Learning complex scattering media for optical encryption

LINA ZHOU, YIN XIAO, AND WEN CHEN*

Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China
**Corresponding author: owen.chen@polyu.edu.hk*

Received XX Month XXXX; revised XX Month, XXXX; accepted XX Month XXXX; posted XX Month XXXX (Doc. ID XXXXX); published XX Month XXXX

Optical encryption has provided a new insight for securing information, however it is always desirable that high security can be achieved to withstand the attacks. In this Letter, we propose a new method via learning complex scattering media for optical encryption. After the recordings through complex scattering media, a designed learning model is trained. The proposed method uses an optical setup with complex scattering media to experimentally record the ciphertexts, and uses a learning model to generate security keys. During the decryption, the trained learning model with its parameters is applied as security keys. In addition, various parameters, e.g., virtual phase-only masks, can be flexibly applied to further enlarge key space. It is experimentally demonstrated that the proposed learning-based encryption approach possesses high security. The proposed method could open up a new research perspective for optical encryption. © 2020 Optical Society of America

http://dx.doi.org/10.1364/OL.99.099999

Optical encryption has become one of the advanced encoding methods owing to its outstanding properties [1–5]. Optical encoding technologies originated from the milestone work done by Refregier and Javidi, who put forward the idea of double random phase encoding (DRPE) [3]. Since then, there is an explosion of optical encryption schemes, such as fractional Fourier transform domain and Fresnel transform domain [6,7]. With a rapid development of optical techniques, different optical cryptosystems are accordingly proposed, such as interferometric imaging, diffractive imaging and ghost imaging [8–12]. However, it has been found that optical encoding schemes cannot withstand some attacks [13–17]. For instance, Carnicer et al. proposed a groundbreaking philosophy of chosen-ciphertext attack to vet the vulnerability of DRPE scheme [13]. Similarly, chosen-plaintext, known-plaintext attack and ciphertext-only attack provided an insight for the cryptoanalysis of optical encryption [14–16]. The major objective in the developed optical cryptoanalyses was to extract an estimated plaintext from the ciphertext by using the estimated security keys. Apart from the aforementioned attacking technologies, another method, called learning-based attack, has also been developed [17]. It can allow a direct extraction of unknown plaintexts from the given ciphertexts without individual retrieval of various security keys or the usage of complex phase retrieval algorithms. The recent progress in optical cryptoanalysis becomes a serious threat to optical encryption schemes, which would request the advances in optical encryption methods or infrastructures.

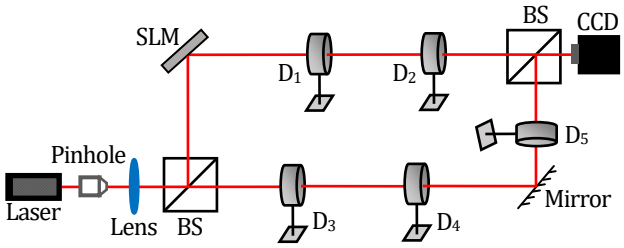

Fig. 1. A schematic experimental setup with complex scattering media: SLM, spatial light modulator (Holoeye LC-R720); D_1, D_2, D_3, D_4 and D_5, diffusers (Thorlabs, DG10-600); BS, beam splitter cube; CCD, charge-coupled device (Thorlabs, DCC1240C).

Inspired by remarkable characteristics of machine learning [18–22], we introduce machine learning into optical cryptography. In this Letter, learning complex scattering media for optical encryption is proposed for the first time to our knowledge. The proposed method uses an optical setup with complex scattering media to experimentally record the ciphertexts, and uses a learning model to generate security keys. It is experimentally demonstrated that the proposed learning-based encryption is feasible and effective, and possesses high security.

A schematic experimental setup is shown in Fig. 1, and complex scattering media can be flexibly designed and applied in the optical setup. The collimated He-Ne laser beam with wavelength of 633.0 nm propagates through a beam splitter to be separated into two beams. One beam illuminates a reflective spatial light modulator (SLM, Holoeye LC-R720). The input grayscale images (i.e., plaintexts) are sequentially embedded into the SLM, which are fashion-product images from fashion MNIST database [18] or handwritten-digit images from MNIST database [19]. Then, the modulated laser beam successively propagates through the

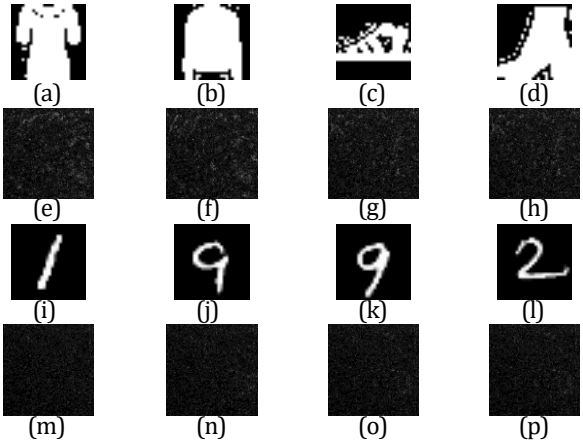

Fig. 2. (a)–(d) and (i)–(l) Typical input images (i.e., plaintexts) embedded into the SLM, and (e)–(h) and (m)–(p) the corresponding speckle patterns (i.e., ciphertexts) experimentally recorded by a CCD.

diffusers D_1 and D_2. The reference beam is sequentially diffracted by the diffusers D_3, D_4 and D_5, and then interferes with the object beam. The interference patterns recorded by a CCD are used as ciphertexts (size of 600×600 pixels) in this study. Figures 2(a)–2(p) show several plaintexts selected from the two databases [18,19] and their corresponding ciphertexts experimentally recorded by the CCD.

It has been demonstrated that various attacks [13–17] pose a great threat to optical cryptosystems. Here, we propose learning-based optical encryption. Using machine learning technologies [18–22], we generate a trained model with its parameters as security keys. The decryption process is not the directly reverse process of encryption. Meanwhile, parameters in the optical setup used to record the ciphertexts can be discarded. The ciphertexts and the plaintext-ciphertext pairs cannot be directly generated by the attackers. The proposed method uses an optical setup with complex scattering media to experimentally record the ciphertexts, and uses a learning model to generate security keys.

Figure 3 shows a designed convolutional neural network (CNN) for the encryption. 5000 images are randomly selected from each database to serve as plaintexts, and their corresponding speckle patterns (i.e., ciphertexts) are recorded by the CCD shown in Fig. 1. The designed learning model is composed of two convolutional layers followed by one pooling layer after each convolutional layer, one reshaping layer and one fully connected layer. The ciphertext is preprocessed by resizing it to reduce the dimensionality and computational complexity. Size of convolutional kernels depends on dimension of the input, and the number of the kernels is determined accordingly. The first convolutional layer is labeled as C_1 with weights and biases respectively denoted as $\mathbf{w}_c$ and $\mathbf{b}_c$. The feature map for C_1 can be described by

$$\mathbf{x}_1 = \sigma[(\mathbf{w}_c * \mathbf{x}_0) + \mathbf{b}_c], \tag{1}$$

where $\mathbf{x}_0$ denotes the ciphertext, $*$ denotes convolution, and σ denotes the activation function used in C_1. After down sampling, the first pooling layer (P_1) is processed by convolution, and the feature map for C_2 is given by

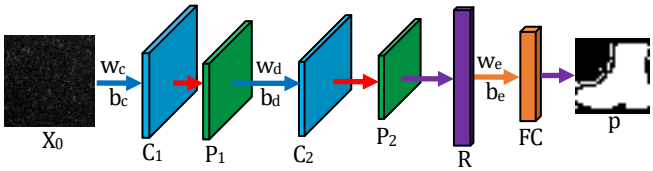

Fig. 3. A framework for the proposed learning-based encryption: $\mathbf{x}_0$, ciphertext; C_1 and C_2, the first and second convolutional layer; P_1 and P_2, the first and second pooling layer; R, shaping layer; FC, fully connected layer; p, plaintext.

$$\mathbf{x}_2 = \sigma[(\mathbf{w}_d * \mathbf{x}_p) + \mathbf{b}_d], \tag{2}$$

where $\mathbf{w}_d$ and $\mathbf{b}_d$ respectively represent weights and biases of the kernels used in C_2, and $\mathbf{x}_p$ denotes feature map of P_1. Subsequently, the pooling layer (P_2) is formed by down sampling. After image resizing, P_2 is reshaped to a one-dimensional vector (R). To achieve a prediction of the plaintext, R is connected to a fully connected layer (FC) with the usage of weights ($\mathbf{w}_e$) and biases ($\mathbf{b}_e$). The feature map of FC is described by

$$\mathbf{x}_{FC} = (\mathbf{w}_e * \mathbf{x}_R) + \mathbf{b}_e, \tag{3}$$

where $\mathbf{x}_R$ denotes feature map of the reshaping layer. Then, the one-dimensional vector of FC is reshaped to a two-dimensional vector, which is the ultimate prediction. Here, $\boldsymbol{\theta}$ is a combination of all the weights and biases.

In this study, the input ciphertext is resized from 600×600 pixels to 100×100. The resized ciphertext convolves with 20 convolutional kernels (size of 5×5) forming the first convolutional layer with size of 96×96×20. Weights $\mathbf{w}_c$ (dimension of 5×5×20) and biases $\mathbf{b}_c$ (dimension of 20×1) in Eq. (1) are randomly initialized. Activation function adopted for each convolutional layer is sigmoid. After down sampling, the first pooling layer is generated with size of 48×48×20. Down-sampling size is 2 for each pooling layer. The pooled data is further processed by convolution with 20 kernels (size of 5×5) forming the second convolutional layer with size of 44 × 44 × 20. Weights $\mathbf{w}_d$ (dimension of 5×5×20) and biases $\mathbf{b}_d$ (dimension of 20×1) in Eq. (2) are randomly initialized. Next, down-sampling processing is adopted again to generate the second pooling layer with size of 22 ×22×20. Subsequently, the second pooling layer is reshaped from a three-dimensional vector to a one-dimensional vector with size of 1×9680. The reshaped data is fully connected to the reshaped plaintext (size of 1×784). Weights $\mathbf{w}_e$ (dimension of 784×9680) and biases $\mathbf{b}_e$ (dimension of 9680×1) in Eq. (3) are randomly initialized. 4800 speckle patterns and their corresponding plaintexts are used as the training data, and other 200 ciphertexts are used to test the learning model. To evaluate the difference between the retrieved plaintexts and original plaintexts, mean squared error (MSE) is calculated. When the MSE value is higher than a preset threshold, the error is back-propagated and then the weights and biases of each layer are updated by stochastic gradient descent [22]. Here, the training epoch is selected to be 5. The momentum is set to be -9.5×10^{-4}, and the learning rate is 10^{-6}. The learning model is trained by using Matlab 2009 running on a PC with Intel Core i7@8GHz, 64GB RAM and Nvidia GTX1080Ti. Total time taken for the model training is about 4.0 hours. After the

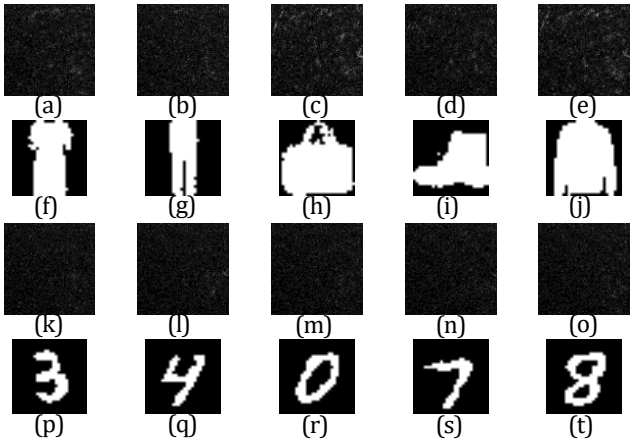

Fig. 4. Decrypted images obtained by using all correct security keys: (a)–(e) and (k)–(o) ciphertexts; (f)–(j) and (p)–(t) the retrieved plaintexts. Peak signal-to-noise ratios (PSNR) for (f)–(j) and (p)–(t) are 25.22 dB, 26.76 dB, 18.61 dB, 22.04 dB, 28.34 dB, 28.14 dB, 23.69 dB, 25.26 dB, 29.20 dB and 22.63 dB, respectively.

training, the unknown plaintext can be retrieved in real time by using the trained learning model with its parameters.

The trained learning model with its parameters, e.g., size of kernels, the number of kernels, activation function, $\mathbf{w}_c$, $\mathbf{w}_d$, $\mathbf{w}_e$ and $\boldsymbol{\theta}$, can be used as security keys. The security keys can also be updated with a new training iteration. It needs to be pointed out that security keys can be obtained by training the designed model using multiple databases. To save time, security keys for each database can be obtained individually. When all security keys are correct, the plaintexts can be fully retrieved as typically shown in Fig. 4. The security keys are further analyzed and demonstrated by using PSNR values. Figures 5(a) and 5(b) show the performance of parameters $\mathbf{w}_e$ used in the decryption process respectively for the two different databases. In this case, all other security keys are assumed to be correct. When eavesdropping percent for parameters $\mathbf{w}_e$ is lower than 99.90%, the decoded images cannot visually render useful information about the plaintexts. Moreover, performance of $\boldsymbol{\theta}$ on PSNR values of the decrypted images has also been studied as shown in Figs. 5(c) and 5(d) respectively for the two different databases. In this case, all other security keys are also assumed to be correct. It is demonstrated in Figs. 5(c) and 5(d) that when eavesdropping percent for $\boldsymbol{\theta}$ is lower than 99.95%, the decoded images cannot visually render useful information about the plaintexts. For the sake of brevity, eavesdropping analysis of other security keys is not presented here. These eavesdropping analyses demonstrate that security of the proposed learning-based optical cryptosystem can be fully guaranteed. Without accurate knowledge about security keys, the plaintexts cannot be effectively extracted. The proposed method uses an optical setup with complex scattering media to experimentally record the ciphertexts, and uses a learning model to generate security keys. Therefore, high security can be achieved.

The higher security can be flexibly achieved by using virtual phase-only masks which serve as supplementary security keys for the decryption. One virtual phase-only mask is further used here and shown in Fig. 6. The axial distance (L_1) between the CCD and virtual phase-only mask is 4.0 cm, and that (L_2) between virtual phase-only mask and the D_2 plane is 2.0 cm. The ciphertexts

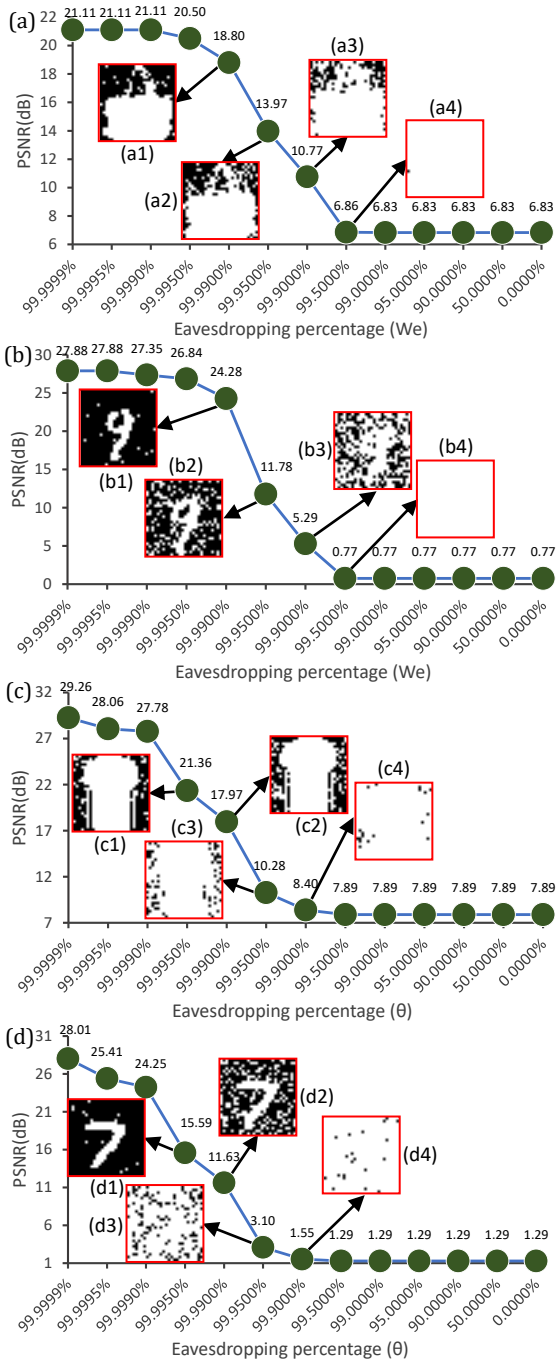

Fig. 5. Eavesdropping analysis of security keys $\mathbf{w}_e$ and $\boldsymbol{\theta}$.

recorded by the CCD are back-propagated to the D_2 plane when virtual phase-only mask is used, and amplitude-only patterns obtained in the D_2 plane rather than the recorded ciphertexts are used as the inputs for the designed learning model. After the training, virtual phase-only mask, axial distances, wavelength, and the trained learning model with its parameters can be used as security keys. The speckle patterns recorded by the CCD are still used as ciphertexts. Figures 7(a)–7(t) show the plaintexts retrieved from their ciphertexts by using all correct security keys. Performance of virtual phase-only mask in the decryption process

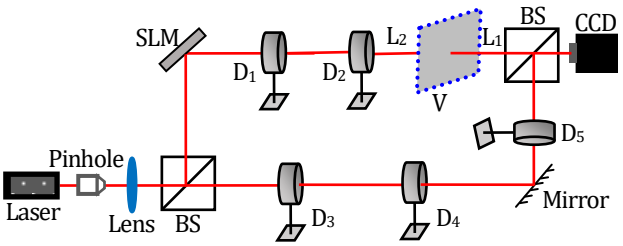

Fig. 6. A schematic experimental setup with complex scattering media: L_1 and L_2, axial distances; V, virtual phase-only mask. A virtual phase-only mask is further used as supplementary security key for the decryption.

is demonstrated in Fig. 8. As shown in Figs. 8(a) and 8(b), when eavesdropping percent for virtual phase-only mask is lower than 99.99%, the decoded images cannot visually render useful information about the plaintexts. For the sake of brevity, eavesdropping analysis of other security keys is not presented here. It is expected that multiple virtual phase-only masks used for the decryption would further enhance the security.

In conclusion, we have proposed a new method for optical encryption by learning complex scattering media. Instead of directly using the parameters in optical setup as security keys, the proposed method uses the trained learning model with its parameters as security keys. In addition to the trained model with its parameters, other parameters, e.g., virtual phase-only mask, can be flexibly used to enlarge key space. The proposed method uses an optical setup with complex scattering media to experimentally record the ciphertexts, and uses a learning model to generate security keys. Therefore, high security is achieved in the proposed method. The proposed method can be theoretically and experimentally implemented, and is not limited to the typical optical setup and the typical number of diffusers presented in this study. The proposed learning-based encryption might open up a new research perspective for optical encryption.

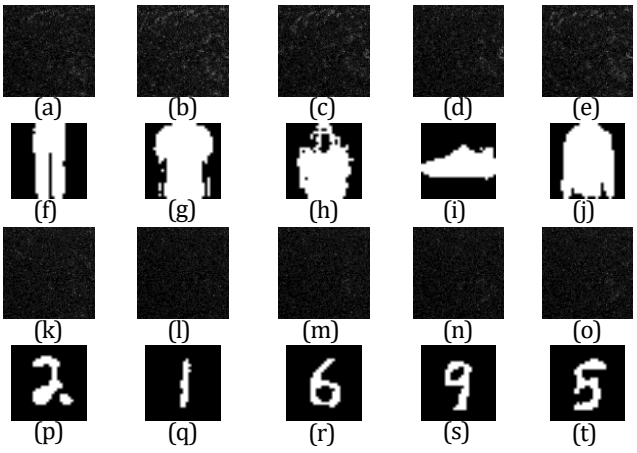

Fig. 7. Decrypted images obtained by using all correct security keys: (a)–(e) and (k)–(o) ciphertexts; (f)–(j) and (p)–(t) the retrieved plaintexts. The PSNR values for (f)–(j) and (p)–(t) are 30.65 dB, 24.63 dB, 16.09 dB, 29.26 dB, 25.42 dB, 23.08 dB, 34.36 dB, 24.63 dB, 29.59 dB and 22.04 dB, respectively.

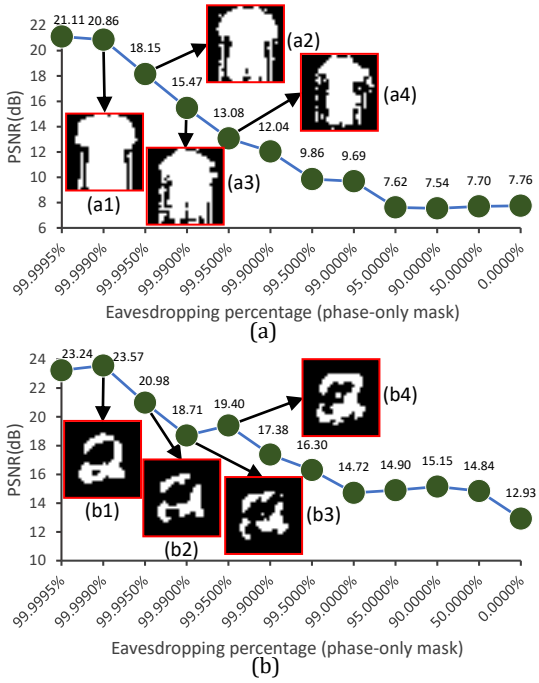

Fig. 8. Eavesdropping analysis of virtual phase-only mask. All other security keys are assumed to be correct.

Funding. Hong Kong Research Grants Council (25201416, C5011-19G).

Disclosures. The authors declare no conflicts of interest.

References

1. B. Javidi, Phys. Today **50**, 27 (1997).
2. O. Matoba, T. Nomura, E. Perez-Cabre, M. S. Millan, and B. Javidi, Proc. IEEE **97**, 1128 (2009).
3. P. Refregier and B. Javidi, Opt. Lett. **20**, 767 (1995).
4. A. Alfalou and C. Brosseau, Adv. Opt. Photon. **1**, 589 (2009).
5. W. Chen, B. Javidi, and X. Chen, Adv. Opt. Photon. **6**, 120 (2014).
6. G. Unnikrishnan, J. Joseph, and K. Singh, Opt. Lett. **25**, 887 (2000).
7. O. Matoba and B. Javidi, Opt. Lett. **24**, 762 (1999).
8. W. Chen, X. Chen, and Colin J. R. Sheppard, Opt. Lett. **35**, 3817 (2010).
9. P. Clemente, V. Durán, V. Torres-Company, E. Tajahuerce, and J. Lancis, Opt. Lett. **35**, 2391 (2010).
10. Y. Zhang and B. Wang, Opt. Lett. **33**, 2443 (2008).
11. E. G. Johnson and J. D. Brasher, Opt. Lett. **21**, 1271 (1996).
12. Y. Shi, T. Li, Y. Wang, Q. Gao, S. Zhang, and H. Li, Opt. Lett. **38**, 1425 (2013).
13. A. Carnicer, M. Montes-Usategui, S. Arcos, and I. Juvells, Opt. Lett. **30**, 1644 (2005).
14. X. Peng, H. Wei, and P. Zhang, Opt. Lett. **31**, 3261 (2006).
15. X. Peng, P. Zhang, H. Wei, and B. Yu, Opt. Lett. **31**, 1044 (2006).
16. M. Liao, W. He, D. Lu, and X. Peng, Sci. Rep. **7**, 41789 (2017).
17. L. Zhou, Y. Xiao, and W. Chen, Opt. Express **27**, 26143 (2019).
18. H. Xiao, K. Rasul, and R. Vollgraf, arXiv preprint arXiv:1708.07747 (2017).
19. L. Deng, IEEE Signal Process. Mag. **29**, 141 (2012).
20. Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
21. K. Zhang, W. Zuo, S. Gu, and L. Zhang, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3929 (2017).
22. I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton, Proceedings of the 30th International Conference on Machine Learning, PMLR, **28**, 1139 (2013).

References with titles

1. B. Javidi, "Securing information with optical technologies," Phys. Today **50**, 27–32 (1997).
2. O. Matoba, T. Nomura, E. Perez-Cabre, M. S. Millan, and B. Javidi, "Optical techniques for information security," Proc. IEEE **97**, 1128–1148 (2009).
3. P. Refregier and B. Javidi, "Optical image encryption based on input plane and Fourier plane random encoding," Opt. Lett. **20**, 767–769 (1995).
4. A. Alfalou and C. Brosseau, "Optical image compression and encryption methods," Adv. Opt. Photon. **1**, 589–636 (2009).
5. W. Chen, B. Javidi, and X. Chen, "Advances in optical security systems," Adv. Opt. Photon. **6**, 120–155 (2014).
6. G. Unnikrishnan, J. Joseph, and K. Singh, "Optical encryption by double-random phase encoding in the fractional Fourier domain," Opt. Lett. **25**, 887–889 (2000).
7. O. Matoba and B. Javidi, "Encrypted optical memory system using three-dimensional keys in the Fresnel domain," Opt. Lett. **24**, 762–764 (1999).
8. W. Chen, X. Chen, and Colin J. R. Sheppard, "Optical image encryption based on diffractive imaging," Opt. Lett. **35**, 3817–3819 (2010).
9. P. Clemente, V. Durán, V. Torres-Company, E. Tajahuerce, and J. Lancis, "Optical encryption based on computational ghost imaging," Opt. Lett. **35**, 2391–2393 (2010).
10. Y. Zhang and B. Wang, "Optical image encryption based on interference," Opt. Lett. **33**, 2443–2445 (2008).
11. E. G. Johnson and J. D. Brasher, "Phase encryption of biometrics in diffractive optical elements," Opt. Lett. **21**, 1271–1273 (1996).
12. Y. Shi, T. Li, Y. Wang, Q. Gao, S. Zhang, and H. Li, "Optical image encryption via ptychography," Opt. Lett. **38**, 1425–1427 (2013).
13. 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**, 1644–1646 (2005).
14. X. Peng, H. Wei, and P. Zhang, "Chosen-plaintext attack on lensless double-random phase encoding in the Fresnel domain," Opt. Lett. **31**, 3261–3263 (2006).
15. X. Peng, P. Zhang, H. Wei, and B. Yu, "Known-plaintext attack on optical encryption based on double random phase keys," Opt. Lett. **31**, 1044–1046 (2006).
16. 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**, 41789 (2017).
17. L. Zhou, Y. Xiao, and W. Chen, "Machine-learning attacks on interference-based optical encryption: experimental demonstration," Opt. Express **27**, 26143–26154 (2019).
18. H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," arXiv preprint arXiv:1708.07747 (2017).
19. L. Deng, "The MNIST database of handwritten digit images for machine learning research [best of the web]," IEEE Signal Process. Mag. **29**, 141–142 (2012).
20. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
21. 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, 3929–3938 (2017).
22. I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton, "On the importance of initialization and momentum in deep learning," Proceedings of the 30th International Conference on Machine Learning, PMLR, **28**, 1139–1147 (2013).