Poster: 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.*

1 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 (e.g., [2, 3]) However, they do not provide a *formal* privacy guarantee.

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.

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

2 Differentially Private Imaging

Latent Space and Image Encoding It has been widely observed that there is linearity and continuity in the latent space of GAN [4] with vector arithmetic phenomenon such as addition and subtraction invariance [5]. 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. Figure 2 illustrates the optimization pipeline.

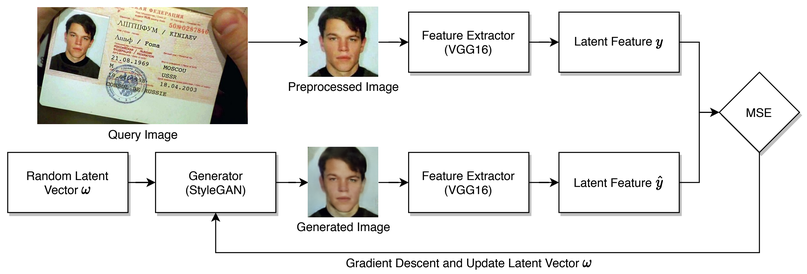

Figure 2: An optimization pipeline for GAN inversion.

Privacy Mechanism A key issue in using the Laplace mechanism for ε-differential privacy is determining the sensitivity. Inspired by [6], 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.

Figure 3: Clipping in the latent space. The three outputs (from left to right) are clipping at $0\%/100\%$, $15\%/85\%$, and $30\%/70\%$, respectively.

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*. 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 [7] and above, we have

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

□

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 $X'^{(i)} \leftarrow g(Z''^{(i)})$;

3 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. Experimental results 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 paper [8].

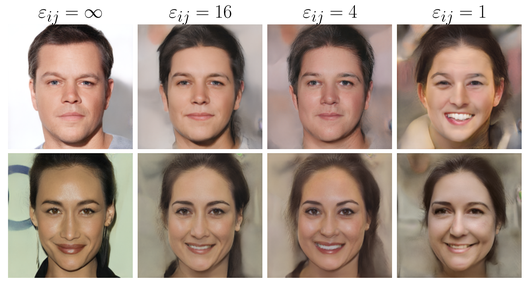

Figure 4: Experimental results with different privacy budgets. In our experiment, latent codes are under a 25%/75% clipping setting [8] and the number of latent components is $18 \times 512 = 9216$, i.e., privacy budget $\varepsilon = \sum \varepsilon_{ij} = 9216 \cdot \varepsilon_{ij}$.

References

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

[2] 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*.

[3] 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.

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

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

[6] 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.

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

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