Image Pixelization with Differential Privacy

Liyue Fan

Image Pixelization with Differential Privacy

Liyue Fan ✉

University at Albany, SUNY
Albany, NY, USA 12222
liyuefan@albany.edu

Abstract. Ubiquitous surveillance cameras and personal devices have given rise to the vast generation of image data. While sharing the image data can benefit various applications, including intelligent transportation systems and social science research, those images may capture sensitive individual information, such as license plates, identities, etc. Existing image privacy preservation techniques adopt deterministic obfuscation, e.g., pixelization, which can lead to re-identification with well-trained neural networks. In this study, we propose sharing pixelized images with rigorous privacy guarantees. We extend the standard differential privacy notion to image data, which protects individuals, objects, or their features. Empirical evaluation with real-world datasets demonstrates the utility and efficiency of our method; despite its simplicity, our method is shown to effectively reduce the success rate of re-identification attacks.

Keywords: Image Privacy; Differential Privacy

1 Introduction

There is a massive amount of image data captured by personal and commercial cameras nowadays. Every second 835 photos are uploaded on Instagram [1]. Over 18,000 traffic cameras spanning more than 200 cities in US are reported on TrafficLand [2]. Sharing image data widely would benefit various research communities. For instance, traffic images can be shared with third-party researchers to study vehicle behaviors toward intelligent transportation systems [3]; images uploaded on social media can be utilized by computer vision researchers to test their algorithms for social relation recognition [4] and early screening of mental illnesses [5]. However, publishing the aforementioned image data would raise *privacy concerns*. In fact, traffic cameras can capture the vehicle license plate; and personal images may capture objects or text that may indicate religious belief, health, habits, and location [6].

A number of studies proposed cryptography-based solutions for image sharing [7, 8], retrieval [9, 10], and feature extraction [11, 12] using untrusted service providers. While those solutions secure the image data with encryption, they exhibit a few drawbacks which make them inapplicable in our setting. Firstly, crypto-based image sharing explicitly trusts the data recipients, i.e., does not account for malicious recipients, and usually requires a secure channel to exchange secrets/keys. It can be challenging in both efficiency and security for

sharing data with a wide range of recipients. Secondly, the features computed by the untrusted server also need to be protected, such as shape positions and scale-invariant feature transform (SIFT), as those features often disclose sensitive information. Existing studies resort to more expensive cryptographic tools, such as homomorphic encryption and garbled circuit [11], or multiple independent servers [12], which potentially limit the feasibility of extracting complex features and enabling time-critical applications.

The sanitization of private content in image data has been studied in computer vision. Standard image obfuscation techniques, such as pixelization and blurring, are used by most privacy enhancing approaches to obscure the regions-of-interest (ROIs), including faces and texts. However, recent studies have shown that pixelization [13], blurring [13], and the P3 system [7] are not effective in privacy preservation. Given sufficient training data and the obfuscation technique, various models can be built to associate the obfuscated images to the ground truth, which can be used to decode redacted documents [13], and to re-identify faces and handwritten digits [14]. Therefore, we are in need of image obfuscation methods that can provide rigorous privacy guarantees.

The *goal* of this study is to ensure a rigorous privacy notion, *differential privacy* [15], for image data sharing. By definition, the adversary cannot effectively distinguish between secrets by observing the output of a differentially private mechanism, thus privacy is protected. To our best knowledge, our study is the first attempt of providing differential privacy guarantees for multimedia data publication. The specific contributions of the paper are as follows:

(1) To extend the standard differential privacy notion to image data, we propose the m-neighborhood notion, which allows for the protection of any sensitive information represented by up to m pixels.

(2) Given the high sensitivity of direct image publication, we propose a pixelization-based method with grid cells of b x b pixels, to achieve a utility-privacy trade off. We show that it provides differential privacy guarantees.

(3) We empirically evaluate the utility and efficiency of the differentially private pixelization with real-world image datasets with different resolutions. Two utility metrics are adopted to measure the absolute error and the perceptual quality, respectively. We show that our private method can yield similar output to the non-private pixelization.

(4) We simulate the re-identification attacks via deep learning and the results show that the differentially private pixelization significantly reduces the re-identification risk, even with low privacy requirements, i.e., $\epsilon \geq 0.1$ and $m = 16$.

The rest of the paper is organized as follows: Section 2 reviews recent and related literature; Section 3 and 4 provide the preliminaries and technical details of the differentially private pixelization; Section 5 presents the empirical evaluation; Section 6 concludes the paper and states future directions.

2 Related Work

Image Privacy Classification. Several studies (e.g., [16, 17, 6]) utilized image content features to predict the privacy settings for image sharing on online social networks (OSN). In particular, those studies explored classification models to predict whether an image is *private* or *public*: *private* images or ROIs should not be shared publicly or with OSN providers so as to stop the flow of information. While those studies show promise to understand the sensitivity of image data, the selected features often lack interpretability, e.g., after PCA projection or deep neural network features. Moreover, the classification models may not be perfectly accurate and images classified as *private* will not be shared with the public, preventing further utilization.

Image Obfuscation. Two popular image obfuscation techniques are *pixelization* (also referred to as mosaicing) and *blurring*. Pixelization [13] can be achieved by superposing a rectangular grid over the original image and averaging the color values of the pixels within each grid cell. On the other hand, blurring, i.e., Gaussian blur, removes details from an image by convolving the 2D Gaussian distribution function with the image. YouTube provides its own face blur implementation [18] for video uploads. McPherson et al. [14] studied pixelization and YouTube face blur and concluded the obfuscated images using those methods can be re-identified. In addition, a secure image sharing method named P3 [7] was also studied in [14] which encrypts the significant Discrete Cosine Transform (DCT) coefficients of the image. As YouTube face blur and P3 are not available/applicable in our study, we will focus on the pixelization technique and design a quantifiable privacy model for obfuscating image data.

Differential Privacy. Differential privacy [15] has become the state-of-the-art privacy paradigm for sanitizing statistical databases. While it provides rigorous privacy guarantees for each individual data record in the database, it is challenging to apply the standard differential privacy notion to non-aggregated data. Several variants of the privacy notion have been proposed. For instance, event-level privacy [19] aims to protect the presence of individual events in one person's data when releasing aggregated data. Local privacy [20] enables answering aggregate queries without a trusted data curator. Geo-indistinguishability [21] was proposed to release anonymized locations in a trajectory by sampling according to geo-distance in a randomized fashion. Although briefly mentioned in [22], there have not been any studies on ensuring differential privacy for image data. The goal of our work is to study the feasibility of differential privacy in image data sanitization by proposing an extended privacy model and an efficient mechanism to achieve it.

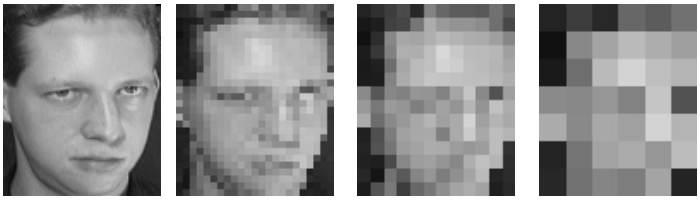
3 Preliminaries

Setting. We consider the problem setting where a data owner wishes to share one or more images with a wide range of untrusted recipients, e.g., researchers

or the greater public. The data owner must sanitize the image data prior to its publication, in order to protect the privacy of individuals or objects captured in the images.

Image Data. In the paper we focus on *grayscale* images: an input image I is regarded as an M x N matrix with integer values between 0 and 255 (0 is black and 255 is white). $I(x, y)$ denotes the "pixel" value at position (x, y) in the matrix. We note that the proposed privacy model and algorithm can be extended to RGB (red-green-blue) and HSV (hue-saturation-value) representations by considering each channel separately. We assume the sensitivity of each image is *independent* of other images to sanitize. Therefore we defer the extension of our study to inter-dependent images, such as a sequence of video frames, to future work in Section 6.

Pixelization. The pixelization technique renders the source image using larger blocks. It is achieved by partitioning the image using a two-dimensional grid, and the average pixel value is released for each grid cell. Similar to [13], we adopt a "square" grid where the pixel width is equal to the pixel height in the grid cells, i.e., each grid cell contains b x b pixels. In general, a smaller b value yields better approximation and visual quality, as is shown in Figure 1.



(a) Image (b) $b = 4$ (c) $b = 8$ (d) $b = 16$

Fig. 1: A Sample AT&T [23] Image and Its Pixelization with Different b Values

Standard Differential Privacy. The widely adopted Differential Privacy [15] definition operates in statistical databases.

Definition 1. [ϵ-Differential Privacy] A randomized mechanism $\mathcal{A}$ gives ϵ-differential privacy if for any neighboring databases D_1 and D_2 differing on at most one record, and for any possible output $\widetilde{D} \in \mathit{Range}(\mathcal{A})$,

$$Pr[\mathcal{A}(D_1) = \widetilde{D}] \leq e^{\epsilon} \times Pr[\mathcal{A}(D_2) = \widetilde{D}] \tag{1}$$

where the probability is taken over the randomness of $\mathcal{A}$.

The parameter ϵ specifies the degree of privacy offered by $\mathcal{A}$, i.e., a smaller ϵ implies stronger privacy and vice versa. It has been shown [15] that ϵ-differential privacy can be achieved with *the Laplace mechanism*, by adding i.i.d. noise $\tilde{N}$ to a function f, i.e., $\tilde{f}(D) = f(D) + \tilde{N}$. Specifically, $\tilde{N}$ is drawn from a Laplace distribution with 0 mean and $\frac{\Delta f}{\epsilon}$ scale, and Δf denotes the *global sensitivity* [15], which captures the maximum difference of f between any neighboring databases.

In this study, we extend the above definition to images, e.g., I_1 and I_2, and define *neighboring* images in the next section.

4 Differentially Private Pixelization

In this section we first propose the notion of neighborhood for image data, and then describe an effective privacy-preserving image publication algorithm.

Privacy Model. The concept of "neighboring images" is the key to the differential privacy notion, which should clearly define the private content under the protection of differential privacy. In this paper, we propose the following notion of image neighborhood.

Definition 2. [m-Neighborhood] Two images I_1 and I_2 are neighboring images if they have the same dimension and they differ by at most m pixels.

Allowing up to m pixels to differ enables us to protect the *presence* or absence of any object, text, or person, represented by those pixels in an image. For instance, each red rectangle in Figure 2a illustrates sensitive information which can be represented by $\sim$ 360 pixels, such as a pedestrian, a van, an object on grass, and a signage. One example neighboring image is shown in Figure 2b, differing only at the left-most pedestrian. By differential privacy, an adversary cannot distinguish between any pair of neighboring images by observing the output image. The privacy of the pedestrian, and any other sensitive information represented by at most m pixels, can thus be protected. The m-Neighborhood notion can also be applied to protect *features* of an object or person. For instance, the rectangle in Figure 2c contains $\sim$ 120 pixels and encloses the area of the eyes which is reportedly the optimal feature for a range of face recognition tasks [24].

When adopting the above definition, the data owner can choose an appropriate m value in order to customize the level of privacy protection, i.e., achieving indistinguishability in a smaller or larger range of neighboring images. We assume that removing those pixels is sufficient to protect the privacy of the underlying information, by definition of differential privacy [15].

(a) PETS [25] (b) Example neighbor image (c) AT&T [23]

Fig. 2: Sample Images and an Example Neighboring Image

Another advantage of our proposed privacy model is that it does not require annotated or detected sensitive regions-of-interest (ROIs). But rather, we

sanitize the given image[1] to protect any ROIs of size m. A straight-forward application of differential privacy is to apply Laplace perturbation to each pixel. As up to m pixels can change and each pixel can change by at most 255, the global sensitivity of direct image perturbation is very high, i.e., $\Delta I = 255m$, leading to high perturbation noise. Therefore, we propose differentially private *pixelization*, which achieves differential privacy while reducing the amount of perturbation noise added to the image.

Differentially Private Pixelization (`Pix`). In a nutshell, our algorithm first performs pixelization on an input image, and applies Laplace perturbation to the pixelized image. Specifically, let c_k denote the k-th grid cell over an M x N image. As shown in Figure 3, there are $\lceil \frac{M}{b} \rceil \lceil \frac{N}{b} \rceil$ cells in total. Let $K = \lceil \frac{M}{b} \rceil \lceil \frac{N}{b} \rceil$. The pixelization of an image I can be denoted as a vector of length K, i.e.,

$$P_b(I) = \{\frac{1}{b^2} \sum_{(x,y)\in c_1} I(x,y), \frac{1}{b^2} \sum_{(x,y)\in c_2} I(x,y), \ldots, \frac{1}{b^2} \sum_{(x,y)\in c_K} I(x,y)\}.$$

The global sensitivity of P_b is thus $\Delta P_b = \max_{I_1,I_2} |P_b(I_1) - P_b(I_2)| = \frac{255m}{b^2}$, as the difference between any two pixels is at most 255 and up to m pixels can differ between any neighboring images I_1 and I_2.

Let $\widetilde{\boldsymbol{N}} = \{\widetilde{N}_1, \widetilde{N}_2, \ldots, \widetilde{N}_K\}$ and each $\widetilde{N}_k$ ($k \in \{1, \ldots, K\}$) is randomly drawn from a Laplace distribution with mean 0 and scale $\frac{255m}{b^2\epsilon}$. The following theorem states the privacy guarantee of the $\widetilde{P}_b$ algorithm, where $\widetilde{P}_b(I) = P_b(I) + \widetilde{\boldsymbol{N}}$, $\forall I$.

c_1	c_2	...	$c_{\lceil \frac{N}{b} \rceil}$
...			$c_{2\lceil \frac{N}{b} \rceil}$
...			...
...			$c_{\lceil \frac{M}{b} \rceil \lceil \frac{N}{b} \rceil}$

Fig. 3: b x b grid cells over an M x N matrix

Theorem 1. *Algorithm $\widetilde{P}_b$ satisfies ϵ-differential privacy.*

Proof. Since the $\Delta P_b = \frac{255m}{b^2}$, by definition [15] applying the Laplace mechanism to P_b achieves differential privacy.

Note that each pixel in $\widetilde{P}_b(I)$ is truncated to the range of $[0, 255]$. This post-processing of $\widetilde{P}_b$ does not affect its privacy guarantee.

5 Experiments

Below we present the empirical evaluation of differentially private pixelization.
Datasets: We considered the Multiple Object Tracking Benchmark [25], which contains video frame sequences widely used in the MOT community. Among those, two datasets adopted in this study are : *PETS* dataset, i.e., PETS09-S2L1, showing walking pedestrians on a university campus with 795 images and 768 x 576 resolution; and *Venice* dataset, i.e., Venice-2, showing walking pedestrians

[1] the *given* image could be an entire image as in Figure 2a, or part of an image, e.g., only face as in Figure 2c

Parameter	Description	Default Value
ϵ	Privacy parameter	0.5
m	Number of different pixels allowed	16
b	Grid cell length	16

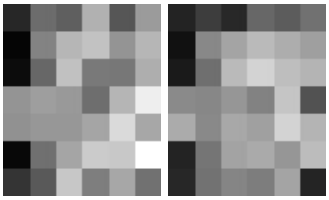
Table 1: Default Parameter Setting

around a large square with 600 images and 1920 x 1080 resolution. Both datasets were converted to grayscale. In addition, we adopted two datasets used in the re-identification attacks via deep learning [14]: *AT&T* [23] database of faces which contains 400 grayscale images of 40 individuals with 92 x 112 resolution; and *MNIST* [26] which contains 60,000 grayscale images of handwritten digits with 28 x 28 resolution.

Setup: We prototyped our method in Python, running on 2.3 GHz i5 Intel Core with 16 GB memory. The parameters take default values in Table 1, unless specified otherwise. The *utility* of our method can be measured by the standard Mean Square Error (MSE), which is defined between the input image and the sanitized image. We also adopted a widely used perceptual quality measure named Structural Similarity (SSIM) [27], which considers the perceived similarity in structural information in addition to luminance and contrast. One example of SSIM's advantage over MSE, is that an image derived by subtracting a certain value from every pixel in the input image would exhibit high structural similarity to the input at a significant absolute error. Due to this consideration, both utility measures were evaluated. In each experiment, we reported the average result among all the images in each dataset.

5.1 Impact of b

We first varied the grid cell length b to empirically evaluate its impact on the utility of the sanitized image. Note that in addition to our differential private method `Pix`, we included the non-private pixelization method, i.e., `Pix_np`, which is parameterized with the same b value, as a reference for utility. Figure 5 and 6 present the utility results measured by MSE and SSIM, respectively. As can be seen, by increasing b, the non-private baseline yields a higher MSE and a lower SSIM in each dataset, as a result of the coarser approximation by pixelization. SSIM drops significantly from $b = 2$ to $b = 6$. On the other hand, our private method generates higher utility images when b increases, approaching the utility of the non-private baseline. This is due to a lower Laplace perturbation error, the magnitude of which is governed by $\frac{255m}{b^2}$. As shown in Figure 4, our private



(a) `Pix` (b) `Pix_np`

Fig. 4: Pixelization output - AT&T - $b = 16$, $m = 16$, $\epsilon = 0.5$

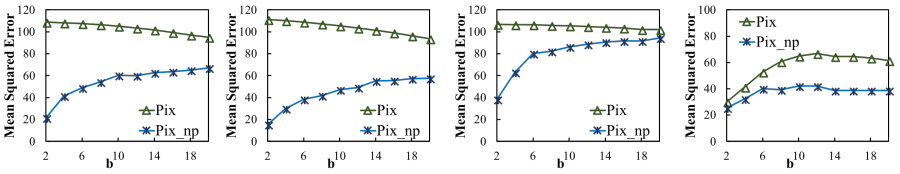


Fig. 5: MSE vs. Varying b

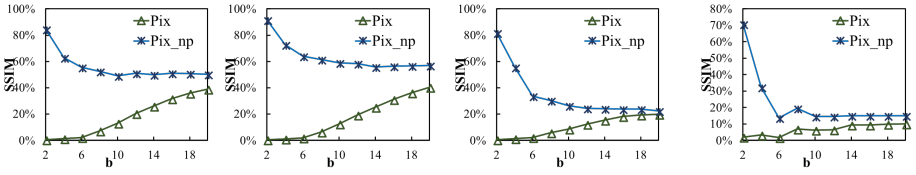


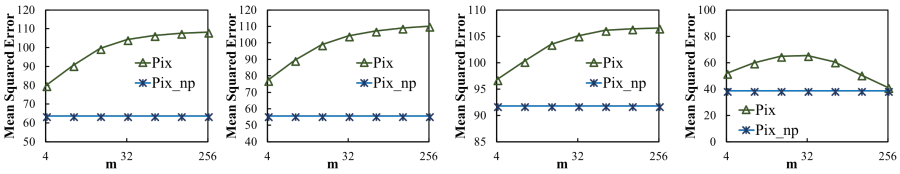
Fig. 6: SSIM vs. Varying b

method outputs an image closely resembles the non-private pixelization, except for a few grid cells. Note that in Figure 5d `Pix` shows an increasing trend in MSE for $2 \leq b \leq 12$. The reason is that *MNIST* depicts white (255) digits on a black (0) background, and when b is small the large Laplace noise does not significantly affect those extreme pixel values.

5.2 Impact of m

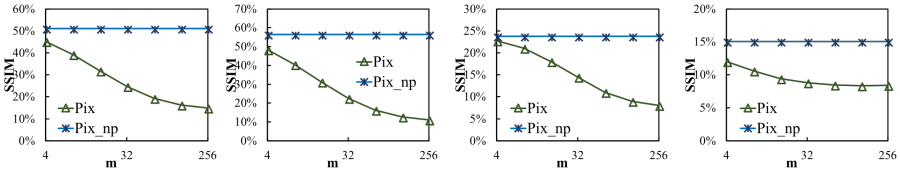
In the following experiment, we varied m, the number of pixels allowed to change between any pair of neighboring images, characterizing the indistinguishability requirements of the differentially private method. Intuitively, a larger m value ensures indistinguishability on a wider range of images, hence stronger privacy. The utility results are depicted in Figure 7 and 8. Note that the non-private method `Pix_np` should not be affected by the variation of m values. As m increases, the utility of our private method `Pix` drops, as the Laplace perturbation noise is larger. This shows the tradeoff between utility and privacy. For the *MNIST* dataset, we observe a lower MSE when $m > 32$ in Figure 7d. The increased Laplace perturbation noise "helped" with sharing images are composed of black and white pixels. However, the increased privacy requirement has a clearer manifest on the perceptual quality, i.e., a steady decreasing trend in Figure 10d, as SSIM captures the image structural information in addition to pixel values.

To further illustrate the utility of the differentially private pixelization, sample images generated under the default parameter setting are provided in Table 4. As can be seen, for images of larger size, e.g., the *PETS* and *Venice* datasets, setting $b = 16$ and $m = 16$ would allow the viewer to recognize the street scene and the number of pedestrians in the sanitized images. For smaller sized images,



(a) PETS (b) Venice (c) AT&T (d) MNIST

Fig. 7: MSE vs. Varying m



(a) PETS (b) Venice (c) AT&T (d) MNIST

Fig. 8: SSIM vs. Varying m

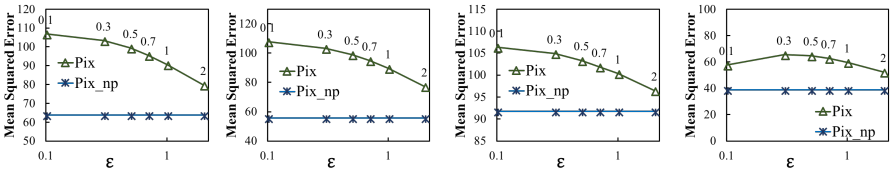
e.g., the *AT&T* and *MNIST* datasets, the pixelization grid size $b = 16$ yields a very coarse approximation, and with $m = 16$ the private perturbation mechanism inflicts a higher visual quality loss, due to smaller image sizes. Therefore, m can be adjusted by the user of our private method depending on the input image size and the privacy requirement. However, we note that when any obfuscation is applied to faces and digits, the goal is usually to reduce the identifiability of the resulting image; the example *AT&T* and *MNIST* images show promising visual results of our method.

5.3 Impact of ϵ

We also studied the impact on utility by varying the privacy parameter ϵ. Intuitively, lower ϵ value ensures stronger privacy, and yields lower utility. As can be seen in Figure 9 and 10, our private method `Pix` shows a lower MSE and a higher SSIM when increasing ϵ. An expected exception is observed for *MNIST* dataset in Figure 9d, where smaller ϵ values, e.g., 0.1, can benefit sharing extreme pixel values. Again, the SSIM measure is shown to be more robust than MSE, exhibiting a consistently increasing trend when ϵ increases in Figure 10d.

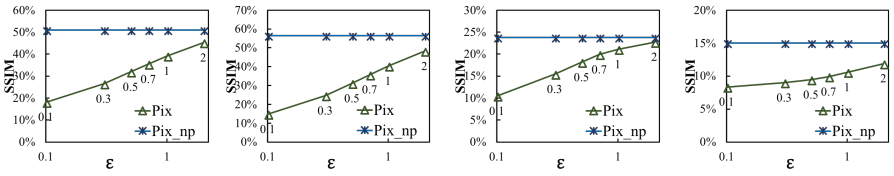
5.4 Runtime

Another important performance index is the efficiency of the proposed method. To this end, we summarized the average runtime to process one image in each dataset in Table 2. As can be seen, our private method is very efficient, taking only 66 milliseconds to sanitize a 1920 x 1080 image. In every dataset, the process time per pixel is around 10^{-5} milliseconds.



(a) PETS (b) Venice (c) AT&T (d) MNIST

Fig. 9: MSE vs. Varying ϵ



(a) PETS (b) Venice (c) AT&T (d) MNIST

Fig. 10: SSIM vs. Varying ϵ

5.5 Mitigation of CNN Attacks

While differential privacy provides a rigorous indistinguishability guarantee, we conducted a study similar to [14] in order to understand whether the differentially private pixelization can mitigate intelligent re-identification attacks. For this study, we partitioned the 10 images for each individual in the *AT&T* dataset (40 individuals in total) by randomly selecting 8 images for training and using the remaining 2 for testing, as in [14]. The *MNIST* dataset is pre-partitioned with 50,000 for training and 10,000 for testing. Assume the adversary has access to the training set obfuscated by a given method, as well as the label of each training image, i.e., individual identity (1-40) and digits (0-9). The goal of the re-identification attack is to breach the privacy of the testing set, i.e., predicting the label for each testing image produced by the same obfuscation method. In this study, we compared our differentially private pixelization with a random guessing baseline and the non-private pixelization method, i.e., mosaicing. Random guessing method predicts the label of a testing image by randomly picking a label, without considering the training set. Our method was applied with the default parameter values, i.e., $b = 16$ and $m = 16$, when varying ϵ. We generated the training set and testing set for each ϵ value.

A convolutional neural network (CNN) was trained for each dataset with the suggested architecture [14]. We reported the classification results[2] of our differentially private method in Table 3. The results for "Mosaicing" were taken from the original study [14]. As can be seen, with the same grid cell length $b = 16$, our differentially private method significantly reduces the attack success rate compared to the non-private method. For the *AT&T* dataset, recall that

[2] Top 1: the label predicted most likely was evaluated.

Dataset	Dimension	Time per image (in ms)
PETS	768 x 576	11.95
Venice	1920 x 1080	66.14
AT&T	92 x 112	0.32
MNIST	28 x 28	0.05

Table 2: Runtime of Differentially Private Pixelization

Dataset	Random Guess	Mosaicing[14]	DP Pixelization ($b = 16$)			
	-	16x16	$\epsilon = 0.1$	0.3	0.5	1
AT&T Top 1	2.50	96.25	3.75	18.75	43.75	77.50
MNIST Top 1	10.00	52.13	16.41	20.41	21.51	22.95

Table 3: Accuracy (in %) of CNN Re-Identification Attacks

with $\epsilon = 0.5$ the differentially private pixelization yields similar output to that of the non-private method as illustrated in Figure 4. But the re-identification risk is lowered by more than 52%, from 96.25% to 43.75%, thanks to the randomized mechanism. As for the *MNIST* dataset, our private method also significantly reduces the success rate of the attack. Dominated by black and white pixels and at a lower resolution, the re-identification risk of *MNIST* images is less sensitive to the privacy parameter ϵ. It is worth mentioning that when $\epsilon = 0.1$, our private method is very hard to breach, and the risk is close to that of random guessing.

6 Conclusion and Discussion

We have presented a private image pixelization method, which was the first attempt at extending differential privacy to image data publication. We proposed the m-neighborhood notion to define the indistinguishability requirement, i.e., roughly the same output for any images differing at up to m pixels. Given the high sensitivity of direct image perturbation, pixelization with grid cells of b x b pixels was adopted to achieve a utility-privacy trade off. We empirically evaluated the utility and efficiency of differentially private pixelization with multiple real-world image datasets, and showed that our private method can yield similar output to that of the non-private pixelization. In addition, an intelligent re-identification attack was simulated and the results showed that differentially private pixelization significantly reduces the attack success even at low privacy requirements, i.e., $\epsilon \geq 0.1$ and $m = 16$. Therefore, we concluded that our method is simple yet powerful.

As a new research endeavor, a number of directions can be explored for future work: 1) the design of post-processing techniques to further improve the utility of the differentially private method, e.g., removing sharp differences; 2) the study of application-specific utility such as crowd and vehicle counting; 3) the

evaluation of human users on the perceived privacy and utility; 4) the extension to correlated images, e.g., video frame sequences.

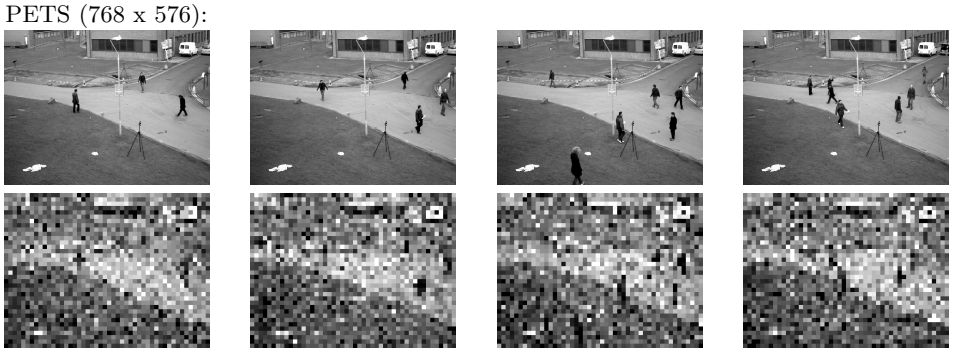
Acknowledgments. This research has been funded by NSF grant CRII-1755884 and a UAlbany FRAP-A Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the sponsors such as NSF.

References

1. InternetLiveStats. http://www.internetlivestats.com [Online; last accessed 21-March-2018].
2. TrafficLand. http://www.trafficland.com [Online; last accessed 21-March-2018].
3. Kamijo, S., Matsushita, Y., Ikeuchi, K., Sakauchi, M.: Traffic monitoring and accident detection at intersections. IEEE Transactions on Intelligent Transportation Systems **1**(2) (Jun 2000) 108–118
4. Sun, Q., Schiele, B., Fritz, M.: A domain based approach to social relation recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (July 2017)
5. Reece, A.G., Danforth, C.M.: Instagram photos reveal predictive markers of depression. EPJ Data Science **6**(1) (Aug 2017) 15
6. Yu, J., Zhang, B., Kuang, Z., Lin, D., Fan, J.: iprivacy: Image privacy protection by identifying sensitive objects via deep multi-task learning. IEEE Transactions on Information Forensics and Security **12**(5) (May 2017) 1005–1016
7. Ra, M.R., Govindan, R., Ortega, A.: P3: Toward privacy-preserving photo sharing. In: Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13), Lombard, IL, USENIX (2013) 515–528
8. He, J., Liu, B., Kong, D., Bao, X., Wang, N., Jin, H., Kesidis, G.: Puppies: Transformation-supported personalized privacy preserving partial image sharing. In: 2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). (June 2016) 359–370
9. Zhang, L., Jung, T., Liu, C., Ding, X., Li, X.Y., Liu, Y.: Pop: Privacy-preserving outsourced photo sharing and searching for mobile devices. In: 2015 IEEE 35th International Conference on Distributed Computing Systems. (June 2015) 308–317
10. Xia, Z., Wang, X., Zhang, L., Qin, Z., Sun, X., Ren, K.: A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. IEEE Transactions on Information Forensics and Security **11**(11) (Nov 2016) 2594–2608
11. Wang, S., Nassar, M., Atallah, M., Malluhi, Q. In: Secure and Private Outsourcing of Shape-Based Feature Extraction. Springer International Publishing, Cham (2013) 90–99
12. Wang, Q., Hu, S., Ren, K., Wang, J., Wang, Z., Du, M.: Catch me in the dark: Effective privacy-preserving outsourcing of feature extractions over image data. In: IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications. (April 2016) 1–9
13. Hill, S., Zhou, Z., Saul, L., Shacham, H.: On the (in) effectiveness of mosaicing and blurring as tools for document redaction. Proceedings on Privacy Enhancing Technologies **2016**(4) (2016) 403–417

14. McPherson, R., Shokri, R., Shmatikov, V.: Defeating image obfuscation with deep learning. CoRR **abs/1609.00408** (2016)
15. Dwork, C., McSherry, F., Nissim, K., Smith, A. In: Calibrating Noise to Sensitivity in Private Data Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg (2006) 265–284
16. Squicciarini, A.C., Caragea, C., Balakavi, R.: Analyzing images' privacy for the modern web. In: Proceedings of the 25th ACM Conference on Hypertext and Social Media. HT '14, New York, NY, USA, ACM (2014) 136–147
17. Spyromitros-Xioufis, E., Papadopoulos, S., Popescu, A., Kompatsiaris, Y.: Personalized privacy-aware image classification. In: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval. ICMR '16, New York, NY, USA, ACM (2016) 71–78
18. Stevens, R., Pudney, I.: Blur select faces with the updated blur faces tool (August 2017) [Online; posted 21-August-2017].
19. Dwork, C., Naor, M., Pitassi, T., Rothblum, G.N.: Differential privacy under continual observation. In: Proceedings of the forty-second ACM symposium on Theory of computing, ACM (2010) 715–724
20. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Local privacy and statistical minimax rates. In: 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. (Oct 2013) 429–438
21. Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K., Palamidessi, C.: Geo-indistinguishability: Differential privacy for location-based systems. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. CCS '13, New York, NY, USA, ACM (2013) 901–914
22. Jana, S., Narayanan, A., Shmatikov, V.: A scanner darkly: Protecting user privacy from perceptual applications. In: 2013 IEEE Symposium on Security and Privacy. (May 2013) 349–363
23. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: Proceedings of 1994 IEEE Workshop on Applications of Computer Vision. (Dec 1994) 138–142
24. Peterson, M.F., Eckstein, M.P.: Looking just below the eyes is optimal across face recognition tasks. Proceedings of the National Academy of Sciences **109**(48) (2012) E3314–E3323
25. Leal-Taixé, L., Milan, A., Reid, I., Roth, S., Schindler, K.: Motchallenge 2015: Towards a benchmark for multi-target tracking. arXiv preprint arXiv:1504.01942 (2015)
26. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11) (1998) 2278–2324
27. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4) (April 2004) 600–612

PETS (768 x 576):



Venice (1920 x 1080):

AT&T (92 x 112):

MNIST (28 x 28):

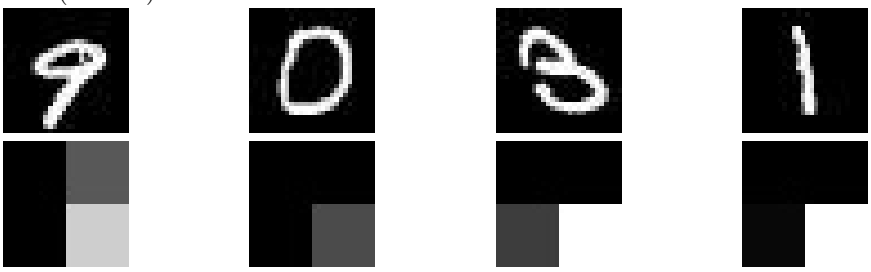


Table 4: First row lists sample images in each dataset and second row is the corresponding differentially private pixelization, under the default parameter setting. Note that when obfuscation is applied to faces in *AT&T* and digits in *MNIST*, the desired outcome is to reduce identifiability.