APPLYING DEEP LEARNING TO KNOWN-PLAINTEXT ATTACK ON CHAOTIC IMAGE ENCRYPTION SCHEMES

Fusen Wang[⋆] *Jun Sang* [⋆] *Chunlin Huang*[⋆] *Bin Cai* [⋆] *Hong Xiang* [⋆] *Nong Sang* [†]

[⋆] Key Laboratory of Dependable Service Computing in Cyber Physical Society of Ministry of Education, Chongqing University, Chongqing, China
[†] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

ABSTRACT

In this paper, we demonstrate that traditional chaotic encryption schemes are vulnerable to the known-plaintext attack (KPA) with deep learning. Considering the decryption process as image restoration based on deep learning, we apply Convolutional Neural Network to perform known-plaintext attack on chaotic cryptosystems. We design a network to learn the operation mechanism of chaotic cryptosystems, and utilize the trained network as the decryption system. To prove the effectiveness, we select three existing chaotic encryption schemes as the attacked targets. The experimental results demonstrate that deep learning can be applied to known-plaintext attack against chaotic cryptosystems successfully. Compared with traditional attack methods for chaotic cryptosystems, the proposed method shows obvious advantages: (1) One neural network may be applied to cryptanalysis of various chaotic cryptosystems, not limited to specific one; (2) the proposed method is significantly convenient and cost-efficient. This paper provides a new idea for the cryptanalysis of chaotic cryptosystems.

Index Terms— Chaotic image encryption, Cryptanalysis, Known-plaintext attack, Convolutional Neural Network

1. INTRODUCTION

Many chaotic image encryption algorithms have been proposed during the past years [1, 2, 3, 4, 5, 6]. Meanwhile, some attack methods against specific chaotic encryption algorithms have been proposed [7, 8, 9]. Dou et al. [7] proposed an effective attack method against the one-dimensional combined chaotic color image encryption algorithm [10] in the case of unknown parameters. Li et al. [8] proposed a chosen-plaintext attack strategy against one-dimensional bit-level chaotic color image encryption algorithm [11] and proved the effectiveness and feasibility of the method. Although the ciphertext image can be accurately restored to the original plaintext image, these cryptanalysis methods also have some shortcomings: (1) It is complicated to design an attack scheme; (2) Usually, one attack method is only designed for a specific chaotic encryption system, which is hard to be applied to other chaotic encryption systems.

In recent years, deep learning has been involved in the domain of cryptography, especially in image encryption [12], [13]. Besides, Hai et al. [14] firstly put forward to apply deep learning to the cryptanalysis of optical encryption schemes for decryption.

Stimulated by Ref. [14], in this paper, we propose a known-plaintext attack method based on deep learning for chaotic cryptosystems. Chaotic image encryption schemes aim to replace the pixel arrangement by permutation operation and change the pixels values by diffusion operation to hide image content. Chaotic image encryption usually contains many nonlinear processes and complex operators, so that each pixel value is correlated with multiple pixels after encryption. It is very difficult to analyze such an image encryption method by mathematical derivation. Neural networks have been widely used to capture the images global and local context dependence in image segmentation, image restoration tasks, etc., so they can also be trained to serve as the inverse operator through implicit expression. Different from Ref. [14], we design a convolutional neural network as the decryption model referring to image restoration network [15] due to their similarity. Although such method cannot accurately recover the ciphertext images to the original ones, we only need to restore them to be sufficiently similar to the original plaintext images since the images allow a certain degree of distortion. Furthermore, such method may be used for more chaotic cryptosystems instead of a specific chaotic cryptosystem.

The main contributions of our work are outlined as follows:

(1) Deep learning is applied to known-plaintext attack on chaotic cryptosystems.

(2) A convolutional neural network is designed to attack various chaotic encryption algorithms. Specifically,

Corresponding author: jsang@cqu.edu.cn

three existing chaotic encryption schemes are selected for experiments and the high-quality decrypted images are obtained.

(3) It is concluded that, different from the traditional known-plaintext attack methods for chaotic cryptosystems, a convolutional neural network may be employed to decrypt different chaotic cryptosystems.

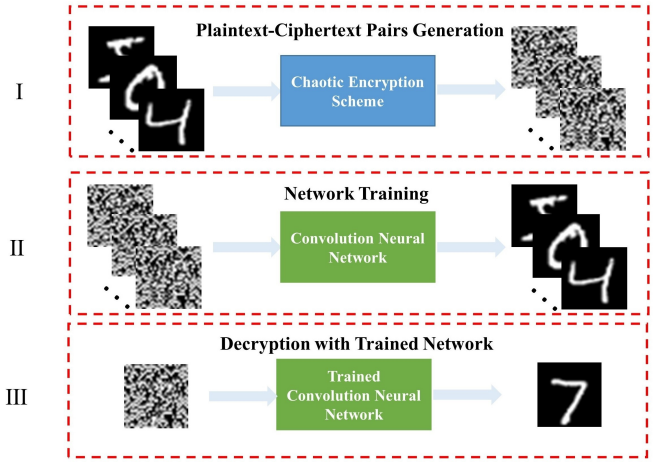


Fig. 1: The overall framework of deep learning-based known-plaintext attack on chaotic cryptosystem.

2. THE PROPOSED DEEP LEARNING-BASED ATTACK METHOD

2.1. Description of the proposed method

The overall framework of our method is shown in Fig. 1, which consists of three parts: (I) Plaintext-Ciphertext Pairs Generation; (II) Network Training; (III) Decryption with Trained Network. For Part I, chaotic encryption algorithms are adopted to encrypt images to generate sufficient "plaintext-ciphertext" image pairs for known-plaintext attack. For Part II, we train the network with the aforementioned "plaintext-ciphertext" image pairs. For Part III, the trained model is regarded as the "equivalent key" of the cryptographic system, which is used to decrypt the subsequent ciphertext image. (Please note that the trained model is limited to cracking ciphertext images under the same encryption algorithm and secret keys, which means that the trained model can be implicitly regarded as the inverse operation of the encryption system)

2.2. Network of the proposed method

Decryption of ciphertext images is quite similar to some classic tasks in deep learning, such as image restoration [15], image denoising [16], and image defogging [17], in which their purposes are always to restore some fuzzy or even invisible images as clearly as possible to recognizable images with the same resolution as the original images. Therefore, we refer to image restoration model [15] and design an Image Decryption neural network with Encoder-Decoder structure (IDEDNet).

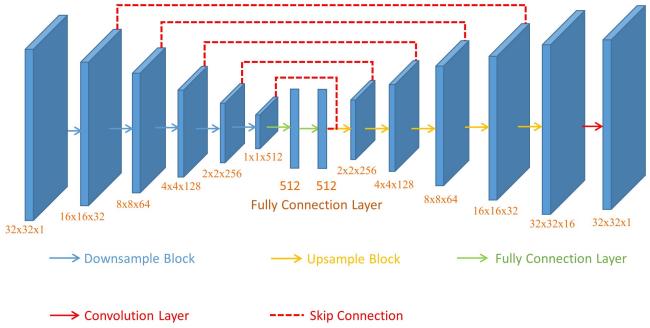


Fig. 2: The architecture of the proposed image decryption encoder-decoder network IDEDNet: HxWxC represents the image height, width, and channel number respectively.

Fig. 2 displays the architecture of our proposed network IDEDNet. The encoding layers (with blue arrows) of the network consist of five down-sampling blocks. Each down-sampling block is composed of Conv-BN-ReLU-AvgPool-Conv-BN-ReLU layers, which represent standard convolutional layer (Conv) with filter size of 3x3, batch normalization (BN), rectified linear unit (ReLU), an average pooling layer with stride 2x2, respectively. Similar to SENet [18], the obtained feature maps go through two fully connected layers (with green arrows) to capture cross-channel dependencies. It can learn the nonlinear relationship between channels.

Then, the fine feature maps are transmitted to five up-sampling blocks to perform per-pixel regression and expand the resolution in the decoder layers (with yellow arrows). Each up-sampling block consists of DConv-Conv-BN-ReLU-Conv-BN-ReLU layers, where DConv represents the deconvolutional layer with filter size of 2x2 and stride 2x2, while the other layers are similar to those in the encoder. Besides, considering that simple deconvolution operation cannot restore the information lost in the previous pooling layer, skip connections [19] (with red dotted lines) are adopted to supplement details of encoder to the estimated plaintext, which can settle the issue of gradient vanishing and achieve better decryption results. Finally, a 1x1 convolutional layer (with a red arrow) is used to generate the estimated decrypted image.

3. IMPLEMENTATION DETAILS

In this section, we describe the implementation details of the proposed method, including the chaotic encryption schemes and the training details.

Table 1: The ciphertext reconstruction result of known-plaintext attack method based on deep learning on MNIST and MNIST-Fashion datasets.

Network	Encryption Scheme	dataset	Training correlation coefficient	Testing correlation coefficient	Epoch	Time/ Epoch
IDEDNet	Song et al. [5]	MNIST and Fashion	97.6%	94.2%	300	5.8s
	Pak et al. [10]		97.7%	94.5%		5.7s
	H. N. Abdullah et al. [4]		98.6%	96.7%		5.8s

3.1. Loss Function

To minimize the information loss between the decrypted image and the plaintext image, the absolute value loss function (L1loss) is used to supervise network training in this paper, which is defined as follows:

$$Loss_1 = \frac{1}{N}\sum_{i=1}^{N} |O(C_i;\theta) - P(C_i)| \tag{1}$$

where N is the number of the training image batch. C_i, $O(C_i;\theta)$, and $P(C_i)$ are the i-th input (ciphertext image), the i-th output (decrypted image) with parameters θ, and i-th ground truth (original plaintext image) of the network, respectively.

3.2. Training details

Adam optimizer is employed with an L2 regularization weight decay rate of 1e-4. The initial learning rate is 1*e-3 and it drops by 10% every 20 epochs when training with a batch of size 64. In addition, a dropout layer with a ratio of 0.5 is added to the partial convolutional layer to prevent the network from overfitting.

3.3. Chaotic encryption systems

To verify our proposed deep learning-based attack method, we apply it to three existing chaotic encryption schemes: (1) A secure image encryption algorithm based on multiple one-dimensional chaotic systems [5] (Song et al.); (2) A new color image encryption using combination of the 1D chaotic map [10] (Pak et al.); (3) Image encryption using hybrid chaotic map [4] (H. N. Abdullah et al.). Please refer to the original paper for specific encryption details and the following three schemes are denoted as the names of the authors.

4. EXPERIMENTS

All experiments are implemented with NVIDIA GTX 1060Ti-Python3.5-Pytorch1.7.

In this paper, we generate 5000 "plaintext-ciphertext" pairs via the above three encryption schemes on the mixed dataset of MNIST [20] and MNIST-Fashion [21], as small size images with 28x28 can reduce the burden on the GPU and speed up network training. The purpose of the mixed datasets of MNIST and MNIST-Fashion is to increase training diversity. Besides, we reshape the images into 32x32 resolution uniformly to facilitate multiple pooling of the network encoder. These "plaintext-ciphertext" pairs are randomly divided for training and testing. 90% of 5000 pieces are used as training sets, while the rest are employed as test sets to detect the generalization ability of our model.

4.1. Evaluation metric

In the experiment, Pearson correlation coefficient [22] was taken as the evaluation metric for two aspects: (1) To estimate the correlation between the plaintext images and the ciphertext images; (2) To evaluate the correlation between the output (decrypted image) and the ground truth (plaintext image). It can be defined by the following equation:

$$E(X) = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H} X(i,j) \tag{2}$$

$$\sigma(X) = \sqrt{\sum_{i=1}^{W}\sum_{j=1}^{H}[X(i,j) - E(X)]^2} \tag{3}$$

$$Corr = \frac{(O - E(O))(P - E(P))}{\sigma(O)\sigma(P)} \tag{4}$$

where W, H represent the width and height of the image. O is the output of the network(decrypted image), and P is the ground truth of the network (plaintext image). $Corr$ is applicable to calculate the similarity between the plaintext image P and the decrypted image O (range of 0$\sim$1).

4.2. Experiment results of three chaotic cryptosystems

Table 1 shows the experimental results of our proposed convolutional neural networks "IDEDNet" to attack the three chaotic encryption schemes (Song et al. [5], Pak et al. [8], H. N. Abdullah et al. [4]) on the mixed dataset of MNIST and MNIST-Fashion, including training correlation coefficient, testing correlation coefficient, epoch and the running time of each Epoch.

The results demonstrate the correctness of our proposed known-plaintext attack method based on deep learning, i.e., different from traditional cryptanalysis methods, one neural network can be applied to various chaotic cryptosystems and achieve superior decryption effects in a short time, which significantly improves the efficiency of subsequent cipher-image cracking.

Figures 3 to 5 present the visualization results, where the first row shows the plaintext images; the second row displays the corresponding ciphertext images; the third row lists the decrypted images on the testing set. We also show the correlation coefficients between the ciphertext image and the plaintext image, and those between the decrypted image and the plaintext image, separately. From the results, we can observe

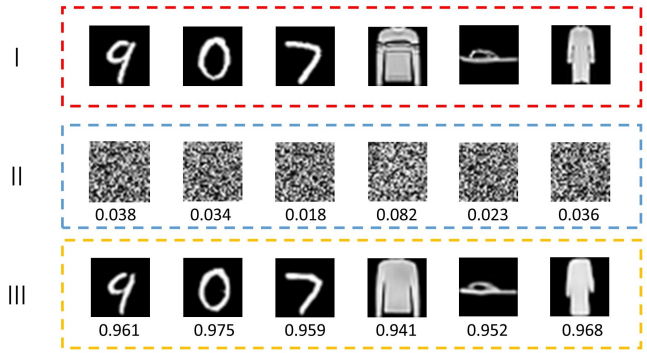


Fig. 3: visualization results on Song et al. [5]. (I) Plaintext image; (II) Ciphertext image; (III) Decrypted image; The number under the image represents the correlation coefficient between the image and the plaintext.

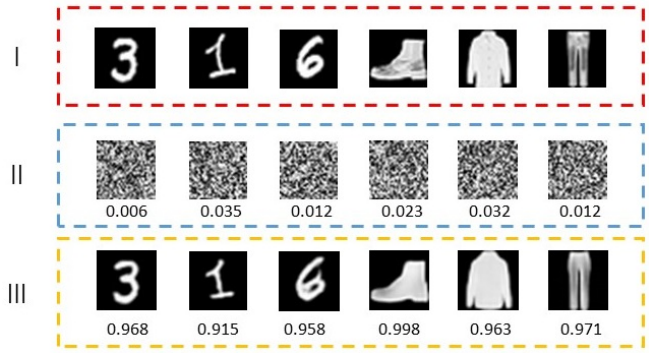


Fig. 4: visualization results on Pak et al. [10]. (I) Plaintext image; (II) Ciphertext image; (III) Decrypted image; The number under the image represents the correlation coefficient between the image and the plaintext.

that the testing ciphertext images can be well reconstructed by the neural network, and the correlation with the plaintext is also close to 1.

Figure 6 presents the curve changes of training Loss and testing evaluation metrics of known-plaintext attack method based on deep learning for three different chaotic encryption schemes on the mixed dataset of MNIST and MNIST-Fashion. We conclude the curve that neural network can be applied to the cryptanalysis of various chaotic cryptosystems and rapidly reach convergence to obtain better reconstruction performance of ciphertext image.

5. CONCLUSION

This paper proposes a deep learning-based known-plaintext attack method for chaotic image encryption schemes, which uses convolutional neural networks to train a large number of "plaintext-ciphertext " pairs generated by chaotic encryption schemes to learn the conversion process between the plaintext and the ciphertext. The trained model is regarded as the "decryption system". In the experiment, the proposed convo-

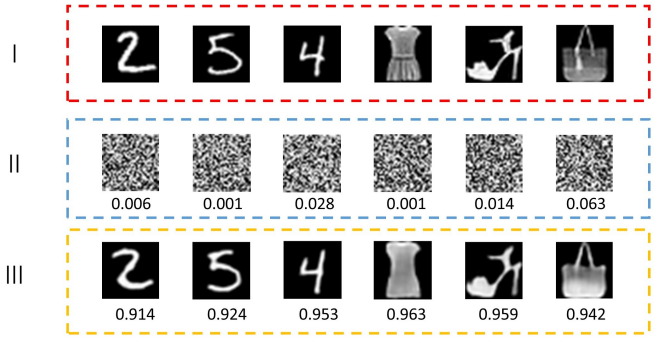
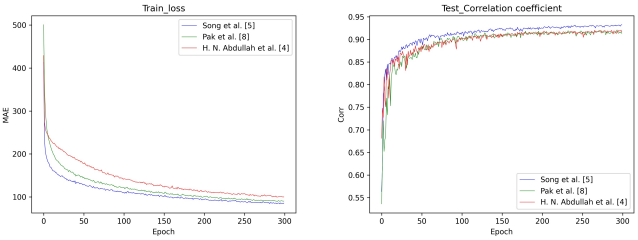


Fig. 5: visualization results on H. N. Abdullah et al. [4]. (I) Plaintext image; (II) Ciphertext image; (III) Decrypted image; The number under the image represents the correlation coefficient between the image and the plaintext.



(a) Results 1 (b) Result 2

Fig. 6: The change curve of training and testing process on the mixed dataset of MNIST and MNIST-Fashion: (a) Training L1Loss, (b) Testing Correlation Coefficient. The blue, green and red line represent the chaotic encryption schemes of Song et al. [5], Pak et al. [10], H. N. Abdullah et al. [4].

lutional neural network is applied for decryption on three existing chaotic encryption systems and excellent reconstruction results of ciphertext images are obtained. In practice, the significance of our work is apparent: (1) Compared with the traditional known-plaintext attack methods specific to a certain chaotic cryptosystem, our method is more cost-effective, flexible, which can avoid spending a lot of time and energy to design the corresponding decryption algorithm according to the encryption algorithm; (2) The chaotic cryptanalysis method based on deep learning can be introduced to other chaotic cryptosystems and even to the field of non-chaotic cryptosystems; (3) In addition, it also proposes a new research direction in the field of multimedia security, i.e., how to prevent cryptography attack methods based on deep learning.

As the ciphertext image reconstruction may be related to the design of the network, the training time, and the complexity of the encryption mechanism, etc., in future work, we will try to do further research on it and also try to extend deep learning-based cryptanalysis to other encryption algorithms.

6. REFERENCES

[1] K. Gupta, R. Gupta, R. Agrawal, and S. Khan, "An ethical approach of block based image encryption using chaotic map," *International Journal of Security and Its Applications*, vol. 9, no. 9, pp. 105–122, 2015.

[2] Y. H. Ail and Z. A. Alobaidy, "Images encryption using chaos and random generation," *Engineering and Technology Journal*, vol. 34, no. 1 Part (B) Scientific, pp. 172–179, 2016.

[3] G. Alvarez and S. Li, "Some basic cryptographic requirements for chaos-based cryptosystems," *Int. J. Bifurcation Chaos*, vol. 16, no. 08, pp. 2129–2151, 2006.

[4] H. N. Abdullah and H. A. Abdullah, "Image encryption using hybrid chaotic map," in *ICCIT*, 2017.

[5] Y. Song, J. Song, and J. Qu, "A secure image encryption algorithm based on multiple one-dimensional chaotic systems," in *ICCC*, 2016.

[6] N. Elabady, H. Abdalkader, M. Moussa, and S. F. Sabbeh, "Image encryption based on new one-dimensional chaotic map," in *ICET*, 2014.

[7] Y. Dou and M. Li, "Cryptanalysis of a new color image encryption using combination of the 1d chaotic map," *Applied Sciences*, vol. 10, no. 6, pp. 2187, 2020.

[8] M. Li, P. Wang, Y. Liu, and H. Fan, "Cryptanalysis of a novel bit-level color image encryption using improved 1d chaotic map," *IEEE Access*, vol. 7, pp. 145798–145806, 2019.

[9] M. Preishuber and T.Hütter, "Depreciating motivation and empirical security analysis of chaos-based image and video encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2137–2150, 2018.

[10] C. Pak and L. Huang, "A new color image encryption using combination of the 1d chaotic map," *Signal Process*, vol. 138, pp. 129–137, 2017.

[11] C. Pak, K. An, P. Jang, J. Kim, and S. Kim, "A novel bit-level color image encryption using improved 1d chaotic map," *Multimed. Tools Appl*, vol. 78, no. 9, pp. 12027–12042, 2019.

[12] X. Li, Y. Jiang, M. Chen, and F. Li, "Research on iris image encryption based on deep learning," *EURASIP J. Image Video Process*, vol. 2018, no. 1, pp. 1–10, 2018.

[13] Y. Qin, C. Zhang, R. Liang, and M. Chen, "Research on face image encryption based on deep learning," in *IOP Conference Series: Earth and Environmental Science*, 2019.

[14] H. Hai, S. Pan, M. Liao, D. Lu, W. He, and X. Peng, "Cryptanalysis of random-phase-encoding-based optical cryptosystem via deep learning," *Opt. Express*, vol. 27, no. 15, pp. 21204–21213, 2019.

[15] R. Gao and K. Grauman, "On-demand learning for deep image restoration," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1086–1095.

[16] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2016, pp. 241–246.

[17] C. Li, T. Fan, X. Ma, Z. Zhang, H. Wu, and L. Chen, "An improved image defogging method based on dark channel prior," in *ICIVC*, 2017.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[20] L. Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Process. Mag*, vol. 29, no. 6, pp. 141–142, 2012.

[21] Han Xiao, Kashif Rasul, and Roland Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[22] B. Mondal and T. Mandal, "A nobel chaos based secure image encryption algorithm," *International Journal of Applied Engineering Research*, vol. 11, no. 5, pp. 3120–3127, 2016.