

Deep-learning-based cryptanalysis of two types of nonlinear optical cryptosystems

Xiao-Gang Wang(汪小刚) and Hao-Yu Wei(魏浩宇)

Citation: Chin. Phys. B, 2022, 31 (**9**): 094202. DOI: 10.1088/1674-1056/ac6331
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

High speed ghost imaging based on a heuristic algorithm and deep learning

Yi-Yi Huang(黄祎祎), Chen Ou-Yang(欧阳琛), Ke Fang(方可), Yu-Feng Dong(董玉峰), Jie Zhang(张杰), Li-Ming Chen(陈黎明), and Ling-An Wu(吴令安)
Chin. Phys. B, 2021, 30 (**6**): 064202. DOI: 10.1088/1674-1056/abea8c

Phase retrieval algorithm for optical information security

Shi-Qing Wang(王诗晴), Xiang-Feng Meng(孟祥锋), Yu-Rong Wang(王玉荣), Yong-Kai Yin(殷永凯), Xiu-Lun Yang(杨修伦)
Chin. Phys. B, 2019, 28 (**8**): 084203. DOI: 10.1088/1674-1056/28/8/084203

Optical encryption of multiple three-dimensional objects based on multiple interferences and single-pixel digital holography

Ying Wang(王莹), Qi Liu(刘琦), Jun Wang(王君), Qiong-Hua Wang(王琼华)
Chin. Phys. B, 2018, 27 (**3**): 034202. DOI: 10.1088/1674-1056/27/3/034202

Optical encryption scheme based on ghost imaging with disordered speckles

Yu-dong Zhang(张玉东), Sheng-mei Zhao(赵生妹)
Chin. Phys. B, 2017, 26 (**5**): 054205. DOI: 10.1088/1674-1056/26/5/054205

An encryption scheme based on phase-shifting digital holography and amplitude-phase disturbance

Hua Li-Li, Xu Ning, Yang Geng
Chin. Phys. B, 2014, 23 (**6**): 064201. DOI: 10.1088/1674-1056/23/6/064201

Deep-learning-based cryptanalysis of two types of nonlinear optical cryptosystems

Xiao-Gang Wang(汪小刚)[1,2,†] and Hao-Yu Wei(魏浩宇)[2]

[1] *Department of Applied Physics, Zhejiang University of Science and Technology, Hangzhou* 310023, *China*
[2] *Department of Optical Engineering, Zhejiang A & F University, Hangzhou* 311300, *China*

(Received 27 December 2021; revised manuscript received 16 March 2022; accepted manuscript online 1 April 2022)

The two types of nonlinear optical cryptosystems (NOCs) that are respectively based on amplitude-phase retrieval algorithm (APRA) and phase retrieval algorithm (PRA) have attracted a lot of attention due to their unique mechanism of encryption process and remarkable ability to resist common attacks. In this paper, the securities of the two types of NOCs are evaluated by using a deep-learning (DL) method, where an end-to-end densely connected convolutional network (DenseNet) model for cryptanalysis is developed. The proposed DL-based method is able to retrieve unknown plaintexts from the given ciphertexts by using the trained DenseNet model without prior knowledge of any public or private key. The results of numerical experiments with the DenseNet model clearly demonstrate the validity and good performance of the proposed the DL-based attack on NOCs.

Keywords: optical encryption, nonlinear optical cryptosystem, deep learning, phase retrieval algorithm

PACS: 42.30.–d, 42.30.Kq, 42.30.Rx, 42.30.Va **DOI:** 10.1088/1674-1056/ac6331

1. Introduction

With the increasing demands for secured transmission and storage, information security has attracted wide attention and many optical encryption techniques have been proposed to secure information.[1,2] In 1995, Refregier and Javidi first proposed double random phase encoding (DRPE),[3] in which an image is encrypted into white noise by using two statistically independent random phase masks (RPMs) that are respectively located at the input image plane and Fourier transform domain. Inspired by the DRPE method, researchers have continuously investigated the DRPE method in different transform domains, such as Fresnel transform domain,[4] fractional Fourier transform domain,[5] and Gyrator transform domain.[6] Since the DRPE-based optical encryption has high flexibility and expansion capability, consequently, many other optical encryption methods have also been developed, which proves to be feasible and effective.[7–25]

In recent years, the vulnerability analysis of optical encryption systems has attracted wide interest from researchers. The DRPE-based optical encryption schemes have been verified to be vulnerable to different kinds of attacks[26–29] due to its inherent properties of linearity and symmetry. It was also found that they are vulnerable to new attack methods using deep learning (DL) techniques.[30–32] In order to remove the inherent symmetry and linearity of DRPE-based cryptosystems, NOC based on phase-truncated Fourier transform (PTFT) was proposed in 2010.[33] However, the PTFT-based nonlinear cryptosystem was still demonstrated to be vulnerable to some specific attacks using iterative amplitude-phase retrieval algorithm (APRA) and has the silhouette problem.[34,35] Meanwhile, several asymmetric encryption schemes based on equal modulus decomposition (EMD) have been proposed in these years,[36,37] which can resist the specific attack and eliminate the silhouette problem. However, it was reported that one can obtain precise results from EMD-based encryption system by performing an attack based on phase retrieval algorithm (PRA).[38] Recently, it has been proved that the learning-based method can also break the optical asymmetric cryptosystems based on PTFT and EMD.[39]

To improve the security level and remove the silhouette problem, amplitude-phase retrieval algorithm (APRA) and phase retrieval algorithm (PRA) have been introduced into nonlinear optical cryptosystems (NOCs).[40,41] The results of analysis show that the APRA-based and PRA-based NOCs can provide a high robustness against various conventional attacks. However, the vulnerabilities of these two types of NOCs against DL-based attacks, to the best of our knowledge, have not been studied before. In this work, a DenseNet model for cryptanalysis is proposed to analyze the securities of the two types of NOCs in the Fourier and Fresnel domains for the first time. In this end-to-end DL-based method, the neural network model is trained by using a set of ciphertexts and the corresponding plaintexts, and the Tensorflow, a popular open-source artificial intelligence library developed by Google, is used for setting up the neural network and conducting the network training. The optimal parameters of this network are determined through a series of forward and backward passes, which minimize a loss function. Numerical experiments show that the trained DenseNet model can be used to retrieve un-

[†]Corresponding author. E-mail: wxg1201@163.com

© 2022 Chinese Physical Society and IOP Publishing Ltd

known plaintexts from the given ciphertexts without using optical keys, which implies that the PRA-based NOCs cannot resist the proposed DL-based method attack.

The rest of this paper is organized as follows. In Section 2, we first describe the basic principles of two types of NOCs. And then the DL-based model for cryptanalysis is proposed in Section 3. In Section 4, we show the results of test simulations and performance of our proposed DL-based attack method. Finally, some conclusions are drawn in Section 5.

2. Principle and method

2.1. Review of Fourier domain APRA-based NOC

In this subsection, we give a brief introduction to the principle of the Fourier domain APRA-based NOC.[40] In the encryption scheme an amplitude and phase mixture retrieval of the Yang–Gu algorithm is used, hereafter referred to as amplitude-phase retrieval algorithm. The flowchart for encryption is shown in Fig. 1(a), where the encryption process is conducted in two procedures. In the first procedure, based on a mixture-type amplitude-phase retrieval, the secret image $f(x,y)$ is encrypted into the amplitude $g'(u,v)$ with a public key $P_1(u,v)$ and a binary phase modulation $\exp[\mathrm{i}\pi\gamma_1(u,v)]$. In the second procedure, the amplitude $g'(u,v)$ is the input image and encrypted into ciphertext $C(x',y')$ in the same way as another public key $P_2(x',y')$ and binary phase modulation $\exp[\mathrm{i}\pi\gamma_2(x',y')]$. Figure 1(b) is a flowchart illustrating the kth iteration in the first procedure of encryption process, which can be given as follows.

(i) In the kth iteration, the plaintext $f(x,y)$ is multiplied by phase function $\exp[\mathrm{i}\alpha_k(x,y)]$, and is then Fourier transformed to the output plane, the complex amplitude function in the output plane can be expressed as

$$F_k(u,v) = \mathrm{FT}\{f(x,y)\exp[\mathrm{i}\alpha_k(x,y)]\}, \tag{1}$$

where FT represents the Fourier transform, (x,y) and (u,v) denote the coordinates in the space domain and frequency domain respectively. The function $\exp[\mathrm{i}\alpha_k(x,y)]$ is a computer-generated RPM at the beginning.

(ii) Spectrum $F_k(u,v)$ is multiplied by the complex conjugate $P_1^*(u,v)$ of the public key $P_1(u,v)$ to obtain the real part

$$g_k(u,v) = \mathrm{Re}[F_k(u,v)\cdot P_1^*(u,v)], \tag{2}$$

where the function $P_1^*(u,v)$ denotes the complex conjugate of the public key $P_1(u,v)$, and $P_1(u,v)$ represents the public random phase keys, which are fixed in the encryption process.

(iii) Function $g_k(u,v)$ is multiplied by the public key $P_1(u,v)$ and is then inverse Fourier transformed to the input plane, the complex function in the input plane can be expressed as

$$I_{k+1}(x,y)\exp[\mathrm{i}\alpha_{k+1}(x,y)] = \mathrm{IFT}\{g_k(u,v)P_1(u,v)\}, \tag{3}$$

where IFT denotes the inverse Fourier transform. The function $I_{k+1}(x,y)\exp[\mathrm{i}\alpha_{k+1}(x,y)]$ is the result of the Fourier transform of the function $g_k(u,v)P_1(u,v)$, where $I_{k+1}(x,y)$ and $\exp[\mathrm{i}\alpha_{k+1}(x,y)]$ represent amplitude and phase function respectively. Note that phase part $\exp[\mathrm{i}\alpha_{k+1}(x,y)]$ is reserved and used to replace $\exp[\mathrm{i}\alpha_k(x,y)]\exp[\mathrm{i}\alpha_k(x,y)]$ in the next iteration while the amplitude part $I_{k+1}(x,y)$ is discarded.

(iv) The convergence of the iteration process is determined by the correlation coefficient (CC) value between $f(x,y)$ and $I_{k+1}(x,y)$.

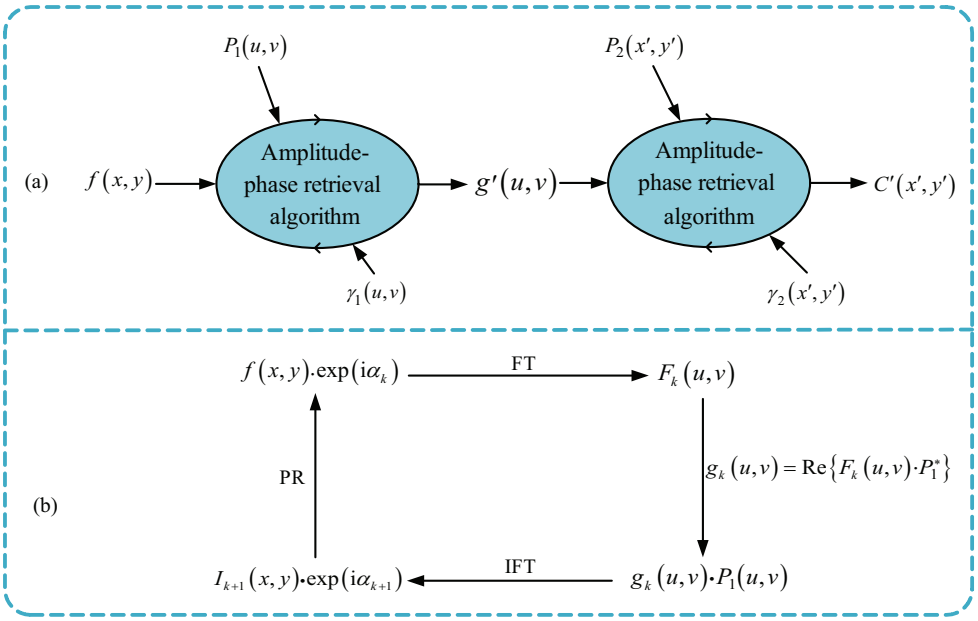


Fig. 1. Flowcharts of (a) encryption process and (b) the k-th iteration in the first procedure of encryption process, with PR denoting operation of phase reservation.

The final retrieved amplitude $g(u,v)$ contains both positive and negative elements, therefore a binary modulation $\exp[\mathrm{i}\pi\gamma_1(u,v)]$ is introduced, and the private modulation $\gamma_1(u,v)$ is given by the following equation:

$$\gamma_1(u,v)=\begin{cases}1, & (u,v)<0,\\ 0, & (u,v)>0.\end{cases} \tag{4}$$

The true amplitude is given by

$$g'(u,v)=g(u,v)\exp[\mathrm{i}\pi\gamma_1(u,v)]. \tag{5}$$

As shown in Fig. 1(a), the amplitude $g'(u,v)$ is then employed as an input signal of the second procedure of encryption process in another FT system with second random phase key (public key) $P_2(x',y')$, and the same iteration algorithm is used to obtain the final encryption result $C'(x',y')$. For the sake of brevity, the second procedure of encryption process is summarized as follows:

$$G'(x',y')=\mathrm{FT}\{g'(u,v)\exp[\mathrm{i}\beta(u,v)]\}, \tag{6}$$

$$C(x',y')=\mathrm{Re}[G'(x',y')\cdot P_2^*(x',y')], \tag{7}$$

$$\gamma_2(x',y')=\begin{cases}1, & C(x',y')<0,\\ 0, & C(x',y')>0,\end{cases} \tag{8}$$

$$C'(x',y')=C(x',y')\exp[\mathrm{i}\pi\gamma_2(x',y')], \tag{9}$$

where (x',y') is the coordinate in the spatial domain, $G'(x',y')$ is the result of the Fourier transform of function $g'(u,v)\exp[\mathrm{i}\beta(u,v)]$. $C(x',y')$ represents the amplitude function that contains both positive and negative elements, The binary modulation $\gamma_2(x',y')$ is generated simultaneously to obtain the ciphertext $C'(x',y')$. For simplicity, we omit the subscripts denoting the iteration number in each procedure for the unknown quantities. As a result, the decryption keys can be given as follows:[40]

$$D_1(u,v)=P_1(u,v)\exp[\mathrm{i}\pi\gamma_1(u,v)], \tag{10}$$

$$D_2(x',y')=P_2(x',y')\exp[\mathrm{i}\pi\gamma_2(x',y')]. \tag{11}$$

The decryption process can be expressed as

$$f(x,y)=\left|\mathrm{IFT}\left\{\left|\mathrm{IFT}[C'(x',y')D_2(x',y')]\right|D_1(u,v)\right\}\right|. \tag{12}$$

In this previously proposed asymmetric cryptosystem based on APRA, two RPMs P_1, P_2 serve as the public encryption keys while the decryption keys D_1 and D_2 are generated by using the public encryption keys. It is alleged that one cannot retrieve the secret image only by using the public keys and the cryptosystem cannot be cracked by using the traditional attack methods.[40]

2.2. Review of Fresnel domain PRA-based NOC

Based on a Gerchberg–Saxton (GS) PRA and DPRE, Rajput and Nishchal proposed a Fresnel domain PRA-based NOC, which could provide a high level of robustness against known-plaintext, chosen-plaintext, and special attacks.[41] Figure 2 displays a flowchart showing the steps of the encryption process. The ciphertext can be obtained by employing the GS PRA twice.

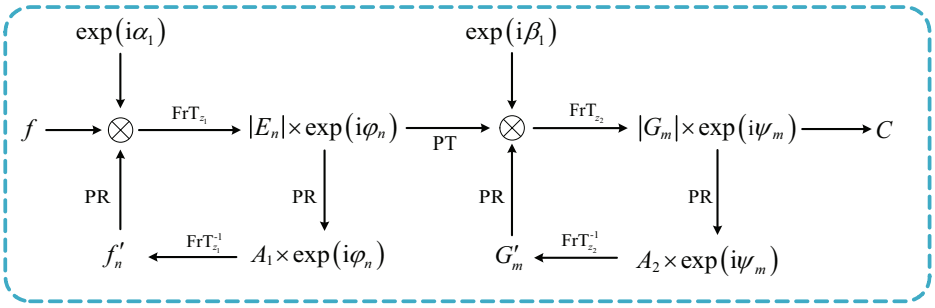


Fig. 2. Flowchart of PRA-based encryption scheme, with PT denoting operation of phase truncation.

(i) In the first level of encryption, the original amplitude image $f(x,y)$ multiplied by an RPM, $\exp[\mathrm{i}\alpha_n(x,y)]$, is used as the input of nth iteration, where the output of the first level is given by

$$\begin{aligned}E_n(u,v)&=\mathrm{FrT}_{z_1}\{f(x,y)\times\exp[\mathrm{i}\alpha_n(x,y)]\}\\&=|E_n(u,v)|\times\exp[\mathrm{i}\varphi_n(u,v)],\end{aligned} \tag{13}$$

where $\mathrm{FrT}_{z_1}\{\ \}$ denotes the free-space Fresnel diffraction for distance z_1.

(ii) During the next iteration, the phase distribution in the input plane is updated by

$$\exp[\mathrm{i}\alpha_{n+1}(x,y)]$$

$$=\mathrm{PR}\left\{\mathrm{FrT}_{z_1}^{-1}\{A_1(u,v)\times\exp[\mathrm{i}\varphi_n(u,v)]\}\right\}, \tag{14}$$

where FrT^{-1} denotes an inverse Fresnel transform and $A_1(u,v)$ represents a random amplitude.

(iii) Repeat steps (i) and (ii) until the correlation coefficient (CC) value reaches the preset threshold value.

The intermediate amplitude value obtained from the first step and another RPM, $\exp[\mathrm{i}\beta_m(u,v)]$, are used as the input of the second level. Suppose that the first level of the algorithm terminates at the N-th iteration, then we will have

$$\begin{aligned}G_m(\xi,\eta)&=\mathrm{FrT}_{z_2}\{|E_N(u,v)|\times\exp[\mathrm{i}\beta_m(u,v)]\}\\&=|G_m(\xi,\eta)|\times\exp[\mathrm{i}\psi_m(\xi,\eta)],\end{aligned} \tag{15}$$

$$G'_m(u,v) = \mathrm{FrT}_{z_2}^{-1}\{A_2(\xi,\eta)\times\exp[\mathrm{i}\psi_m(\xi,\eta)]\} = \left|G'_m(u,v)\right|\times\exp[\mathrm{i}\beta_{m+1}(u,v)], \quad (16)$$

where z_2 is diffraction distance, $A_2(\xi,\eta)$ is another random amplitude mask, and m is the iteration number of the second level of the algorithm. In the next iteration, the phase function $\exp[\mathrm{i}\beta_{m+1}(u,v)]$ is reserved, and the amplitude function $G'_m(u,v)$ is replaced by $E_N(u,v)$.

Suppose that the second level of the algorithm terminates at the M-th iteration, then the ciphertext will be able to be expressed as

$$C(\xi,\eta) = |G_M(\xi,\eta)|. \quad (17)$$

The private decryption keys are given by

$$D_1(u,v) = \exp[\mathrm{i}\varphi_N(u,v)]\times\exp[-\mathrm{i}\beta_M(u,v)], \quad (18)$$

$$D_2(\xi,\eta) = \exp[\mathrm{i}\psi_M(\xi,\eta)]. \quad (19)$$

The decrypted image can be obtained from

$$f(x,y) = \left|\mathrm{FrT}_{-z_1}\{\mathrm{FrT}_{-z_2}[C(\xi,\eta)\times D_2(\xi,\eta)]\times D_1(u,v)\}\right|. \quad (20)$$

3. DenseNet architecture designed for cryptanalysis

The main objective of the NOCs is to break the linearity of conventional systems, where the encryption method and keys are necessary for the recipient to perform the decryption. In those NOCs, the encryption keys are randomly generated by the computer, and the encryption keys, *i.e.* P_1, P_2, A_1, and A_2, are fixed once generated in the iterations. When those encryption keys are used for encrypting different plaintexts, do the NOCs remain unbreakable? In this case, we noticed that the correspondence between cyphertext and plaintext is one-to-one and the encryption keys are the same. Now, a question arises. Can we construct an iteration procedure to asymmetrically map the plaintexts into ciphertexts? Or is it possible to detect the relationships between cyphertext and plaintext by using a series of available "plaintext-ciphertext" pairs? The answer is yes. Since the encryption keys used for encrypting different plaintexts are the same, what we can infer is that the relationship between different ciphertexts is existent. For simplicity, the relationship between the plaintext and the cyphertext in NOCs is described as

$$I(x,y) = \mathfrak{F}[O(x_0,y_0)], \quad (21)$$

where $\mathfrak{F}[\cdot]$ denotes the encryption algorithm that transforms the plaintext in the input plane into the cyphertext in the output plane. For convenience, we use the same notation, *i.e.* $\mathfrak{F}[\cdot]$, for the two different NOCs. The reconstruction process can be given by

$$O_0(x_0,y_0) = \mathcal{R}[I(x,y)], \quad (22)$$

where $\mathcal{R}[\cdot]$ represents the mapping of the ciphertexts I into the plaintexts O.

As is well known, the machine learning algorithms prove to be able to yield high-quality solutions in several computer imaging problems during the past few years. Owing to their nonlinear characteristics, it is reasonable to use deep neural networks to estimate the relationship described by Eqs. (21) and (22). In our proposed method, the proposed DenseNet is designed to achieve the approximate equivalent models of NOCs by using a number of "plaintext-ciphertext" pairs.

Figure 3(a) show the proposed DenseNet architecture for cryptanalysis. It adopts encoder-decoder framework[42] and can be described as follows.

The inputs fed to the designed DenseNet model are ciphertexts with a size of 64×64 pixels. Each ciphertext firstly passes through a standard convolutional layer with 24 filter kernels with a size of 3×3, and is then successively decimated by six "Denseblock+AveragePooling" blocks, which serve as an encoder to extract the feature maps from the input ciphertexts. Each denseblock consists of four composite convolution layers, and each layer connects to every other layers within the same block in a feed-forward fashion.[43] Each composite convolutional layer is comprised of three consecutive operations: batch normalization (BN), rectified linear unit (ReLU) and convolution with small filters (size of 3×3); "AveragePooling" denotes 2×2 average pooling layer with stride 2, which is used for downsampling. After transmitting through another dense block, the data pass through six "Upsampling+Denseblock" blocks. Upsampling represents up-convolutional layer, which serves as a decoder to perform pixel-wise regression. Finally, the feature maps pass onto a standard convolutional layer, of which the filter size is 1×1, and the estimate of the plaintext is then produced. In this method, the skip connections are used to pass the high-frequency information learned in the initial layers down the network toward the output reconstruction.[44]

In this method, L2 regularization with weight decay of 0.001 is applied to all convolutional filters. The same regularization is also used in batch normalization. A small dropout rate of 0.05 is set to prevent overfitting. Owing to the GPU memory constraints, our DenseNet model is trained by using the Adam optimizer with a mini-batch size of 2. The training process starts with an initial learning rate of 0.001 and drops it by a factor of 2 after every 50 epochs. The neural network is trained for 300 epochs with the training samples being shuffled at each epoch. In this approach, mean square error (MSE) is used to monitor the network performance and can be defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2, \quad (23)$$

where n represents the pixel number of the input ciphertext, Y_i denotes the ground truth data, *i.e.* the plaintext, and $\hat{Y}_i$ is the output prediction, respectively.

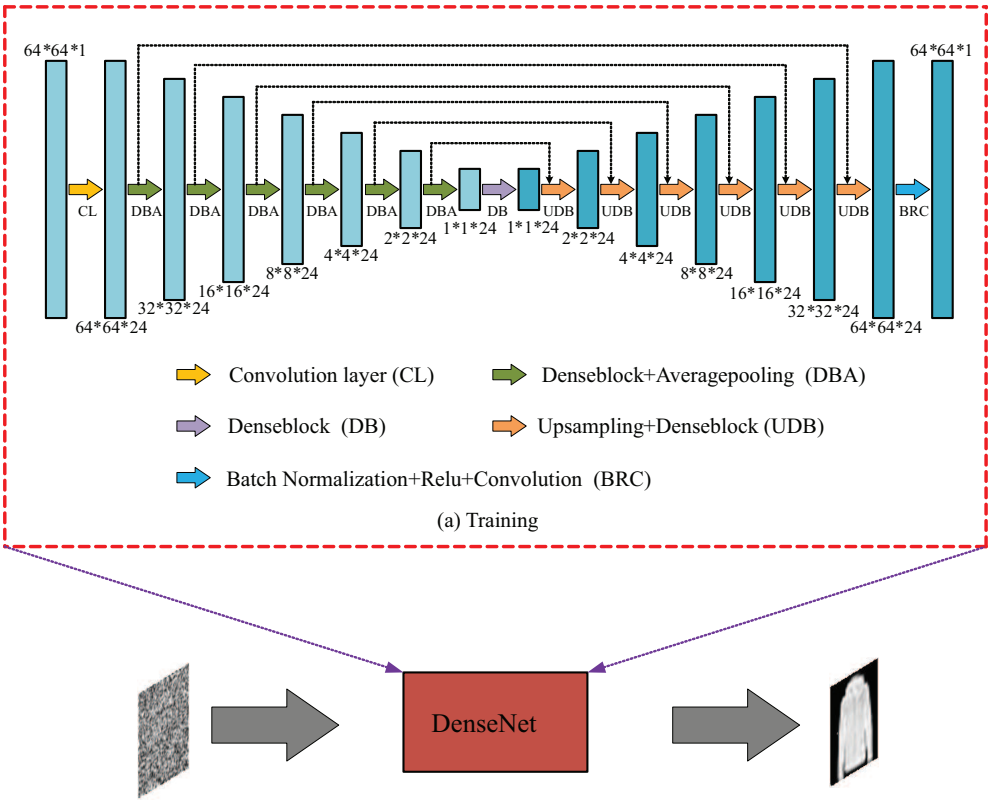


Fig. 3. Designed DenseNet architecture, indicating (a) training process and (b) testing process of the DenseNet model.

4. Numerical experiments and discussion

Numerical experiments have been performed to verify the feasibility of the proposed DL-based attack method. In the Fresnel domain nonlinear cryptosystem, the wavelength of laser light is $\lambda = 632.8$ nm and the two diffraction distances are set to be $z_1 = 0.1$ m, and $z_2 = 0.2$ m. In the iteration processes of the two different types of NOCs based on the APRA and PRA, the correlation coefficient (CC) is employed to evaluate the difference between the original input image and its estimates. The CC thresholds for the iteration processes are both set to be 0.98. The experiments are carried out on a database of 10000 grayscale images with a size of 64×64 pixels, which are selected from the Fashion-MNIST datasets randomly.[45] The chosen 9000 images are used as the training data and the rest 1000 images as the testing data.

The proposed DL-based algorithm is implemented on a GeForce GTX1650 GPU by Tensorflow 1.2, a deep learning framework. A learning rate constant of $\eta = 0.001$ is used for the learning of neural network and the number of training epochs is set to be 300. The execution time for the training of DenseNet model for each NOC is about 24.0 h. The MSE curves versus the epoch of the two DenseNets respectively corresponding to the two types of NOCs are shown in Fig. 4, where the MSEs of two networks drop to 0.01 after 300 epochs.

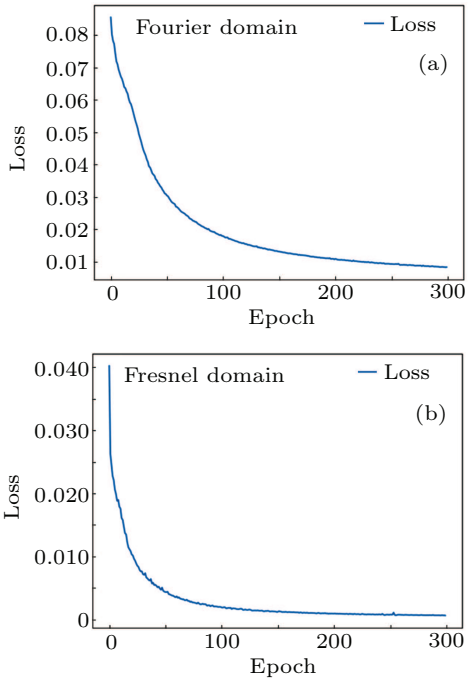


Fig. 4. MSE curves in training processes for (a) APRA-based NOC and (b) PRA-based NOC.

To verify the effectiveness of the designed DenseNet model, test sets are fed into each of the pre-trained DenseNets for ARPA-based and PRA-based NOCs. The reconstruction results using the proposed DenseNets are illustrated in Figs. 5(I) and 5(II) separately, where the input testing data,

ground truth and the results derived from pre-trained DenseNet are respectively shown in Figs. 5(a)–5(c). As can be seen from Fig. 5(c), all the retrieved images can be recognized by visual inspection. The average CC values of the retrieved images in Figs. 5(I) and 5(II) are 0.9702 and 0.9675, respectively. The successful decryption of the encrypted images without knowing the correct keys proves the good performance of our proposed DL-based method.



Fig. 5. Images retrieved from test data. (I) APRA-based NOC. (II) PRA-based NOC. (a) Testing ciphertext, (b) ground truth, and (c) retrieved images by trained DenseNet.

To further verify the effectiveness of the proposed method for different databases, 500 images are then randomly selected for predicting from MNIST database. Figure 6 shows the unknown plaintext images retrieved by utilizing the proposed DenseNet model, which has been pre-trained on the MNIST database. The ciphertext images, ground truth, retrieved images are shown in Figs. 6(a)–6(c), respectively. Likewise, the encrypted images can be retrieved with high quality. The average CC values of the retrieved images in Figs. 6(I) and 6(II) are 0.9819 and 0.9535, respectively.

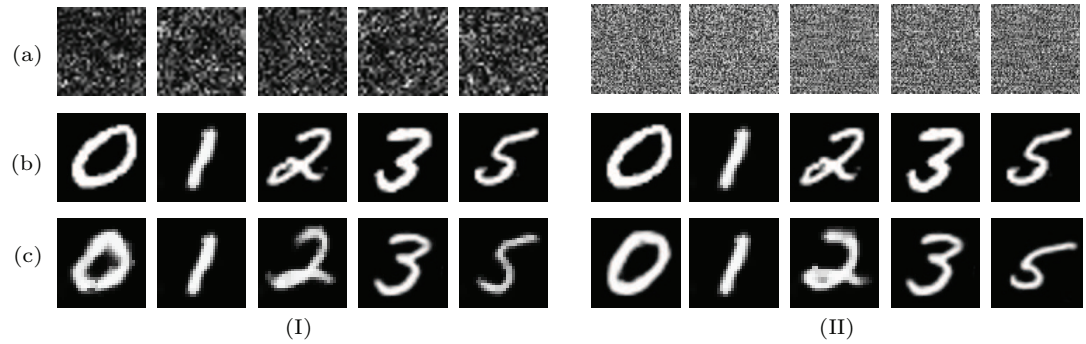


Fig. 6. Predicted results of MNIST database. (I) APRA-based NOC. (II) PRA-based NOC. (a) Ciphertext images, (b) plaintext images, and (c) retrieved images by using pre-trained DenseNet.

To demonstrate the robustness of the attack method based on the DenseNet, we investigate the influence of occlusion on the performance of the system. The simulation results are shown in Fig. 7, where panels (I) and (II) represent the test results corresponding to the APRA-based NOC and the PRA-based NOC, respectively. Figure 7(a) shows the ciphertext with 5% occlusion, and the plaintexts retrieved by pre-trained DenseNet model are shown in Fig. 7(b). The average CC values of retrieved images are 0.9342 and 0.8942 with respect to APRA-based NOC and PRA-based NOC, respectively. Figure 7(c) is the ciphertext with 10% occlusion, and the plaintexts attacked by DenseNet are shown in Fig. 7(b), the average CC values of the retrieved images corresponding to the two types of NOCs are 0.8423 and 0.8179, respectively. It can be found that the secret images can still be recovered and recognized when there are few data lost in the ciphertext.

Furthermore, we investigate the effect of noise pollution on the decryption. The zero-mean Gaussian noise is added into the ciphertexts and observe their retrieved images. The simulation results are shown in Fig. 8, where panels (I) and (II) represent the results from the attack on APRA-based NOC and PRA-based NOC, respectively. Figure 8(a) shows ciphertexts polluted by Gaussian noise with standard deviation of 0.01, and the derived images are shown in Fig. 8(b), where the corresponding average CC values of the two sets are 0.9423 and 0.9156. The ciphertexts polluted by Gaussian noise with a standard deviation of 0.02 are shown in Fig. 8(c), and the two sets of retrieved images are shown in Fig. 8(d), with the average CC values being 0.8953 and 0.8625, respectively, which demonstrates the good performance against the Gaussian noise.

Compared with conventional Unet and convolution neural network (CNN), the proposed DenseNet has more direct connections between layers, strengthening feature propagation, encouraging feature reuse and substantially reducing the number of parameters. The previously proposed neural networks

(CNN and Resnet)[20–22] have been further used to attack the two types of NOCs. However, the results obtained under the same training conditions, *i.e.*, the same quantity of data, training time and epoch, are not satisfactory. The results are given in Fig. 9, which shows the better performance of our proposed neural network.

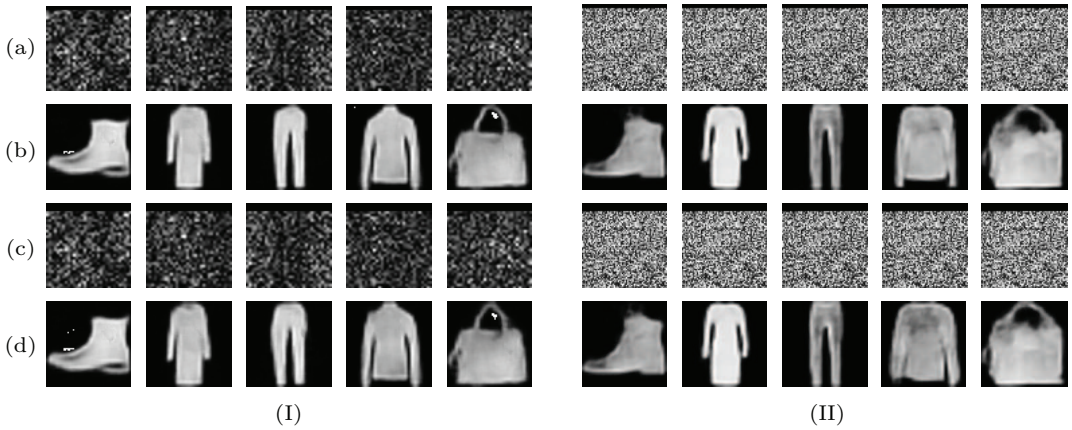


Fig. 7. Robustness test for occlusion attack. (I) APRA-based NOC. (II) PRA-based NOC. (a) Ciphertexts with 5% occlusion, (b) retrieved images from panel (a), (c) ciphertexts with 10% occlusion, and (d) retrieved images from panel (c).

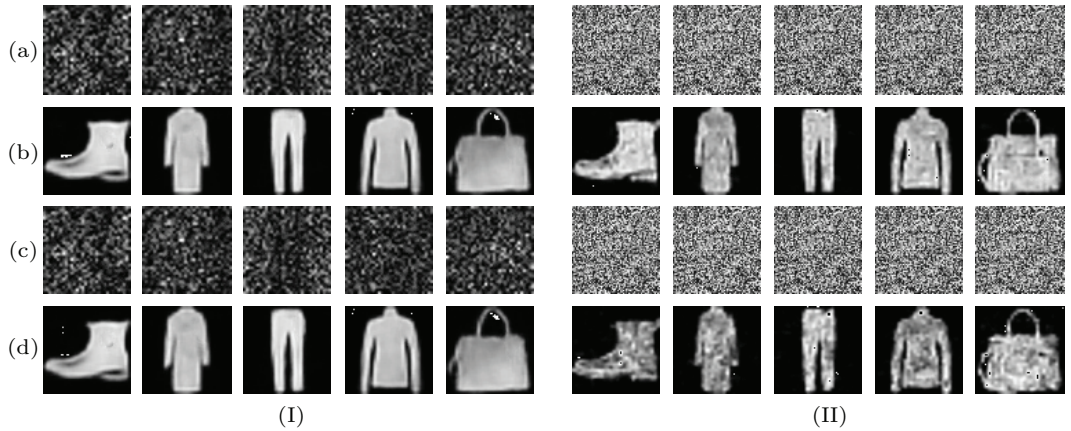


Fig. 8. Robustness test of DenseNet on Gaussian noise with respect to (I) APRA-based NOC and (II) PRA-based NOC, showing (a) ciphertexts with 0.01 variance Gaussian, (b) retrieved images from panel (a), (c) ciphertexts with 0.02 variance Gaussian, and (d) retrieved images from panel (c).

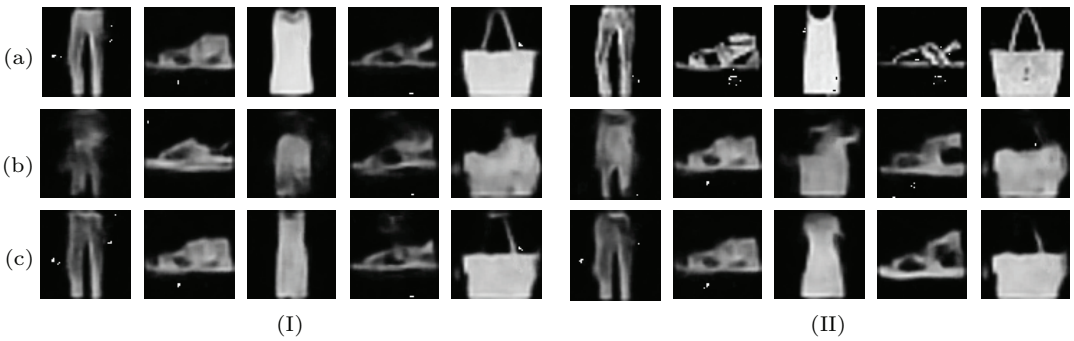


Fig. 9. Attack results of NOCs using different neural networks. (I) APRA-based NOC and (II) PRA-based NOC, showing (a) decrypted images using DenseNet, (b) retrieved image by using CNN, and (c) retrieved images using Resnet.

The simulation results show that it is possible for us to estimate the relationship between the outputs and inputs by learning models designed for cryptoanalysis. The proposed DenseNet can be considered as a black box, which can achieve the approximate equivalent models of NOCs by using a series of "plaintext-ciphertext" pairs without additional conditions. The unknown plaintexts can be retrieved by using the trained DenseNet model without decryption keys and complex phase retrieval algorithms.

5. Conclusions

In this paper, we proposed a novel DL-based method for cryptanalysis of two types of NOCs based on DL for the first time. The designed DenseNet model is built up according

to the densely connected convolutional neural network architecture and trained by a series of "ciphertext-plaintext" pairs. With pre-trained DenseNet, secret images can be successfully derived from ciphertexts without using encryption key or decryption key. The robustness of the DenseNet model against occlusion and noise pollutionis also investigated. The results of numerical experiments show that the two types of NOCs are vulnerable to our well-designed DL-based attack. It can also be seen that the performance obtained in this work is better than that obtained by the previously proposed methods.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant Nos. 61975185 and 61575178), the Natural Science Foundation of Zhejiang Province, China (Grant No. LY19F030004), and the Scientific Research and Development Fund of Zhejiang University of Science and Technology, China (Grant No. F701108L03).

References

[1] Chen W, Javidi B and Chen X 2014 *Adv. Opt. Photon.* **6** 120
[2] Wang S Q, Meng X F, Wang Y R, Yin Y K and Yang X L 2018 *Chin. Phys. B* **28** 084203
[3] Refregier P and Javidi B 1995 *Opt. Lett.* **20** 767
[4] Situ G H and Zhang J J 2004 *Opt. Lett.* **29** 1584
[5] Liu Z J and Liu S T 2007 *Opt. Commun.* **275** 324
[6] Abuturab M R 2013 *Opt. Lasers Eng.* **51** 317
[7] Hua L L, Xu N and Yang G 2014 *Chin. Phys. B* **23** 064201
[8] Zhou N R, Dong T J and Wu J H 2010 *Opt. Commun.* **283** 3037
[9] Yuan S, Yang Y R, Liu X M, Zhou X and Wei Z Z 2018 *Opt. Lasers Eng.* **100** 105
[10] Li T and Shi Y S 2017 *Chin. Phys. Lett.* **34** 034203
[11] Wei H Y and Wang X G 2021 *Opt. Laser Technol.* **142** 107257
[12] Chen J X, Zhang Y, Li J C and Zhang L B 2018 *Opt. Lasers Eng.* **101** 51
[13] Sui L S, Duan K K and Liang J L 2015 *Opt. Commun.* **343** 140
[14] Xiong Y, Du J and Quan C 2020 *Opt. Lasers Eng.* **134** 106289
[15] Wang Q, Xiong D, Alfalou A and Brosseau C 2019 *Opt. Commun.* **415** 56
[16] Qu G, Meng X F, Yang X L, Wu H Z, Wang P W, He W Q and Chen H Y 2021 *Opt. Lasers Eng.* **137** 106376
[17] Zhang Y D and Zhao S M 2017 *Chin. Phys. B* **26** 054205
[18] Wang Y, Liu Q, Wang J and Wang Q H 2018 *Chin. Phys. B* **27** 034202
[19] Gan W W, Chen L F and Liu Y Q 2021 *Opt. Commun.* **482** 126609
[20] Zhang L H, Zhang Z S, Ye H L, Wang Z R, Wang K M and Zhang D W 2020 *Appl. Phys. B* **126** 251
[21] Shan M G, Liu L, Liu B and Zhong Z 2021 *Opt. Lasers Eng.* **145** 106662
[22] Lv W J, Sun X K, Yang D Y, Zhu Y P and Shi Y S 2021 *Opt. Lasers Eng.* **141** 106574
[23] Liu H C and Chen W 2020 *Opt. Lasers Eng.* **130** 106094
[24] Zhou L N, Xiao Y and Chen W 2020 *Opt. Lett.* **45** 5279
[25] Zhang L H, Xiong R, Chen J and Zhang D W 2019 *Appl. Phys. B* **126** 704
[26] Carnicer A, Montes-Usategui M, Arcos S and Juvells I 2005 *Opt. Lett.* **30** 1644
[27] Peng X, Wei H Z and Zhang P 2006 *Opt. Lett.* **31** 3261
[28] Peng X, Zhang P, Wei H Z and Yu B 2006 *Opt. Lett.* **31** 1044
[29] Jiao S M, Li G W, Zhou C Y, Zou W B and Li X 2018 *J. Opt. Soc.* **35** A1
[30] Hai H, Pan S X, Liao M H, Lu D J, He W Q and Peng X 2019 *Opt. Express* **27** 21204
[31] Zhou L N, Xiao Y and Chen W 2019 *Opt. Express* **27** 26143
[32] Chen L F, Peng B Y, Gan W W and Liu Y Q 2020 *Opt. Express* **28** 28154
[33] Qin W and Peng X 2010 *Opt. Lett.* **35** 118
[34] Wang X G and Zhao D M 2012 *Opt. Commun.* **285** 1078
[35] Wang X G and Zhao D M 2013 *Opt. Lett.* **38** 3684
[36] Fatima A, Mehra I and Nishchal N K 2016 *J. Opt.* **18** 085701
[37] Chen H, Tanougast C, Liu Z J and Sieler L 2017 *Opt. Lasers Eng.* **93** 1
[38] Wang Y, Quan C and Tay C J 2016 *Appl. Opt.* **55** 679
[39] He W Q, Pan S X, Liao M H, Lu D J, Xing Q and Peng X 2021 *Opt. Lasers Eng.* **138** 106415
[40] Liu W, Liu Z J and Liu S T 2013 *Opt. Lett.* **38** 1651
[41] Rajput S K and Nishchal N K 2014 *Appl. Opt.* **53** 418
[42] Badrinarayanan V, Kendall A and Cipolla R 2017 *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 2481
[43] Huang G, Liu S, van der Maaten L and Weinberger K Q 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18–23, 2018, Salt Lake City, UT, USA, p. 2752
[44] Mao X J, Shen C and Yang Y B 2016 arXiv: 1603.09056
[45] Xiao H, Rasul K and Vollgraf R 2017 arXiv: 1708.07747