2021 20th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

DP-Face: Privacy-Preserving Face Recognition Using Siamese Network

Nazhao Yan, Hang Cheng*, Meiqing Wang
School of Mathematics and Statistics
Fuzhou University
Fuzhou Fujian, China
Email: zoeyyannazhao@gmail.com, hcheng@fzu.edu.cn, mqwang@fzu.edu.cn

Qinjian Huang, Fei Chen
School of Computer Science and Big Data
Fuzhou University
Fuzhou Fujian, China
Email: 821403552@qq.com,chenfei314@fzu.edu.cn

***Abstract*—With the rapid development of deep learning, face recognition technology based on deep learning has been widely developed in recent years. However, during the training of the deep learning model, there is a risk of privacy leakage. If an attacker obtains private data, such as tags of the training data, the face images may be restored, and private information is leaked. To protect the private information of the face recognition model, we introduce differential privacy technology to propose a privacy-preserving face recognition scheme using the Siamese Network framework called DP-Face. Unlike other privacy-preserving face recognition methods, we can adjust the balance between privacy and utility through privacy budgets according to actual needs. Experimental results show that the effectiveness and privacy of the proposed DP-Face can be well guaranteed.**

***Keywords*-privacy-preserving; differential privacy; face recognition; deep learning**

I. Introduction

In recent years, following dramatic advances in artificial intelligence technologies, biometric recognition technology is rapidly integrating into people's lives. As one of the most popular biometric technologies, face recognition can identify individuals without their knowledge. In 2015, the FaceNet constructed by Schroff et al. achieved a recognition rate of 99.6%, which implies the reliability and practicability of face recognition technology [1]. Besides, Alipay provides facial payment services to more than 243 million users in China in 2020.

However, more and more security problems have gradually been exposed as deep learning is widely used in face recognition. Generally speaking, the training process of a deep learning model requires a large amount of representative data, which may contain sensitive information of the data owner such as age and gender. For the sake of security, the model should not disclose such private sensitive information. In a recent work [2], the attacker only needs to obtain the label of the data in the training set and get access to the model, and then the face image can be recovered from the model. In this case, the capability of effective face recognition is desired when keeping users' privacy confidential to the attacker. Obviously, it can be solved by employing the signal processing technologies in the encrypted domain [3], [4].

In this paper, we employ differential privacy technology to present a privacy-preserving face recognition (DP-Face). This scheme assumes that only one party holds sensitive private data, and we mainly focus on privacy leakage caused by model output. Our DP-Face uses Siamese Network to construct a similarity measurement network for better recognition accuracy. Meanwhile, to protect the privacy of training data, we introduce the differential privacy strategy. In addition, DP-Face can achieve the trade-off between the privacy and utility of the model by adjusting the privacy budget. The experiments illustrate that our DP-Face can well finish the task of face recognition while protecting private data from being leaked.

II. Preliminary

In this section, we review the basic concept of differential privacy technology and Siamese Network, which serve as the cornerstone for constructing the proposed DP-Face.

A. Differential Privacy

Differential privacy [5] ensures that inserting or deleting records in the data set will not affect the output results. And also, the protection model does not care about the background knowledge of the attacker. Even if the attacker has all records except a certain record, it cannot infer the undisclosed record. The formal definition of differential privacy is given below.

Definition 1 ((ε, δ)-DP). A randomized algorithm $\mathcal{M} : \mathcal{D} \rightarrow \mathcal{R}$ satisfies (ε, δ)-differential privacy if for any neighbouring datasets $d, d' \in \mathcal{D}$ and for any all output sets $S \subseteq \mathcal{R}$ it holds that

$$Pr[\mathcal{M}(d) \in S] \leq e^{\varepsilon} Pr[\mathcal{M}(d') \in S] + \delta, \qquad (1)$$

where d and d' have at most one record difference. Note that the randomness of the algorithm $\mathcal{M}$ means that for a specific input, the output of the algorithm is not a fixed value but obeys a certain distribution. The parameter ε represents the privacy budget. And the smaller the ε, the higher the degree of privacy guarantee.

The typical mechanisms for achieving differential privacy are the Laplace mechanism and Gaussian mechanism [6]. Among them, the Laplace mechanism provides a strict $(\varepsilon, 0)$-DP, and the Gaussian mechanism provides a relaxed (ε, δ)-DP [7]. In DP-Face, we use the Gaussian mechanism, which is defined as follows:

$$\mathcal{M}(x) \triangleq f(x) + N\left(0, \sigma^2 s_f^2\right), \qquad (2)$$

2473-3636/21/$31.00 ©2021 IEEE
DOI 10.1109/DCABES52998.2021.00030

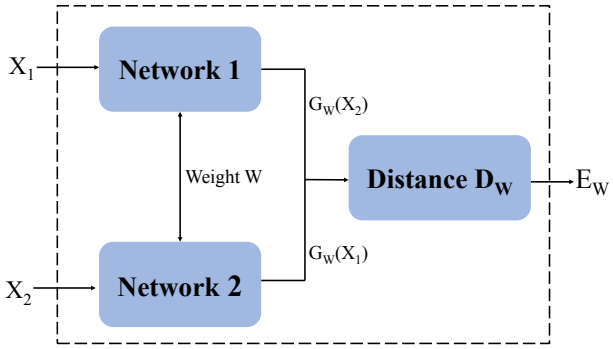

Figure 1. Architecture of Siamese Network.

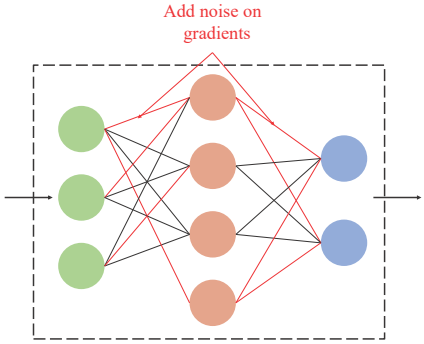

Figure 2. Construction of the backbone network of Siamese Network.

where s_f indicates the sensitivity related to the function f.

Definition 2 (l_2**-Sensitivity).** For any neighbouring datasets $d, d' \in \mathcal{D}$, given a function $f : \mathcal{D} \to \mathcal{R}$, the l_2-sensitivity Δf is defined as:

$$\Delta f = \max_{d,d'} ||f(d) - f(d')||_2. \quad (3)$$

B. Siamese Network

The goal of the Siamese Network [8], [9] is to calculate the similarity of two similar images. It has two identical sub-networks, both of which have the same parameters and weights. The special characteristic of this network is that its training samples take image pairs as input, and the features are extracted by two sub-networks respectively, and finally, the feature vector pairs of the samples are obtained. The architecture of the Siamese Network is shown in Figure 1.

Here, (X_1, X_2) represents the input image pair, $(G_W(X_1), G_W(X_2))$ is the output feature pair of network 1 and 2. Then, the similarity E_W of the sample pair will be predicted through a reasonable similarity measure.

In Siamese Network, the commonly used loss function is the contrastive loss function, which is defined as:

$$L(W) = \frac{1}{2}(1-Y)D_W^2 + \frac{1}{2}Ymax(0, m - D_W)^2, \quad (4)$$

where $D_W = \|G_W(X_1) - G_W(X_2)\|_2$ denotes the Euclidean distance between the outputs, Y denotes the label, and m is the marginal value greater than 0.

III. Privacy-Preserving Face Recognition Using Siamese Network

In this section, we present the proposed Privacy-Preserving Face Recognition with Siamese Network (DP-Face) model and the privacy Analysis of DP-Face.

A. DP-Face Framework

The Siamese Network in our DP-Face scheme is designed on a pre-trained network. Pre-trained VGG16, Inception, or ResNet in ImageNet is often selected as a candidate backbone network for Siamese Network. Meanwhile, The output of the fully connected layer of the backbone network is used as a feature of the input image.

Note that the gradient contains private information about the dataset in the deep learning model. It can be ensured that the subsequent updating operation of the parameter value will not leak user information as long as the gradient is disturbed. Therefore, instead of adding noise to the final parameters, we add noise proportional to the training data on the gradient of the Wasserstein distance. The specific details of DP-Face are shown in Figure 2 and Algorithm 1.

Algorithm 1: Differentially private Stochastic Gradient Descent in DP-Face

Input: Examples $x_1, x_2, \ldots, x_N$, learning rate α_t, number of iterations n, group size L, gradient norm bound C, loss function $\mathcal{L}(\theta) = \frac{1}{N}\mathcal{L}(\theta, x_i)$.

Output: θ_n and compute the overall privacy cost (ε, δ) using a privacy accounting method.

Initialize model parameters θ_0

for $t = 1$ *to* n **do**

Take a random sample L_t with sampling probability L/N

Compute gradient

For each $i \in L_t$, compute $g_t(x_i) \leftarrow \nabla_{\theta_t}\mathcal{L}(\theta_t, x_i)$

Clip gradient

$\overline{g}_t(x_i) \leftarrow g_t(x_i)/max(1, \frac{\|g_t(x_i)\|_2}{C})$

Add noise

$\widetilde{g_t} \leftarrow \frac{1}{L}(\sum_i \overline{g}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I}))$

Descent

$\theta_{t+1} \leftarrow \theta_t - \alpha_t \widetilde{g_t}$

end

B. Privacy Analysis of DP-Face

To prove that DP-Face satisfies differential privacy, we analyze the privacy loss of our DP-Face model.

Definition 3 (Privacy Loss). For the given neighbouring datasets d and d', assuming aux is the auxiliary input and $o(o \in \mathcal{R})$ is the output, the privacy loss at o can be defined as:

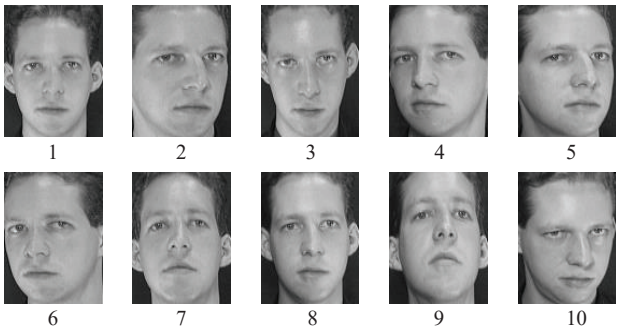

Figure 3. Sample images from ORL Database of Faces.

$$c(o;\mathcal{M},aux,d,d') \triangleq log\frac{Pr[\mathcal{M}(aux,d)=o]}{Pr[\mathcal{M}(aux,d')=o]} \quad (5)$$

The privacy loss is introduced to describe the distribution difference between two data. Besides, the privacy loss random variable $c(o;\mathcal{M},aux,d,d')$ is used to describe the privacy budget of $\mathcal{M}(d)$ in **Definition 1**.

Definition 4 (Log moment generation function). For the randomized algorithm $\mathcal{M}$, define the λ-th moment $\beta_{\mathcal{M}}(\lambda;aux,d,d')$ as the log of moment generation function evaluated at λ:

$$\beta_{\mathcal{M}}(\lambda;aux,d,d') \triangleq \log \mathbb{E}_{o\sim\mathcal{M}(aux,d)}\left[\exp\left(\lambda c\left(o;\mathcal{M},aux,d,d'\right)\right)\right]. \quad (6)$$

Definition 5 (Moment Accountant). The moment accountant is defined as:

$$\beta_{\mathcal{M}}(\lambda) \triangleq \max_{aux,d,d'} \beta_{\mathcal{M}}(\lambda;aux,d,d'), \quad (7)$$

In our DP-Face, we use the moment accountant to track the privacy loss caused by the published model and provide a privacy loss boundary with high accuracy.

The following theorems and lemma can be proved in the works [5], [10], which ensure our DP-Face scheme satisfies (ε,δ)-DP guarantee.

Theorem 1 (Composability). Suppose that a randomized algorithm $\mathcal{M}$ consists of a sequence of adaptive mechanisms $\mathcal{M}_1,\cdots,\mathcal{M}_k$, where $\mathcal{M}_i:\prod_{j=1}^{i-1}\mathcal{R}_j\times\mathcal{D}\rightarrow\mathcal{R}_i$. For any λ and $\alpha_{\mathcal{M}}(\lambda)$,

$$\alpha_{\mathcal{M}}(\lambda)\leq\sum_{i=1}^{k}\alpha_{\mathcal{M}_i(\lambda)}.$$

Theorem 2 (Tail bound). For any $\varepsilon>0$, the randomized algorithm $\mathcal{M}$ is (ε,δ)-differentially private for

$$\delta=\min_{\lambda} exp(\alpha_{\mathcal{M}}(\lambda)-\lambda\varepsilon).$$

Lemma 1. For any $\delta\in(0,1)$, $\sigma>\frac{\sqrt{2\ln(1.25/\delta)}\Delta f}{\varepsilon}$, the noise $Y\sim N(0,\delta^2)$ satisfies (ε,δ)-DP.

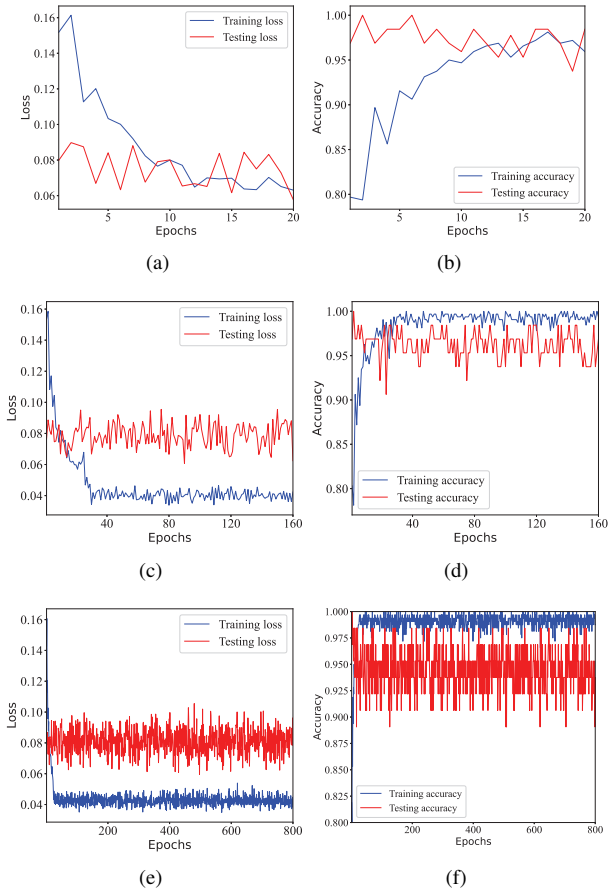

Figure 4. The loss and accuracy of DP-Face at different noise levels: (a) Loss at large noise level; (b) Accuracy at large noise level; (c) Loss at medium noise level; (d) Accuracy at medium noise level; (e) Loss at small noise level; (f) Accuracy at small noise level.

IV. Performance of DP-Face

To evaluate the performance of DP-Face, we utilize Tensorflow and Keras in Python 3 on one server, which is equipped with an 8-core Intel Core i7-9700 CPU @3.00GHz and 8GB of RAM running Windows 10-64bit. All the experiments were conducted with the ORL database created by Olivetti Research Laboratory in Cambridge, England.It contains 40 directories, each of which represents a different person, and contains ten face images. All the images are in PGM format, and each image is sized at 92×112, with 256-level grey per pixel. The partial face images are displayed in Figure 3.

A. Data Processing

The images in the ORL dataset are divided into the training set and the test set, where the training set is the first 37 directories of the 40 directories in the ORL and the last three directories for the test set. Considering the training set (370 face images in total) is limited, we use the pre-trained ResNet50 as the backbone network of the Siamese Network for feature extraction.

B. Performance Evaluation

In the plaintext model corresponding to our DP-Face, for the output of ResNet50, we first calculate the Euclidean

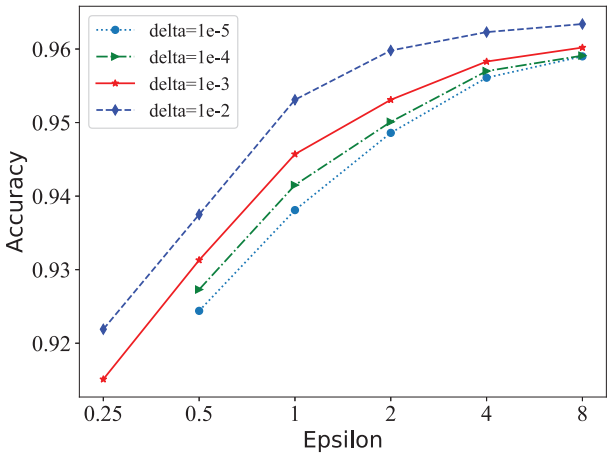

Figure 5. Effects of different parameters on classification accuracy.

distance between two output feature, then use the Dense layer and the Activation layer to construct the similarity calculation network. Here, the activation function is ReLU and the loss function is contrastive loss function. Using the batchsize with 64, we can reach the accuracy of 98.44% in the plaintext domain.

To better observe the effect of DP-Face, we conducted the experiments to evaluate the loss and accuracy of the results under different noise levels. Specifically, the noise is set to $0.05, 0.10, 0.20$, respectively. The higher the value, the more noise is added. Figure 4 shows the training and testing loss/accuracy as a function of the number of epochs in each plot. The toss accuracy performance varies with respect to the levels of the noises. As the noise increases, the accuracy of DP-Face will decrease, whereas the more the loss is generated. In the case of high noise, medium noise, and small noise, the accuracy was 95.31%, 96.87%, and 98.40%, respectively. In other words, when the added noise range is $0.20 \sim 0.05$, the accuracy of DP-Face remains above 95%, which means our DP-Face can complete face recognition tasks with high accuracy.

Besides, we verify the effect of differential privacy budget ε and a fixed noise scale δ for accuracy of DP-Face with the Gaussian mechanism. Here, privacy budget ε ranges from 0.1 to 10 and the noise scale δ is $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, respectively. From Figure 5, we can obtain the accuracy of the different (ε, δ), for example, when $\varepsilon = 0.25$ and $\delta = 0.01$, the accuracy of our DP-Face is 92.19%. In addition, it can be observed that for a fixed privacy budget ε, changing the δ value has a small impact on accuracy, but for a fixed noise scale δ value, changing the ε value has a large impact on accuracy. Plus, for a fixed δ, when more noise is added, the intensity of privacy protection will be greater; that is, the privacy budget ε will be smaller. However, it is not hard to find from Figure 5 that when the privacy budget is smaller, the model accuracy is lower. It means that the availability of the face recognition scheme will also decrease. Nevertheless, for the above (ε, δ), the accuracy of our DP-Face remains above 90%. Therefore, this framework can ensure that the accuracy of face recognition remains at a high level while protecting the privacy of face images.

V. Conclusion

In this paper, differential privacy technology and Siamese Network are used to construct a privacy-preserving face recognition scheme, which preserves the privacy of face training data in a differentially private case. The DP-Face model mitigates information leakage by adding designed noise to the gradients during the deep learning process. In addition, experiments show that the DP-Face can still converge under the constraints of noise-added training data and achieve high-accuracy face recognition.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (No. 62172098), the Natural Science Foundation of Fujian Province (No. 2020J01497).

References

[1] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[2] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1322–1333.

[3] H. Cheng, X. Liu, H. Wang, Y. Fang, M. Wang, and X. Zhao, "Securead: A secure video anomaly detection framework on convolutional neural network in edge computing environment," *IEEE Transactions on Cloud Computing*, 2020.

[4] H. Cheng, H. Wang, X. Liu, Y. Fang, M. Wang, and X. Zhang, "Person re-identification over encrypted outsourced surveillance videos," *IEEE Transactions on Dependable and Secure Computing*, 2019.

[5] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[6] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," *arXiv preprint arXiv:1905.02383*, 2019.

[7] Z. Bu, J. Dong, Q. Long, and W. J. Su, "Deep learning with gaussian differential privacy," *Harvard data science review*, vol. 2020, no. 23, 2020.

[8] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 539–546.

[9] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 378–383.

[10] F. Fioretto and P. Van Hentenryck, "Privacy-preserving federated data sharing." in *AAMAS*, 2019, pp. 638–646.