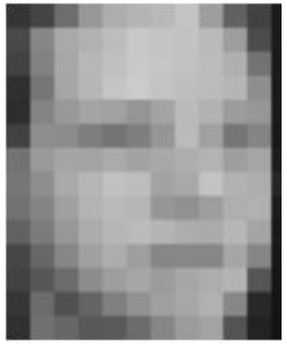
DP-Shield: Face Obfuscation with Differential Privacy

Muhammad Saleem
Department of Computer Science, UNC Charlotte
msaleem2@uncc.edu

Saleem, Reilly and Fan. "DP-Shield: Face Obfuscation with Differential Privacy". In EDBT'22.

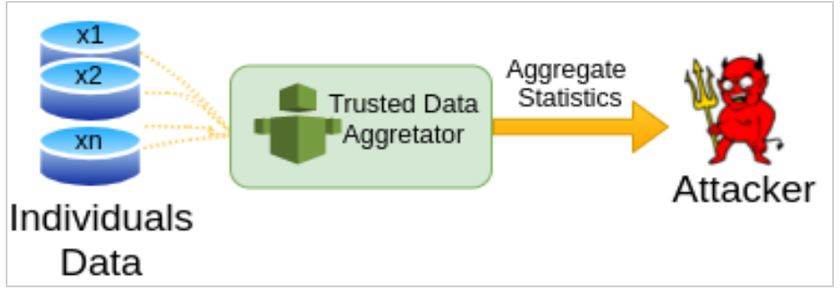
Motivation

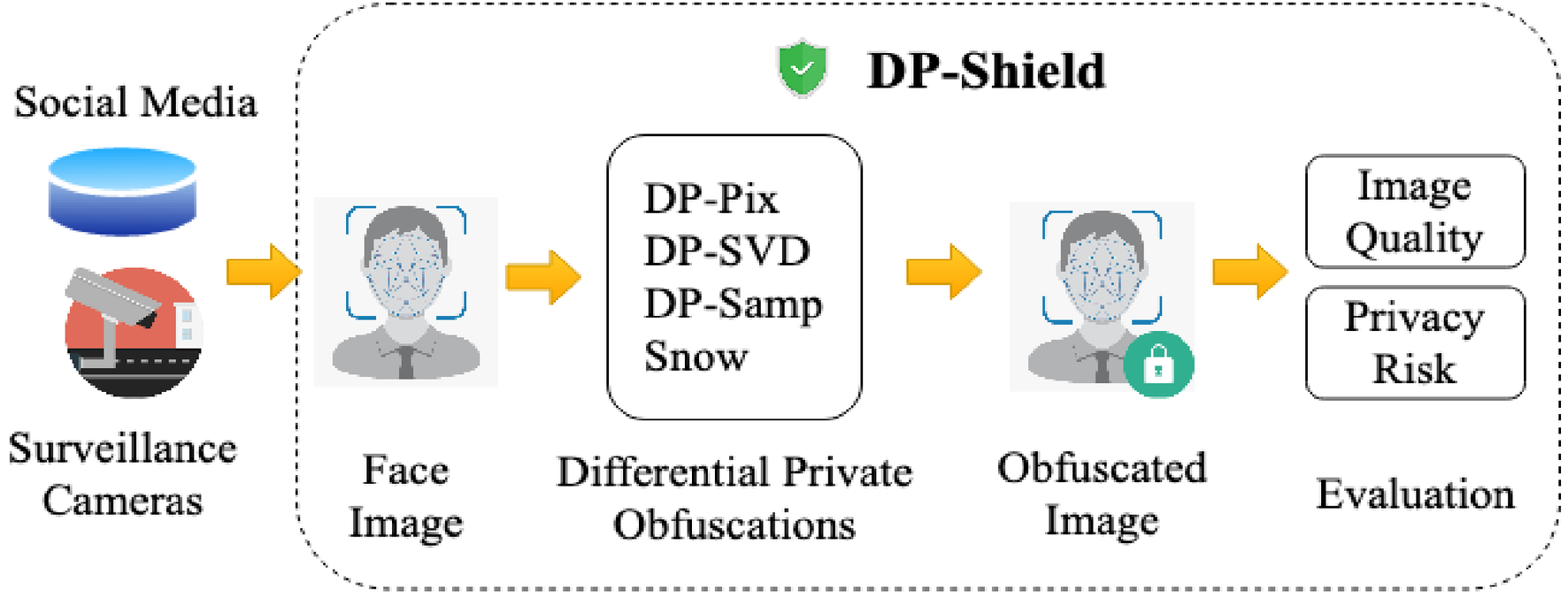
- Immense amount of image data gathered from various sources
  - Surveillance cameras, social media, sensors, …
- Various research investigations have been made possible by image data e.g., Computer Vision, Social Science, ...
- Images may reveal sensitive information about:
  - Individuality
  - Lifestyle
- Standard approaches to obfuscating images
  - For example, pixelization, gaussian noise, blurring, GAN inpainting, mosaiced, redaction etc.
    - Visually hide sensitive information
    - Do not provide quantifiable privacy guarantees
    - Prone to re-identification attacks
  - CNN model re-identify up to 96% obfuscated faces [McPherson et al. 2016]



DP-SHIELD Interactive Framework

- Differential privacy (DP) [Dwork et al. 2006] is the state-of-the-art notion for sanitizing aggregate databases queries.
  - Provides protection to Indvidual's data
  - Quantifying privacy leakage and privacy to be bounded
  - Robust against inference attacks [Fan 2018] [Fan 2019]





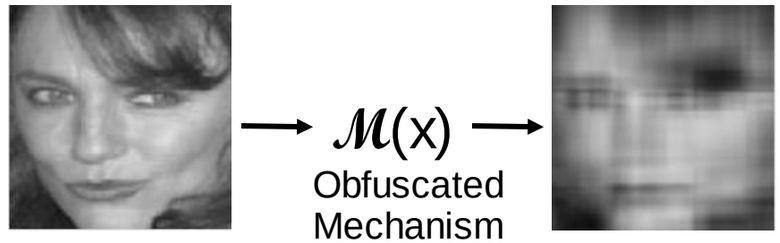
Evaluation Methodology

- Face Datasets
  - VGGFace2 [Cao et al. 2018]
  - CASIA-WEBFACE [Yi et al. 2014]

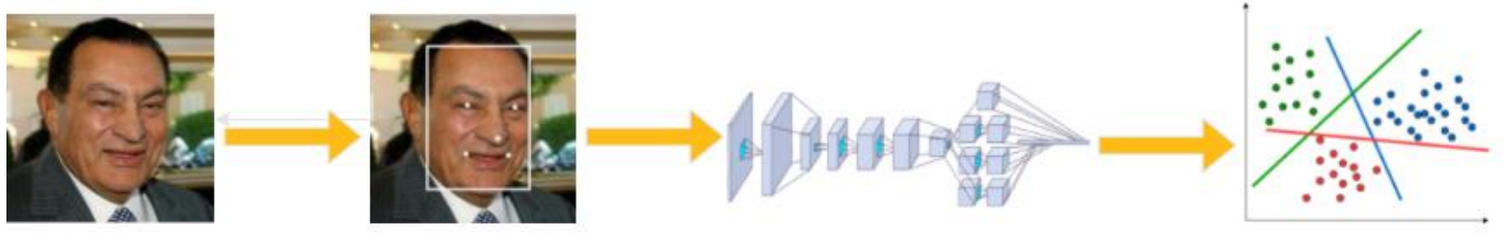
VGGFace2

CASIA-WEBFACE

- Image Quality Measures
  - Structural Similarity (SSIM) [Wang et al. 2004] and Mean Squared Error

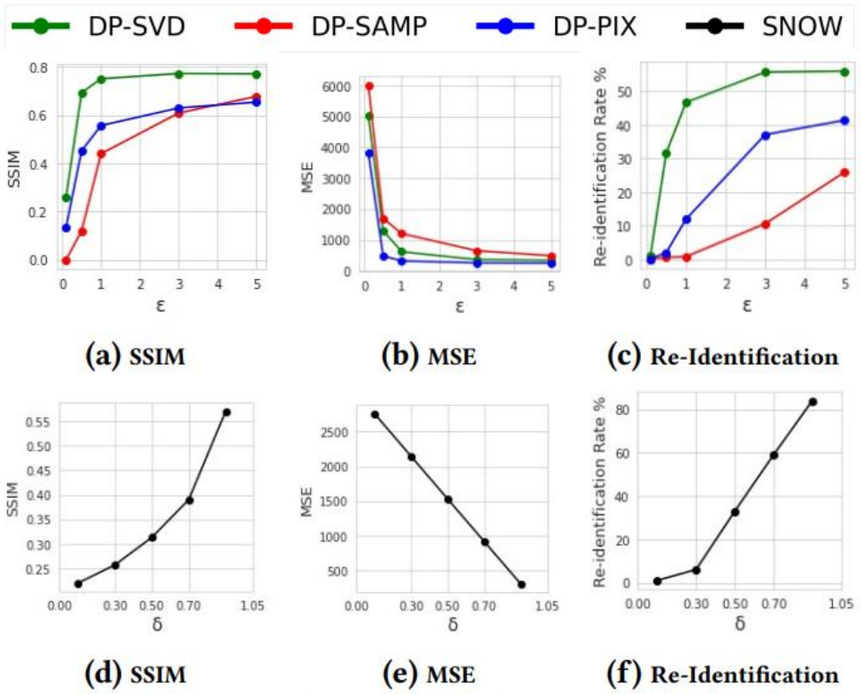


- Empirical Privacy Protection
  - FaceNet with Inception ResNet (V1) network [Schroff et al. 2016]
  - Train SVC classifier on the FaceNet embeddings to re-identify obfuscated image



Evaluation

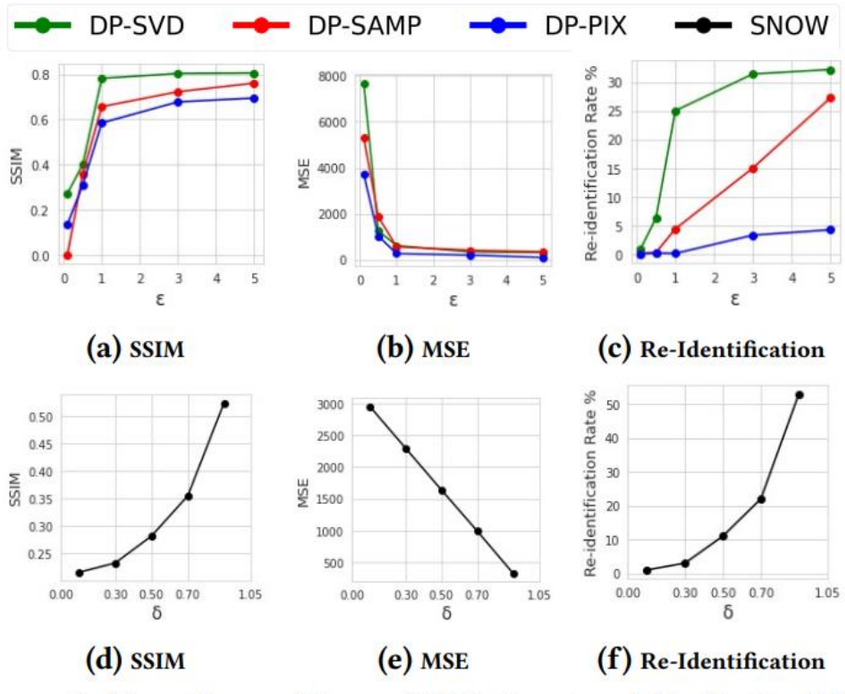
- ϵ and δ indicate level of privacy protection, and lower values indicate stronger privacy protection
  - Strong DP protection may lead to low image quality (low SSIM and high MSE)
- Privacy protection (Re-identification Rate)
  - Weaker DP protection means higher privacy risks and obfuscated image may disclose identity of individual.
  - Method design also matters in providing privacy protection



(a) SSIM (b) MSE (c) Re-Identification

(d) SSIM (e) MSE (f) Re-Identification

Figure 2: Mean Squared Error (MSE), Structural Similarity (SSIM) and Re-Identification Rate results on VGGFace2 dataset.



(a) SSIM (b) MSE (c) Re-Identification

(d) SSIM (e) MSE (f) Re-Identification

Figure 3: Mean Squared Error (MSE), Structural Similarity (SSIM) and Re-Identification Rate results on CASIA-WebFace dataset.

Questions?

Contact Muhammad (msaleem2@uncc.edu)