AnonymousNet: Natural Face De-Identification with Measurable Privacy

Tao Li
Department of Computer Science
Purdue University
taoli@purdue.edu

Lei Lin
Goergen Institute for Data Science
University of Rochester
lei.lin@rochester.edu

Abstract

With billions of personal images being generated from social media and cameras of all sorts on a daily basis, security and privacy are unprecedentedly challenged. Although extensive attempts have been made, existing face image de-identification techniques are either insufficient in photo-reality or incapable of balancing privacy and usability qualitatively and quantitatively, i.e., they fail to answer counterfactual questions such as “is it private now?”, “how private is it?”, and “can it be more private?” In this paper, we propose a novel framework called AnonymousNet, with an effort to address these issues systematically, balance usability, and enhance privacy in a natural and measurable manner. The framework encompasses four stages: facial attribute estimation, privacy-metric-oriented face obfuscation, directed natural image synthesis, and adversarial perturbation. Not only do we achieve the state-of-the-arts in terms of image quality and attribute prediction accuracy, we are also the first to show that facial privacy is measurable, can be factorized, and accordingly be manipulated in a photo-realistic fashion to fulfill different requirements and application scenarios. Experiments further demonstrate the effectiveness of the proposed framework.

1. Introduction

The deployment of internet of things devices, such as surveillance cameras and sensors, are rising dramatically in recent years. Particularly, the popularity of smartphones allows billions of photos being uploaded to social networks and shared among people on a daily basis. Although the blooming development has advanced machine learning applications to bring convenience and enhance user experience, it may capture confidential information incidentally and increases the risks of privacy leakage. To protect privacy, the most straightforward approach is access control [51], such as the “Restrict Others” setting in Facebook [31]. Cryptography techniques such as encryption and secure multi-party computation [28] can also be applied to mitigate privacy threats. In computer vision community, privacy-enhancing technologies are mainly obfuscation-based; for example, obfuscating sensitive information like faces and numbers in an image by using traditional approaches including blurring, pixelation, and masking (see Figure 2). However, there are at least two drawbacks with these traditional approaches. First, researchers have shown these techniques are vulnerable. The faceless person recognition system proposed in [43] can be trained with limited training samples and then identify the target in an obfuscated image by using body features. Another study also shows that deep learning models can successfully identify faces in images encrypted with these techniques with high accuracies [38]. Second, images processed with these techniques result in unsatisfying perception in general. A study shows that both blurring and blocking will impact image perception scores, and even lower scores are observed for images obfuscated by blocking [31].

On the other hand, new techniques and mechanisms are being applied to enhance image obfuscation. A game the-
ory framework called Adversarial Image Perturbation has been proposed to determine effective image obfuscation approach when the choice of countermeasures is unknown [44]. Recently, the generative adversarial networks (GAN) can generate realistic natural images following original input data distribution via adversarial training [11], therefore it has become more and more popular for novel image obfuscation techniques. Wu et al. [60] developed a GAN with two newly designed modules verificator and regulator for face de-identification. Considering subjects in social media photos appear in diverse activities and head orientations, Sun et al. [52] proposed a two-stage model to inpaint the head region conditioned on facial landmarks.

Note that there exists a tradeoff between privacy protection and dataset usability [63, 14]. High obfuscation levels fail to preserve utility for various tasks while low obfuscation levels lead to recognition of private information. Unfortunately, current methods are unable to find a way to quantify this matter; neither can they be adapted with correspondence to various privacy metrics nor real-world scenarios under different requirement settings. To tackle these, we propose the AnonymousNet, a four-stage frameworks consisted of facial semantic extraction powered by a deep Convolutional Neural Network [23]; a attribute selection method with regards to privacy metrics such as $k$-anonymity [54], $l$-diverse [37], and $t$-closeness [24]; a Generative Neural Network [11] for photo-realistic image generator; and a universal adversarial perturbation [40] to mitigate potential security and privacy threats.

The rest of the paper is organized as follows: Section 2 provides necessary background in privacy-preserving data mining and reviews recent advances in facial image editing; Section 3 formalizes the face de-identification problem and introduce privacy metrics; Section 4 outlines the four-stage AnonymousNet framework, including facial feature extraction, semantic-based attribute obfuscation, de-identified face generation, and adversarial perturbation; Section 5 details experiment settings and evaluate the results; we conclude this paper in Section 6 by discussions of future research directions.

2. Related Work

Facial Landmark Detection  Photo privacy protection includes two important aspects: sensitive content detection and obfuscation method [30]. Facial information is one of the highest sensitive contents. Various head poses, orientations, lighting conditions and so on make photo privacy protection more challenging. Facial landmark detection has been an active research field because it is the first important step for other facial-related applications. In the past few decades, numerous algorithms like support vector machine [46] and random forecast [2] have been proposed. Despite these models have made significant progress, they rely on handcrafted local features and may not operate under a wide range of conditions like illumination, occultation, poses.

Recently deep learning models have made huge advances to deal with issues like occultation and variability in facial landmark detection. [53] proposed a cascaded convolutional network for facial keypoint detection which can utilize the texture context information over the entire face
and encode geometric constraints implicitly. Further, a deep multi-task learning framework is presented that combines facial landmark detection with correlated tasks like head pose estimation [64]. Dong et al. (2018) adopted a generative adversarial network to transform facial images into style-aggregated ones, which are then deployed together to train a facial landmark detector [8].

**Image Inpainting** Previous study shows that comparing with image blurring and blocking, image inpainting provides a more effective and better user experience [30]. Face replacement has become a popular image inpainting technology for privacy protection. [4] created a large face library from public internet. Given an input face, the most similar one in the library will be chosen for face replacement. [39] applied a 2D morphable model to adjust the shape of a source face to match the target face. [33] proposed a framework to generate a personalized 3D head model from one frontal face. The 3D head model can then be rendered at any pose to swap with the face in the target image. Figure 2 compares several classic image processing methods for face identity obfuscation.

**GAN-based Face Generation** Generative Adversarial Network (GAN) is a system of two neural networks who contests with each other under a zero-sum game setting. GAN was first introduced by Goodfellow et al. [11] in 2014. Since then, great progresses have been made, as shown in Figure 3: in 2015, Radford et al. [45] designed deep convolutional generative adversarial networks (DCGANs) which demonstrated the adversarial pair (both the generator and discriminator) can learn a hierarchy of representations from object parts to scenes; Liu et al. [34] proposed coupled generative adversarial network (CoGAN) in 2016, which is capable of learning a joint distribution with only samples from marginal distributions and without tuple of corresponding images from different domains; Karras et al. [20] described a new training method for GAN in 2017, whose main idea is to train generator and discriminator progressively by starting from a low resolution and adding new layers of the model with fine details incrementally; and more recently, Karras et al. [21] proposed a style-based generator architecture (StyleGAN) that is able to learn high-level facial attributes in a automated and unsupervised manner, generates images with stochastic variations, and achieves the state-of-the-art.

**Privacy-Preserving Data Mining** Theories and practices of privacy-preserving techniques have been studied extensively in the database and data mining communities; to name a few: [1, 6, 56]. To measure the disclosure risk of an anonymized table of a database, Samarati & Sweeney (1998) [50] introduced the $k$-anonymity property such that each record in the database is indistinguishable with at least $k - 1$ records. Although being sufficient to protect against identity disclosure, $k$-anonymity is limited to prevent attribute disclosure. In regards to that, Machanavajjhala et al. (2006) [37] introduced a new privacy property called $l$-diversity, which requires protected attributes to have at least $l$ well-represented values in each equivalence class. Newton et al. (2005) [42] introduced the first privacy-enabling algorithm, $k$-Same, to the context of image databases. Gross et al. (2005) [13] demonstrated a tradeoff between disclosure risk (i.e., image obfuscation level) and classification accuracy. To tackle this, they introduced $k$-Same-Select to balance privacy and usability. Zhang (2018) [63] further designed a OBFUSCATE function that adds random noises to existing samples or creates new samples, in an effort to hide sensitive information in the dataset while preserving model accuracy.

3. Preliminary

3.1. Formulation of Face De-Identification

We first formally define the problem of face de-identification, extending notations and definitions from Newton et al. [42]. This process is important as it helps us to precisely define the methods and to build a solid foundation for following discussions of theoretical properties.

**Definition 3.1. (Face Image).** A face image (or simple ‘face’ or ‘image’) is a 3D matrix $I$ of $m$ columns, $n$ rows, and $c$ channels. $c$ is usually 3 in common color spaces (e.g., RGB and YUV). Each cell in $I$ stores a color coding for a pixel, ranging from 0 to 255 inclusively. A face image contains a normalized image of only one person’s face.

**Definition 3.2. (Face Set).** A face set is a set of $M$ face images: $\{\Gamma_i : i = 1, \ldots, M\}$.

**Definition 3.3. (Person-Specific Face Set).** Let $H$ be a face set of $M$ face images, $\{\Gamma_1, \ldots, \Gamma_M\}$. $H$ is said to be...
Definition 3.4. (Face De-Identification Function). Let $\mathbf{H}$ and $\mathbf{H}_d$ be person-specific face set.

$$f : \mathbf{H} \rightarrow \mathbf{H}_d$$

is called face de-identification function if it attempts to obscurce the identity of the original face image.

Definition 3.5. (De-Identified Face). Given $\Gamma \in \mathbf{H}$ and de-identification function $f$, $\Gamma_d \in \mathbf{H}_d$ is a $f$ de-identified face of $\Gamma$ if

$$\Gamma_d = f(\Gamma)$$

Figure 2 illustrates several canonical image processing methods for face de-identification, including blurring, pixelation, masking, deep Convolutional Neural Network-based style transfer [19] (abstract painting style, portrait painting style, and cartoon style [15]), and Generative Adversarial Network-based image inpainting [61].

3.2. Privacy Metrics

The motivation of privacy metrics is to provide qualitative and quantitative measurement of the degree of privacy enjoyed by specific users (personal images in our case) and the proper amount of privacy protection offered with correspondence to trade-offs in usability. This section, we only discuss metrics that are used in our framework. [57] provides a more comprehensive list for interested readers.

$k$-Anonymity $k$-anonymity is a widely applied metric to evaluate a dataset’s level of anonymity [50]. It requires that each record in the dataset is indistinguishable with at least $k - 1$ other records with respect to quasi-identifiers, which refer to attributes that can potentially be taken together to identify an individual like zip code or birth date [24]. In the case of a face dataset, these quasi-identifiers may include semantic attributes like eyeglasses, pointy nose, and oval face, and so on. If a dataset satisfies the condition of $k$-Anonymity, with only quasi-identifiers of one individual known, the true record can only be chosen with a probability of $1/k$.

However, there are some scenarios that $k$-anonymity cannot provide enough protection. For example, for $k$ subjects that have the same values for the quasi-identifiers, if they all have the same sensitive attribute like heart disease, an adversary can be certain that the target identify in the $k$ subjects must have heart disease. The $k$-Anonymity thus fails to protect sensitive information from the homogeneity attack [24].

$l$-Diversity $l$-diversity is proposed to address the limitations of $k$-anonymity [37]. The basic concept is that for the equivalence class representing a set of records with the same values for the quasi-identifiers, it should have at least $l$ “well-represented” values for the sensitive attribute.

The most straightforward definition of “well-represented” values is to ensure the equivalence class has $l$ distinct values for the sensitive attribute. In this definition, the frequencies of $l$ distinct values are not considered. An adversary may conclude that the sensitive attribute of a targeted identity has the value with the highest frequency. Therefore there is a stronger definition of $l$-diversity named Entropy $l$-diversity, which is defined as follows:

$$\text{Entropy}(E) \geq \log l$$

$$\text{Entropy}(E) = -\sum_{s \in S} p(E, s) \log p(E, s)$$

where $E$ is the equivalence class, $S$ is the value set of the sensitive attribute, and $p(E, s)$ is the fraction of records in $E$ that have sensitive value $s$.

$t$-Closeness Adversaries sometimes have knowledge of the global distribution of sensitive attributes, for example, the distributions of facial attributes are easy to obtain (see Figure 6). To prevent privacy disclosure by an adversary with such knowledge, [24] introduced $t$-closeness, which updates $k$-anonymity with correspondence to the distribution of sensitive values, requiring that the distribution $S_E$ of sensitive values in any equivalence class $E$ must be close to their distribution $S$ in the entire database, i.e.,

$$\forall E : d(S, S_E) \leq t$$

where $d(S, S_E)$ is the distance between distribution $S$ and $S_E$ measured by the Earth Mover Distance [47] and $t$ is the privacy threshold at which $d(S, S_E)$ should not exceed.

Randomness Randomization is another approach to protect data privacy. It is realized by adding random noise to existing samples [63]. Given an individual sample, we can randomly select part of its features with a ratio of $\gamma$, and add Gaussian noise $N(0, \sigma)$. Instead of random feature selection, we can also identify sensitive features first, and add Gaussian noise $N(0, \sigma)$. Another randomization approach is to introduce new samples into the dataset [63]. To generate a new sample, we can randomly pick an original sample from the dataset, revise its feature value and add small amount of noise at the same time. For example, the value $x_i$ of pixel $i$ in an original image can be replaced with $255 - x_i$ with Gaussian noise $N(0, \sigma)$ added in the new sample. From a broader perspective, adversarial perturbation can also be consider as a randomization method.
4. The AnonymousNet

In an effort to obfuscate facial identities and generate photo-realistic alternatives, balance privacy and usability qualitatively and quantitatively, answer counterfactual questions such as “is it private now?”, “how private is it?”, and “can it be more/less private?”, and finally achieve controllable and measurable privacy, we propose the AnonymousNet framework, which encompasses four stages: facial feature extraction, semantic-based attribute obfuscation, de-identified face generation, and adversarial perturbation, as detailed below.

Stage-I: Facial Attribute Prediction

We adopt GoogLeNet [55] for facial attribute extraction, which consists of 22 layers with 9 Inception blocks and excelled in ImageNet Large-Scale Visual Recognition Challenge 2014 [49]. Unlike most other classifiers in ImageNet [7], GoogLeNet is not trained for one label, but rather is fed with 40 facial attributes at the same time and outputs multiple classification results accordingly. GoogLeNet further leverages a trick of using \(1 \times 1\) kernels to increase depth while reducing dimension and reserving computational resources. Figure 4 outlines the facial attribute prediction pipeline, where labelled images are first fed into the model, features are extracted from the fully connected (FC) layer, and then 40 random forest classifiers [32] are trained and facial attributes are subsequently obtained.

Stage-II: Privacy-Aware Face Obfuscation

Provided with semantic information of each facial image as well as attribute distribution over the entire database (see Figure 6), we are one step closer towards our goal: face de-identification with privacy guarantees. We first consider a toy example: database consist of 2 identities, both of which share a common attribute <Male>. In this case, modifying the gender attribute will not change the level of privacy, since altering this attribute will not change the possibility of guessing the identity given gender. Consider another example that for the same database and each entity has three boolean attributes: <Male, BigNose, BlackHair> and one identity has black hair and the other does not. For this case, both of them should be updated to either black hair or non-black hair since the identity will be revealed if the hair color is known. These two example provide us insights of how to select facial attributes such that \(k\)-anonymity and \(l\)-diversity are satisfied.

However, as discussed in Section 3.2, this may not be enough to protect privacy, since sensitive information that reveals identities can still be revealed by exploiting global distributions of the attributes [24]. Here, we propose the Privacy-Preserving Attribute Selection (PPAS) Algorithm, a method to select and update facial attributes such that the distribution \(S_E\) of any attribute \(E\) is close to its real world distribution \(\bar{S}\) subject to constraint defined in Equation 5. Unlike normal \(t\)-closeness, we further introduce a stochastic perturbation in the attribute selection process working toward \(\epsilon\)-differential privacy [9]. Our approach is formalized in Algorithm 1 (for binary attributes).

![Algorithm 1: The PPAS algorithm.](#)
Stage-III: Natural and Directed De-Identification

To obfuscate facial images while preserve visual reality, we adopt a generative adversarial network (GAN) [11], which is designed as two players, $D$ and $G$, playing a min-max game with adversarial loss:

$$L_{adv} = \mathbb{E} \log(D(x)) + \mathbb{E} \log(1 - D(G(x)))$$  \hspace{1cm} (6)

where generator $G$ is trained to fool discriminator $D$, who tries to distinguish real images from adversarial ones. GAN has been successful in many applications [21, 29] yet is notoriously difficult to train [27]. As the face de-identification task can be categorized as a image-to-image translation problem, we customize the GAN model based on StarGAN [5], which has been widely demonstrated to have high usability along with realistic results, and the keys of which are attribute classification loss $L_{cls}$ and image reconstruction loss $L_{rec}$. $L_{adv}$, $L_{cls}$ and $L_{rec}$ forms the final objective function:

$$L = \lambda_1 L_{adv} + \lambda_2 L_{cls} + \lambda_3 L_{rec}$$  \hspace{1cm} (7)

Stage-IV: Adversarial Perturbation

Adding a Gaussian noise to an image is generally considered as a simple and effective way to trick deep neural network-based detectors which have been showcased to be vulnerable against this attack. There are also approaches that carefully craft perturbations for each data point. For example, [12] propose a fast gradient sign method to perturb an input through one step of gradient ascent. Furthermore, [40] shows there exists a universal adversarial perturbation that can cause images misclassified with high probability by state-of-the-art deep neural networks. The basic idea is formulated as follows: Suppose $\mu$ is the a distribution of images in $\mathbb{R}^d$, $\hat{k}$ is a classifier that the output given an image $x$ is $\hat{k}(x)$. The universal perturbation vector $v \in \mathbb{R}^d$ that can fool the classifier $\hat{k}$ should satisfy:

$$\|v\|_p \leq \xi$$  \hspace{1cm} (8)

$$\mathbb{P}_{x \sim \mu}(\hat{k}(x + v) \neq \hat{k}(x) \geq 1 - \delta)$$  \hspace{1cm} (9)

where $\xi$ limits the size of the universal perturbation vector, $\delta$ quantifies the failure rate of all the adversarial samples.

In this work, we introduce a universal perturbation vector as identified through an iterative approach. For each iteration $i$, we apply DeepFool [41] to identify the minimal perturbation to let $\hat{k}$ misclassify each input, and update the universal perturbation corresponding to hyperparameter $\epsilon_i$ to the total perturbation $v$. It is shown that the algorithm works on a small portion of images sampled from the training dataset, and the universal perturbation generalizes well with respect to the data and the network architectures. Figure 5 outlines the proposed perturbation pipeline.

Figure 5: An example of adversarial perturbation. In stage-IV, we introduce a small universal perturbation adjusted by parameter $\epsilon$ to synthesized images, tricking malicious detectors while preserving perceptual integrity.

5. Experiment

5.1. Dataset

We use Large-scale CelebFaces Attributes (CelebA) Dataset [36] for facial attribute estimation, which contains 202,599 images and 10,177 identities. Each image has 40 attributes labels of boolean values, and their distributions have been shown in Figure 6. Following the same protocol as described in [36], we split the dataset into three folders: 160,000 images (8,000 identities) for training; 20,000 images (1,000 identities) for validation; and the rest 20,000 images (1,000 identities) for testing.

5.2. Image Preprocessing

Before feeding the data into our deep models, we perform data preprocessing for each images in the datasets, including steps in order: face detection, landmark detection, alignment, and image cropping. To obtain face landmarks, we deploy a Deep Alignment Network (DAN) [22], which is a deep neural network of multiple stages and has demonstrated convincing performance even in extreme pose or lighting conditions in the wild. Based on the 68 landmarks provided by DAN for each image, we align and center the face in the image by calibrating positions of both left and right eyes. Subsequently, we crop the images to a size of $256 \times 256$. Figure 7 illustrates the preprocessing pipeline.

Figure 7: Image preprocessing pipeline. We deploy a Deep Alignment Network (DAN) [22] to obtain facial landmarks, based on which we accordingly align faces and crop images.
5.3. Training

Attribute Estimation. As discussed in Section 4 earlier, a GoogLeNet [55] is deployed for attribute estimation, using the same training settings as [26]. We adopt sigmoid cross-entropy as the loss function:

$$L = -\frac{1}{n} \sum [y \ln a + (1 - y) \ln(1 - a)]$$  \hspace{1cm} (10)

where $y$ is label and $a$ is output. When training, we use a base learning rate of $10^{-5}$, which is reduced by a polynomial decay with a gamma of 0.5. Momentum is set to be 0.9 and the weight decay is $2 \times 10^{-4}$. $6 \times 10^5$ iterations with a batch size of 64 are conducted for the training. After extracting deep features from the FC layer, we train 40 random forest classifiers for attribute estimation and achieve an accuracy that is comparable to current state-of-the-art [48].

Attribute Translation. After obtaining facial attributes that satisfies privacy constraints computed from previous steps, we employ StarGAN [5] for face attributes translation and use CeleA [36] as the training set. We follow the settings in [5], where Wasserstein loss [3] is adopted to expedite the training process and the definition is as below:

$$L_{adv} = \mathbb{E}_x[D_{src}(X)] - \mathbb{E}_{x,c}[D_{src}(G(x, c))]$$  \hspace{1cm} (11)

$$= -\lambda_g \mathbb{E}_{\hat{x}}[(||\nabla_{\hat{x}} D(\hat{x})||^2 - 1)^2]$$  \hspace{1cm} (12)

where $\hat{x}$ is uniformly sampled between a pair of original and synthetic images and we use $\lambda = 10$ here. In terms of network architecture, we adopt a generator network of two convolutional layers for downsampling, six residual blocks [16], and two transposed convolutional layers for upsampling; and we use PatchGANs [18] as the discriminator network for binary classification (i.e., whether certain image patches are fake or real).

5.4. Evaluation

Figure 1 compares our approach with canonical face obfuscation methods, showcasing a significant improvement in photo-reality and a success in face de-identification. Figure 8 illustrates experimental results in pairs, where left is the original image and right is the result generated by our framework. It demonstrates that face identities are preserved in a perceptually natural manner, and in the meantime, each pair of images still shares certain common attributes in correspondence with various privacy policies and application scenarios (for example, pair (1, 4) differs in hair color, lip color, gender, and age, but shares the same eye sizes; pair (1, 1) only differs in gender and leaves other semantic attributes identical). Also, artifacts introduced by the proposed the adversarial perturbation are shown to be visually negligible, even though they have been upscaled for illustrative purposes. Table 1 compare image quality under widely used perceptual metrics. Being aware that these model-based metrics fail to capture many nuances of human perception [62] and a smaller or larger value does not necessarily imply higher or lower image quality [25], we list the results here only for reference purpose.

<table>
<thead>
<tr>
<th></th>
<th>Blurring</th>
<th>Pixelation</th>
<th>Masking</th>
<th>Inpainting</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM [58]</td>
<td>0.8281</td>
<td>0.8226</td>
<td>0.8762</td>
<td>0.8020</td>
<td>0.7984</td>
</tr>
<tr>
<td>MS-SSIM [59]</td>
<td>0.8842</td>
<td>0.8784</td>
<td>0.8204</td>
<td>0.7659</td>
<td>0.8650</td>
</tr>
</tbody>
</table>

Table 1: Image quality comparison under different metrics.

6. Conclusion and Future Work

To naturally obfuscate face identities while preserves privacy in a controllable and measurable manner, we proposed the AnonymousNet framework, which consists of four stages: facial feature extraction, semantic-based at-

![Figure 6: Facial attributes and their distributions in the CeleA dataset [36].](image-url)
Figure 8: Some experimental results. In each pair, left is the original image and right is the synthesized result with an altered identity. The results show that face identities are preserved in a perceptually natural manner, and in the meantime, each pair of images still shares certain common attributes in correspondence with various privacy policies and application scenarios. Furthermore, artifacts introduced by the proposed adversarial perturbation are shown to be visually negligible (the perturbation has been intentionally upscaled here for illustrative purposes).

tribute obfuscation, de-identified face generation, and adversarial perturbation. This framework successfully overcomes the shortages of existing methods - being able to generate photo-realistic images with fake identity and capable of balancing privacy and usability both qualitatively and quantitatively. It also answers questions such as “is it private now?”, “how private is it?”, and “can it be more/less private?” counterfactually. Considering the threats from adversaries, especially the malicious detectors that are prevalent in today’s Internet, we further introduced a universal adversarial perturbation so as to trick other deep neural networks as much as possible. Experimental results support the effectiveness of our approach by showing photo-realistic results with negligible artifacts.

In the future, we would like to evaluate perturbation performance among different deep neural network-based detectors qualitatively and quantitatively, which is ignored here due limitations in space and computational resources. We are also interested in robustness, scalability, and extensibility of this framework under various real-world settings. In the experiments, we find that different facial attributes vary in “distinguish power”, i.e., attributes such as Age and Gender are perceptually more powerful in helping distinguish an identity than Cheekbones, Height, which align with our intuitions. This advises a future research direction that a user study can be made to explore these differences qualitatively and quantitatively, and in the end, figure out “the crux of facial indistinguishability”.

Acknowledgment

The authors especially thank Professor Chris Clifton for insightful discussions in differential privacy and privacy metrics in the context of facial images. We are also grateful for Professor Fengqing Maggie Zhu and anonymous reviewers from the CV-COPS’19 program committee for their comments and helpful suggestions. The idea of this paper was born when the first author writing [35] during his research internship at ObEN, Inc. He thanks colleagues at ObEN for the wonderful times they shared.

References


