Evaluating Provably Private Obfuscations for Eye and Face Images

Nick Reilly, Dr. Liyue Fan

Department of Computer Science, UNC Charlotte

Problem

Privacy of image data is a significant concern in today's society. A substantial amount of image data is continuously being collected, with much of it containing sensitive biometrics, such as face and iris, that can be used maliciously if in the hands of an adversary. Therefore, it is imperative to develop image privacy solutions which do not disclose sensitive information about participants. Many image obfuscation methods exist and are widely used (blurring, covering sensitive regions,) but these are primitive approaches, prone to inference attacks, and do not quantify privacy leakage.

The goal of this project is to provide a comparative analysis of state-of-the-art differential privacy mechanisms for protecting eye-tracking images and the creation of software tools to facilitate the adoption of image privacy.

Methodology and Data

In this project, we propose the use of differentially private mechanisms, namely Snow[1] and SamplingDP[2], to perform private image obfuscation and analyze the efficacy of these mechanisms against inference attacks and practical utility. Performance is evaluated using practical metrics for quantifying privacy and utility on the CASIA-IrisV2 dataset: for privacy evaluation, we will perform iris re-identification attacks on obfuscated images; for utility evaluation, we will consider specific tasks, such as gaze estimation, as well as perceptual quality measures.

Differential Privacy (DP). Differential privacy is a state-of-the-art notion for quantifying privacy leakage in sensitive datasets and has been adopted in large-scale by organizations such as Google, Apple, and the Census Bureau. The goal of differential privacy is to guarantee that the privacy of any individual participating in a dataset will not be at risk, regardless of any data that is available or may become available.

Privacy Mechanism - Snow

Snow Mechanism. Given grayscale image $I(x)$, where x denotes the index of each pixel in the image, and parameter p, we randomly select a subset of pixels $\mathcal{S}$ from I of size $p \cdot I_{width} \cdot I_{height}$. We create a new image $I'(x)$ such that:

$$I'(x) = \begin{cases} 127 & x \in \mathcal{S} \\ I(x) & x \notin \mathcal{S} \end{cases}$$

The fundamental idea of Snow is the introduction of noise to an image by randomly flipping the intensity of pixels to a constant value. The mechanism is simple to implement with low computation time, making it an appealing choice for researchers.

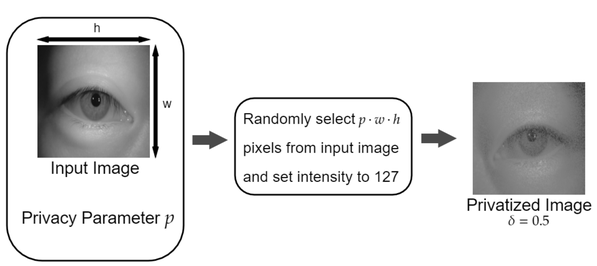

Fig. 1: Snow Mechanism

Privacy Mechanism - SamplingDP

SamplingDP Mechanism. Given grayscale image I, privacy budget ϵ, and selection rule k, SamplingDP strategically generates a private image that maximizes utility and exhausts the privacy budget. SamplingDP consists of four broad phases:

1. Select k representative intensities $\rightarrow$ 2. Budget allocation $\rightarrow$

3. Sample x_i pixels from each intensity $\rightarrow$ 4. Generate private image through interpolation

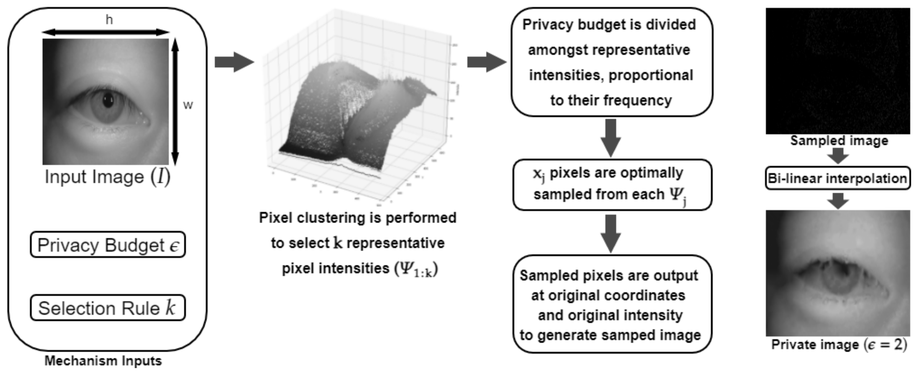

Fig. 2: SamplingDP Mechanism

Privacy Guarantees

Privacy Parameters. Two parameters exist that are central to the idea of differential privacy, ϵ and δ. ϵ is known as the privacy budget and bounds the sensitivity of the mechanism. The strongest form of privacy is ϵ-DP; however, achieving this level of privacy is not always possible and the addition of a δ value is required, establishing the need for (ϵ, δ)-DP. A mechanism that satisfies (ϵ, δ)-DP will achieve (ϵ)-DP with probability $(1-\delta)$. Lower values of these parameters give larger privacy guarantees.

Privacy Guarantee of Snow. We define two neighboring grayscale images, $\mathcal{X}$ and $\mathcal{X}'$ that differ by at most one pixel. The Snow mechanism $(\mathcal{M})$ satisfies $(0, \delta)$-DP if for every $\mathcal{X}$, $\mathcal{X}'$, and $O \in \text{range}(\mathcal{M})$:

$$P[\mathcal{M}(\mathcal{X}) = O] \leq P[\mathcal{M}(\mathcal{X}') = O] + \delta$$

It can be shown that constraining ϵ to 0 allows us to describe δ in terms of p: $\delta = 1 - p$. Thus, we see that for any desired δ, Snow can achieve $(0, \delta)$-DP.

Privacy Guarantee of SamplingDP. The privacy guarantee of SamplingDP lies in the pixel sampling phase. Given two neighboring images $\mathcal{I}$ and $\mathcal{I}'$ that differ by at most one pixel, it can be shown that the pixel sampling algorithm $(\mathcal{A})$ provides ϵ-DP if for any $\mathcal{I}$, $\mathcal{I}'$, and $O \in \text{range}(\mathcal{A})$:

$$\forall \theta_i \in \Psi, e^{-\epsilon(\theta_i)} \leq \frac{P[\mathcal{A}(\mathcal{I}(\theta_i)) = O(\theta_i)]}{P[\mathcal{A}(\mathcal{I}'(\theta_i)) = O(\theta_i)]} \leq e^{\epsilon(\theta_i)}$$

Where Ψ denotes the set of representative intensities, θ_i denotes the ith intensity of Ψ, and $\epsilon(\theta_i)$ denotes the privacy budget allocated to the θ_i.

Metrics

We employ various metrics to quantify the privacy and utility achieved by the mechanisms and use them to compare obfuscated image to their source. **Structural Similarity (SSIM)** measures percieved image quality between two images and uniquely considers texture/luminance. **Gaze Error** measures the difference in the predicted gaze of the two images. **Confident Pupil Rate** measures the proportion of obfuscated images who's pupil was confidently localized by a convolutional neural network. We employ an iris authentication algorithm to attack the obfuscated images and measure privacy. **Correct Recognition** Rate measures the proportion of obfuscated images that were correctly re-identified to a held-out target image belonging to the same participant.

Results

SamplingDP					
ϵ	**RMSE**	**SSIM**	**Gaze Error (°)**	**Confident Pupil Rate**	**Correct Recogniton Rate**
0.1	17.594	0.789	16.109	4.47%	0%
0.3	13.79	0.801	9.741	21.9%	0%
0.5	12.578	0.807	6.242	38.27%	0.26%
0.7	11.796	0.81	5.876	50.55%	0.26%
1	11.147	0.813	5.088	56.63%	0.58%
3	9.464	0.831	3.3	74.78%	1.05%
5	8.847	0.84	2.797	81.11%	1.84%
7	8.459	0.847	2.378	82.89%	4.83%
10	8.093	0.854	2.21	85.78%	10.68%

Table 1: SamplingDP Utility and Attack Evaluations

Snow					
δ	**RMSE**	**SSIM**	**Gaze Error (°)**	**Confident Pupil Rate**	**Correct Recogniton Rate**
0.33	36.436	0.368	4.424	30.09%	0%
0.4	34.365	0.366	2.256	61.95%	0%
0.5	31.372	0.373	1.908	80.09%	0%
0.55	29.76	0.38	1.582	83.63%	0%
0.6	28.056	0.39	1.285	86.5%	0%
0.65	26.243	0.404	1.179	88.27%	0.79%
0.7	24.301	0.421	0.953	91.59%	3.68%
0.75	22.182	0.444	0.802	92.48%	6.87%
0.8	19.839	0.474	0.546	94.47%	11.7%
0.85	17.182	0.517	0.507	95.8%	23.34%

Table 2: Snow Utility and Attack Evaluations

In both tables, blue and orange columns highlight utility and privacy metrics respectively. We see that as ϵ increases, utility is monotonically increasing while privacy monotonically decreases. It is important to note that the privacy parameters between the two mechanisms are not directly comparable (i.e. $\epsilon = 0.5 \neq \delta = 0.5$). We note that the values of δ required to achieve any utility with Snow are widely regarded as poor privacy, while SamplingDP is able to achieve good utility with respectable values of ϵ.

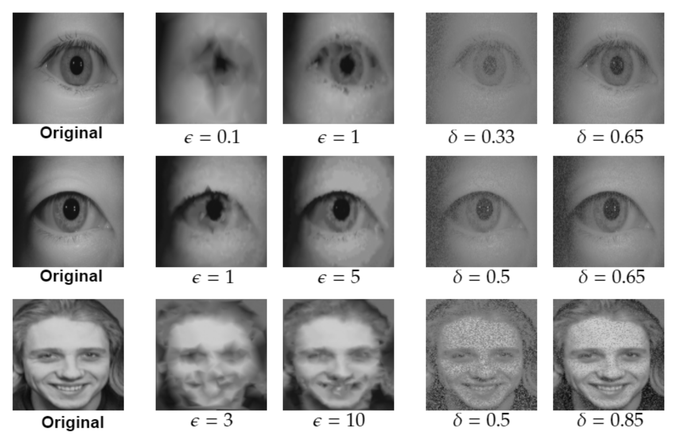

Fig. 3: Visual Results. second/third columns: SamplingDP, fourth/fifth: Snow.

[1] Brendan John et al. "Let It Snow: Adding Pixel Noise to Protect the User's Identity". In: ETRA '20 Adjunct (2020). DOI: 10.1145/3379157.3390512. URL: https://doi.org/10.1145/3379157.3390512.

[2] Han Wang, Shangyu Xie, and Yuan Hong. *VideoDP: A Universal Platform for Video Analytics with Differential Privacy*. 2019. arXiv: 1909.08729 [cs.CR].