# Learnable Pixel-Based Encryption for Privacy-Preserving Image Classification

Ijaz Ahmad, Seokjoo Shin*

Dept. of Computer Engineering, Chosun University

ahmadijaz@chosun.kr, *sjshin@chosun.ac.kr (Corresponding author)

# 개인정보 보호 이미지 분류를 위한 학습 가능한 픽셀 기반 암호화

아흐마드 이자즈, 신석주
조선대학교 컴퓨터공학과

## Abstract

When outsourcing Deep Learning (DL) computations to a third-party, it is necessary to protect the privacy sensitive data during transmission as well as when performing computations on them. Learnable encryption techniques have been proposed that protect identifiable information in images and can enable various computer vision applications in the encrypted domain. However, there are known attacks against simple learnable encryption methods that preserve a DL model performance while for the secure methods the model accuracy is significantly reduced. In this work, we propose a novel transformation function to mitigate the aforementioned limitations of the learnable encryption schemes. Our method performs substitution in such a way that the diffusion property of encryption is ensured. The proposed method allows lossless construction of privacy-preserving deep learning models for the classification tasks. The analysis on CIFAR10 dataset shows that compared to the conventional secure learnable encryption schemes our method reduced the error in accuracy from 29.6% to 7.11% while providing a necessary level of security.

## Ⅰ. Introduction

Cloud computational resources provide a cost-effective solution to the training of powerful Deep Learning (DL) models. However, when data is outsourced to the avail of such third-party owned resources it is necessary to protect the data during transmission and computations. In the latter case the data protection is required to hide identifiable information in images and also to protect the data ownership. Learnable encryption is one of the approaches towards privacy-preserving deep learning. Several such algorithms have been proposed that find a better tradeoff between security and the data utility [1]. Simple learnable encryption methods such as [2], [3], are successful in preserving DL model performance; however, these methods are not secure. On the other hand, for secure learnable encryption methods such as [4], the DL model performance is severely suffered.

The conventional learnable encryption methods [2], [3] are vulnerable to chosen-plaintext attack as demonstrated in [5]. Though [3] was made resistant against such an attack in [6] by modifying the selected pixels using a sequence of random values; however, their secure key can still be recovered simply by encrypting a blank image. In this paper, we propose a novel transformation function for learnable encryption methods. In our method, the substitution of pixel values is not only dependent on the secret key but on previously encrypted pixel values as well. Thereby, guarantees the diffusion property of encryption. The main advantage of the proposed method is the lossless construction of DL model, which makes it compatible with the state-of-the-art DL models.

## Ⅱ. Proposed Learnable Encryption Method

Following CPBE, we use a sequence of values $[0,255]$. For this purpose, we utilize Logistic map [7] given by

$$d_i = \mu d_{i-1}(1 - d_{i-1}), \quad (1)$$

where, $\mu$ is the control parameter of the chaotic map. For $\mu \in [3.57,4]$, the system is chaotic. The Lyapunov exponent for $\mu = 3.57$ is 0.0012 and for $\mu = 4$ is 0.0693. Both the parameter $\mu$ and the initial value $d_0$ serve as the secret key in the substitution stage. The choice of Logistic map is based on its low complexity. Nonetheless, it can be readily replaced with a more complex chaotic map.

To achieve the diffusion property in learnable encryption methods, we have adopted the pixel substitution function proposed in [8]. The $i^{\text{th}}$ pixel value $c_i$ in the cipher image is obtained as

$$c_i = s_i \oplus \{[p_i + s_i] \,\%\, 256\} \oplus c_{i-1}, \quad (2)$$

where $p_i$ is the current pixel value in the plain image and $c_{i-1}$ is the previously encrypted pixel value. $s_i$ is the $i^{\text{th}}$ element in the sequence obtained as:

$$s_i = \left(\left\lfloor \left(\frac{d_i}{2}\right) \times 10^{14} \right\rfloor\right) \% \, 256. \quad (3)$$

To obtain a learnable cipher image, proposed method modifies each pixel value $p_{(x,y)}$ in the plain image by performing a negative and positive transformation based on (2) as:

$$c_{(x,y)} = \begin{cases} p_{(x,y)} \oplus c_{(x-1,y)} & k_i = 0 \\ s_i \oplus \{[p_{(x,y)} + s_i] \,\%\, 256\} \oplus c_{(x-1,y)} & k_i = 1' \end{cases} \quad (4)$$

where $k_i \in K$ (for $i = 1,2,\ldots,N$ and $N$ being the number of pixels in the plain image), is a binary uniformly distributed key. This binary key adds an additional layer of randomness to the Logistic map to compensate for its simplicity by utilizing only half of the values.

In the conventional methods [3], [6], for $k_i = 0$, (4) is an identity map. For $k_i = 1$, (4) gives $p_{(x,y)} \oplus 255$ in [3] and $p_{(x,y)} \oplus s_{(i)}$ in [6]. Each pixel is independently encrypted thus lacks the diffusion property [3], [6]. Conversely, proposed method exhibits the diffusion property as the cipher image depends on both the keystream elements and all of the previous pixel values.

## III. Results

This section presents our experimental setup and analysis results. The classifier performance was compared in both plain and encrypted domain. The learnable cipher images were obtained from Learnable Encryption (LE) [2], Extended Learnable Encryption (ELE) [4], Pixel-based Encryption (PBE) [3], Chaotic PBE (CPBE) [6] and the proposed learnable encryption methods. The CPBE utilizes the whole range **[0,255]**; therefore, we considered encrypting all pixels (CPBE 100%) and half of the pixels (CPBE 50%). Similarly, we implemented two variations of the proposed method: 'Proposed 50%' where for $k_i = 0$ in (4), the transformation function was set to identity map and 'Proposed 100%' which utilized the transformation as it is in (4). For visual analysis, Fig. 1. shows example cipher images of the learnable encryption methods [3], [6] that are closely related to the proposed method. The proposed method masked the image contents well.

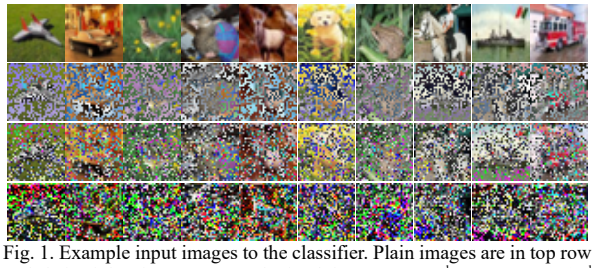

Fig. 1. Example input images to the classifier. Plain images are in top row and their cipher images were obtained from PBE (2nd row), CPBE (3rd row) and proposed (4th row) learnable encryption methods. The image contents are better masked in the proposed method.

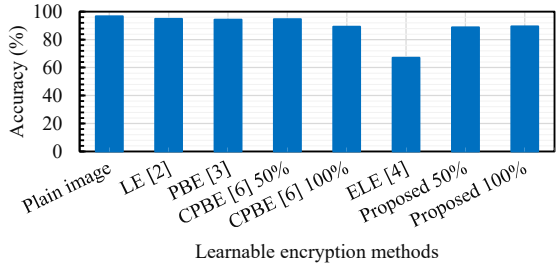


Fig. 2. Performance analysis of the DL-based classifier on different learnable cipher images.

For a fair comparison, we have used the same classification DL model with the same training parameters as in [4]. The simulations were carried out on CIFAR10 dataset [9], which consists of 60K true color images of 32×32 resolution. The images belong to 10 classes and are divided as 50K for training and 10K for test sets. In addition, we have separated 10% of training images as our validation set. Fig. 2. shows the privacy-preserving classification accuracy of different learnable encryption methods. The performance is reported as percentage accuracy. The methods [2], [3], [6] have preserved the model performance; however, they have weaker security. On the other hand, the most secure method ELE has the worst accuracy. Compared to which the proposed method improved the accuracy by 7%. In addition, the CPBE performance degraded by 5% when all the pixels were encrypted (CPBE 100%). However, the proposed method has achieved the same accuracy regardless of the number of pixels being encrypted.

## III. Conclusion

This study proposed a learnable encryption method for privacy-preserving classification task that allowed lossless construction of DL models. The simulation analysis showed that a better tradeoff between security and DL model performance has been achieved.

Integration of the proposed transformation function with block-based learnable encryption methods to satisfy other requirements such as compression, can be an interesting research direction.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (RS-2023-00278294).

## REFERENCES

[1] R. El Saj, E. Sedgh Gooya, A. Alfalou, and M. Khalil, "Privacy-Preserving Deep Neural Network Methods: Computational and Perceptual Methods—An Overview," Electronics, vol. 10, no. 11, p. 1367, Jun. 2021, doi: 10.3390/electronics10111367.

[2] M. Tanaka, "Learnable Image Encryption," ArXiv180400490 Cs, Mar. 2018, Accessed: Dec. 20, 2021. [Online]. Available: http://arxiv.org/abs/1804.00490

[3] W. Sirichotedumrong, Y. Kinoshita, and H. Kiya, "Pixel-Based Image Encryption Without Key Management for Privacy-Preserving Deep Neural Networks," IEEE Access, vol. 7, pp. 177844–177855, 2019, doi: 10.1109/ACCESS.2019.2959017.

[4] K. Madono, M. Tanaka, M. Onishi, and T. Ogawa, "Block-wise Scrambled Image Recognition Using Adaptation Network," 2020, doi: 10.48550/ARXIV.2001.07761.

[5] A. H. Chang and B. M. Case, "Attacks on Image Encryption Schemes for Privacy-Preserving Deep Neural Networks," ArXiv200413263 Cs, Apr. 2020, Accessed: Dec. 20, 2021. [Online]. Available: http://arxiv.org/abs/2004.13263

[6] I. Ahmad and S. Shin, "A Pixel-based Encryption Method for Privacy-Preserving Deep Learning Models," 2022, doi: 10.48550/ARXIV.2203.16780.

[7] R. L. Devaney, An introduction to chaotic dynamical systems, 2nd ed. in Addison-Wesley studies in nonlinearity. Redwood City, Calif: Addison-Wesley, 1989.

[8] I. Ahmad and S. Shin, "A novel hybrid image encryption-compression scheme by combining chaos theory and number theory," Signal Process. Image Commun., vol. 98, p. 116418, Oct. 2021, doi: 10.1016/j.image.2021.116418.

[9] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," p. 60.