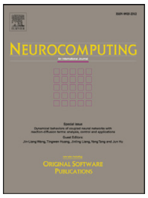
Neurocomputing 384 (2020) 94–103

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



AutoGAN-based dimension reduction for privacy preservation

Hung Nguyen[a], Di Zhuang[a], Pei-Yuan Wu[b,*], Morris Chang[a]

[a] University of South Florida, USA
[b] National Taiwan University, Taiwan

ARTICLE INFO

Article history:
Received 1 March 2019
Revised 23 September 2019
Accepted 1 December 2019
Available online 4 December 2019

Communicated by Dr. Yongsheng Dong

Keywords:
Generative adversarial nets
Auto-encoder
Neural-network
Privacy preservation
Dimension reduction
Access control

ABSTRACT

Protecting sensitive information against data exploiting attacks is an emerging research area in data mining. Over the past, several different methods have been introduced to protect individual privacy from such attacks while maximizing data-utility of the application. However, these existing techniques are not sufficient to effectively protect data owner privacy, especially in the scenarios that utilize visualizable data (e.g. images, videos) or the applications that require heavy computations for implementation. To address these problems, we propose a new dimension reduction-based method for privacy preservation. Our method generates dimension-reduced data for performing machine learning tasks and prevents a strong adversary from reconstructing the original data. We first introduce a theoretical approach to evaluate dimension reduction-based privacy preserving mechanisms, then propose a non-linear dimension reduction framework motivated by state-of-the-art neural network structures for privacy preservation. We conducted experiments over three different face image datasets (AT&T, YaleB, and CelebA), and the results show that when the number of dimensions is reduced to seven, we can achieve the accuracies of 79%, 80%, and 73% respectively and the reconstructed images are not recognizable to naked human eyes.

1. Introduction

Machine Learning (ML) is an important aspect of modern applications that rely on big data analytics (e.g., an on-line system collecting data from multiple data owners). However, these applications are progressively raising many different privacy issues as they collect different types of data on a daily basis. For example, many types of data are being collected in smart cities such as patient records, salary information, biological characteristics, Internet access history, personal images and so on. These types of data then can be widely used in daily recommendation systems, business data analysis, or disease prediction systems which in turn affect the privacy of individuals who contributed their sensitive data. Considering a multi-level access control system of a company using biometric recognition (e.g., face recognition, fingerprint) for granting permission to access data resources, the company staff members may concern their biological information being vulnerable to adversaries. Even though the utility of these biometric features can be effectively used in machine learning tasks for authentication purpose, leaking this information might lead to privacy breaches. For example, an adversary could utilize them to determine the members' identities.

Several tools and methods have been developed to preserve the privacy in machine learning applications, such as homomorphic encryption [1–3], secure multi-party computing [4,5], differential privacy (DP) [6–10], compressive privacy [11–17] and so on. Typically, differential privacy-based methods aim at preventing leaking individual information caused by queries. However, they are not designed to serve large number of queries since they require adding huge amount of noise to preserve privacy, thus significantly decreasing the ability to learn meaningful information from data. On the other hand, homomorphic encryption-based methods can be used to privately evaluate a function over encrypted data by a third party without accessing to plain-text data, hence the privacy of data owners can be protected. However, due to the high computational cost and time consumption, they may not work with a very large dataset, normally required in ML applications.

In this study, we consider an access control system collecting dimension-reduced face images of staff members to perform authentication task and to provide permission for members who would like to access company's data resources (Fig. 1). We propose a non-linear dimension reduction framework to decrease data dimension for the authentication purpose mentioned above and to protect against an adversary from reconstructing member images. Firstly, we introduce ϵ-DR Privacy as a theoretical tool for dimension reduction privacy evaluation. It evaluates the reconstruction

[*] Corresponding author.
E-mail addresses: nsh@mail.usf.edu (H. Nguyen), zhuangdi1990@gmail.com (D. Zhuang), peiyuanwu@ntu.edu.tw (P.-Y. Wu), morrisjchang@gmail.com (M. Chang).

https://doi.org/10.1016/j.neucom.2019.12.002
0925-2312/© 2019 Elsevier B.V. All rights reserved.

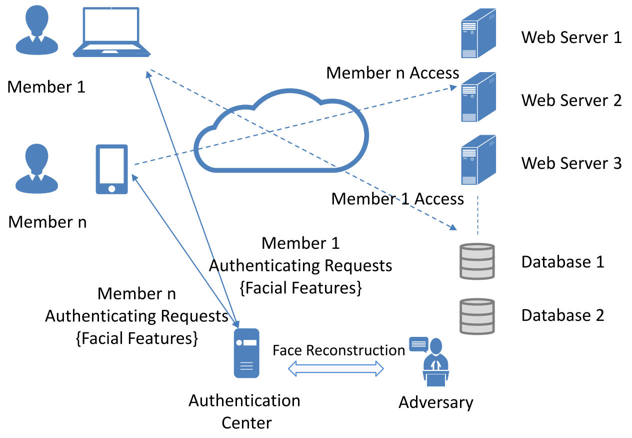


Fig. 1. Attack model.

distance between original data and reconstructed data of a dimension reduction (DR) mechanism. This approach encourages a DR mechanism to enlarge the distance as high distance yields high level of privacy. While other methods such as differential privacy-based methods rely on inference uncertainty to protect sensitive data, ϵ-DR Privacy is built on reconstruction error to evaluate privacy. Therefore, unlike differential privacy methods, ϵ-DR Privacy is not negatively impacted by the number of queries. Secondly, as detailed in Section 3, we recommend a privacy-preserving framework Autoencoder Generative Adversarial Nets-based Dimension Reduction Privacy (AutoGAN-DRP) for enhancing data owner privacy and preserving data utility. The *utility* herein is evaluated via machine learning task performance (e.g., classification accuracy).

Our dimension reduction (DR) framework can be applied to different types of data and used in several practical applications without heavy computation of encryption and impact of query number. The proposed framework can be applied directly to the access control system mentioned above. More elaboratively, face images are locally collected, nonlinearly compressed to achieve DR, and sent to the authentication center. The server then performs classification tasks on the dimension-reduced data. We assume the authentication server is semi-honest, that is to say it does not deviate from authenticating protocols while being curious about a specific member's identity. Our DR framework is designed to resist against reconstruction attacks from a strong adversary who obtains the training dataset and the transformation model.

During the stage of experiments, we implemented our framework to evaluate dimension-reduced data in terms of accuracy of the classification tasks, and we attempted to reconstruct original images to examine the capacity of adversaries. We performed several experiments on three facial image datasets in both gray-scale and color, i.e., *the Extended Yale Face Database B* [18], *AT&T* [19], and *CelebFaces Attributes Dataset (CelebA)* [20]. The experiment results illustrate that with only seven reduced dimensions our method can achieve accuracies of 93%, 90%, and 80% for AT&T, YaleB, and CelebA respectively. Further, our experiments show that at the accuracies of 79%, 80% and 73% respectively, the reconstructed images could not be recognized by human eyes. In addition, the comparisons shown in Section 6 also illustrate that AutoGAN-DRP is more resilient to reconstruction attacks compared to related works.

Our work has two main contributions:

- To analytically support privacy guarantee, we introduce ϵ-DR Privacy as a theoretical approach to evaluate privacy preserving mechanism.
- We propose a non-linear dimension reduction framework for privacy preservation motivated by Generative Adversarial Nets [21] and Auto-encoder Nets [22].

The rest of our paper is organized as follows. Section 2 summarizes state-of-the-art privacy preservation machine learning (PPML) techniques and reviews knowledge of deep learning methods including generative adversarial neural nets and Auto-encoder. Section 3 describes the privacy problem through a scenario of a facial recognition access control system, introduces the definition of ϵ-DR Privacy to evaluate DR-based privacy preserving mechanisms, and presents our framework AutoGAN-based Dimension Reduction for Privacy Preservation. Section 4 presents and discusses our experiment results over three different face image datasets. Section 5 compares AutoGAN-DRP to a similar work GAP in terms of reconstruction error and classification accuracy. Section 6 demonstrates reconstructed images over AutoGAN-DRP and other privacy preservation techniques (i.e., Differential Privacy and Principle Component Analysis). Finally, the conclusion and future work are mentioned in Section 7.

2. Related work

2.1. Literature review

Cryptographic approach: This approach usually applies to the scenarios where the data owners do not wish to expose their plain-text sensitive data while asking for machine learning services from a third-party. The most common tool used in this approach is fully homomorphic encryption that supports multiplication and addition operations over encrypted data, which enabling the ability to perform a more complex function. However, the high cost of the multiplicative homomorphic operations renders it difficult to be applied on machine learning tasks. In order to avoid multiplicative homomorphic operations, additive homomorphic encryption schemes are more widely used in privacy preserving machine learning (PPML). However, the limitation of the computational capacity in additive homomorphic schemes narrows the ability to apply on particular ML techniques. Thus, such additive homomorphic encryption-based methods in [1,2,23,24] are only applicable to simple machine learning algorithms such as decision tree and naive bayes. In Hesamifard's work [3],the fully homomorphic encryption is applied to perform deep neural networks over encrypted data, where the non-linear activation functions are approximated by polynomials.

In secure multi-party computing (SMC), multiple parties collaborate to compute functions without revealing plain-text to other parties. A widely-used tool in SMC is garbled circuit [4], a cryptographic protocol carefully designed for two-party computation, in which they can jointly evaluate a function over their sensitive data without the trust of each other. In [25], Mohammad introduced a SMC protocol for principle component analysis (PCA) which is a hybrid system utilizing additive homomorphic and garbled circuit. In secret sharing techniques [5], a secret **s** is distributed over multiple pieces **n** also called *shares*, where the secret can only be recovered by a sufficient amount of **t** *shares*. A good review of secret sharing-based techniques and encryption-based techniques for PPML is given in [26]. Although these encryption-based techniques can protect the privacy in particular scenarios, their computational cost is a significant concern. Furthermore, as [26] elaborated, the high communication cost also poses a big concern for both techniques.

Non-cryptographic approach: Differential Privacy (DP) [27] aims to prevent membership inference attacks. DP considers a scenario that an adversary infers a member's information based on the difference of outputs of a ML mechanism before and after the member join a database. The database with the member's information and without the member's information can be considered as two neighbor databases which differ by at most one element. DP adds noise to the outputs of the ML mechanism to result in similar outputs from the two neighbor databases. Thus, adversaries cannot differentiate the difference between the two databases. A mechanism M satisfies ϵ-differential privacy if for any two neighbor databases D and D', and any subset S of the output space of M satisfies $Pr[M(D) \in S] \leq e^{\epsilon} Pr[M(D') \in S]$. The similarity of query outputs protects a member information from such membership inference attacks. The *similarity* is guaranteed by the parameter ϵ in a mechanism in which the smaller ϵ provides a better level of privacy preservation. [6,7,28,29] propose methods to guarantee ϵ-differential privacy by adding noise to outcome of the weights $w^* = w + \eta$, where η drawn from Laplacian distribution and adding noise to the objective function of logistic regression or linear regression models. [8,9] satisfy differential privacy by adding noise to the objective function while training a deep neural network using stochastic gradient descent as the optimization algorithm.

In addition, there are existing works proposing differential privacy dimension reduction. One can guarantee ϵ-differential privacy by perturbing dimension reduction outcome. Principal component analysis (PCA) whose output is a set of eigenvectors is a popular method in dimension reduction. The original data is then represented by its projection on those eigenvectors, which keeps the largest variance of the data. One can reduce the data dimension by eliminating insignificant eigenvectors which contain less variance, and apply noise on the outcome to achieve differential privacy[10]. However, the downside of these methods is that they are designed for specific mechanisms and datasets and not working well with the others. For example, record-level differential privacy is not effectively used with image dataset as shown in [30]. Also, the amount of added noise is accumulative based on the number of queries so that this approach usually leads to low accuracy results with a high number of queries.

Similar to our work, Generative Adversarial Privacy (GAP) [12] is a perturbation method utilizing the minimax algorithm of Generative Adversarial Nets to preserve privacy and to keep utility of image datasets. GAP perturbs data within a specific l_2 distance constraint between original and perturbed data to distort private class labels and at the same time preserve non-private class labels. However, it does not protect the images themselves, and an adversary can visually infer private label (e.g., identity) from images. In contrast, our method protects an image by compressing it into a few dimension vector and then transferring without clearly exposing the original image.

2.2. Preliminaries

To enhance the distance between original and reconstructed data in our DR system, we utilize the structure of Generative Adversarial Network (GAN) [21] for data perturbation and deep Auto-encoder [22] for data reconstruction. The following sections briefly review Auto-encoder and GAN.

2.2.1. Auto-encoder

Auto-encoder is aimed at learning lower dimension representations of unsupervised data. Auto-encoder can be used for denoising and reducing data dimension. It can be implemented by two neural network components: *encoder* and *decoder*. The *encoder* and *decoder* perform reverse operations. The input of the *encoder* is the original data while the output of the *decoder* is expected to be similar to the input data. The middle layer extracts latent representation of original data that could be used for dimension reduction. An Auto-encoder training process can be described as a minimization problem of the auto-encoder's loss function $\mathcal{L}(\cdot)$:

$$\mathcal{L}(x, g(f(x))) \tag{1}$$

where x is input data, f(·) is an encoding function, and g(·) is a decoding function.

2.2.2. GAN

Generative Adversarial Nets is aimed at approximating distribution p_d of a dataset via a generative model. GAN simultaneously trains two components *generator G* and *discriminator D*, and the input of G is sampled from a prior distribution $p_z(z)$ through which G generates fake samples similar to the real samples. At the same time, D is trained to differentiate between fake samples and real samples, and send feedback to G for improvement. GAN can be formed as a two-player minimax game with value function V(G,D):

$$\min_G \max_D V(G, D) = E_{x \sim p_d}[log(D(x))] + E_{z \sim p_z}[log(1 - D(G(z)))] \tag{2}$$

The two components, *Generator* and *Discriminator* can be built from neural networks (e.g., fully connected neural network, convolutional neural network). The goal of G is to reduce the accuracy of D. Meanwhile, the goal of D is to differentiate fake samples from real samples. These two components are trained until the discriminator cannot distinguish between generated samples and real samples.

3. Methodology

In this section, we first describe the problem and threat model, then we introduce a definition of DR-Privacy and our dimensionality reduction method (AutoGAN-DRP).

3.1. Problem statement

We introduce the problem through the practical scenario mentioned in Section 1. Fig. 1 briefly describes the entire system in which staff members (clients) in a company request access to company resources, such as websites and data servers through a face recognition access control system. For example, if member n requests to access web server 2, the local device first takes a facial photo of the member by an attached camera, locally transforms it into lower dimension data, and sends to an authentication center. The authentication server then obtains the low dimensional data and determines member access eligibility by using a classifier without clear face images of the requesting member. We consider that the system has three levels of privileges (i.e., single level, four-level, eight-level) corresponding to three groups of members. We assume the authentication server is semi-honest (it obeys work procedure but might be used to infer personal information). If the server is compromised, an adversary in the authentication center can reconstruct the face features to achieve plain-text face images and determine members' identity.

3.2. Threat model

In the above scenario, we consider that a strong adversary who has access to the model and training dataset attempts to reconstruct the original face images for inferring a specific member's identity. Our attack model can be represented in Fig. 1. The adversary utilizes training data and facial features to identify a member identity by reconstructing the original face images using a reconstructor in an auto-encoder. Rather than using fully connected neural network, we implement the auto-encoder by convolutional neural network which more effective for image datasets. Our goal is to design a data dimension reduction method for reducing data dimension and resisting full reconstruction of original data.

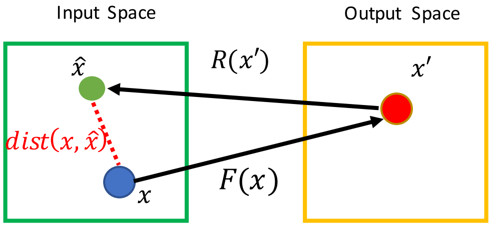


Fig. 2. DR projection and reconstruction.

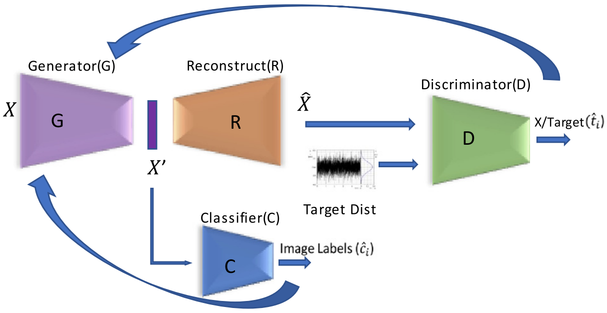


Fig. 3. AutoGAN-DRP.

3.3. ϵ-Dimension Reduction privacy (ϵ-DR Privacy)

We introduce the Dimension Reduction Privacy (DR-Privacy), and define a formal definition of the ϵ-DR Privacy to mathematically quantify/evaluate the mechanisms designed to preserve the DR-Privacy via dimension reduction. The DR-Privacy aims to achieve privacy-preserving via dimension reduction, which refers to transforming the data into a lower dimensional subspace, such that the private information is concealed while the underlying probabilistic characteristics are preserved, which can be utilized for machine learning purposes. To quantify the DR-Privacy and guide us to design such DR functions, we define ϵ-DR Privacy as follows.

Definition 1 ((. ϵ-DR Privacy)) A Dimension Reduction Function $F(\cdot)$ satisfies ϵ-DR Privacy if for each i.i.d. m-dimension input sample x drawn from the same distribution D, and for a certain distance measure $dist(\cdot)$, we have

$$\mathbb{E}[dist(x,\hat{x})] \geq \epsilon \tag{3}$$

where $\mathbb{E}[\cdot]$ is the expectation, $\epsilon \geq 0$, $x' = F(x)$, $\hat{x} = R(x')$, and $R(\cdot)$ is the Reconstruction Function.

For instance, as shown in Fig. 2, given original data x, our framework utilizes certain dimension reduction function $F(x)$ to transform the original data x into the transformed data x'. The adversaries aim to design a corresponding reconstruction function $R(x')$ such that the reconstructed data $\hat{x}$ would be closed/similar to the original data x. DR-Privacy aims to design/develop such dimension reduction functions, that the distance between the original data and its reconstructed data would be large enough to protect the privacy of the data owner.

3.4. AutoGAN-based dimension reduction for privacy preserving (autoGAN-DRP)

We propose a deep learning framework for transforming face images to low dimensional data which is hard to be fully reconstructed. The framework can be presented in Fig. 3. We leverage the structure of an auto-encoder [22] which contains encoder and decoder (in this work, we called them generator and reconstructor) in order to reduce data dimension. More specifically, the low dimensional representations are extracted from the middle layer of the auto-encoder (the output of the generator). The dimension-reduced data can be sent to the authentication server as an authentication request. We consider an adversary as a reconstructor implemented by a decoder. To resist against fully reconstructing images, the framework utilizes a discriminator in GAN [21] to direct reconstructed data to a designated target distribution with an assumption that the target distribution is different from our data distribution. In this work, the target distribution is sampled from Gaussian distribution and the mean is the average of training data. After projecting data into a lower dimension domain, the re-constructor is only able to partially reconstruct the data. Therefore, the adversary might not be able to recognize an individual's identity. To maintain data utility, we also use feedback from a classifier. The entire framework is designed to enlarge the distance between original data and its reconstruction to preserve individual privacy and retain significant data information. The dimension-reduced transformation model is extracted from the framework and provided to clients for reducing their face image dimensions. The classification model will be used in an authentication center that classifies whether a member's request is valid to have access (1) or not (0).

We formulate the problem as follows: Let X be the public training dataset. (x_i, y_i) is the ith sample in the dataset in which each sample x_i has d features and a ground truth label y_i. The system is aimed at learning a dimension reduction transformation $F(\cdot)$ which transforms the data from d dimensions to d' dimensions in which $d' \ll d$. Let X' be the dataset in lower dimension domain. The dimension-reduced data should keep significant information to work with different types of machine learning tasks and should resist against the reconstruction or inference from data owner information.

Our proposed framework is designed to learn a DR function $F(\cdot)$ that projects data onto low dimension space and preserves privacy at certain value of ϵ. The larger distance implies higher level of privacy. Fig. 3 presents our learning system in which the dimension-reduced data X' is given by a generator G. Since X' is expected to be accurately classified by a classifier C, the generator improves by receiving feedback from the classifier via the classifier's loss function $\mathcal{L}_C$. We use a binary classifier for single-level authentication system and multi-class classifiers for multi-level authentication system. The classifier loss function is defined as the cross entropy loss of the ground truth label y and predicted label $\hat{y}$ as follows.

$$\mathcal{L}_C = -\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}\log(\hat{y}_{ij}) \tag{4}$$

where m denotes the number of classes and n denotes the number of samples.

To evaluate data reconstruction and enlarge the reconstruction distance, a re-constructor R is trained as a decoder in an auto-encoder and sends feedback to the generator via its loss function $\mathcal{L}_R$. The re-constructor plays its role as an aggressive adversary attempting to reconstruct original data by using known data. The loss function of R is the mean square error of original training data (x) and reconstructed data $(\hat{x})$, as displayed in (5).

$$\mathcal{L}_R = \sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{5}$$

To direct the reconstructed data to a direction that reveals less visual information, the generator is trained with a discriminator D as a minimax game in GAN. The motivation is to direct reconstructed data to a certain target distribution (e.g., normal distribution). To ensure a distance, the target distribution should be different to training data distribution. The discriminator aims to differentiate the reconstructed data from samples of the target distribution. The loss function of D $(\mathcal{L}_D)$ can be defined as a cross-entropy loss of ground truth labels (0 or 1) t and prediction labels $\hat{t}$ shown in (6).

$$\mathcal{L}_D = -\sum_{i=1}^{n} (t_i \log(\hat{t}_i) + (1 - t_i) \log(1 - \hat{t}_i)) \tag{6}$$

The optimal generator parameter θ^* is given by the optimization problem of the generator loss function $\mathcal{L}_G$:

$$\underset{\theta}{minimize}\ \mathcal{L}_G(\theta) = \alpha \min_{\phi} \mathcal{L}_C - \beta \min_{\omega} \mathcal{L}_D - \gamma \min_{\varphi} \mathcal{L}_R + \mathcal{C}(\epsilon) \tag{7}$$

where θ, ϕ, ω, and φ are the model parameters of the generator, classifier, discriminator, and re-constructor respectively. α, β, and γ are weights of components in the objective function of the generator and can be freely tuned. $\mathcal{C}(\epsilon)$ is a constraint function with respect to hyper-parameter ϵ, as to be elaborated in the following section.

3.5. Optimization with constraint

In order to meet a certain level of reconstruction distance, we consider the constrained problem:

$$\begin{aligned} &\underset{\theta}{minimize}\ \mathcal{L}_G(\theta) \\ s.t \quad &\mathbb{E}_{x \sim p_d}[dist(x, \hat{x})] \le \epsilon \end{aligned} \tag{8}$$

The optimization problem above can be approximated as an unconstrained problem [31]:

$$\underset{\theta}{minimize}\ (\mathcal{L}_G(\theta) + \gamma \mathcal{C}(\epsilon)) \tag{9}$$

where γ is a penalty parameter and $\mathcal{C}$ is a penalty function

$$\mathcal{C}(\epsilon) = \max(0, \mathbb{E}_{x \sim p_d}[dist(x, \hat{x})] - \epsilon) \tag{10}$$

Note that $\mathcal{C}$ is nonnegative, and $\mathcal{C}(\theta) = 0$ iff the constraint in (8) is satisfied.

3.6. Training algorithms

Algorithm 1 describes the training process of AutoGAN-DRP. The framework contains four components, and they are trained one by one (lines 4–15) within one global training step. After sampling batches from target distribution and data for inputs of the models (lines 2–3), we then train the four components. First, the reconstructor is trained in n_r iterations while other components' parameters are fixed (lines 4–6). Second, the discriminator is trained (lines 7–9). Third, the classifier is trained in n_c iterations (lines 10–12). Fourth, the generator is trained in n_g iterations (lines 13–15). After training each component in their number of local training steps, the above training process is repeated until it reaches the number of global training iterations (lines 1–16). In our setting, the numbers of local training iterations (n_c, n_r, n_d, n_g) are much smaller than the number of global iterations n.

4. Experiments and discussion

In this section, we demonstrate our experiments over three popular supervised face image datasets: *the Extended Yale Face Database B* [18], *AT&T* [19], and *CelebFaces Attributes Dataset (CelebA)* [20]. To comprehensively evaluate our method performance, we also conduct experiments with different generator and re-constructor structures, different types of classifications (binary and multi-class classification), different numbers of reduced dimensions. The effectiveness of the method is then evaluated in terms of utility and privacy.

Algorithm 1 Algorithm for stochastic gradient descent training of ϵ -DR Privacy.

```
Input: Training dataset $X$.
    Parameter: learning rate $\alpha_r, \alpha_d, \alpha_c, \alpha_g$,training steps $n_r, n_d, n_c, n_g$
    A constraint for $\epsilon$-DR
Output: Transformation Model
    Initialization.
1:  for $n$ global training iterations do
2:      Randomly sample a mini batch from target distribution and label $\boldsymbol{t}$.
3:      Randomly sample mini batch of data $\boldsymbol{x}$ and corresponding label $\boldsymbol{y}$
4:      for $i = 0$ to $n_r$ iterations do
5:          Update the Reconstruction:
            $\varphi_{i+1} = \varphi_i - \alpha_r \nabla_{\varphi} \mathcal{L}_R(\varphi_i, \boldsymbol{x})$
6:      end for
7:      for $j = 0$ to $n_d$ iterations do
8:          Update the Discriminator parameter:
            $\omega_{j+1} = \omega_j - \alpha_d \nabla_{\omega} \mathcal{L}_D(\omega_j, \boldsymbol{x}, \boldsymbol{t})$
9:      end for
10:     for $k = 0$ to $n_c$ iterations do
11:         Update the Classifier parameter:
            $\phi_{k+1} = \phi_k - \alpha_c \nabla_{\phi} \mathcal{L}_C(\phi_k, \boldsymbol{x}, \boldsymbol{y})$
12:     end for
13:     for $l = 0$ to $n_g$ iterations do
14:         Update the Generator parameter:
            $\theta_{l+1} = \theta_l - \alpha_g \nabla_{\theta} \mathcal{L}_G(\theta_l, \boldsymbol{x}, \boldsymbol{t}, \boldsymbol{y})$
15:     end for
16: end for
17: return
```

4.1. Experiment setup

The Extended Yale Face Database B (YaleB) contains 2470 grayscale images of 38 human subjects under different illumination conditions and their identity label. In this dataset, the image size is 168 × 192 pixels. The AT&T dataset has 400 face images of 40 subjects. For convenience, we resize each image of these two dataset to 64 × 64 pixels. CelebA is a color facial image dataset containing 202,599 images of 10,177 subjects. 1709 images of the first 80 subjects are used for our experiment. Each image is resized to 64 × 64 × 3 pixels. All pixel values are scaled to the range of [0,1]. We randomly select 10% of each subject's images for validation and 15% for testing dataset.

The generator and re-constructor in Fig. 3 are implemented by three different structures. Specifically, we follow the architecture of recent powerful models VGG19, VGG16 [32] and a basic convolutional network (CNN). We modify the models to adapt to our data size (64 × 64). Discriminator and Classifier are built on fully connected neural network and convolutional network respectively. Leaky ReLU is used for activation function in hidden layers. We use linear activation function for generator's output layers and softmax activation functions for other components' output layers. Each component is trained in 5 local iterations (n_r, n_g, n_d, n_c), and the entire system is trained in 500 global iterations (n). The target distribution is drawn from Gaussian distribution (with the covariance value of 0.5 and the mean is the average of the training data). Table 1 provides detail information of neural networks' structures and other implementation information.

To evaluate the reliability, we test our framework with different levels of authentication corresponding to binary classification (single-level) and multi-class classification (multi-level). For the single-level authentication system, we consider half of the subjects

Table 1
Implementation information.

	VGG16			VGG19			CNN		
	Hidden layers	Units	Parameter	Hidden layers	Units	Parameter	Hidden layers	Units	Parameter
Generator	Conv block	64 × 2	16,295,623	Conv block	64 × 2	21,605,319	Conv	256	16,451,847
	Max_pooling			Max_pooling			BatchNorm		
	Conv_block	128 × 2		Conv_block	128 × 2		Conv	512	
	Max_pooling			Max_pooling			BatchNorm		
	Conv_block	256 × 3		Conv block	256 × 4		Conv	1024	
	Max_pooling			Maxpooling			BatchNorm		
	Conv_block	512 × 3		Conv block	512 × 4		Dense	1024	
	Max_pooling			Max pooling					
	Conv_block	512 × 3		Conv block	512 × 4				
	Max_pooling			Max pooling					
	Dense	1024		Dense	1024				
	Dense	1024		Dense	1024				
Reconstructor	Dense	1024	10,184,000	Dense	1024	13,281,472	Dense	1024	18,048,256
	Dense	1024		Dense	1024		BatchNorm		
	Dense	1024		Dense	1024		Reshape		
	Reshape			Reshape			Conv	1024	
	Conv_block-T	512 × 3		Conv_block-T	512 × 4		BatchNorm		
	Up_sampling			Up_sampling			Conv	512	
	Conv_block-T	512 × 3		Conv_block-T	512 × 4		BatchNorm		
	Up_sampling			Up_sampling			Conv	256	
	Conv_block-T	256 × 3		Conv_block-T	256 × 4				
	Up_sampling			Up_sampling					
	Conv_block-T	128 × 2		Conv_block-T	128 × 2				
	Up_sampling			Up_sampling					
	Conv_block-T	64 × 2		Conv_block-T	64 × 2				
Classifier	Dense	2048	12,636,168	Dense	2048	12,636,168	Dense	2048	12,636,168
	BatchNorm			BatchNorm			BatchNorm		
	Dropout			Dropout			Dropout		
	Dense	2048		Dense	2048		Dense	2048	
	BatchNorm			BatchNorm			BatchNorm		
	Dropout			Dropout			Dropout		
	Dense	2048		Dense	2048		Dense	2048	
	BatchNorm			BatchNorm			BatchNorm		
	Dropout			Dropout			Dropout		
	Dense	2048		Dense	2048		Dense	2048	
	BatchNorm			BatchNorm			BatchNorm		
	Dropout			Dropout			Dropout		
Discriminator	Conv	128	5,084,737	Conv	128	5,084,737	Conv	128	5,084,737
	Dropout			Dropout			Dropout		
	Conv	256		Conv	256		Conv	256	
	Dropout			Dropout			Dropout		
	Flatten			Flatten			Flatten		
	Dense	1024		Dense	1024		Dense	1024	
	Dense	1024		Dense	1024		Dense	1024	

Shared parameters: optimizer Adam, learning rate 0.0001, 7 dimensions
Hardware: GPU Testla T4 16Gb, CPU Xeon Processors @2.3Ghz
Software: Tensorow 2.0 beta. The number of trainable parameters are reported by model.summary() from Keras library.

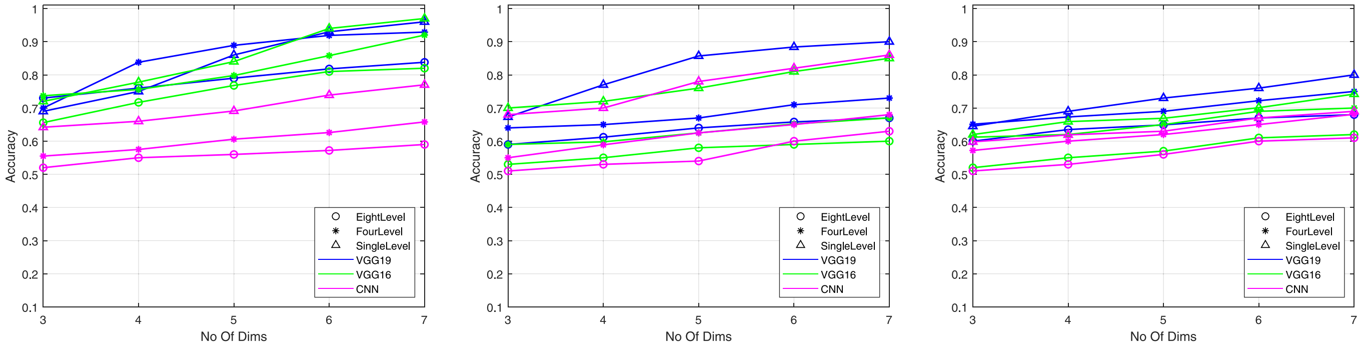
in the dataset are valid to access company's resources while the rest are invalid. We randomly divide the dataset into two groups of subjects and labels their images to (1) or (0) depending on their access permission. For the cases of multi-level authentication system, we divide the subjects into four groups and eight groups. Therefore, the authentication server becomes four-class and eight-class classifier respectively.

4.2. Utility

We use accuracy metric to evaluate the utility of dimension-reduced data. The testing dataset is tested with the classifier extracted from our framework. Different structures of Generator and re-constructor are applied including VGG19, VGG16, basic CNN on different privilege levels which correspond to multi-class classification. Fig. 4 illustrates the accuracies for different dimensions from three to seven over the three facial datasets. Overall, the accuracies improve when the number of dimension increases. The accuracies on the two gray image datasets (AT&T and Yale_B) reaches 90% and higher when using VGG with only seven dimensions. This accuracy figure for Celeba is smaller, but it still reaches 80%. In general, VGG19 structure performs better than using VGG16 and basic CNN in terms of utility due to the complexity (Table 1) and adaptability to image datasets of VGG19. As the dimension number is reduced from 4096 (64 × 64) to 7, we can achieve a compression ratio of 585 yet achieve accuracy of 90% for the two gray datasets and 80% for the color dataset. This implies our method could gain a high compression ratio and maintain a high utility in terms of accuracy. During conducting experiments we also observe that the accuracy could be higher if we keep the original resolution of images. However, for convenience and reducing the complexity of our structure, we resize images to the size of 64 × 64 pixels.

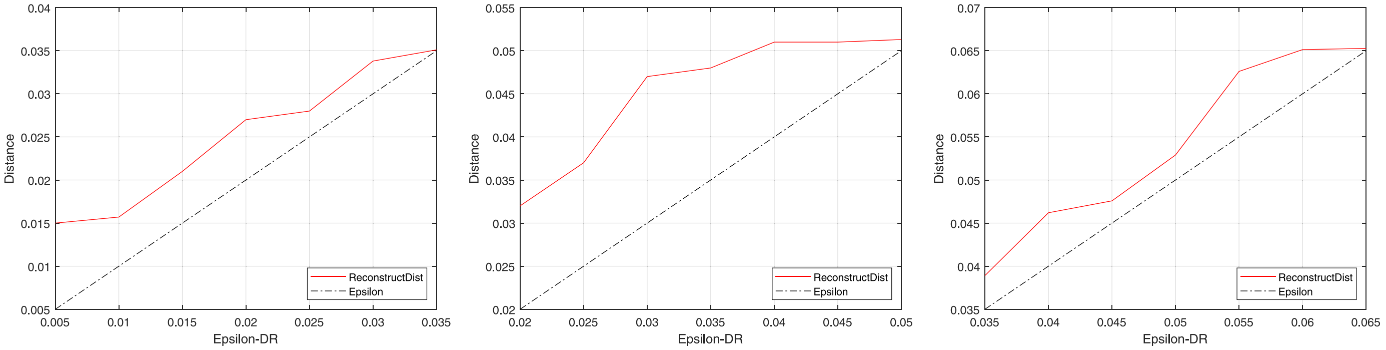
4.3. Privacy

In this study, the Euclidean distance is used to measure the distance between original and reconstructed images: $dist(x, \hat{x}) = ||x - \hat{x}||^2$. Fig. 5 illustrates the average distances between original images and reconstructed images on testing data with different ϵ constraints (other setting parameters: seven dimensions,



(a) AT&T (b) Yale_B (c) CelebA

Fig. 4. Accuracy for Different Number of Reduced Dimensions.



(a) AT&T (b) Yale_B (c) CelebA

Fig. 5. Average Distance Measurement Result { 7 dimensions, Single-Level}.

single-level authentication, and VGG19 structure). The achieved distances (red lines) are larger than the hyper-parameter ϵ (black dotted lines) where ϵ is less than 0.035 for AT&T, 0.052 for YaleB and 0.067 for CelebA. Thus, our framework can satisfy ϵ-DR with ϵ of above values. Due to the fact that the re-constructor obtained some information (we consider the adversary can reach the model and the training data), we can only set the distance constraint ϵ within a certain range as shown in 5. The intersection between the red line and the dotted black line points out the largest distance our framework can achieve. Since the mean of the target distribution is set to be the same as the mean of training dataset, reconstructed images will be close to the mean of training dataset which we believe it will enlarge the distance and expose less individual information. Thus, the range of epsilon can be estimated base on the expectation of the distance between testing samples and the mean of training data. In addition, the first section of Table 2 demonstrates some samples and their corresponding reconstructions in single-level authentication and seven dimensions with different achieved accuracies and distances. The reconstructed images could be nearly identical, thus making it visually difficult to recognize the identity of an individual.

5. Comparison to GAP [12]

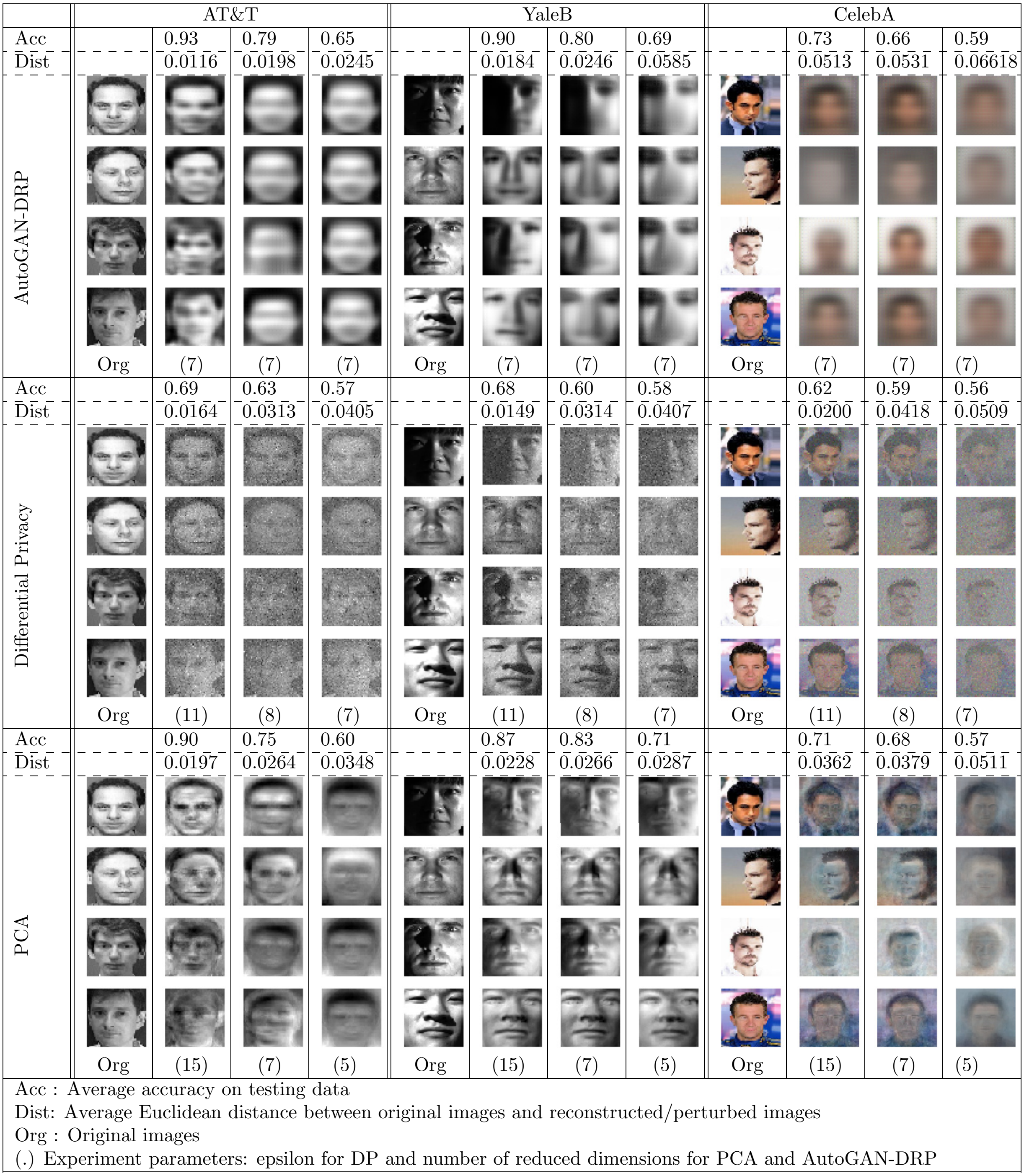
In this section, we compare the proposed framework with GAP, which shares many similarities. At first, we attempt to visualize AutoGAN-DRP and GAP by highlighting their similarities and differences. Then, we exhibit our experiment results of the two methods on the same dataset.

In terms of similarities, AutoGAN-DRP and GAP are utilizing minimax algorithms of Generative Adversarial Nets, applying the state-of-the-art convolution neural nets for image datasets, considering l_2 norm distance (i.e., distortion in GAP, privacy measurement in AutoGAN-DRP) between the original images and reconstructed images. Specifically, both GAP and AutoGAN-DRP consider the reconstruction distance between original and reconstructed images. In GAP this *distortion* refers to the Euclidean between original and privatized images, and AutoGAN-DRP denotes the *distance* as the Euclidean distance between original and reconstructed images. In this context, the distance and distortion refer to the same measurement and have the same meaning. To be consistent, we use the term *distance* to present this measurement in the rest of this section.

However, there are also distinctions between GAP and AutoGAN-DRP. In GAP, the adversary aims to identify a private label (e.g., gender) which should be kept secret while AutoGAN-DRP aims to visually protect the owner's face images by enlarging the reconstruction distance. Thus, instead of considering a private label in loss function of the generator in GAP, AutoGAN-DRP is aimed at driving the reconstructed data into a target distribution using a discriminator.

Fig. 6 illustrates the visualization of AutoGAN-DRP and GAP. In AutoGAN-DRP, privacy is assessed based on how well an adversary can reconstruct the original data and measured by the distance between original and reconstructed data. The dimension-reduced data is reconstructed using the state-of-the-art neural network (an Auto-encoder). The larger the distance is, the more privacy can be achieved. Further, if the reconstructed images are blurry, privacy can be preserved since it is hard to visually determine an individual identity. The data utility is quantified by the accuracy of the classification tasks over dimension-reduced data which captures the most significant data information. Meanwhile, GAP perturbs images with a certain distortion constraint to achieve privacy. It evaluates data utility by the classification accuracy of non-private label and assesses privacy by the classification accuracy of private label. Similar to AutoGAN-DRP, the high distortion is most likely to yield high level of privacy. In GAP, however, high distortion might dramatically reduce the classification accuracy of non-private label.

Table 2
Sample visualization of AutoGAN, DP, PCA over three datasets.



	AT&T				YaleB				CelebA			
Acc		0.93	0.79	0.65		0.90	0.80	0.69		0.73	0.66	0.59
Dist		0.0116	0.0198	0.0245		0.0184	0.0246	0.0585		0.0513	0.0531	0.06618
AutoGAN-DRP	Org	(7)	(7)	(7)	Org	(7)	(7)	(7)	Org	(7)	(7)	(7)
Acc		0.69	0.63	0.57		0.68	0.60	0.58		0.62	0.59	0.56
Dist		0.0164	0.0313	0.0405		0.0149	0.0314	0.0407		0.0200	0.0418	0.0509
Differential Privacy	Org	(11)	(8)	(7)	Org	(11)	(8)	(7)	Org	(11)	(8)	(7)
Acc		0.90	0.75	0.60		0.87	0.83	0.71		0.71	0.68	0.57
Dist		0.0197	0.0264	0.0348		0.0228	0.0266	0.0287		0.0362	0.0379	0.0511
PCA	Org	(15)	(7)	(5)	Org	(15)	(7)	(5)	Org	(15)	(7)	(5)

Acc : Average accuracy on testing data
Dist: Average Euclidean distance between original images and reconstructed/perturbed images
Org : Original images
(.) Experiment parameters: epsilon for DP and number of reduced dimensions for PCA and AutoGAN-DRP

This might be caused by the high correlation between private and non-private labels. This difference enables AutoGAN-DRP to preserve more utility than GAP at the same distortion level, as the experiment result (depicted in Fig. 7) reveals.

In the experiment, we reproduce a prototype of Transposed Convolutional Neural Nets Privatizer (TCNNP) in GAP using materials and source code provided by [12]. We also modify our framework to make it as similar to TCNNP as possible. Specifically, a combination of two convolutional layers with ReLU activation function and two fully connected neural network layers are used for implementing the Generator similar to TCNNP. Our Classifier is constructed on two convolutional layers and two fully connected hidden layers similar to the Adversary in GAP. We also test our framework on GENKI, the same dataset with GAP. The utility is evaluated by the accuracy of facial expression classification (a binary classification). It should be noted that our framework have been shown to work on different datasets with multi-class classification, which is more challenging and comprehensive. Fig. 7 shows the accuracy results of GAP and AutoGAN-DRP for GENKI dataset. AutoGAN-DRP achieves distances ranging from 0.037 to 0.039 for different dimensions from one to seven. At the same range of distance (distortion per pixel), GAP achieves accuracy of only 72% while AutoGAN-DRP gains accuracy rates starting from 77% to 91% for different number of dimensions. It becomes evident that our

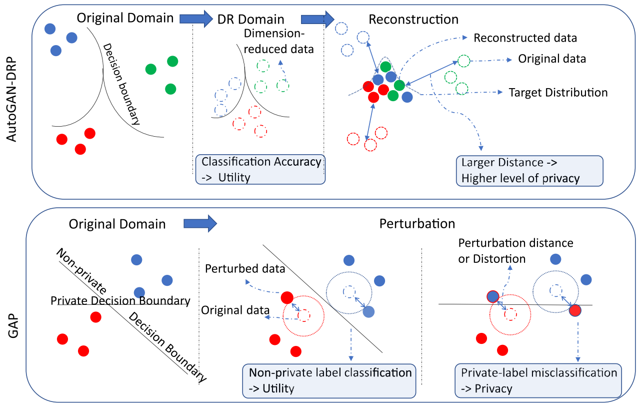

AutoGAN-DRP
Original Domain
DR Domain
Reconstruction
Decision boundary
Dimension-reduced data
Reconstructed data
Original data
Target Distribution
Classification Accuracy -> Utility
Larger Distance -> Higher level of privacy
GAP
Original Domain
Perturbation
Non-private Decision Boundary
Private Decision Boundary
Perturbed data
Original data
Perturbation distance or Distortion
Non-private label classification -> Utility
Private-label misclassification -> Privacy

Fig. 6. AutoGAN-DRP Vs GAP Explanation.

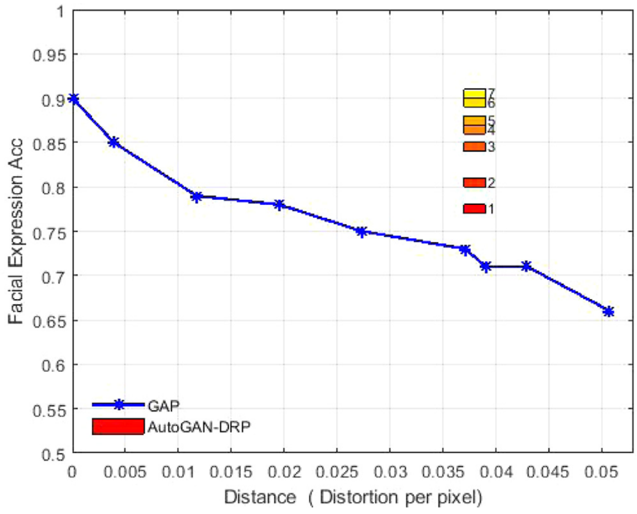


Fig. 7. GENKI Facial Expression Accuracy Vs Distance using GAP and AutoGAN-DRP.

method can achieve higher accuracy than that of GAP at the same distortion level.

6. Visual comparison to privacy preserving techniques using Differential Privacy (DP) [27] and Principle Component Analysis (PCA) [33]

In this section, we compare AutoGAN-DRP with other privacy preserving methods in terms of ability to visually identify client's identities. We choose the widely used tool for privacy preserving Differential Privacy (DP) [27] and another privacy preservation method utilizing dimensionality reduction technique (i.e., Principle Component Analysis [33]).

In these experiments, we implement AutoGAN-DRP following VGG19 structure for the Generator and Re-constructor, and other setting parameters (e.g., number of hidden layers, learning rate, optimization) are shown in Table 1. The images are reduced to seven dimensions for different values of ϵ-DR to achieve different distances and accuracies. The datasets are grouped into two groups corresponding to a binary classifier.

For implementing DP, we first generate a classifier on the authentication server by training the datasets with a VGG19 binary classifier (the structure of hidden layers is similar to our Generator in Table 1). The testing images are then perturbed using differential privacy method. Specifically, Laplace noise is added to the images with the sensitivity coefficient of 1 (it is computed by the maximum range value of each pixel [0,1]) and different DP epsilon parameters (this DP epsilon is different from our ϵ-DR). The perturbed images are then sent to the authentication server and fed to the classifier. We visually compare the perturbed images of this method with AutoGAN.

In addition, we follow instruction in FRiPAL [11] in which the clients reduce image dimension using Principle Component Analysis (PCA) and send reduced features to the server. FRiPAL claims that by reducing image dimension, their method can be more resilient to reconstruction attacks. The experiments are conducted with different number of reduced dimension. The images are reconstructed using *MoorePenrose inverse* method with assumption that an adversary has assess to the model. The classification accuracy is evaluated using a classifier which has similar structure to AutoGAN's classifier.

Table 2 shows image samples and results over the three datasets. Overall, AutoGAN-DRP is more resilient to reconstruction attacks compared to the other two techniques. For instance, at the accuracy of 79% on AT&T dataset, 80% on YaleB, and 73% on CelebA, we cannot distinguish entities from the others. For DP method, the accuracy decreases when the DP epsilon decreases (adding more noise), and the perturbed images become harder to recognize. However, at a low accuracy 57%, we are still able to distinguish identities by human eyes. The reason is that DP noise does not focus on the important visual pixels. For PCA, the accuracy also goes down when the number of dimensions decreases and the distances increase. Since PCA transformation is linear and deterministic, the original information can be significantly reconstructed using the inverse transformation deriving from the model or training data. Thus, at the accuracy of 75% on AT&T, 71% on YaleB, and 68% on CelebA, we still can differentiate individuals. Overall, our proposed method shows the advantage in securing the data while retaining high data utility.

7. Conclusion

In this paper, we introduce a mathematical tool ϵ-DR to evaluate privacy preserving mechanisms. We also propose a non-linear dimension reduction framework. This framework projects data onto lower dimension domain in which it prevents reconstruction attacks and preserves data utility. The dimension-reduced data can be used effectively for the machine learning tasks such as classification. In our future works, we plan to extend the framework to adapt with different types of data, such as time series and categorical data. We will apply different metrics to compute the distance other than l_2 norm and investigate the framework on several applications in security systems and data collaborative contributed systems.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript:

Hung Nguyen - University Of South Florida
Di Zhuang - University Of South Florida
Morris - Chang University Of South Florida
Pei-Yuan Wu - National Taiwan University

Acknowledgment

This work is sponsored by DARPA Brandeis program under agreement number N66001-15-C-4068. The views, options, and/or findings contained in this article/presentation are those of the author/presenter and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neucom.2019.12.002.

References

[1] R. Bost, R. Popa, S. Tu, S. Goldwasser, Machine learning classification over encrypted data, Ndss '15 (February) (2015) 1–31, doi:10.14722/ndss.2015.23241.

[2] F. Emekci, O.D. Sahin, D. Agrawal, A. El Abbadi, Privacy preserving decision tree learning over multiple parties, Data Knowl. Eng. 63 (2) (2007) 348–361, doi:10.1016/j.datak.2007.02.004.

[3] E. Hesamifard, H. Takabi, M. Ghasemi, CryptoDL : Deep Neural Networks over Encrypted Data, (2017). arXiv:1711.05189v1

[4] A.C.-C. Yao, How to generate and exchange secrets, Proceedings of the 27th Annual Symposium on Foundations of Computer Science(sfcs 1986)(1) (1986) 162–167. 10.1109/SFCS.1986.25 http://ieeexplore.ieee.org/document/4568207/

[5] A. Shamir, How to share a secret, Commun. ACM (1979) 612–613, doi:10.1007/978-3-642-15328-0_17.

[6] K. Chaudhuri, C. Monteleoni, Privacy-preserving logistic regression, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), Advances in Neural Information Processing Systems 21, Curran Associates, Inc., 2009, pp. 289–296.

[7] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, M. Winslett, Functional Mechanism: Regression Analysis under Differential Privacy (2012) 1364–1375. arXiv:1208.0219

[8] N. Phan, Y. Wang, X. Wu, D. Dou, Differential privacy preservation for deep auto-Encoders: an application of human behavior prediction, Aaai (2016) 1309–1316.

[9] M. Abadi, A. Chu, I. Goodfellow, H.B. McMahan, I. Mironov, K. Talwar, L. Zhang, Deep Learning with Differential Privacy(Ccs) (2016). 10.1145/2976749.2978318 arXiv:1607.00133

[10] L.O.-M. Xiaoqian Jiang, Z. Ji, S. Wang, N. Mohammed, S. Cheng, Differential-Private data publishing through component analysis, Trans. data Priv. 6 (1) (2013) 19–34 arXiv:NIHMS150003, doi:10.1080/10810730902873927.Testing.

[11] D. Zhuang, S. Wang, J.M. Chang, Fripal: Face recognition in privacy abstraction layer, in: Proceedings of the IEEE Conference on Dependable and Secure Computing, IEEE, 2017, pp. 441–448.

[12] X.C.L.S.R.R. Chong Huang, Peter Kairouz, Generative adversarial privacy, Proceedings of the Privacy in Machine Learning and Artificial Intelligence Workshop, ICML 2018 (2018).

[13] X. Chen, P. Kairouz, R. Rajagopal, Understanding compressive adversarial privacy, CoRR abs/1809.08911 (2018) arXiv:1809.08911.

[14] S. Zhou, K. Ligett, L. Wasserman, Differential privacy with compression, CoRR (2009).

[15] S.Y. Kung, Compressive privacy: from informationestimation theory to machine learning [lecture notes], IEEE Signal Process. Mag. 34 (1) (2017) 94–112, doi:10.1109/MSP.2016.2616720.

[16] S. Kung, A compressive privacy approach to generalized information bottleneck and privacy funnel problems, J. Frankl. Inst. 355 (2017), doi:10.1016/j.jfranklin.2017.07.002.

[17] K. Xie, X. Ning, X. Wang, J. Wen, X. Liu, S. He, D. Zhang, An efficient privacy-preserving compressive data gathering scheme in wsns, in: G. Wang, A. Zomaya, G. Martinez, K. Li (Eds.), Algorithms and Architectures for Parallel Processing, Springer International Publishing, Cham, 2015, pp. 702–715.

[18] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[19] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of the IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142, doi:10.1109/ACV.1994.341300.

[20] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in: Proceedings of International Conference on Computer Vision (ICCV), 2015.

[21] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks(2014) 1–9. 10.1001/jamainternmed.2016.8245arXiv:1406.2661

[22] P. Baldi, Autoencoders, unsupervised learning, and deep architectures, ICML Unsupervised Transf. Learn. (2012) 37–50 arXiv:1509.02971, doi:10.1561/2200000006.

[23] R. Bost, R. Ada Popa, S. Tu, S. Goldwasser, Machine learning classification over encrypted data, 2015, doi:10.14722/ndss.2015.23241.

[24] E. Hesamifard, H. Takabi, M. Ghasemi, Cryptodl: deep neural networks over encrypted data, CoRR abs/1711.05189 (2017) arXiv:1711.05189.

[25] M. Al-Rubaie, P. Wu, J.M. Chang, S. Kung, Privacy-preserving PCA on horizontally-partitioned data, in: IEEE Conference on Dependable and Secure Computing, Taipei, 2017, pp. 280–287, doi:10.1109/DESEC.2017.8073817.

[26] T. Pedersen, Y. Saygn, E. Sava, Secret Sharing vs. Encryption-based Techniques For Privacy Preserving Data Mining, in: Proc. of UNECE/Eurostat Work Session on SDC, 2007.

[27] C. Dwork, Differential privacy, Proc. 33rd Int. Colloq. Autom. Lang. Program. (2006) 1–12 arXiv:1011.1669v3, doi:10.1007/11787006_1.

[28] Z. Wu, Z. Wang, Z. Wang, H. Jin, Towards privacy-preserving visual recognition via adversarial training: a pilot study, CoRR abs/1807.08379 (2018) arXiv:1807.08379.

[29] M. Yang, T. Zhu, B. Liu, Y. Xiang, W. Zhou, Machine learning differential privacy with multifunctional aggregation in a fog computing architecture, IEEE Access 6 (2018) 17119–17129, doi:10.1109/ACCESS.2018.2817523.

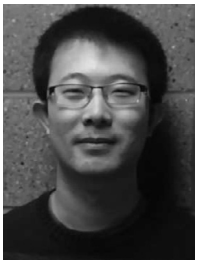
[30] B. Hitaj, G. Ateniese, F. Perez-Cruz, Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning (2017). 10.1145/3133956.3134012arXiv:1702.07464

[31] P.A. Jensen, Algorithms for constrained optimization, (https://www.me.utexas.edu/~jensen/ORMM/supplements/units/nlp_methods/const_opt.pdf).

[32] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, 2015.

[33] K.E. Svante Wold, P. Geladi, Principal component analysis, Chemometr. Intell. Laborat. Syst. 2 (1987) 37–52 arXiv:1409.1556.

Hung Nguyen received the M.Sc. degree and he is currently pursuing his Ph.D. degree in Department of Electrical Engineering, University of South Florida, FL, USA. His current research interests include machine learning techniques, artificial intelligence, cyber security, and privacy enhancing technologies. He is a student member of IEEE.



Di Zhuang received the B.E. degree in computer science and information security from Nankai University, China. He is currently pursuing his Ph.D. degree in electrical engineering with University of South Florida, Tampa. His research interests include cyber security, social network science, privacy enhancing technologies, machine learning and big data analytics. He is a student member of IEEE.

Pei-Yuan Wu is an assistant professor at National Taiwan University since 2017. He was born in Taipei, Taiwan, R.O.C., in 1987. He received the B.S.E. degree in electrical engineering from National Taiwan University in 2009, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University in 2012 and 2015, respectively. He joined Taiwan Semiconductor Manufacturing Company from 2015 to 2017. He was a recipient of the Gordon Y.S. Wu Fellowship in 2010, Outstanding Teaching Assistant Award at Princeton University in 2012. His research interest lies in artificial intelligence, signal processing, estimation and prediction, and cyber-physical system modeling.

J. Morris Chang is a professor in the Department of Electrical Engineering at the University of South Florida. He received his Ph.D. degree from the North Carolina State University. He received the University Excellence in Teaching Award at Illinois Institute of Technology in 1999. His research interests include: cyber security, wireless networks, and energy efficient computer systems. In the last six years, his research projects on cyber security have been funded byDARPA. He is a handling editor of Journal of Microprocessors and Microsystems and an editor of IEEE IT Professional.