Differentially Private Imaging

Tao Li and Chris Clifton
Department of Computer Science, Purdue University, West Lafayette, Indiana, USA
{taoli,clifton}@purdue.edu

Abstract*—There is growing concern about image privacy due to the popularity of social media and photo devices, along with increasing use of face recognition systems. However, established image de-identification techniques are either too subject to re-identification, produce photos that are insufficiently realistic, or both. To tackle this, we present a novel approach for image obfuscation by manipulating latent spaces of an unconditionally trained generative model that is able to synthesize photo-realistic facial images of high resolution. This manipulation is done in a way that satisfies the formal privacy standard of local differential privacy. To our knowledge, this is the first approach to image privacy that satisfies ε-differential privacy *for the person.

I. Introduction

Image obfuscation techniques have been used to protect sensitive information, such as human faces and confidential texts. However, recent advances in machine learning, especially deep learning, make standard obfuscation methods such as pixelization and blurring less effective at protecting privacy; it has been showed that over 90% of blurred faces can be re-identified by commerical face recognition systems [1].

Many attempts have been made to obfuscate images and some privacy guarantees are provided. A pixelization method proposed in [2] satisfies pixel-wise ϵ-differential privacy. However, the utility of the pixelized images is far from satisfactory, as the images appear like traditional pixelization or blurring techniques. A more serious problem is that this provides differential privacy for *pixels*, not for the individuals pictured in the image, and thus is subject to re-identification of the individuals in the image [3]. Other obfuscation methods have been proposed recently to balance privacy and utility. For example, [4] adopted generative adversarial networks (GANs) for facial image obfuscation by identifying a face and accordingly inpainting it with a synthesized face alternative. This unfortunately has the potential to lose important characteristics of the original image; [5] leveraged a conditional GAN to manipulate facial attributions in accordance with different privacy requirements. These approaches suffer the common failing that they do not provide a *formal* privacy guarantee. As such, they may be subject to re-identification or re-construction attacks.

In this paper, we show how differential privacy can be provided at the level of the individual in the image. The key idea is that we transform the image into a semantic latent space. We then add random noise to the latent space representation in a way that satisfies ε-differential privacy. We then generate a new image from the privatized latent space representation. This ensures a formal privacy guarantee, while providing an image that preserves important characteristics of the original, and some level of photo-realism. A key to formal privacy methods is randomness: the same image privatized twice will not look the same (as demonstrated in fig. 1, which includes multiple images of some authors.), and it is a key component to prevent reconstruction attacks. We show that randomized manipulations in the latent semantic space can be expected to provide realistic images (e.g., fig. 2). The method guarantees similarity-based indistinguishability among images, providing privacy guarantees in worst-case scenarios and boosting the utility of the obfuscated image.

Fig. 1: Can you identify the authors? These are images of the authors, with noise added that satisfies differential privacy sufficient to prevent identification attacks.

II. Differentially Private Imaging

From the above, it should now seem obvious how we can get differential privacy: Add noise to the latent vector in a way that satisfies differential privacy. This leaves three questions to address in this section: What mechanism do we use to add noise? How much noise do we need to add? And how to obtain the latent vector in the first place? Answering these questions requires a better understanding of the latent space.

a) Latent Space and Image Encoding: It has been widely observed that there is linearity and continuity in the latent space of GAN [7] with vector arithmetic phenomenon such as addition and subtraction invariance [8]. Given a facial image, the problem of finding its corresponding latent representation can be considered as an optimization problem [1] where we search the latent space to find a latent vector, from which the reconstructed image is close enough (and hopefully identical) to the query image.

b) Privacy Mechanism: A key issue in using the Laplace mechanism for ε-differential privacy is determining the sensitivity. Inspired by [9], we use the maximum observed sensitivity to *clip* images in the latent space. Any values that fall outside the observed bounds are clipped to the observed bounds, guaranteeing that the range of the input to the differential privacy mechanism is known, allowing us to determine sensitivity.

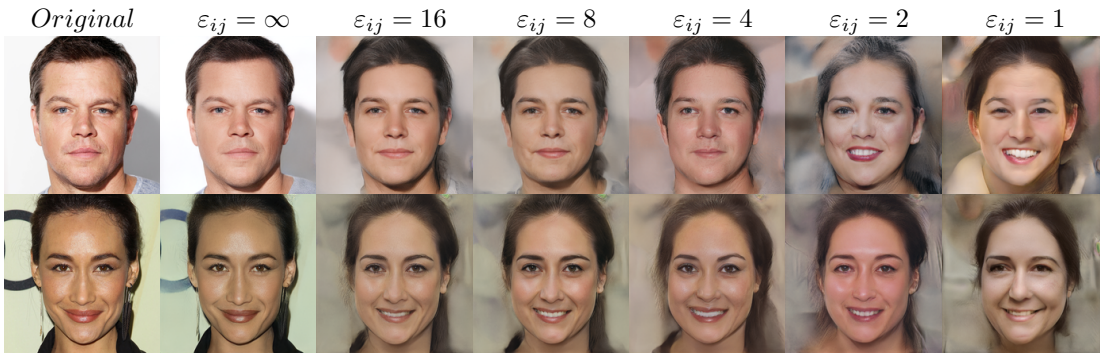

Fig. 2: Experimental results with different privacy budgets. In our experiment, latent codes are under a 25%/75% clipping setting [6] and the number of latent components is $18 \times 512 = 9216$, i.e., privacy budget $\varepsilon = \sum \varepsilon_{ij} = 9216 \cdot \varepsilon_{ij}$.

c) Algorithm: The idea of providing ε-local differential privacy is that the privacy budget ε is divided among the various components in the latent space. Each is used, along with the sensitivity derived from the clipping values for that component (based solely on the public training data), to determine a random draw of Laplace noise for that component, which is again clipped (a postprocessing step). This gives an ε-differentially private version of the image *in the latent space*. We then use this latent vector, with no reference to the original image or the latent space transformation of the original image, to generate an image using the previously described generative network. Algorithm 1 outlines the approach.

Theorem 1. Algorithm 1 provides ε-local differential privacy.

Proof. $\mathcal{M}$ is the randomized mechanism in algorithm 1. Using the notations in [10] and above, we have

$$\begin{aligned}
\frac{\Pr[\mathcal{M}(v, f, \varepsilon) = s]}{\Pr[\mathcal{M}(v', f, \varepsilon) = s]} &= \frac{\Pr[Lap(S_L \cdot w_f/\varepsilon)] = s - f(v)}{\Pr[Lap(S_L \cdot w_f/\varepsilon)] = s - f(v')} \\
&= \frac{\frac{S_L \cdot w_f}{\varepsilon} \cdot \exp(-\frac{|s - f(v)|\varepsilon}{S_L \cdot w_f})}{\frac{S_L}{\varepsilon} \cdot \exp(-\frac{|s - f(v')|\varepsilon}{S_L \cdot w_f})} \\
&= \exp(\frac{\varepsilon |f(v') - f(v)|}{S_L \cdot w_f}) \leq \exp(\varepsilon) \\
&= \exp(\frac{\varepsilon |f(v') - f(v)|}{S_L}) \leq \exp(\varepsilon \cdot w_f)
\end{aligned}$$

□

III. Conclusion

In this work, we provide the first meaningful formal definition of ε-differential privacy for images by leveraging the latent space of images and Laplace mechanism. A practical framework is presented to tackle real world images. Experimental results (e.g., fig. 2) show that the proposed mechanism is able to preserve privacy in accordance with privacy budget ε while maintain high perceptual quality for sufficiently large values of ε. We leave more analysis and results in the full version of this paper in [6].

Algorithm 1: DP Imaging with Laplace Mechanism

Require: Input image $X^{(i)}$;
Require: Encoder $f : \mathcal{X} \to \mathcal{Z}$;
Require: Generator $g : \mathcal{Z} \to \mathcal{X}$;
Require: Latent space sensitivities S_{Lj};
Require: Latent space weights w_j s.t. $\sum w_j = 1$;
Require: Privacy parameter ε;
Require: Laplace distribution $Lap(0, \lambda)$;
Require: Clipping function $f_c(i, j, \alpha)$;
latent vector $Z^{(i)} \leftarrow f(X^{(i)})$;
for each latent semantics $Z_j^{(i)}$ **do**
  obtain a random δ from $Lap(S_{Lj} \cdot w_j/\varepsilon)$;
  $Z_j'^{(i)} \leftarrow Z_j^{(i)} + \delta$;
  $Z_j''^{(i)} \leftarrow f_c(Z_j'^{(i)})$;
end for
desired noisy image $X'^{(i)} \leftarrow g(Z''^{(i)})$;

References

[1] T. Li and M. S. Choi, "DeepBlur: A simple and effective method for natural image obfuscation," *arXiv preprint arXiv:2104.02655*, 2021. 1

[2] L. Fan, "Image pixelization with differential privacy," in *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 2018, pp. 148–162. 1

[3] R. Gross, L. Sweeney, F. De la Torre, and S. Baker, "Model-based face de-identification," in *CVPR'06*. 1

[4] Q. Sun, L. Ma, S. Joon Oh, L. Van Gool, B. Schiele, and M. Fritz, "Natural and effective obfuscation by head inpainting," in *CVPR'18*. 1

[5] T. Li and L. Lin, "Anonymousnet: Natural face de-identification with measurable privacy," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019, pp. 56–65. 1

[6] T. Li and C. Clifton, "Differentially private imaging via latent space manipulation," *arXiv preprint arXiv:2103.05472*, 2021. 2

[7] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *CVPR'19*. 1

[8] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015. 1

[9] R. Chetty and J. N. Friedman, "A practical method to reduce privacy loss when disclosing statistics based on small samples," in *AEA Papers and Proceedings*, vol. 109, 2019, pp. 414–20. 1

[10] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *ACM CCS'14*. 2