Supplementary Material for "Privacy-Preserving Face Recognition with Learnable Privacy Budgets in Frequency Domain"

Jiazhen Ji[1], Huan Wang[2], Yuge Huang[1], Jiaxiang Wu[1], Xingkun Xu[1], Shouhong Ding[1], ShengChuan Zhang[2], Liujuan Cao[2], and Rongrong Ji[2]

[1] Youtu Lab, Tencent
[2] Xiamen University
{royji, yugehuang, willjxwu, xingkunxu, ericding}@tencent.com, hanawh@stu.xmu.edu.cn, {zsc_2016, caoliujuan, rrji}@xmu.seu.cn

This supplementary material provides additional details on the following:

- Detail proof of Lemma 1.
- Accuracy of different choice of ϵ.
- Visualization of the recovery images of different ϵ under black-box attack.
- Performance training with bigger training sets and testing with un-saturated testing sets.
- Parameters reuse under different privacy budgets.
- Robustness of our method.
- Discussion of the situation about choosing the dataset.

1 Proof of Lemma 1

Proof. To prove Lemma 1, we just need to prove that with these settings, Equation 4 satisfied. It is equivalent to prove

$$\prod_{i,j,k} Pr(\mathcal{K}(X^1_{i,j,k}) = E_{i,j,k}) \leq e^{\epsilon d(X_1,X_2)} \prod_{i,j,k} Pr(\mathcal{K}(X^2_{i,j,k}) \tag{1}$$

$$\prod_{i,j,k} \frac{e^{\frac{|X^1_{i,j,k}-E_{i,j,k}|}{\sigma_{i,j,k}}}}{2\sigma_{i,j,k}} \leq e^{\epsilon \max_{i,j,k}(d_{i,j,k}(x_1,x_2))} \prod_{i,j,k} \frac{e^{\frac{|X^2_{i,j,k}-E_{i,j,k}|}{\sigma_{i,j,k}}}}{2\sigma_{i,j,k}} \tag{2}$$

$$\prod_{i,j,k} e^{\frac{|X^2_{i,j,k}-E_{i,j,k}|-|X^1_{i,j,k}-E_{i,j,k}|}{\sigma_{i,j,k}}} \leq e^{\epsilon \max_{i,j,k}(d_{i,j,k}(x_1,x_2))} \tag{3}$$

$$\sum_{i,j,k} \frac{|X^2_{i,j,k}-E_{i,j,k}|-|X^1_{i,j,k}-E_{i,j,k}|}{\sigma_{i,j,k}} \leq \epsilon \max_{i,j,k}(d_{i,j,k}(x_1,x_2)) \tag{4}$$

Because

$$\sum_{i,j,k} \frac{|X^2_{i,j,k}-E_{i,j,k}|-|X^1_{i,j,k}-E_{i,j,k}|}{\sigma_{i,j,k}} \leq \sum_{i,j,k} \frac{|X^2_{i,j,k}, X^1_{i,j,k}|}{\sigma_{i,j,k}} \tag{5}$$

And

$$\sum_{i,j,k} \frac{|X^2_{i,j,k}, X^1_{i,j,k}|}{\sigma_{i,j,k}} = \sum_{i,j,k} \epsilon_{i,j,k} d_{i,j,k}(x_1, x_2) \tag{6}$$

So, it is less or equal to

$$\sum_{i,j,k} \epsilon_{i,j,k} \max_{i,j,k}(d_{i,j,k}(x_1, x_2)) = \epsilon \max_{i,j,k}(d_{i,j,k}(x_1, x_2)) \tag{7}$$

By now, we have proved that Equation 4 satisfied.

2 Performance for Different Privacy Budgets

To show the performance of different privacy budgets, we first test the accuracy of different privacy budgets. The results have been shown in Fig. 1. As you can see, with the lower choice of ϵ, the accuracy also becomes lower. However, the degree of its decline is still within the acceptable range. Then we perform the black-box attack on the model we trained with different ϵ to show the privacy of different ϵ. We visualized the results of the black-box attack in Fig. 2. It is visible that as the epsilon decreases, the recovered image becomes increasingly blurred and invisible.

Method (%)	LFW	CFP-FP	AgeDB-30	CALFW	CPLFW
ArcFace (Baseline)	99.60	98.32	95.88	94.16	92.68
Ours ($\epsilon_{mean} = 4$)	99.53	98.14	95.33	93.98	92.5
Ours ($\epsilon_{mean} = 2$)	99.52	97.36	94.46	93.6	90.86
Ours ($\epsilon_{mean} = 1$)	99.42	97.36	94.416	93.75	91.23
Ours ($\epsilon_{mean} = 0.5$)	99.48	97.20	94.36	93.47	90.6

Fig. 1. Comparison of the face recognition accuracy among different privacy budgets.

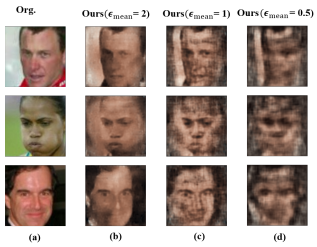

Fig. 2. Visualization of different privacy budget under black-box attack.

3 Performance training with bigger training sets and testing with un-saturated testing sets.

Part of the results training with MS1M are shown in Tab.1. ArcFace reported rank-1 accuracy 98.14% and verification TAR 98.34% at $1e^{-6}$ FAR on MegaFace with ResNet50 trained from MS1M. In comparison, our method maintains a good accuracy, *i.e.*, 97.01% and 97.53% respectively.

Method (%)	LFW	IJB-B(1e-4)	IJB-C(1e-4)	MegaFace (Id, R)	MegaFace (Ver, R)
ArcFace (Baseline)	99.76	93.37	95.12	98.14	98.34
Ours ($\epsilon_{mean} = 0.5$)	99.55	91.83	93.70	97.01	97.53

Table 1. Comparison of recognition accuracy on different benchmarks.

4 Parameters reuse under different privacy budgets.

Since the average privacy budget is the only hyper-parameter in the perturbation, our method only needs to learn a budget allocation module. As a result, our method does not need to retrain the face recognition model when changing the privacy budget. In our experience, this allocation module is relatively stable under different privacy budgets. As shown in Tab.2, the performance is similar when the same privacy budget is used in the inference phase; even the allocation modules are trained with different budgets.

Method (%)	LFW	CFP-FP	AgeDB-30	CALFW	CPLFW
($\epsilon_{train} = 2$, $\epsilon_{test} = 2$)	99.52	97.36	94.46	93.60	90.86
($\epsilon_{train} = 0.5$, $\epsilon_{test} = 2$)	99.53	97.24	94.56	93.56	90.98

Table 2. Comparison of accuracy with different privacy budgets in training and testing.

Method (%)	LFW	CFP-FP	AgeDB-30	CALFW	CPLFW
Baseline	99.70 (0.07)	98.48 (-0.04)	95.80 (-0.03)	94.23 (0.26)	93.48 (0.18)
Ours ($\epsilon_{mean} = 0.5$)	99.51 (0.03)	96.77 (-0.43)	94.16 (-0.21)	93.4 (-0.07)	90.97 (0.37)

Table 3. Accuracy after adding noise to inputs.

5 Robustness of our method.

To verify the robustness of our method, we perform random sampling using the parameters $N(0, 0.01)$ and multiplying it with the range of original images and frequency domain representations as to their additional noise and testing their recognition accuracy, respectively. Tab.3 shows in parentheses the change in accuracy after adding noise compared to before. Even with a very small privacy budget, the accuracy degradation of our method compared to the baseline is only slightly more and still within acceptable limits.

6 Reasons for using VGGFace2 as our training set.

Our research aims to help protect privacy issues in face recognition, so it is inevitable that relevant face recognition datasets are used in the training process. Most of the currently publicly available face recognition datasets haven't been approved by IRB. Considering the amount of data, the quality of the data and the level of acceptance by researchers we had to use VGGFace2 as our dataset. However, I believe our work will be helpful to subsequent researchers in face recognition when using relevant datasets, i.e., using the privacy-preserving datasets that transformed by our method.