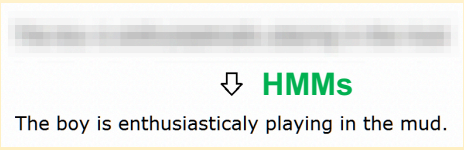

Blurring


Pixelization

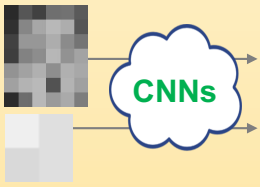

Blacking

However, machine learning models can adapt to standard obfuscation. For example:

- Hill et. al [3]



- McPherson et. al [4]



up to 96% accurate

- Oh et. al [5]

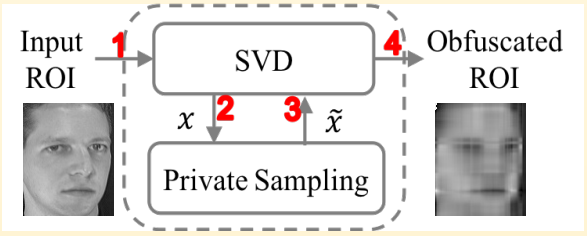
Prior Research:

- Sun et. al [6] and Ren et. al [7] adopt GANs to modify identities, but do not provide formal privacy.
- Fan [8] achieves rigorous ϵ –**Differential Privacy** but low utility, due to an *overly strong* privacy model.

Method

Objective 1: Quantifiable Privacy for ROIs

Objective 2: Privacy Utility Trade-off



Metric Privacy ($\epsilon \times d_{\mathcal{X}}$-privacy) [9] for any secret pair x and x':

$$K(x)(Z) \leq e^{\epsilon \times d_{\mathcal{X}}(x,x')} K(x')(Z), \quad \forall \text{ output } Z$$

- Privacy based on "similarity" → Utility friendly
- Standard DP is a special instance of Metric Privacy [9]

Results:

Row 1 – original images; Row 2 – our method, ϵ = 0.1; Row 3 – our method, ϵ = 0.3; Row 4 – our method, ϵ = 1; Row 5 – [8], ϵ = 1.

Acknowledgements: This research is supported in part by NSF grant CNS-1755884. Any opinions, findings, and conclusions or recommendations do not necessarily reflect the views of the National Science Foundation.

References

1. A. Frome et al., "Large-scale privacy protection in Google Street View," 2009 IEEE 12th International Conference on Computer Vision, Kyoto, 2009, pp. 2373-2380.
2. D. Aitkenhead. `I've done really bad things': The undercover cop who abandoned the war on drugs. The Guardian, 2016.
3. Hill, S., Zhou, Z., Saul, L., & Shacham, H. (2016). On the (In)effectiveness of Mosaicing and Blurring as Tools for Document Redaction, Proceedings on Privacy Enhancing Technologies, 2016(4).
4. Richard McPherson, Reza Shokri, and Vitaly Shmatikov. Defeating image obfuscation with deep learning. CoRR, abs/1609.00408, 2016.
5. Oh, Seong Joon, et al. "Faceless person recognition: Privacy implications in social media." European Conference on Computer Vision. Springer, Cham, 2016.
6. Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. Natural and effective obfuscation by head inpainting. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
7. Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo. Learning to anonymize faces for privacy preserving action detection. In ECCV, pages 620–636, 2018.
8. Liyue Fan. Image pixelization with differential privacy. In Data and Applications Security and Privacy XXXII, pages 148–162, Springer Cham, 2018.
9. Chatzikokolakis, Konstantinos, et al. "Broadening the scope of differential privacy using metrics." International Symposium on Privacy Enhancing Technologies Symposium. Springer, Berlin, Heidelberg, 2013.