

# Difficulty in estimating visual information from randomly sampled images

1<sup>st</sup> Masaki Kitayama  
Tokyo Metropolitan University,  
Tokyo, Japan

2<sup>nd</sup> Nobutaka Ono  
Tokyo Metropolitan University,  
Tokyo, Japan

3<sup>rd</sup> Hitoshi Kiya  
Tokyo Metropolitan University,  
Tokyo, Japan

**Abstract**—In this paper, we evaluate dimensionality reduction methods in terms of difficulty in estimating visual information on original images from dimensionally reduced ones. Recently, dimensionality reduction has been receiving attention as the process of not only reducing the number of random variables, but also protecting visual information for privacy-preserving machine learning. For such a reason, difficulty in estimating visual information is discussed. In particular, the random sampling method that was proposed for privacy-preserving machine learning, is compared with typical dimensionality reduction methods. In an image classification experiment, the random sampling method is demonstrated not only to have high difficulty, but also to be comparable to other dimensionality reduction methods, while maintaining the property of spatial information invariant.

## I. INTRODUCTION

Recently, it has been very popular to utilize cloud servers to carry out machine learning algorithms instead of using local servers. However, since cloud servers are semi-trusted, private data, such as personal information and medical records, may be revealed in cloud computing. For the reason, privacy-preserving machine learning has become an urgent challenge [1]–[5]. In this paper, we focus on dimensionality reduction methods in terms of two issues: difficulty in estimating visual information on original images from dimensionally reduced ones, and performance that reduced data can maintain in an image classification experiment. In machine learning, dimensionality reduction is used for not only reducing the number of random variables, but also protecting visual information for privacy-preserving machine learning. However, dimensionality reduction methods have never been evaluated in terms of above the two issues at same time.

For such a reason, difficulty in estimating visual information is discussed. In particular, the random sampling method that was proposed for privacy-preserving machine learning [6], is compared with typical dimensionality reduction methods such as random projection and PCA [7], [8]. In an image classification experiment, the random sampling method is demonstrated not only to maintain high difficulty, but also to have close machine learning performance to that of the random projection method.

## II. LINEAR DIMENSIONALITY REDUCTIONS

Let us consider a projection from a vector  $x \in \mathbb{R}^D$  to a low-dimensional vector  $y \in \mathbb{R}^K$  ( $K < D$ ). If the projection can be represented by using a matrix  $P \in \mathbb{R}^{K \times D}$  as

$$y = Px \quad , \quad (1)$$

it is a linear dimensionality reduction and  $P$  is called a projection matrix. In machine learning,  $P$  is used for reducing the number of random variables for avoiding negative effects of high-dimensional data. The random projection method [7] and principal component analysis (PCA) are typical linear dimensionality reduction methods. The random projection is a method that does not use any statistics of dataset, but PCA is not. For the random projection, elements of  $P$  have a normal distribution with an average value of 0 and a variance of  $\sqrt{1/K}$ . Therefore, the random projection is not required to calculate any statistics of dataset for designing a projection matrix  $P$ .

## III. RANDOM SAMPLING

The random sampling method was proposed as a dimensionality reduction method for privacy preserving machine learning [9]. It is also expected to be applied to deep convolutional neural network, due to the property of spatial information invariant [6].

Let us consider applying the random sampling method to a pixel vector  $x \in \mathbb{R}^D$  of an image to create  $y \in \mathbb{R}^K$  ( $K < D$ ). Next, let  $\{\phi(i) \mid i = 1, \dots, K\}$  denote  $K$  indexes selected from  $D$  pixel indexes, where  $\phi(i) \neq \phi(i')$  if  $i \neq i'$ , randomly generated with a seed. By using  $\phi(i)$ , the random sampling operation can be written as

$$y = (x_{\phi(1)}, x_{\phi(2)}, \dots, x_{\phi(K)})^T \quad , \quad (2)$$

where  $x_{\phi(i)}$  is the  $\phi(i)$ -th element of  $x$ . Here, if we define a matrix  $P \in \mathbb{R}^{K \times D}$  with elements  $p_{ij}$  ( $i = 1, \dots, K, j = 1, \dots, D$ ) defined by

$$p_{i,j} = \begin{cases} 1 & (j = \phi(i)) \\ 0 & (\text{otherwise}) \end{cases} \quad , \quad (3)$$

the random sampling is reduced to the form of Eq.(1). That is, the random sampling is a linear dimensionality reduction, and is a method that does not use the statistics of dataset as well as the random projection.

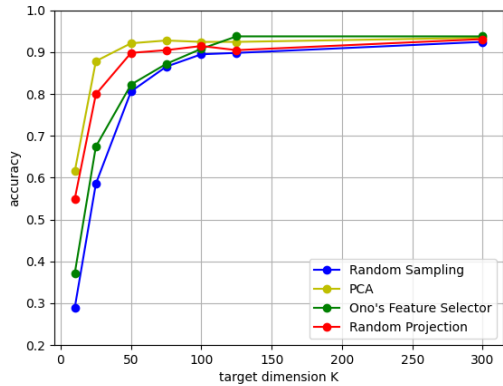


Fig. 1. Classification accuracy with various dimensionality reduction methods (Linear SVM).

#### IV. VISUAL INFORMATION ESTIMATION

Assuming that an attacker knows a projection matrix  $P$  used in dimensionality reduction, difficulty in estimating visual information on plain images is discussed. The attacker's goal is to create  $Q$  to approximately reconstruct the target image  $x$  from the low dimensional vector  $y$  as

$$x' = Qy \quad (4)$$

In this paper, two attacks are considered to estimate  $Q$ .

##### A. Attack With Pseudo-Inverse Matrix

An attacker can use a pseudo-inverse matrix ( $Q_{\text{pinv}}$ ) of projection matrix  $P$  to estimate visual information on original images, where  $Q_{\text{pinv}}$  is designed by using an algorithm with the singular-value decomposition of  $P$  [10].

##### B. Regression Attack With Attacker's Dataset

An attacker first prepares his own dataset ( $X_{\text{attack}}$ ) and a dataset ( $Y_{\text{attack}}$ ) projected from  $X_{\text{attack}}$  by using  $P$ , and then designs a linear reconstruction matrix ( $Q_{\text{reg}}$ ) that regresses  $X_{\text{attack}}$  from  $Y_{\text{attack}}$  in accordance with the least squares method. In general, the effectiveness of this attack depends on the relation between the distribution of  $X_{\text{attack}}$  and that of target images. Therefore, in this paper, we classify  $X_{\text{attack}}$  into two types in accordance with the distribution of  $X_{\text{attack}}$ .

- type 1:  $X_{\text{attack}}$  consists of images with the same class-labels and distribution as those of the target images.
- type 2:  $X_{\text{attack}}$  consists of images with class-labels and a distribution that are different from those of the target images.

#### V. EXPERIMENT

Face-image classification experiments were carried out for evaluating the random sampling method in terms of both classification accuracy and difficulty in estimating visual information. The dataset was Extended Yale

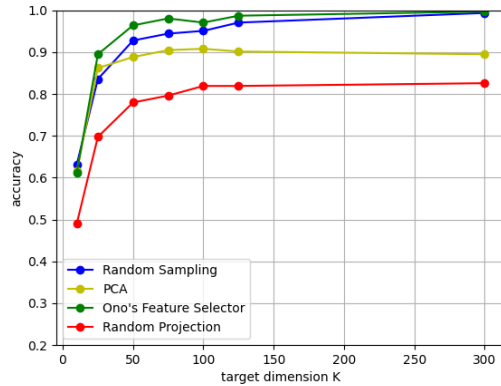


Fig. 2. Classification accuracy with various dimensionality reduction methods (Random Forest).

Database B [11], which contains 38 individuals and 64 frontal facial images with  $168 \times 192$  pixels per each person. Each image was normalized to a size of  $28 \times 28$ , so it had  $D = 784$  dimension as a vector, we splitted the dataset into two datasets:  $X_{\text{main}}$  and  $X_{\text{sub}}$ , each of which had 19 classes and 1216 images, without duplication of classes. Moreover,  $X_{\text{main}}$  was divided into  $X_{\text{train}}$  (912 images) for training and  $X_{\text{test}}$  (304 images) for testing.

We also used the CIFAR-10 [12] dataset:  $X_{\text{CIFAR-10}}$  for evaluating difficulty in estimating visual information on  $X_{\text{test}}$ . This dataset consists of 60k images with 10 classes such as dogs and ships, whose distribution is different from  $X_{\text{train}}$ ,  $X_{\text{test}}$  and  $X_{\text{sub}}$ .

Finally, each vector  $x \in \{X_{\text{train}}, X_{\text{test}}, X_{\text{sub}}, X_{\text{CIFAR-10}}\}$  was projected to  $y \in \{Y_{\text{train}}, Y_{\text{test}}, Y_{\text{sub}}, Y_{\text{CIFAR-10}}\}$  with a target dimension ( $K$ ) by using the random sampling and three dimensionality reduction methods: the random projection, PCA, and a feature selection algorithm proposed by Ono [13]. PCA and Ono's method require calculating the statistics of  $X_{\text{train}}$ , but the random sampling and random projection do not.

##### A. Machine Learning Performance

We trained a random forest classifier and SVM with the linear kernel by using  $Y_{\text{train}}$ , and tested by using  $Y_{\text{test}}$ . Figures 1 and 2 show the comparison of the dimensionality reduction methods in term of classification accuracy. Under the use of SVM, the random sampling had a similar performance to Ono's method. For the random forest, the random sampling also has almost the same accuracy as that of Ono's method. As a result, the random sampling was demonstrated to be comparable with other dimensionality reduction methods, while maintaining the property of spatial-information invariant.

##### B. Robustness Against Visual Information Estimation

Assuming that an attacker knows only projection matrix  $P$  used in dimensionality reduction, difficulty in estimating

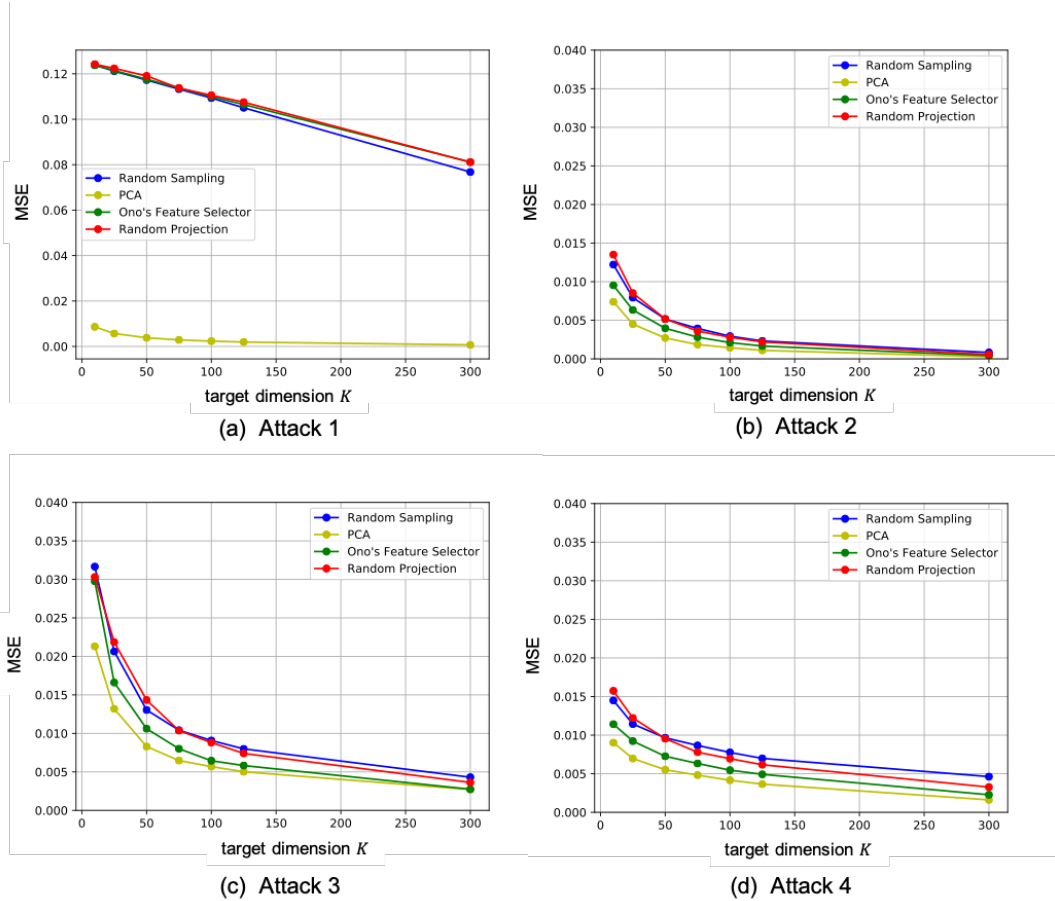


Fig. 3. Mean squared error between original image and reconstructed image.

visual information on  $X_{\text{test}}$  was evaluated. We assumed the following four attacks to estimate  $X_{\text{test}}$  from  $Y_{\text{test}}$ .

- Attack 1: Attack using a pseudo-inverse matrix of  $P$ .
- Attack 2: Regression attack with  $X_{\text{train}}$  (type 1 in section IV-B).
- Attack 3: Regression attack with  $X_{\text{CIFAR-10}}$  (type 2 in section IV-B).
- Attack 4: Regression attack with  $X_{\text{sub}}$ .

In attack 4, the attacker does not have facial images of the people included in  $X_{\text{test}}$ , but knows the conditions under which  $X_{\text{test}}$  was taken.

Figure 3 shows MSE values between  $X_{\text{test}}$  and  $X'_{\text{test}}$ . From the figure, the random sampling was relatively robust compared with the other dimensionality reduction methods. The absolute values of MSE of the random sampling were large for attacks 1 and 3, but were small for attacks 2 and 4 as well as the other methods.

We defined accuracy reduction ratio (ARR) as another criterion. First, we trained a logistic regression classifier ( $\theta$ ) by using  $X_{\text{train}}$ , and then the ARR is defined as

$$\text{ARR} = \frac{\text{ACC}_{\theta}(X_{\text{test}}) - \text{ACC}_{\theta}(X'_{\text{test}})}{\text{ACC}_{\theta}(X_{\text{test}})}, \quad (5)$$

where  $\text{ACC}_{\theta}(X)$  is the function which returns the accuracy when dataset  $X$  is applied to  $\theta$ . A high ARR value indicates that  $X'_{\text{test}}$  has low class-specific visual information, i.e., the dimensionality reduction is robust against the attack.

Figure 4 shows the comparison of ARR values. From the figure, the random sampling was demonstrated to be robust against attack 4. It indicates that attack 4 did not effectively recover the class-specific information of the target images.

Figure 5 shows an original images and examples of reconstructed images. Although the images reconstructed by attack 4 can be easily interpreted as human faces, the facial features were different from the original person.

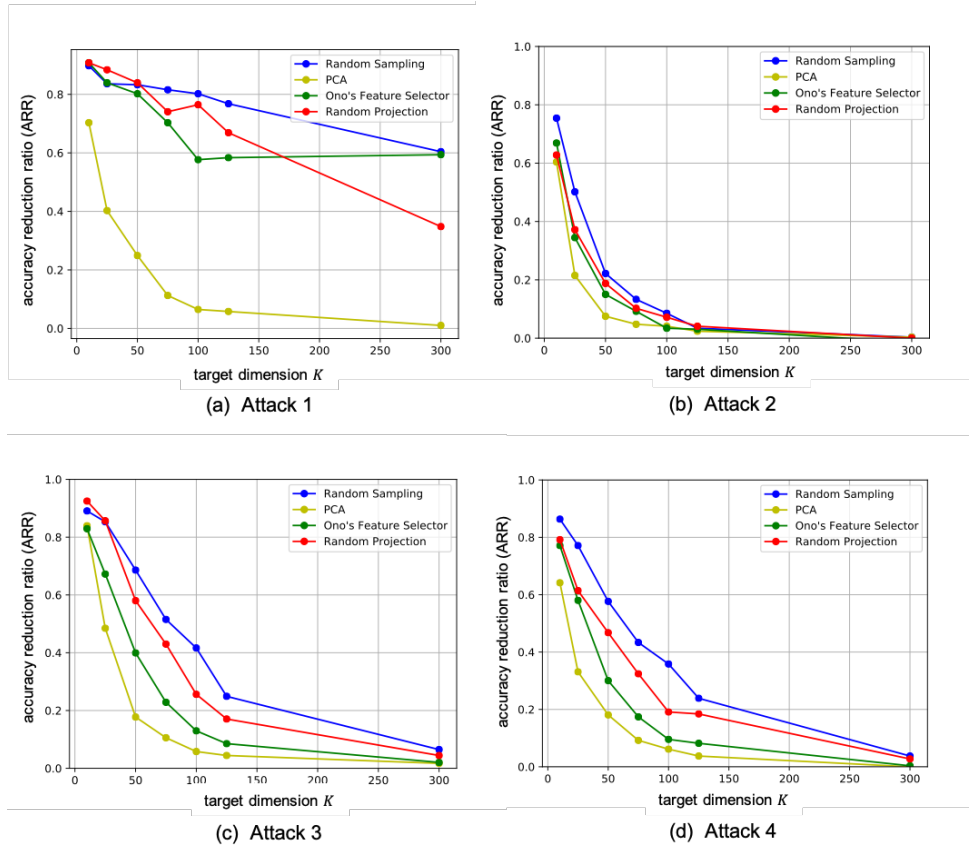


Fig. 4. Comparison of accuracy reduction ratio values.

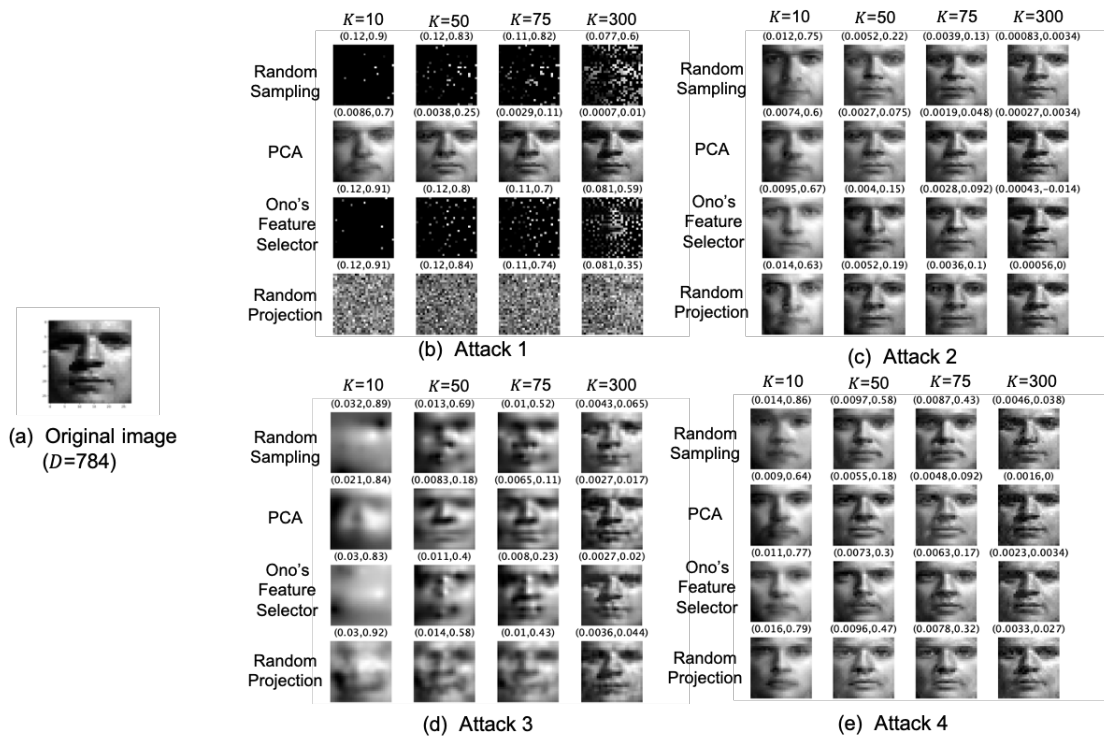


Fig. 5. Examples of reconstructed images ( $D=784$ ) with MSE (left) and ARR (right) values.

## VI. CONCLUSION

In this paper, we compared the random sampling method with typical linear dimensionality reduction methods in terms of both machine learning performance and difficulty in estimating visual information in face classification experiments. The random sampling was demonstrated not only to have the property of spatial position invariant which is useful for privacy-preserving learning, but also to maintain comparable performance to other dimensionality reduction methods and difficulty in visual information estimation under practical conditions.

## REFERENCES

- [1] Wenjie Lu, Shohei Kawasaki, and Jun Sakuma, "Using fully homomorphic encryption for statistical analysis of categorical, ordinal and numerical data.," *IACR Cryptology ePrint Archive*, vol. 2016, pp. 1163, 2016.
- [2] Ayana Kawamura, Yuma Kinoshita, Takayuki Nakachi, Sayaka Shiota, and Hitoshi Kiya, "A privacy-preserving machine learning scheme using etc images," *arXiv:2007.08775*. [Online]. Available: <https://arxiv.org/abs/1911.00227>, Nov. 2020.
- [3] Warit Sirichotedumrong, Takahiro Maekawa, Yuma Kinoshita, and Hitoshi Kiya, "Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain," in *Proc. IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 674–678.
- [4] Warit Sirichotedumrong, Yuma Kinoshita, and Hitoshi Kiya, "Pixel-based image encryption without key management for privacy-preserving deep neural networks," *IEEE Access*, vol. 7, pp. 177844–177855, 2019.
- [5] Warit Sirichotedumrong and Hitoshi Kiya, "A gan-based image transformation scheme for privacy-preserving deep neural networks," *arXiv:2006.01342*. [Online]. Available: <https://arxiv.org/abs/2006.01342>, Jun. 2020.
- [6] MaungMaung AprilPyone and Hitoshi Kiya, "Encryption inspired adversarial defense for visual classification," To be appeared in *IEEE International Conference on Image Processing 2020*. *arXiv:2005.07998*. [Online]. Available: <https://arxiv.org/abs/2005.07998>, May 2020.
- [7] Ella Bingham and Heikki Mannila, "Random projection in dimensionality reduction: applications to image and text data," in *Proc. of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 245–250.
- [8] Svante Wold, Kim Esbensen, and Paul Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [9] Ayana Kawamura, Kenta Iida, and Hitoshi Kiya, "A dimensionality reduction method with random sampling for privacy-preserving machine learning," Tech. Rep., IEICE, vol.119, no.335, (no.SIS2019-26), 2019, (in Japanese).
- [10] David A Harville, "Matrix algebra from a statistician's perspective," 1998.
- [11] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
- [12] Alex Krizhevsky, Geoffrey Hinton, et al., "Learning multiple layers of features from tiny images," 2009.
- [13] Nobutaka Ono, "Dimension reduction without multiplication in machine learning," Tech. Rep., IEICE, Vol. 119, no. 440, (no.SIP2019-106), 2020, (in Japanese).