Research Article | Vol. 27, No. 18 | 2 Sep 2019 | OPTICS EXPRESS 26143
Optics EXPRESS

Machine-learning attacks on interference-based optical encryption: experimental demonstration

LINA ZHOU,[1] YIN XIAO,[1] AND WEN CHEN[1,2,*]

[1]*Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China*
[2]*The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China*
owen.chen@polyu.edu.hk

Abstract: Optical techniques have boosted a new class of cryptographic systems with some remarkable advantages, and optical encryption not only has spurred practical developments but also has brought a new insight into cryptography. However, this does not mean that it is elusive for the opponents to attack optical encryption systems. In this paper, for the first time to our knowledge, we experimentally demonstrate the machine-learning attacks on interference-based optical encryption. Using machine-learning models that are trained by a series of ciphertext-plaintext pairs, an unauthorized person is capable to retrieve the unknown plaintexts from the given ciphertexts without the usage of various different optical encryption keys existing in interference-based optical encryption. In comparison with conventional cryptanalytic methods, the proposed machine-learning-based attacking method can estimate transfer function or point spread function of interference-based optical encryption systems without subsidiary conditions. Simulations and optical experiments demonstrate feasibility and effectiveness of the proposed method, and the proposed machine-learning-based attacking method provides a versatile approach to analyzing the vulnerability of interference-based optical encryption.

1. Introduction

Information security is of growing technical importance in various applications, e.g., secured communication channels, secured identification and secured data storage [1–5]. In recent years, optical techniques [6–15] have been developed as potential alternatives in many particular applications. Because of unique capabilities of optical techniques (e.g., parallel processing and multi-dimensional processing), there is an increasing interest in exploiting optical techniques to implement more advanced schemes for securing information [6,10–15]. Current optical encryption systems begin with the pioneering work done by Refregier and Javidi [15]. Double random phase encoding (DRPE) [15] was first proposed to encode an image (i.e., plaintext) by using two statistically-independent random phase-only masks which are respectively located at the input image plane and Fourier transform domain. Using the DRPE, the plaintext can be converted into stationary white noise [15], and no information can be visually observed from the recorded speckle pattern (i.e., ciphertext). Inspired by the DRPE scheme, many researchers have further developed the DRPE scheme by implementing it in different domains, e.g., Fresnel transform domain and fractional Fourier transform domain [16–25]. The DRPE-based optical encryption possesses high feasibility, high flexibility and high capability in different domains. Subsequently, many other optical techniques and algorithms [6,10,26–28] have also been developed and are proven to be feasible and effective for optical encryption, e.g., ghost imaging-based encryption, diffractive imaging-based encryption and interference-based encryption. For instance, interference-based optical setups are simple and flexible to be carried out for optical encryption and decryption [27].

#369695 https://doi.org/10.1364/OE.27.026143
Journal © 2019 Received 11 Jun 2019; revised 14 Jul 2019; accepted 22 Jul 2019; published 29 Aug 2019

Optics EXPRESS

Nowadays, many studies are mainly focused on employing more optical techniques for designing optical encryption systems which are essential for the development of cryptography. Much effort has been devoted to advancing research on the design of optical security systems, and vulnerability analysis of the designed optical encryption systems does not attract sufficient attention. Whether the designed optical encryption systems can withstand the attacks from unauthorized persons is still a serious concern. In essence, cryptography and cryptoanalysis are mutually beneficial, and can form a close relationship for common developments. Cryptographic techniques claimed to be secure should withstand various attacks via the cryptoanalysis. Conversely, the cryptoanalysis can stimulate the development of more advantageous and secure schemes. Hence, cryptoanalysis of optical encryption schemes [29–33] is also of high importance. Carnicer et al. [29] first reported that the DRPE-based optical encryption is vulnerable to chosen-ciphertext attack. Peng et al. [30] proposed an analogous method called chosen-plaintext attack to retrieve the plaintext from the stationary white noise. Liao et al. [32] analyzed optical encryption systems using ciphertext-only attack (COA) to retrieve the plaintext. Recently, Liu et al. [33] improved the COA method by using multiple phase retrieval algorithms to develop a hybrid iterative phase retrieval algorithm. Conventional methods for analyzing the vulnerability of optical cryptosystems focused on applying elaborately-designed information to retrieve or estimate various optical encryption keys. Conventional methods for attacking optical cryptosystems are also limited by a need of complex phase retrieval algorithms along with the requirement of some preconditions. Moreover, in many applications, the retrieval or estimate of various different optical encryption keys is difficult and time-consuming, which restricts their applications and is unfavorable in practice. Therefore, it is desirable to develop a new method for the cryptoanalysis of optical encryption systems, which is capable to extract the unknown plaintexts from the given ciphertexts without the usage of various different optical encryption keys and various complex phase retrieval algorithms.

In this paper, we experimentally demonstrate for the first time to our knowledge that interference-based optical encryption cannot withstand the proposed machine-learning attacks, and the estimate of various different optical encryption keys or the usage of complex phase retrieval algorithms are not requested. Machine-learning method is a powerful tool which can discover and emulate senior representations of the relationships between the given data by using the specially treated neural networks [34]. After training, the trained machine-learning model can be considered as a black box which contains an explicit illustration of the given data. Without definite representations of the parameters, the trained learning model can make predictions of original objects leading to the decreased time for extracting effective information [34]. Due to remarkable features of machine-learning models, machine-learning-based attacking method is proposed here and is used to extract the unknown plaintexts from the given ciphertexts without the usage of various different optical encryption keys. Simulations and optical experiments simultaneously demonstrate that the proposed machine-learning attacks are feasible and effective, and it is expected that the proposed method can be a promising strategy for the cryptoanalysis of various interference-based optical encryption systems.

2. Theoretical demonstration

Vulnerability of interference-based optical encryption is analyzed here, and the schematic setup is shown in Fig. 1. The optical encryption procedure is briefly described as follows: random mask $M_1(x,y)$ is bonded with the plaintext $f(x,y)$ at the object beam arm, and another random mask $M_2(x,y)$ is placed at the reference beam arm. The two modulated beams form a pattern by using a beam splitter which is recorded by a CCD camera. The interference procedure and the pattern $H(\mu,\nu)$ can be described by

Optics EXPRESS

$$H(\mu,\nu) = \mathrm{FrT}_{d,\lambda}[f(x,y)M_1(x,y)] + \mathrm{FrT}_{d,\lambda}[M_2(x,y)], \quad (1)$$

where FrT denotes free-space wave propagation in the Fresnel domain [6–10], d denotes an axial distance between the random mask and CCD camera, and λ denotes laser wavelength. The ciphertext $I(\mu,\nu)$ recorded at the CCD plane, i.e., an intensity pattern, can be described by

$$I(\mu,\nu) = \left|\mathrm{FrT}_{d,\lambda}[f(x,y)M_1(x,y)] + \mathrm{FrT}_{d,\lambda}[M_2(x,y)]\right|^2. \quad (2)$$

Optical encryption using interference principle was claimed to be secure in previous studies, and it was demonstrated that in the case of minimal deviations it is difficult to visually recognize original image [6,27]. However, its vulnerability has not been effectively analyzed previously. Recently, machine-learning methods are increasingly popular in many disciplines owing to their awesome performances of end-to-end learning, which can refrain from complicated extractions of inner parameters [34–40]. In this paper, machine-learning scheme is proposed and applied for the cryptoanalysis of interference-based optical encryption. By training a designed machine-learning model using sufficient pairs of the ciphertexts and plaintexts, the unknown plaintexts can be retrieved in real time from the given ciphertexts by using the trained learning model without the usage of various different optical encryption keys and various complex phase retrieval algorithms.

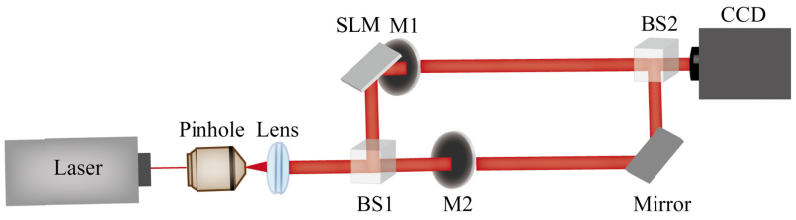

Fig. 1. Schematic setup for interference-based optical encryption. SLM: spatial light modulator; BS: beam splitter; CCD: charge-coupled device; M: random mask.

In interference-based optical encryption, a laser beam launched by a He-Ne laser source (wavelength of 633.0 nm) is first expanded by propagating through a microscope objective, and then the expanded light is collimated by a collimating lens. After that, the laser beam propagates through a beam splitter and is divided into two coherent beams. One laser beam illuminates on the SLM (Holoeye LC-R 720, reflective) and then immediately propagates through a random mask M1 which will form an object beam. The SLM acts as an object, i.e., plaintext. The images used as plaintexts are handwritten-digit patterns (8-bit grayscale images with 28×28 pixels) from the MNIST database [38] which is widely used for machine learning. In addition, 8-bit grayscale fashion images from the fashion MNIST database [39] have also been used as plaintexts in this study. In Fig. 1, another laser beam propagates through random mask M2 which will form a reference beam. The object beam and the reference beam are further combined by using a beam splitter. Hence, when a series of plaintexts are sequentially embedded into the SLM, a series of ciphertexts, i.e., intensity patterns, can be correspondingly recorded by a CCD camera. Instead of making great effort to retrieve or estimate various different security keys existing in interference-based optical encryption by using complex phase retrieval algorithms, we demonstrate via simulations and optical experiments that machine-learning attacks can be proposed and designed to retrieve the unknown plaintexts from the given ciphertexts by using a trained machine-learning model. For the complex optical encryption setups, the proposed method can also be applied to retrieve the unknown plaintexts from the given ciphertexts by using a trained machine-learning model.

2.1. The designed CNN architecture

Here, machine-learning attacks are proposed and designed, and convolutional neural network (CNN) is newly designed and applied to verify the vulnerability of interference-based optical encryption as schematically illustrated in Fig. 2.

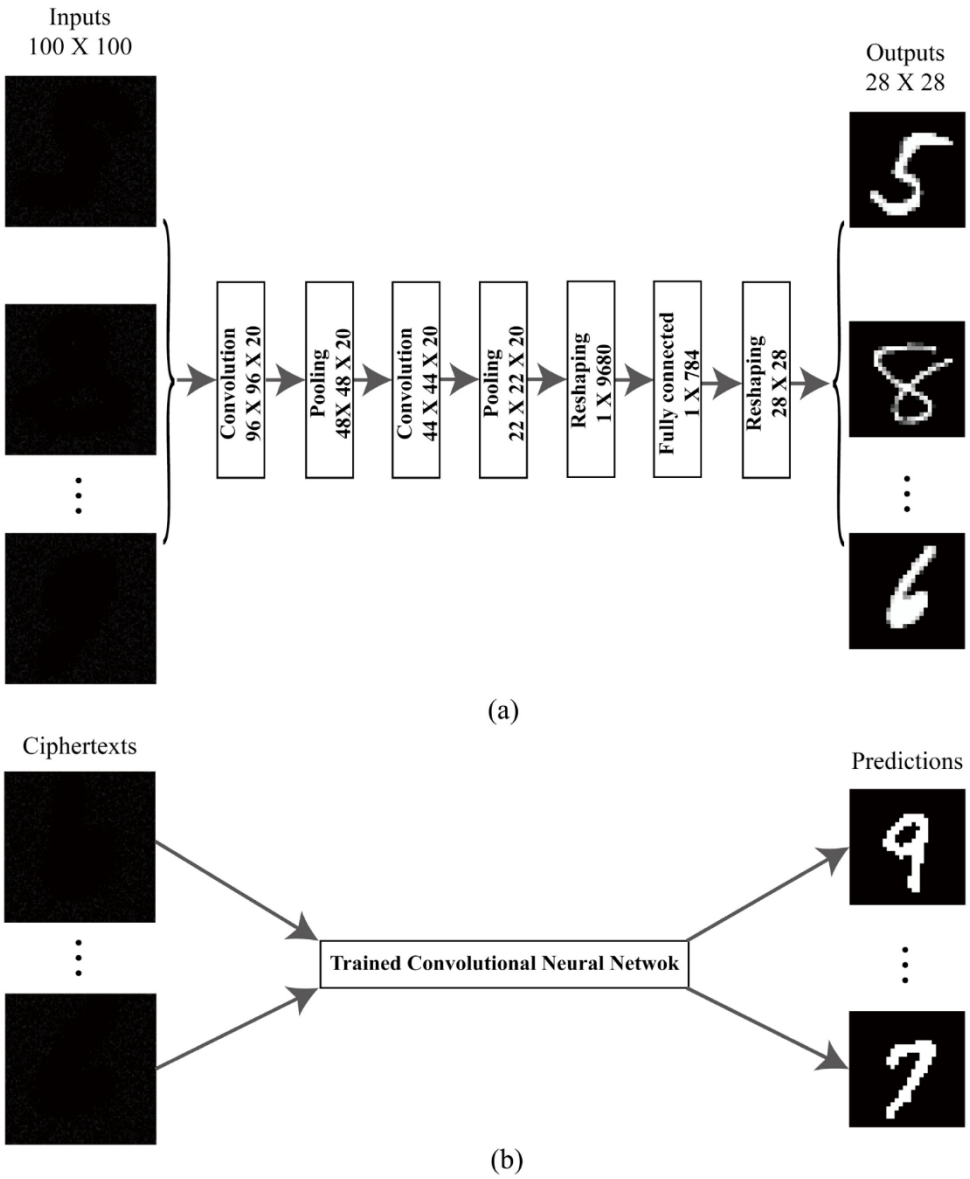

Fig. 2. Schematic of the designed CNN architecture for attacking interference-based optical encryption. The inputs (i.e., ciphertexts) are resized from 512×512 pixels to 100×100 pixels to lower computational load. After two convolutions and two pooling layers, the input is reshaped and fully connected to the ground truth. Using sufficient pairs of ciphertexts and plaintexts fed to the learning model, the CNN model is trained to predict unknown plaintexts from the given ciphertexts. (a) Training phase: pairs of ciphertexts and plaintexts obtained from interference-based optical encryption are respectively fed to the inputs and outputs of a designed CNN model. (b) Testing phase: typical examples show that the trained CNN model can be used in real time to predict the unknown plaintexts from the given ciphertexts.

Optics EXPRESS

The CNN architecture designed for the proposed machine-learning attacks is shown in Fig. 2(a), and the designed CNN architecture is described as follows: The input (i.e., ciphertext) with a dimension of 100×100 pixels is resized from the recorded pattern to lower computational load. Then, it convolves with 20 kernels of size 5×5 forming the first convolution layer (size of 96×96×20). The activation function used in the two convolution layers is the sigmoid function which is frequently applied in machine learning because of boundedness. In the first pooling layer, an action of down-sampling is taken to further reduce the computational load, and size of the first pooling layer is 48×48×20. The pooled data is sent to the second convolution layer which adopts the same number and size of kernels. Size of the second convolution layer is 44×44×20 followed by the second pooling layer with size of 22×22×20. After two rounds of convolution and down-sampling processing, the first reshaping layer reshapes the second pooling layer (size of 22×22×20) to a 1×9680 vector, and then the reshaped vector is transported to the fully connected layer with size of 1×784. Before the output layer, the second reshaping layer reshapes the 1×784 vector to an 28×28 image. A number of the recorded ciphertexts (i.e., intensity patterns) and the corresponding ground truths (i.e., plaintexts) are fed to the designed CNN model. After training, the trained learning model can be used to retrieve the unknown plaintexts from the given ciphertexts, and typical testing examples are shown in Fig. 2(b). Here, the CNN architecture is implemented by using Matlab platform on a PC with Nvidia Geforce GTX1080Ti GPU.

2.2. Simulations

Both MNIST database [38] and fashion MNIST database [39] are used to verify feasibility and effectiveness of the proposed machine-learning attacks on interference-based optical encryption. Figure 3 shows simulation results about the retrieval of the unknown plaintexts by using the trained CNN model without the usage of various different optical encryption keys. Several different ciphertexts are used as typical examples for the testing, as respectively shown in Figs. 3(a), 3(c), 3(e), 3(g), 3(i) and 3(k). It can be seen in Figs. 3(a), 3(c), 3(e), 3(g), 3(i) and 3(k) that the plaintexts are encoded into noisy patterns, and no information about the plaintexts can be visually recognized from the ciphertexts. By using the trained machine-learning models, the plaintexts are correspondingly extracted without the usage of various different optical encryption keys as respectively shown in Figs. 3(b), 3(d), 3(f), 3(h), 3(j) and 3(l). The retrieved plaintexts are of high quality, and performance of the proposed method is further evaluated by using peak signal-to-noise ratio (PSNR). The PSNR values of the retrieved plaintexts in Figs. 3(b), 3(d), 3(f), 3(h), 3(j) and 3(l) are 25.13 dB, 25.84 dB, 25.02 dB, 17.13 dB, 14.78 dB and 18.06 dB, respectively. The PSNR values mean that the unknown plaintexts are fully extracted. In comparison with conventional cryptanalytic methods, the proposed machine-learning attacks on interference-based optical encryption are capable to extract the unknown plaintexts from the given ciphertexts without the usage of various different optical encryption keys and various complex phase retrieval algorithms, and are superior to conventional cryptanalytic methods. In this theoretical demonstration, interference-based optical cryptosystem is proven to be vulnerable to the proposed machine-learning attacks.

To verify the robustness and applicability of the proposed machine-learning attacking method, the trained learning model is also applied to attack different databases. For the objects (e.g., double digits, lowercase and uppercase letters) which are not from the database used in the training phase, the trained learning model can also retrieve the unknown plaintexts from the correspondingly given ciphertexts. Several typical ciphertexts (i.e., for other objects) obtained by using the interference-based optical cryptosystem are respectively shown in Figs. 3(m), 3(o) and 3(q). Figures 3(n), 3(p) and 3(r) show the retrieved plaintexts obtained from the machine-learning model trained by the MNIST database, and PSNR values of the retrieved images in Figs. 3(n), 3(p) and 3(r) are 26.07 dB, 29.59 dB and 32.78 dB,

respectively. These retrieved images effectively illustrate that the proposed machine-learning attacks are also applicable for other objects (i.e., from different databases) which are not used in the training phase. It can be ascribed to two main reasons for the availability and universality of the trained learning model for other images which are different from those used in the training phase. One reason is that the designed learning model is trained to learn the relationship between the input ciphertexts and output plaintexts without individual retrieval of optical security keys. The other reason is that the ciphertexts obtained by the optical cryptosystem share some similarities and correlations. In addition, when the database with more complex objects is trained, the proposed method is also feasible and applicable.

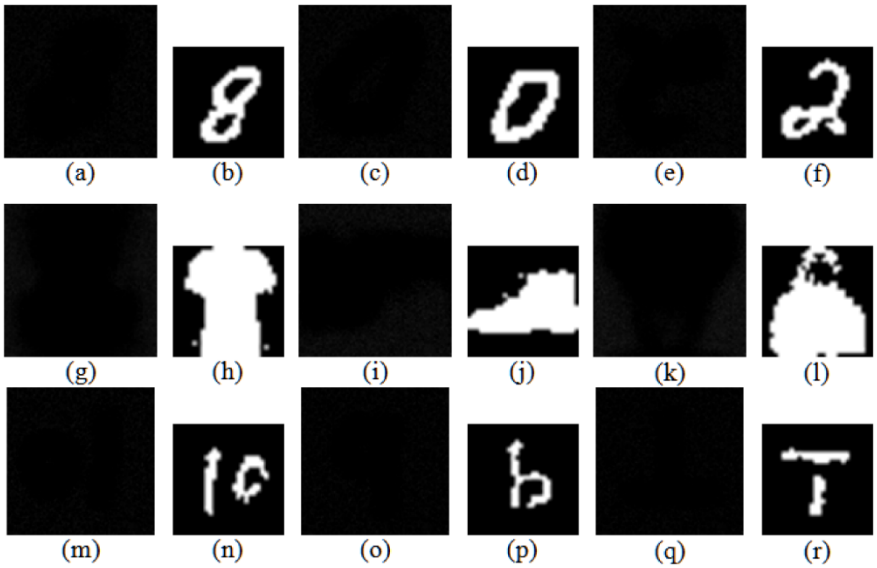

Fig. 3. Simulation results of the proposed learning attack on interference-based optical encryption. Testing phase: (a), (c), (e), (g), (i), (k), (m), (o) and (q) the ciphertexts obtained by using the interference-based optical encoding (further sent to the trained learning model). (b), (d), (f), (h), (j) and (l) The unknown plaintexts retrieved by using the trained learning model respectively corresponding to (a), (c), (e), (g), (i) and (k). (n), (p) and (r) The retrieved plaintexts (i.e., from different databases) obtained by using the learning model trained by the MNIST database.

3. Experimental demonstration and discussion

Experimental verification of the proposed machine-learning attacks on interference-based optical encryption is also conducted. The experimental setup is schematically shown in Fig. 1. The laser beam is launched by a He-Ne laser source (Newport R-30993, wavelength of 633.0 nm). It is expanded by a microscope objective (Newport M-40X, 0.65NA), and then collimated by a collimating lens with a focal length of 50.0 mm. The collimated laser beam illuminates on a SLM. The plaintexts used in the experiments are 8-bit grayscale handwritten-digit images from the MNIST database [38] and 8-bit fashion images from the fashion MNIST database [39]. A series of plaintexts are sequentially embedded into the SLM by using a programmable controller (i.e., via Labview). The masks used in the experiments are diffusers (Thorlabs, N-BK7) to scatter the object beam and reference beam [41]. Hence, a series of ciphertexts, i.e., intensity patterns, are correspondingly recorded by a CCD camera (Thorlabs, DCC3240M) with pixel number of 1280×1024 and pixel size of 5.30 μm.

Optics EXPRESS

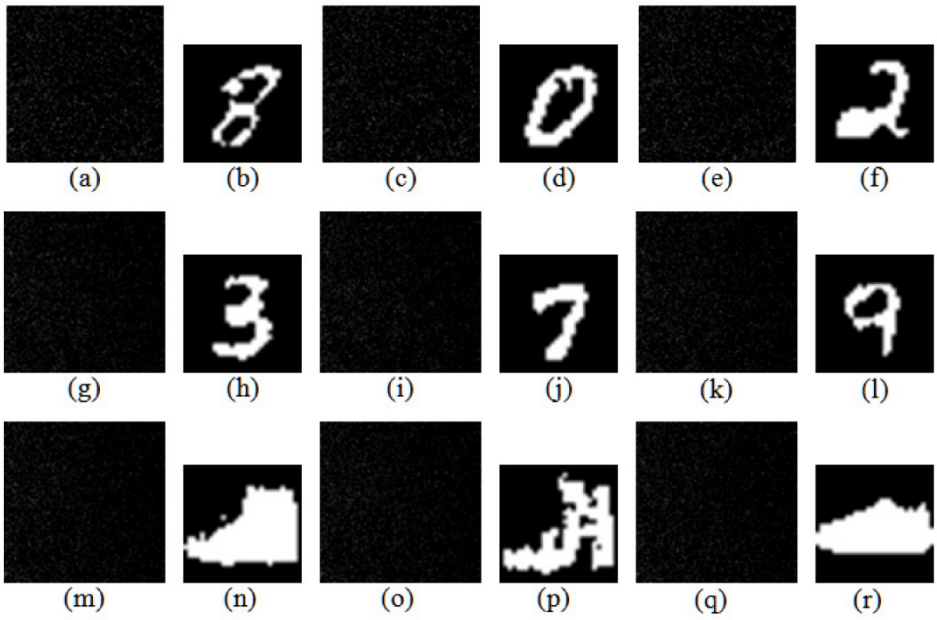

Fig. 4. Experimental results of the proposed machine-learning attacks to the interference-based optical encryption. Testing phase: (a), (c), (e), (g), (i), (k), (m), (o) and (q) ciphertexts recorded by using interference-based optical encryption setup. (b), (d), (f), (h), (j), (l), (n), (p) and (r) The unknown plaintexts retrieved by using the trained learning model respectively corresponding to (a), (c), (e), (g), (i), (k), (m), (o) and (q).

The 5000 handwritten-digit images from the MNIST database [38] and also 5000 fashion images from the fashion MNIST database [39] are encoded by the interference-based optical encryption setup, and their corresponding ciphertexts are sequentially recorded by CCD camera. Here, a window with 600×600 pixels is cropped as region of interest to reduce the computational load, and to improve efficiency of the designed machine-learning model the recorded ciphertexts are further resized to 100×100 pixels. For both MNIST database and fashion MNIST database, 4800 pairs of ciphertexts and plaintexts are used to respectively train their learning models, and another 200 ciphertexts, i.e., their plaintexts treated as unknown, are respectively recorded for the testing. Architecture of the designed learning models for the experiments is shown in Fig. 2(a). It is worth noting that value of the momentum m for updating the SGD is set as -9.5×10^{-5}. After the training of 4 hours, the learning models are well trained to be used for retrieving the unknown plaintexts from the given ciphertexts in the testing phase. Typical retrieval examples from the experimentally-obtained ciphertexts are shown in Fig. 4, which use the trained learning models to implement machine-learning attacks on the interference-based optical encryption. Figures 4(a), 4(c), 4(e), 4(g), 4(i), 4(k), 4(m), 4(o) and 4(q) show the recorded ciphertexts by using those plaintexts respectively from the MNIST database [38] and the fashion MNIST database [39]. It can be seen that the plaintexts are encoded into noisy patterns by using interference-based optical cryptosystem. By using their trained learning models, it is found in the testing phase that the experimentally-obtained ciphertexts are successfully attacked, and the retrieved plaintexts are shown in Figs. 4(b), 4(d), 4(f), 4(h), 4(j), 4(l), 4(n), 4(p) and 4(r) respectively. Performance of the proposed machine-learning attacks is also evaluated by using the PSNR, which are 25.58 dB, 27.64 dB, 26.08 dB, 20.89 dB, 22.74 dB, 28.91 dB, 19.50 dB, 12.64 dB, and 26.07 dB, respectively. The experimental results demonstrate that the unknown plaintexts can be extracted by using the trained learning models without the usage of various different optical encryption keys and various complex phase retrieval algorithms. Hence, the proposed

Optics EXPRESS

machine-learning attacks are also verified to be effective by using the experimental results for analyzing the vulnerability of interference-based optical encryption.

It has been illustrated previously that more random masks [6,10,11,27] used in optical encryption setup can enhance the security, and here we further study the applicability of the proposed machine-learning attacks on interference-based optical encryption with multiple diffusers. In Fig. 5, two diffusers are cascaded and used at the object beam arm as a typical example. At the object beam arm, the random mask M1 is bonded with the plaintext, and the random mask M2 is placed about 5.0 cm away from random mask M1. The axial distance between the random mask M2 and CCD is 5.0 cm. The images from the MNIST database and fashion MNIST database are used as the plaintexts, and a series of ciphertexts, i.e., intensity patterns, are sequentially recorded by the CCD camera. The number of images chosen from each database is 5000, and 4800 pairs of ciphertexts and plaintexts are selected to train the designed learning models for each database. Another 200 recorded ciphertexts are used for the testing. Time used for training each database is about 4 hours. After the training of each database, two learning models are trained and can be used to correspondingly retrieve the unknown plaintexts from the given ciphertexts as shown in Figs. 6(a), 6(c), 6(e), 6(g), 6(i), 6(k), 6(m), 6(o) and 6(q).

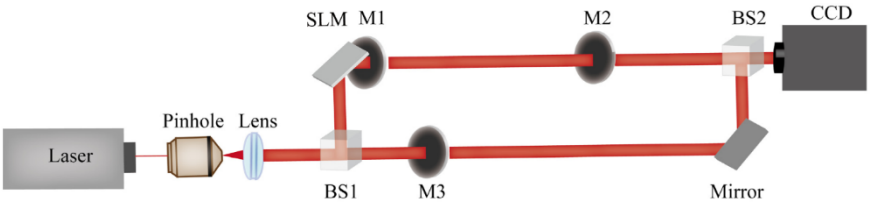

Fig. 5. Experimental setup for interference-based optical encryption with cascaded random masks at the object beam arm.

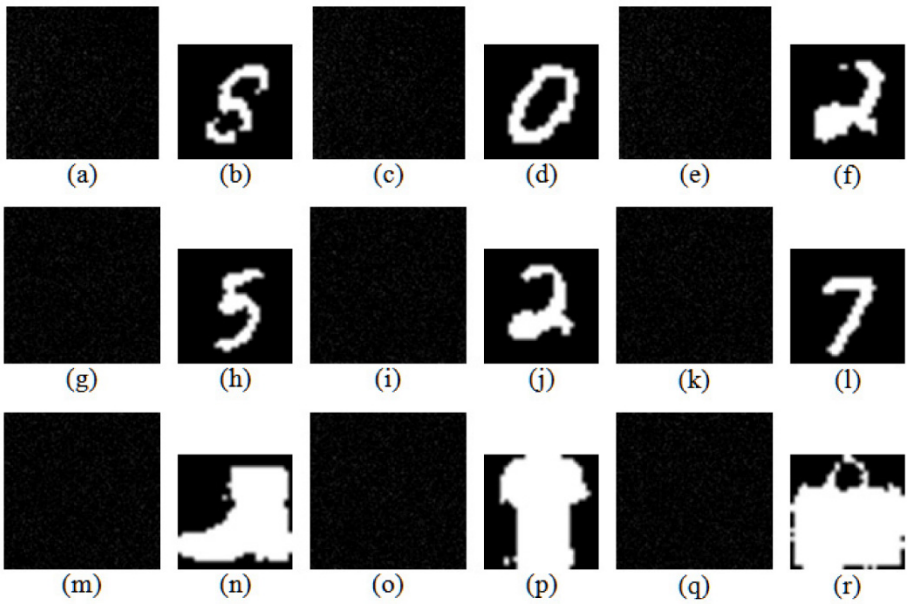

Fig. 6. Experimental results of the proposed machine-learning attacks on the interference-based optical encryption. Testing phase: (a), (c), (e), (g), (i), (k), (m), (o) and (q) ciphertexts recorded by using the interference-based optical encryption system. (b), (d), (f), (h), (j), (l), (n), (p) and (r) The unknown plaintexts retrieved by using the trained learning model respectively corresponding to (a), (c), (e), (g), (i), (k), (m), (o) and (q).

Optics EXPRESS

The unknown plaintexts can be correspondingly retrieved as respectively shown in Figs. 6(b), 6(d), 6(f), 6(h), 6(j), 6(l), 6(n), 6(p) and 6(r) by using their trained learning models. The PSNR values for Figs. 6(b), 6(d), 6(f), 6(h), 6(j), 6(l), 6(n), 6(p) and 6(r) are 15.10 dB, 25.95 dB, 22.96 dB, 31.12 dB, 24.63 dB, 30.68 dB, 23.40 dB, 20.39 dB and 20.50 dB, respectively. It is demonstrated that without the extraction or estimate of various different optical encryption keys, unauthorized persons can retrieve the unknown plaintexts from the given ciphertexts by using the trained learning models.

Interference-based optical encryption using cascaded random masks at the reference beam arm is also experimentally studied to verify feasibility and effectiveness of the proposed machine-learning attacks. Optical setup for interference-based optical cryptosystem with cascaded random masks at the reference beam arm is schematically shown in Fig. 7. After the training, some recorded ciphertexts are further tested as shown in Figs. 8(a), 8(c), 8(e), 8(g), 8(i), 8(k), 8(m), 8(o) and 8(q). The unknown plaintexts are successfully retrieved by using their correspondingly trained machine-learning models as shown in Figs. 8(b), 8(d), 8(f), 8(h), 8(j), 8(l), 8(n), 8(p) and 8(r), and PSNR values of the retrieved plaintexts in Figs. 8(b), 8(d), 8(f), 8(h), 8(j), 8(l), 8(n), 8(p) and 8(r) are 26.21 dB, 24.56 dB, 26.29 dB, 24.67 dB, 33.41 dB, 27.88 dB, 19.50 dB, 20.16 dB, and 19.30 dB, respectively. It is also illustrated that the proposed machine-learning attacks can perform well in making predictions of the unknown plaintexts from the given ciphertexts without the usage of various different security keys existing in the interference-based optical encryption.

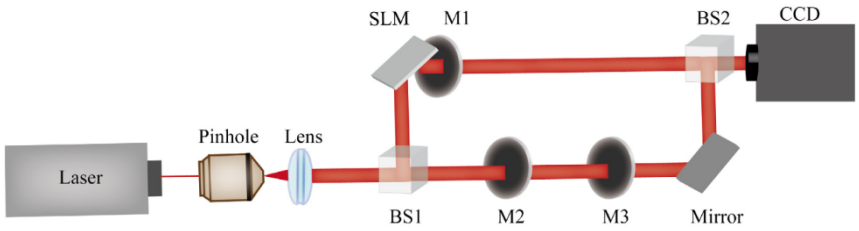

Fig. 7. Experimental setup for interference-based optical encryption with cascaded random masks at the reference beam arm.

The aforementioned results demonstrate that the proposed machine-learning attacks are also applicable to analyze different optical encryption setups, e.g., multiple masks or diffusers used in the optical paths. The proposed machine-learning attacks can retrieve unknown plaintexts from the given ciphertexts without knowledge of various parameters, e.g., the diffusers and axial positions of the diffusers. In comparison with the existing attacking methods, the proposed machine-learning attacks do not need to individually retrieve each security key. For complex optical encryption systems, it could be difficult and time-consuming for conventional attacking methods, however the trained machine-learning models can still extract the unknown plaintexts from the given ciphertexts. Hence, it is believed that the proposed machine-learning attacks are able to analyze the security of various interference-based optical encryption systems and others.

The proposed machine-learning attacks are verified to be feasible and applicable for analyzing the vulnerability of interference-based optical cryptosystem. Although interference-based optical encryption with more random masks [6,10,11,27] is demonstrated to be of the higher security, the proposed machine-learning attacks still have the capability to extract the unknown plaintexts from the given ciphertexts without the usage of various different optical encryption keys. The proposed learning method can function as a black box which can effectively estimate transfer function or point spread function of the interference-based optical encryption systems without subsidiary conditions. The proposed machine-learning attacks provide new avenues for the cryptoanalysis of interference-based optical encryption.

Optics EXPRESS

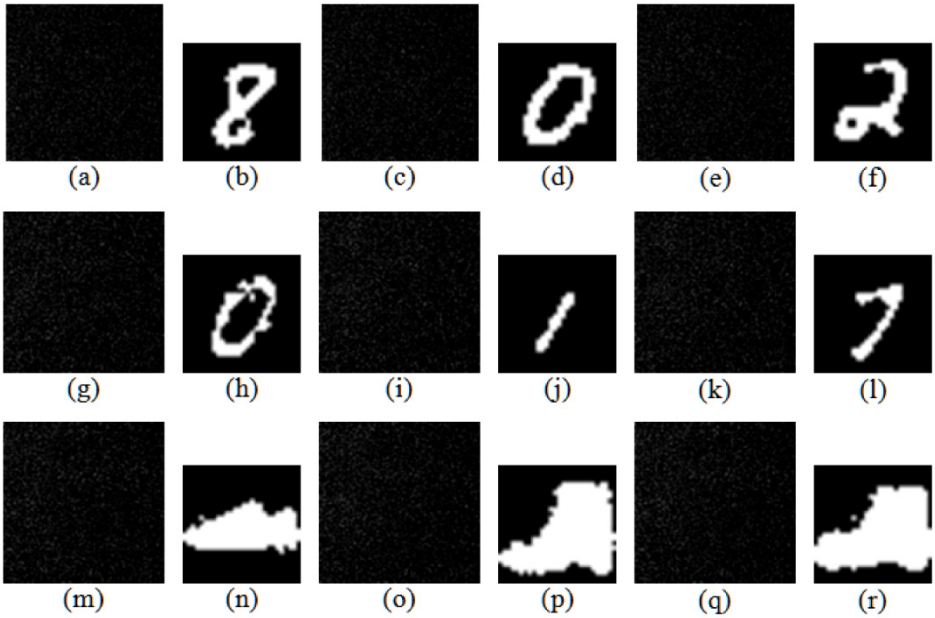

Fig. 8. Experimental results of the proposed machine-learning attacks on interference-based optical encryption. Testing phase: (a), (c), (e), (g), (i), (k), (m), (o) and (q) ciphertexts recorded by using the interference-based optical encryption system. (b), (d), (f), (h), (j), (l), (n), (p) and (r) The unknown plaintexts retrieved by using the trained learning model respectively corresponding to (a), (c), (e), (g), (i), (k), (m), (o) and (q).

4. Conclusions

We experimentally demonstrate for the first time to our knowledge that interference-based optical encryption is vulnerable to the proposed machine-learning attacks. The designed learning-based architecture is validated to have the ability to retrieve the unknown plaintexts from the given ciphertexts without the usage of various different optical encryption keys and various complex phase retrieval algorithms. Moreover, the retrieved plaintexts are of high quality, which also avoids extra effort to further identify the retrieved plaintexts. Feasibility and effectiveness of the proposed machine-learning attacks are verified by simulations and optical experiments. It is also validated that the proposed machine-learning attacks can work effectively to extract the unknown plaintexts from the given ciphertexts obtained by using interference-based optical encryption with cascaded random masks. Due to the distinctive advantages of the designed machine-learning attacks, it is believed that the proposed machine-learning method can also be applied to attack other optical encryption systems. It is expected that the proposed machine-learning attacks can provide a promising strategy for the cryptoanalysis of optical encryption systems, and can lead to the further investigation of more advanced and secure interference-based optical encryption systems.

Funding

National Natural Science Foundation of China (NSFC) (61605165); Hong Kong Research Grants Council (25201416); Shenzhen Science and Technology Innovation Commission through Basic Research Program (JCYJ20160531184426473).

References

1. B. Javidi, “Securing information with optical technologies,” Phys. Today **50**(3), 27–32 (1997).
2. F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, “Information hiding-a survey,” Proc. IEEE **87**(7), 1062 (1999).
3. R. C. Merkle, “Secure communications over insecure channels,” Commun. ACM **21**(4), 294–299 (1978).

Optics EXPRESS

4. K. Morita, H. Yoshimura, M. Nishiyama, and Y. Iwai, "Protecting Personal Information using Homomorphic Encryption for Person Re-identification," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)* (IEEE, 2018), pp. 166–167.
5. A. Surekha, P. R. Anand, and I. Indu, "E-Payment Transactions Using Encrypted QR Codes," Int. J. Appl. Eng. Res. **10**(77), 461 (2015).
6. W. Chen, B. Javidi, and X. Chen, "Advances in optical security systems," Adv. Opt. Photonics **6**(2), 120–155 (2014).
7. O. Matoba and B. Javidi, "Encrypted optical memory system using three-dimensional keys in the Fresnel domain," Opt. Lett. **24**(11), 762–764 (1999).
8. W. Chen and X. Chen, "Space-based optical image encryption," Opt. Express **18**(26), 27095–27104 (2010).
9. P. C. Mogensen and J. Glückstad, "A phase-based optical encryption system with polarisation encoding," Opt. Commun. **173**(1–6), 177–183 (2000).
10. A. Alfalou and C. Brosseau, "Optical image compression and encryption methods," Adv. Opt. Photonics **1**(3), 589–636 (2009).
11. W. Chen, X. Chen, and C. J. Sheppard, "Optical image encryption based on diffractive imaging," Opt. Lett. **35**(22), 3817–3819 (2010).
12. Y. Shi, T. Li, Y. Wang, Q. Gao, S. Zhang, and H. Li, "Optical image encryption via ptychography," Opt. Lett. **38**(9), 1425–1427 (2013).
13. Y. Zhang, C. H. Zheng, and N. Tanno, "Optical encryption based on iterative fractional Fourier transform," Opt. Commun. **202**(4–6), 277–285 (2002).
14. J. F. Barrera, A. Mira, and R. Torroba, "Optical encryption and QR codes: secure and noise-free information retrieval," Opt. Express **21**(5), 5373–5378 (2013).
15. P. Refregier and B. Javidi, "Optical image encryption based on input plane and Fourier plane random encoding," Opt. Lett. **20**(7), 767–769 (1995).
16. O. Matoba and B. Javidi, "Encrypted optical storage with wavelength-key and random phase codes," Appl. Opt. **38**(32), 6785–6790 (1999).
17. G. Unnikrishnan, J. Joseph, and K. Singh, "Optical encryption by double-random phase encoding in the fractional Fourier domain," Opt. Lett. **25**(12), 887–889 (2000).
18. G. Situ and J. Zhang, "Double random-phase encoding in the Fresnel domain," Opt. Lett. **29**(14), 1584–1586 (2004).
19. G. Unnikrishnan and K. Singh, "Double random fractional Fourier domain encoding for optical security," Opt. Eng. **39**(11), 2853–2859 (2000).
20. R. Tao, Y. Xin, and Y. Wang, "Double image encryption based on random phase encoding in the fractional Fourier domain," Opt. Express **15**(24), 16067–16079 (2007).
21. Z. Liu and S. Liu, "Double image encryption based on iterative fractional Fourier transform," Opt. Commun. **275**(2), 324–329 (2007).
22. L. Chen and D. Zhao, "Optical image encryption with Hartley transforms," Opt. Lett. **31**(23), 3438–3440 (2006).
23. Z. Liu, Q. Li, J. Dai, X. Sun, S. Liu, and M. A. Ahmad, "A new kind of double image encryption by using a cutting spectrum in the 1-D fractional Fourier transform domains," Opt. Commun. **282**(8), 1536–1540 (2009).
24. M. R. Abuturab, "Color image security system based on discrete Hartley transform in gyrator transform domain," Opt. Lasers Eng. **51**(3), 317–324 (2013).
25. N. Singh and A. Sinha, "Chaos based multiple image encryption using multiple canonical transforms," Opt. Laser Technol. **42**(5), 724–731 (2010).
26. P. Clemente, V. Durán, V. Torres-Company, E. Tajahuerce, and J. Lancis, "Optical encryption based on computational ghost imaging," Opt. Lett. **35**(14), 2391–2393 (2010).
27. E. Tajahuerce, O. Matoba, S. C. Verrall, and B. Javidi, "Optoelectronic information encryption with phase-shifting interferometry," Appl. Opt. **39**(14), 2313–2320 (2000).
28. Y. Zhang and B. Wang, "Optical image encryption based on interference," Opt. Lett. **33**(21), 2443–2445 (2008).
29. A. Carnicer, M. Montes-Usategui, S. Arcos, and I. Juvells, "Vulnerability to chosen-cyphertext attacks of optical encryption schemes based on double random phase keys," Opt. Lett. **30**(13), 1644–1646 (2005).
30. X. Peng, H. Wei, and P. Zhang, "Chosen-plaintext attack on lensless double-random phase encoding in the Fresnel domain," Opt. Lett. **31**(22), 3261–3263 (2006).
31. X. Peng, P. Zhang, H. Wei, and B. Yu, "Known-plaintext attack on optical encryption based on double random phase keys," Opt. Lett. **31**(8), 1044–1046 (2006).
32. M. Liao, W. He, D. Lu, and X. Peng, "Ciphertext-only attack on optical cryptosystem with spatially incoherent illumination: from the view of imaging through scattering medium," Sci. Rep. **7**(1), 41789 (2017).
33. X. Liu, J. Wu, W. He, M. Liao, C. Zhang, and X. Peng, "Vulnerability to ciphertext-only attack of optical encryption scheme based on double random phase encoding," Opt. Express **23**(15), 18955–18968 (2015).
34. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**(7553), 436–444 (2015).
35. K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 3929–3938.
36. N. M. Nasrabadi, "Pattern recognition and machine learning," J. Electron. Imaging **16**(4), 049901 (2007).

Optics EXPRESS

37. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems* (NIPS, 2012), pp. 1097–1105.
38. L. Deng, "The MNIST database of handwritten digit images for machine learning research [best of the web]," IEEE Signal Process. Mag. **29**(6), 141–142 (2012).
39. H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," arXiv preprint arXiv:1708.07747 (2017).
40. I. Sutskever, J. Martens, G. Dahl, and G. E. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on Machine Learning* (ICML, 2013), pp. 1139–1147.
41. X. D. Chen, *Computational Methods for Electromagnetic Inverse Scattering* (Wiley-IEEE, 2018).