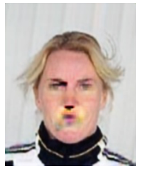
Generative Modeling for Context-Aware Local Privacy

Wei-Ning Chen, Wei-Chen Chen, Dennis Rich

Problem Setup

data X → Privatization Channel W → perturbed data $\tilde{X}$



Goal: erase sensitive info S (smiling or not) from X while minimizing distortion $d(X, \tilde{X})$

Contribution

Related works

- Local differential privacy [1]: need data distribution, poor privacy-utility tradeoff
- Detect and perturb [2]: add random noise to sensitive parts. May not achieve minimum distortion.

Our method

- Decentralized: trust no one even data collector
- Data driven: **do not** need data distribution
- Optimize distortion given privacy constraint

Learning Algorithm

Objectives

$$\min_{W} \quad \mathbb{E}_{P_{X,\tilde{X}}} d(X, \tilde{X})$$
$$\text{s.t.} \quad \underbrace{\| P_{\tilde{X}|S=s_1} \circ W - P_{\tilde{X}|S=s_2} \circ W \|_{\mathrm{TV}} \leq \epsilon}_{\text{Bayesian error of sensitive info. } \geq 1/2-\epsilon}.$$

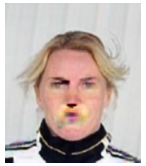
Reformulate as GAN (dual form)

$$\min_{W} \mathbb{E} d(X, \tilde{X}) + \lambda \left(\max_{f_\omega} \mathbb{E}_{P_{\tilde{X}|s_1}} f_\omega(\tilde{X}) - \mathbb{E}_{P_{\tilde{X}|s_2}} f_\omega(\tilde{X}) \right)$$

- λ: privacy level
- W: privatization channel (generator)
- f_ω: discriminator

Training Architecture

→ Auto-encoder →



Privatizer

CNN

{smiling, not smiling}

Sensitive label

Discriminator

Loss function: pixel-wise MSE $-\lambda$ CNN's cross-entropy

Results

Ground Truth

$\lambda = 0.03$

$\lambda = 0.04$

$\lambda = 0.05$

Privacy

Utility

Conclusion and Future Works

- Proposed a data driven framework for context-aware privacy
- Achieved better privacy-utility trade-off with theoretical guarantee on privacy
- Can easily incorporate other utility measures
- What next: characterize the fundamental limits on privacy-utility trade-off curves

[1] John Duchi et. al, "Local privacy and statistical mini-max rates", FOCS 2013
[2] Hsiang Hsu et. al, "Discovering information-leaking samples and features ", NeurIPS 2019
[3] video available at https://youtu.be/URXQr8STJL0