https://doi.org/10.1016/j.optlaseng.2021.106570

(Manuscript submitted to the OLEN)

Learning-based optical authentication in complex scattering media

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

Abstract

In recent years, optical encryption has become a promising method for securing information. However, optical cryptosystems have been demonstrated to be vulnerable to the attacking algorithms. Therefore, it is highly desirable that new optical cryptosystems can be continuously developed to achieve the higher security and withstand the attacks. In this paper, learning-based optical authentication in complex scattering media is proposed and experimentally verified. The ciphertext is generated by using an optical setup in complex scattering media, and a learning model is developed and trained to generate security keys. Moreover, other parameters, e.g., virtual phase-only masks, can be flexibly designed and used to further enlarge key space. The training pairs fed to the designed leaning model consist of specially processed images (i.e., acting as plaintexts) and speckle patterns (i.e., ciphertexts) recorded by CCD. The retrieved plaintexts obtained from the trained learning model cannot visually render recognizable information, and can be effectively authenticated by using nonlinear correlation algorithm. Optical experiments are conducted to verify effectiveness and applicability of the proposed learning-based optical authentication in complex scattering media to achieve the higher security and withstand the attacks. © Elsevier

Keywords: Optical authentication, Learning-based optical security, Optical encoding, Complex scattering media, Experimental demonstration

1. Introduction

It has been well recognized that information security is essential for data storage and transmission in various application areas [1]. Owing to remarkable advantages of optical means including parallel operation and multi-dimensional freedom, optical cryptography provides a promising and significant approach for securing information [2–4]. Double random phase encoding (DRPE) proposed by Refregier and Javidi [5] was the first technique for securing information in an optical way. The DRPE-based scheme has been developed and extended from Fourier domain to fractional Fourier domain, Fresnel domain and Gyrator domain etc [6–9]. Apart from DRPE schemes, other optical techniques have been continuously developed to enhance the security, e.g., interference-based, diffractive-imaging-based, computer-generated-hologram-based, ghost-imaging-based and compressive sensing-based [10–14]. Later on, sparse representation based optical authentication [15] has been further developed to establish an additional security layer. However, various cryptoanalyses were developed to analyze the vulnerability of optical cryptosystems [16–20]. The attacking algorithms originated from the work done by Carnicer et al. [16] who demonstrated the vulnerability of DRPE-based schemes by using chosen-ciphertext attack. Different attacking algorithms have sprung up, e.g., chosen-plaintext attack, ciphertext-only attack and known-plaintext attack [17–21]. A recent study of learning-based attack demonstrated how to attack optical cryptosystems without direct retrieval of security keys and the usage of complex phase retrieval algorithms [22]. Therefore, it is highly desirable that more secure and advanced optical cryptosystems can be continuously developed to withstand the attacks.

In this paper, learning-based optical authentication in complex scattering media is proposed and experimentally verified to achieve the higher security and withstand the attacks. It is an exemplification in this study to use an interferometric setup to demonstrate feasibility and effectiveness of the proposed method. The proposed method is implemented by using multiple cascaded diffusers in an optical setup, and a series of speckle patterns (i.e., ciphertexts) are recorded. Convolutional neural network (CNN) [23,24] is designed and trained by using the plaintexts and the corresponding ciphertexts (i.e., a series of speckle patterns) respectively as outputs and inputs. Here, the plaintexts sent to the designed learning model are obtained as follows: grayscale images from the MNIST or fashion MNIST databases [25,26] are Hadamard transformed, and then the generated Hadamard spectrum is further processed by using a sparsity constraint [27,28] followed by inverse Hadamard transform. In the proposed method, security keys consist of the trained learning model with its training parameters. In addition, it is also demonstrated that other parameters, e.g., virtual phase-only masks, can be flexibly designed and applied to further enlarge key space. By using the trained learning model, an arbitrarily given ciphertext can be processed to retrieve its corresponding plaintext information. Since the plaintext retrieved from the trained learning model cannot visually render recognizable information, nonlinear correlation algorithm is further applied for optical authentication. When correct security keys are used, the retrieved plaintext can be effectively authenticated with only one sharp peak in the generated nonlinear correlation map. The proposed learning-based optical authentication approach can be applied and integrated into various optical cryptosystems to open up a novel research perspective for optical security.

2. Experimental setup and principles

2.1. Experimental setup

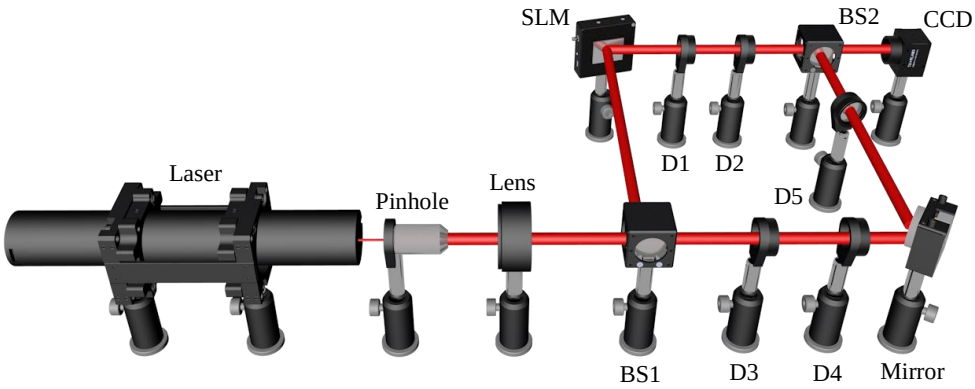


Fig. 1. A schematic optical setup for the proposed learning-based optical authentication in complex scattering media. SLM: spatial light modulator; BS1 and BS2: beam splitter; D1-D5: diffusers; CCD: charge-coupled device.

An interference-based optical setup in Fig. 1 is used to demonstrate feasibility and effectiveness of the proposed method. A schematic experimental setup for the proposed learning-based optical authentication in complex scattering media is shown in Fig. 1. Complex scattering media used in the optical setup can be flexibly designed by using multiple cascaded diffusers. A laser beam with wavelength of 632.8 nm is expanded and collimated by a pinhole and a lens to generate a plane wave, and then the collimated light passes through the first beam splitter (BS1) to be split into two laser beams. One laser beam (i.e., object beam) is reflected by a spatial light modulator (SLM, Holoeye LC-R 720), and then is scattered by two cascaded diffusers D1 and D2 (Thorlabs, DG10-600). Another laser beam i.e., reference

beam, is scattered by three cascaded diffusers D3, D4 and D5. Finally, the two laser beams interfere with each other, and a speckle pattern is recorded by CCD (Thorlabs, DCC1240). Handwritten-digit images from the MNIST database and fashion-product images from the fashion MNIST database [25,26] are selected and used, and the image size is 64×64 pixels. In the proposed method, the plaintexts are obtained as follows: grayscale images randomly selected from the databases are pre-processed by Hadamard transform, and then the generated Hadamard spectrum is further processed by using a sparsity constraint followed by inverse Hadamard transform. Figures 2(a)–2(d) show several typical grayscale images respectively from the MNIST database and fashion MNIST database, and the corresponding ciphertexts recorded by CCD are shown in Figs. 2(e)–2(h), respectively. Here, a novel method, called learning-based optical authentication in complex scattering media, is proposed and experimentally verified. The complex scattering media can be flexibly designed by using a different number of diffusers, and the optical setup to be used for recording the ciphertexts can also be flexibly selected in practice. It is worth noting that the proposed method can be numerically or experimentally carried out and verified. In this study, the proposed method is experimentally demonstrated by using an interference-based optical setup with multiple cascaded diffusers, as schematically shown in Fig. 1.

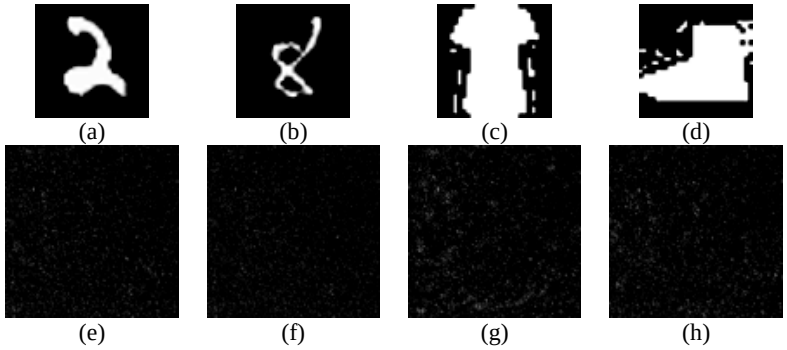


Fig. 2. (a)-(d) Typical grayscale images (64×64 pixels) selected from the MNIST database and fashion MNIST database, and (e)-(h) ciphertexts (600×600 pixels) recorded by CCD respectively corresponding to (a)-(d).

2.2. Learning-based optical authentication

A novel learning-based optical authentication in complex scattering media is proposed and illustrated in Fig. 3, and the process is as follows: from each database, i.e., MNIST database and fashion MNIST database [25,26], 5000 grayscale images are randomly selected. Each grayscale image, denoted as P, is further processed by using Hadamard transform described by

$$\mathrm{F}=\mathbf{H}\mathrm{P}, \tag{1}$$

where **H** denotes a standard Hadamard matrix [29], and F denotes Hadamard spectrum. A sparsity constraint is applied to compress the generated Hadamard spectrum F in order to generate a sparse Hadamard spectrum F_s. To be specific, sparse Hadamard spectrum F_s is obtained by reserving certain amount of the randomly selected Hadamard spectrum coefficients and setting other Hadamard spectrum coefficients to be zero. The Hadamard matrix **H** is composed of orthogonal bases, and a plaintext can be obtained by using inverse Hadamard transform $\mathbf{H}^{-1}$ [29] described by

$$\mathrm{P}_R=\mathbf{H}^{-1}\mathrm{F}_s, \tag{2}$$

where P_R denotes the plaintext in this study. With pairs of the plaintexts P_R and the corresponding speckle patterns recorded by CCD, a designed learning model can be well trained. The trained learning model with its training parameters is used as security keys. When correct security keys are used, an arbitrarily given ciphertext sent to the trained learning model can be processed and the retrieved plaintext does not visually render recognizable information. To authenticate the retrieved plaintext, a nonlinear correlation algorithm [15] is employed. Information authentication is implemented by using nonlinear correlation between Hadamard spectrum of the retrieved plaintext and Hadamard spectrum of original plaintext P_R. The retrieved plaintext can be effectively authenticated by nonlinear correlation with only one sharp peak in the generated correlation map, when correct security keys are used. It means that the receiver is an authorized person or possesses correct security keys. When the generated nonlinear correlation distribution is noisy with multiple sharp peaks, it means that the receiver is an unauthorized person or has used wrong security keys.

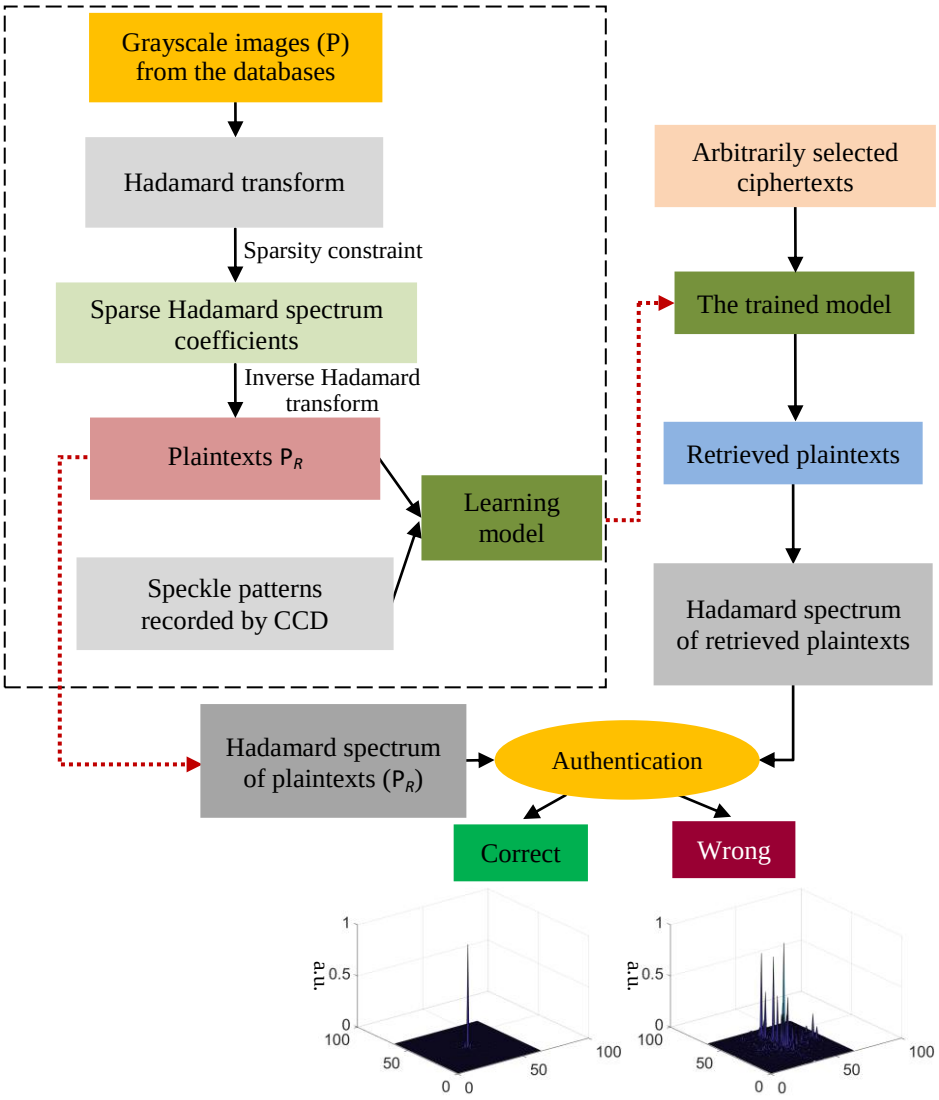


Fig. 3. A flow chart for the proposed learning-based optical authentication approach.

A typical example using the MNIST database is shown in Figs. 4(a)–4(d). In this case, only 10% of Hadamard spectrum coefficients are randomly selected to be reserved and others are discarded. Figures 4(e)–4(h) show a typical example using the fashion MNIST database. It can be seen that the plaintexts P_R in Figs. 4(d) and 4(h) cannot visually render any recognizable information about images P in Figs. 4(a) and 4(e), respectively.

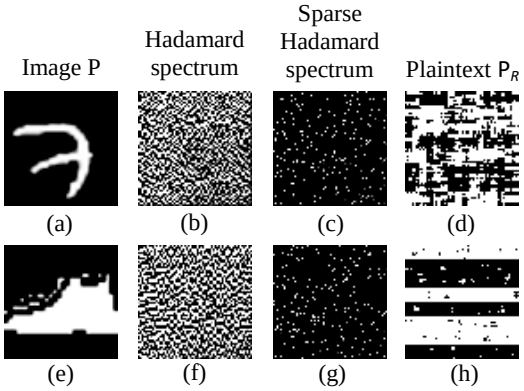


Fig. 4. (a) and (e) Typical images respectively from the MNIST database and fashion MNIST database, (b) and (f) Hadamard spectrums respectively corresponding to (a) and (e), (c) and (g) sparse Hadamard spectrums using a sparsity constraint with a sampling rate of 10% respectively corresponding to (b) and (f), and (d) and (h) the generated plaintexts P_R respectively corresponding to (a) and (e).

2.3. A designed CNN model

Here, a CNN model is designed and applied to demonstrate feasibility and effectiveness of the proposed learning-based optical authentication in complex scattering media, as shown in Fig. 5. The designed CNN model for the proposed learning-based optical authentication is shown in Fig. 5(a), which is described as follows: the input speckle pattern (i.e., ciphertext) with 600×600 pixels is first resized to 100×100 pixels in order to reduce the computational load. Subsequently, the resized ciphertext convolves with 20 convolutional kernels (size of 5×5) to extract effective information, forming the first convolutional layer (C_1) with size of 96×96×20. Weights w_a and bases b_a for C_1 are randomly initialized, and sizes of w_a and b_a are 5×5×20 and 20×1, respectively. Then, the extracted feature map in C_1 is down-sampled to the first pooling layer (P_1) with a dimension of 48×48×20. The first pooling layer is further convolved with 20 convolutional kernels with size of 5×5, generating the second convolutional layer (C_2) with size of 44×44×20. Weights w_b and bases b_b for C_2 are also randomly initialized, and size of w_b and b_b are 5×5×20 and 20×1, respectively. Following C_2, the second pooling layer P_2 (22×22×20) is generated by down-sampling. In this study, two convolutional layers and two pooling layers are used. More convolutional layers and pooling layers can be adopted for feature map extraction in practice. Then, the second pooling layer (P_2) is reshaped to be a one-dimensional vector with size of 1×9680. The reshaped data (R) is fully connected to a one-dimensional vector with size of 1×4096. Weights w_e and bases b_e for the fully connect layer (FC) are also randomly initialized, and sizes of w_e and b_e are 4096×9680 and 4096×1, respectively. Finally, FC is reshaped to be a two-dimensional vector with a dimension of 64×64. It is worth noting that the activation function used is sigmoid. 4600 pairs of speckle patterns and the corresponding plaintexts P_R are fed to the designed CNN model. Mean squared error (MSE) is calculated to monitor the training process. When the MSE is higher than a preset threshold, the calculated error is back propagated and the weights and bases are updated by stochastic gradient descent [30,31]. Here, there are 5 training iterations. The learning rate is set to be 10^{-6}, and the momentum is set to be -9.5×10^{-4}.

The training process is implemented by using Matlab 2009 on a PC with Nvidia Geforce GTX1080Ti GPU and Intel Core i7@8GHz, 64GB RAM. It takes about 23.0 hours to train the learning model for each database. Finally, the designed learning model can be well trained, and other 400 pairs are used to test the trained learning model. Different from traditional optical cryptosystems, the proposed method exploits the trained learning model with the training parameters as security keys. Although there are no explicit physical implications of the training parameters, they play an important role in the plaintext retrieval. To be specific, these training parameters, e.g., weights and biases, can be considered as matrices to be transmitted. Only when correct security keys are known and applied, plaintexts can be effectively retrieved. Figure 5(b) shows a typical example for a given ciphertext sent to the trained CNN model. It can be seen that the retrieved plaintext does not visually render any recognizable information, and will be further authenticated in this study.

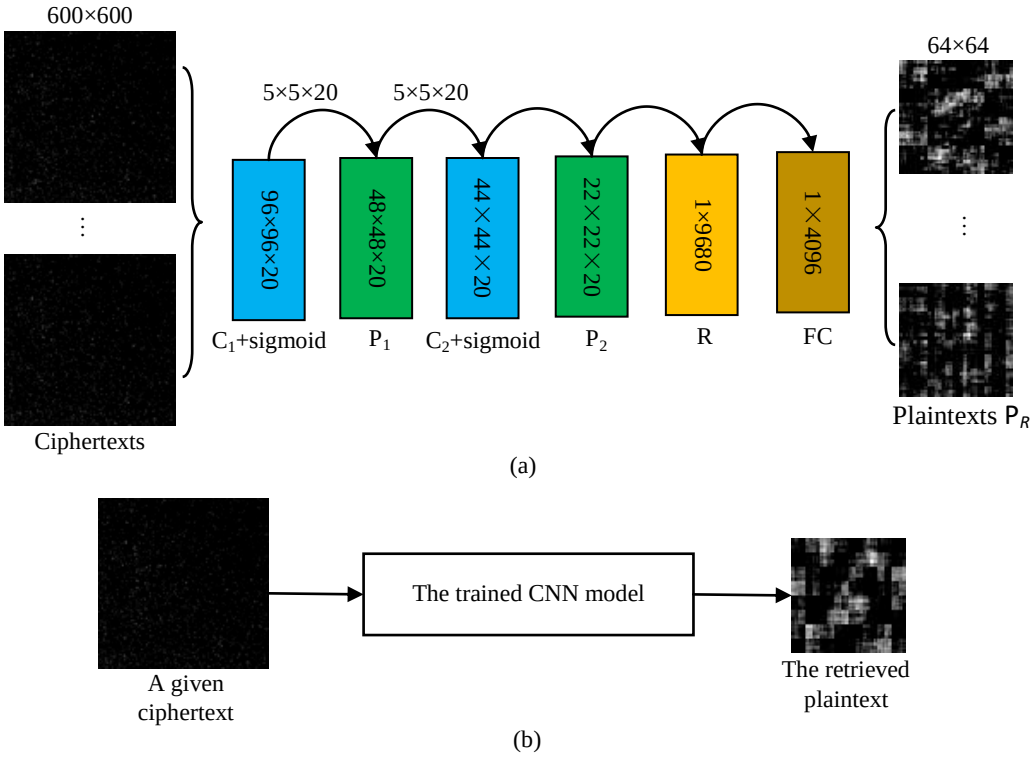


Fig. 5. The designed CNN model for the proposed learning-based optical authentication in complex scattering media. There are two convolutional layers, two pooling layers, one reshaping layer and one fully connected layer. (a) Training phase: the ciphertexts are resized from 600×600 pixels to 100×100 pixels, and then are used as inputs. The outputs are the corresponding plaintexts P_R. With pairs of ciphertexts and the corresponding plaintexts fed to the designed CNN model, the designed learning model can be fully trained. (b) Testing phase: the trained CNN model can be used to make a prediction in real time.

3. Experimental results and discussion

Figures 6(a)–6(n) show several plaintext retrieval results when different sampling rates are used as sparsity constraints in Hadamard transform domain to randomly select Hadamard spectrum coefficients. The typical ciphertexts in Figs. 6(a) and 6(h) are recorded by using the optical setup in Fig. 1. Figure 6(a) shows the ciphertext when the MNIST database is used, and Fig. 6(h) shows the ciphertext when the fashion MNIST database is used. These two

ciphertexts are respectively sent to the correspondingly trained CNN models. Figures 6(b)–6(g) show the retrieved plaintexts corresponding to Fig. 6(a), when different sampling rates of 5%, 10%, 20%, 30%, 40% and 70% are respectively applied in Hadamard domain. Figures 6(i)–6(n) show the retrieved plaintexts corresponding to Fig. 6(h), when different sampling rates of 5%, 10%, 20%, 30%, 40% and 70% are respectively applied in Hadamard domain. Peak signal-to-noise ratios (PSNRs) for Figs. 6(b)–6(g) are 19.99 dB, 21.17 dB, 22.18 dB, 22.80 dB, 24.18 dB and 23.80 dB, respectively. The PSNRs for Figs. 6(i)–6(n) are 8.15 dB, 11.56 dB, 12.15 dB, 10.68 dB, 11.99 dB and 13.88 dB, respectively. When the higher sampling rate is used, more effective information about the plaintexts can be visually rendered as shown in Figs. 6(g) and 6(n).

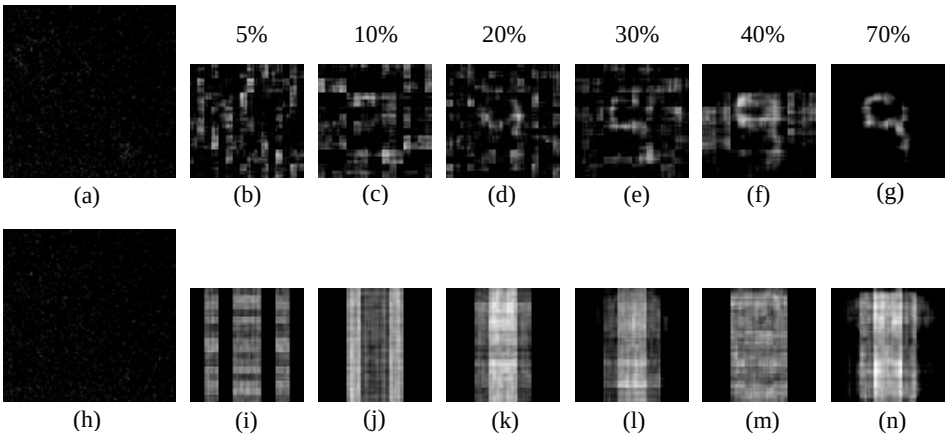


Fig. 6. (a) and (h) Typical ciphertexts obtained when the MNIST database and fashion MNIST database are respectively used, (b) and (i) the retrieved plaintexts when the sampling rate of 5% is used in Hadamard domain, (c) and (j) the retrieved plaintexts when the sampling rate of 10% is used, (d) and (k) the retrieved plaintexts when the sampling rate of 20% is used, (e) and (l) the retrieved plaintexts when the sampling rate of 30% is used, (f) and (m) the retrieved plaintexts when the sampling rate of 40% is used, and (g) and (n) the retrieved plaintexts when the sampling rate of 70% is used.

Although the retrieved plaintexts in Figs. 6(b)–6(d) and 6(i)–6(m) cannot visually render recognizable information, they can be further authenticated in order to establish an additional security layer. Typical examples are used to demonstrate validity of the proposed learning-based optical authentication approach, as shown in Fig. 7. The images shown at the first row in Fig. 7 are obtained by using the MNIST database, and the images shown at the second row in Fig. 7 are obtained by using the fashion MNIST database. For each database, the designed learning model is trained with the corresponding training data. The training data consist of the ciphertexts recorded by CCD and the plaintexts P_R obtained by using Hadamard transform with a sampling rate of 5%. By using correct security keys, the given ciphertexts sent to the correspondingly trained learning model can be decoded. The ciphertexts in Figs. 7(a) and 7(e) are sent to the correspondingly trained learning models, and the retrieved plaintexts are shown in Figs. 7(b) and 7(f), respectively. It can be seen in Figs. 7(b) and 7(f) that the retrieved plaintexts are not recognizable. The retrieved plaintexts are further authenticated by using a nonlinear correlation algorithm [15]. The generated nonlinear correlation distributions are shown in Figs. 7(c) and 7(g), respectively. It can be seen that only one sharp peak is generated in each nonlinear correlation distribution, verifying feasibility and effectiveness of the proposed method. When the receiver is an authorized person or possesses correct security keys, there is only one sharp peak in the generated nonlinear correlation maps. When wrong security keys, e.g., wrong weights and biases, are used in the proposed learning-based authentication method, the generated nonlinear correlation maps are noisy with multiple sharp

peaks, as shown in Figs. 7(d) and 7(h). In this study, it has been illustrated that ciphertexts are generated by using an optical setup in complex scattering media, and the trained learning model with its training parameters is used as security keys rather than parameters in the optical setup. Therefore, conditions for the attacking algorithms [16–22] become invalid, and withstanding the potential attacks can be realized.

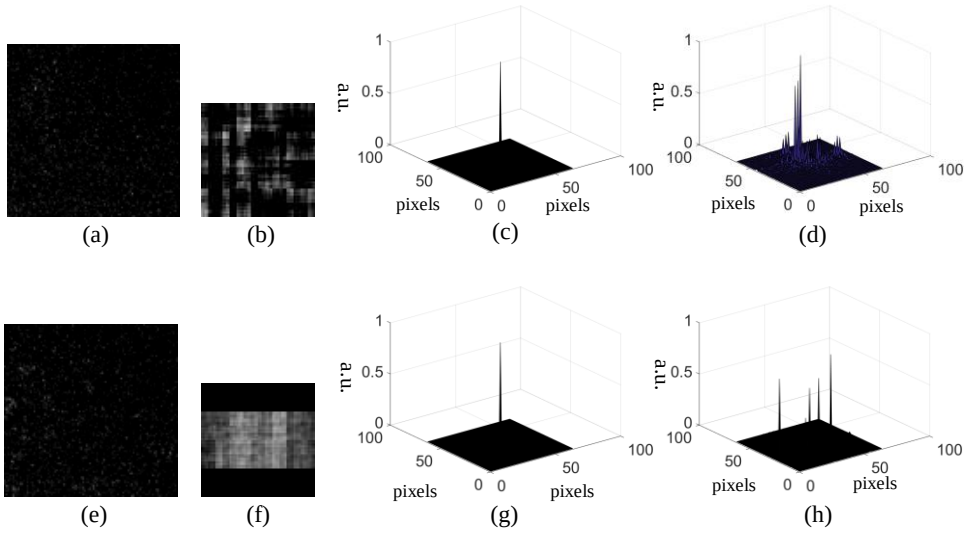
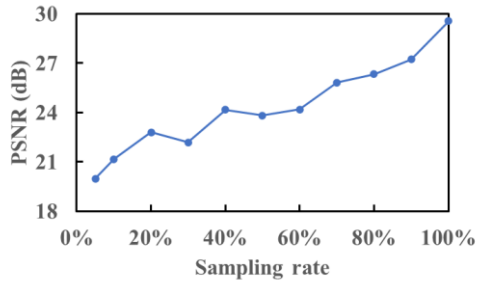


Fig. 7. Experimental results obtained by using the proposed learning-based optical authentication approach: (a) and (e) Ciphertexts obtained when the MNIST database and fashion MNIST database are respectively used, (b) and (f) the retrieved plaintexts obtained by the correspondingly trained learning model with correct security keys, (c) and (g) the generated nonlinear correlation maps respectively corresponding to (b) and (f), and (d) and (h) the generated nonlinear correlation maps obtained when wrong security keys, i.e., wrong weights and biases, are used.

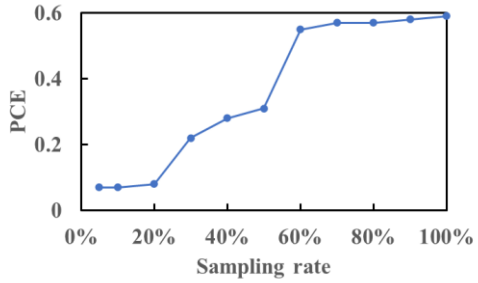
To quantitatively evaluate the quality, PSNR and peak-to-correlation energy (PCE) are calculated. The PCE is defined as a ratio between the maximum correlation peak and the total energy of correlation distribution. The higher value of PCE means the better correlation between two objects, and the higher value of PSNR denotes the higher quality of the retrieved images. Here, PSNR is calculated to evaluate quality of the retrieved plaintexts with respect to different sampling rates, and the PCE is calculated to analyze the generated nonlinear correlation distributions with respect to different sampling rates. Figures 8(a)–8(d) show the curves of PSNRs and PCEs for the two different databases with respect to different sampling rates of 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. Figures 8(a) and 8(c) show that PSNR values of the retrieved plaintexts increase when the higher sampling rates are used. Figure 8(b) shows the PCE of the generated nonlinear correlation between Hadamard spectrums of original plaintexts P_R (generated by using the MNIST database) and Hadamard spectrums of the correspondingly retrieved plaintexts. Figure 8(d) shows the PCE of the generated nonlinear correlation between Hadamard spectrums of original plaintexts P_R (generated by using the fashion MNIST database) and Hadamard spectrums of the correspondingly retrieved plaintexts. As can be seen in Figs. 8(a)–8(d), the higher sampling rate can lead to the plaintext retrieval with the larger PSNR, and the higher sampling rate can lead to the generation of nonlinear correlation distribution with the sharper peak.

Eavesdropping analysis is also carried out to vet the security of the proposed learning-based optical authentication approach. For the sake of brevity, the weights w_e used in the FC layer are analyzed here, and other security keys are assumed to be correct. The sampling rate is 5% which is used to compress the Hadamard spectrum in order to generate plaintexts P_R.

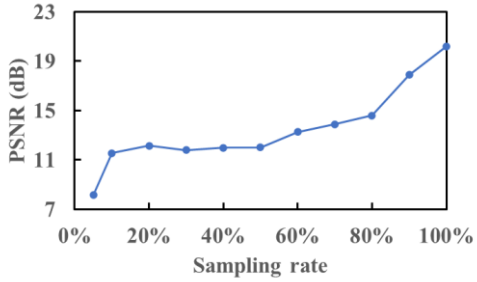
Figures 9(a)–9(d) show PSNR and PCE values with respect to different eavesdropping percentages of weights w_e in the FC layer. Although different PSNR values are obtained with respect to different eavesdropping percentages of weights w_e in the FC layer in Figs. 9(a) and 9(b), all the retrieved plaintexts cannot visually render recognizable information since only 5% of Hadamard spectrum coefficients are used to generate plaintexts P_R. As can be seen in Figs. 9(c) and 9(d), when the eavesdropping percentage of weights w_e in the FC layer is lower than 99.99999%, PCE values of the generated nonlinear correlation distributions are low. It is demonstrated that the PSNR values cannot be used to evaluate optical authentication in this study, and the PCE values can be calculated and effectively applied.



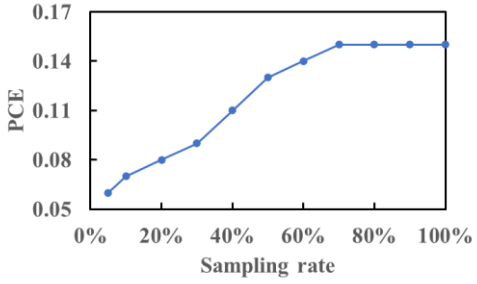
(a)



(b)



(c)



(d)

Fig. 8. PSNRs of the retrieved plaintexts and PCEs of the generated nonlinear correlation distribution with respect to different sampling rates of 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. (a) and (c) PSNR curves respectively for the MNIST database and fashion MNIST database, and (b) and (d) PCE curves respectively for the MNIST database and fashion MNIST database.

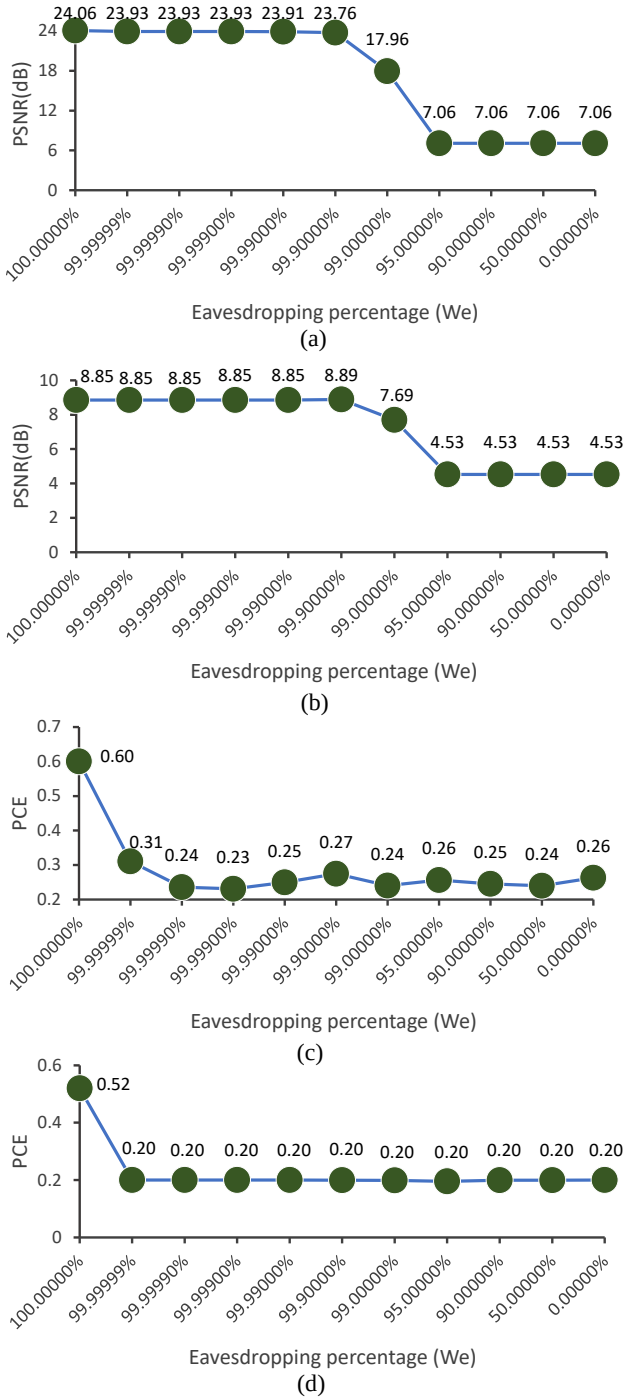


Fig. 9. PSNR of the retrieved plaintexts and PCE of the generated nonlinear correlation distributions with respect to different eavesdropping percentages of weights w_e in the FC

layer. (a) and (c) PSNR curves respectively for the MNIST database and fashion MNIST database, and (b) and (d) PCE curves respectively for the MNIST database and fashion MNIST database. The sampling rate, i.e., only 5% of Hadamard spectrum coefficients, is fixed here and used to generate plaintexts P_R.

To further enhance security of the proposed learning-based optical authentication in complex scattering media, virtual phase-only masks can be flexibly designed and applied as additional security keys. In Fig. 10, ciphertexts recorded by CCD are back-propagated to the first virtual phase-only mask (V1) plane with an axial distance of 1.0 mm, and then are further back-propagated to the second virtual phase-only mask (V2) plane with an axial distance of 1.0 mm. Then, amplitude-only patterns just before mask V2 are reserved and used as inputs for the designed learning model. 4600 pairs of the training data are used to generate a trained learning model, and other 400 pairs are used for testing. The trained learning model with its training parameters, virtual phase-only mask V1, axial distances and wavelength can be used as security keys. When back propagation with a certain axial distance is further conducted before virtual mask V2, the second virtual phase-only mask V2 can be further used as security key. Note that the speckle patterns recorded by CCD are still employed as ciphertexts rather than the amplitude-only patterns sent to the designed learning model.

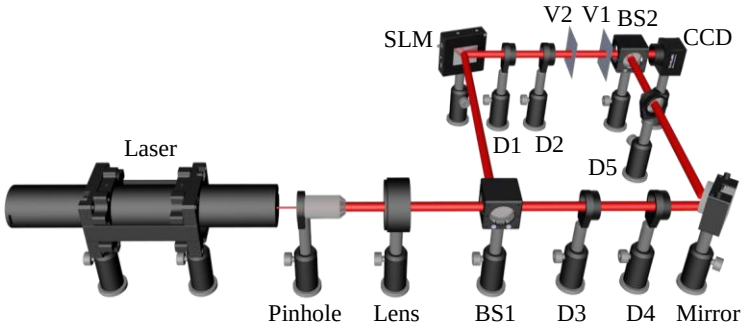


Fig. 10. A schematic optical setup for the proposed learning-based optical authentication approach with virtual phase-only masks. V1 and V2: virtually random phase-only masks.

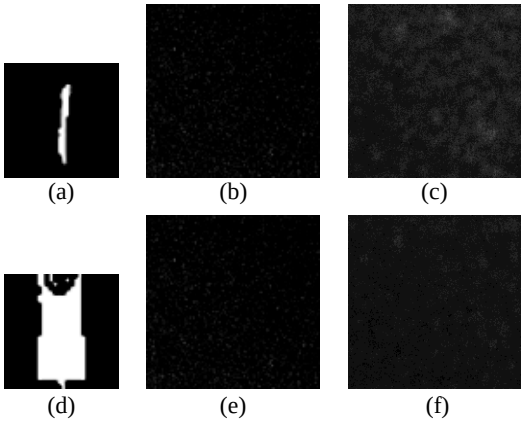


Fig. 11. (a) and (d) Typical images (64×64 pixels) respectively selected from the MNIST database and the fashion MNIST database, (b) and (e) the corresponding ciphertexts (600×600 pixels) recorded by CCD, and (c) and (f) the corresponding amplitude-only patterns just before virtual mask V2 using the optical setup in Fig.10.

Figures 11(a)–11(f) show two typical examples, when the MNIST database and the fashion MINST database are respectively used based on the optical setup in Fig. 10. The generated amplitude-only patterns just before virtual mask V2 and the corresponding plaintexts P_R are sent to a designed learning model for the training. Figures 12(a)–12(n) show several plaintext retrieval results by using different sampling rates. Figures 12(a) and 12(h) show the generated amplitude-only patterns just before virtual mask V2 when the MNIST database and the fashion MNIST database are respectively used. Figures 12(b)–12(g) and 12(i)–12(n) show the retrieved plaintexts, when different sampling rates of 5%, 10%, 20%, 30%, 40% and 70% are used, respectively. The PSNRs for Figs. 12(b)–12(g) are 18.89 dB, 19.00 dB, 23.18 dB, 23.73 dB, 23.59 dB and 23.62 dB, respectively. The PSNRs for Figs. 12(i)–12(n) are 7.93 dB, 8.82 dB, 14.07 dB, 14.15 dB, 15.89 dB and 17.18 dB, respectively.

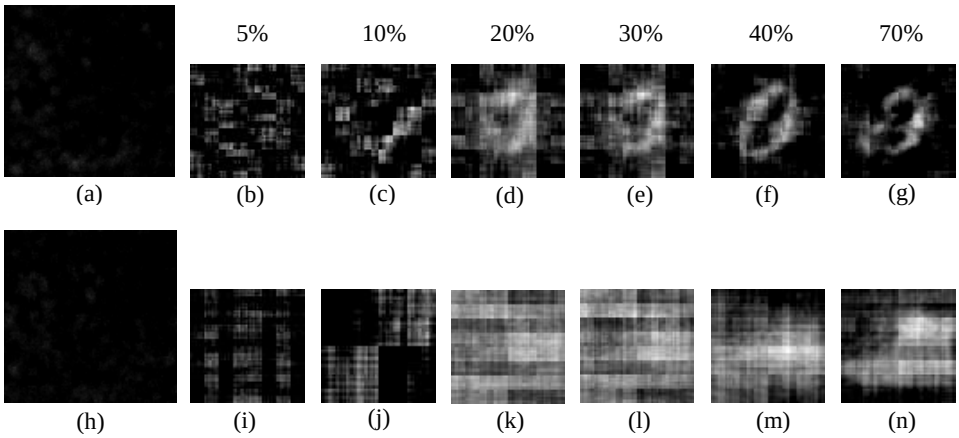


Fig. 12. (a) and (h) The generated amplitude-only patterns just before the virtual mask V2 when the MNIST database and the fashion MNIST database are respectively used for the training, (b) and (i) the retrieved plaintexts when the sampling rate of 5% is used, (c) and (j) the retrieved plaintexts when the sampling rate of 10% is used, (d) and (k) the retrieved plaintexts when the sampling rate of 20% is used, (e) and (l) the retrieved plaintexts when the sampling rate of 30% is used, (f) and (m) the retrieved plaintexts when the sampling rate of 40% is used, and (g) and (n) the retrieved plaintexts when the sampling rate of 70% is used.

Although the retrieved plaintexts in Figs. 12(b), 12(c) and 12(i)–12(m) do not visually render recognizable information, nonlinear correlation can be carried out between Hadamard spectrums of the retrieved plaintexts and Hadamard spectrums of original plaintexts P_R for optical authentication. Figures 13(a) and 13(e) show two typical amplitude-only patterns just before virtual mask V2 when the two databases are respectively used. Since only 5% of Hadamard spectrum coefficients are reserved, the retrieved plaintexts are noisy as shown in Figs. 13(b) and 13(f). It can be seen in Figs. 13(c) and 13(g) that there is only one sharp peak in each generated nonlinear correlation map, indicating that the receiver is an authorized person or possesses correct security keys. To verify the proposed method, wrong security keys, i.e., wrong weights and biases, have also been used to retrieve the plaintexts, and the corresponding nonlinear correlation distributions are shown in Figs. 13(d) and 13(h). It can be seen in Figs. 13(d) and 13(h) that the generated nonlinear correlations maps are noisy with multiple sharp peaks, indicating that the receiver is an unauthorized person or does not have correct security keys.

The PSNR values are calculated to quantitatively evaluate quality of the retrieved plaintexts with respect to different sampling rates, and PCE values are calculated to analyze the generated nonlinear correlation maps with respect to different sampling rates. Figures 14(a)–14(d) show the curves of PSNRs and PCEs respectively for the MNIST database and

the fashion MNIST database with respect to different sampling rates of 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. For the MNIST database, Fig. 14(a) shows that PSNR values increase with the larger sampling rates, and Fig. 14(b) shows the PCE of the generated nonlinear correlation between Hadamard spectrum of original plaintexts P_R and Hadamard spectrum of the retrieved plaintexts with respect to different sampling rates. It is illustrated that the higher sampling rate leads to the sharper peak in the generated nonlinear correlation map. Similarly, for the fashion MNIST database, PSNR values and PCE values with respect to different sampling rates are shown in Figs. 14(c) and 14(d), respectively.

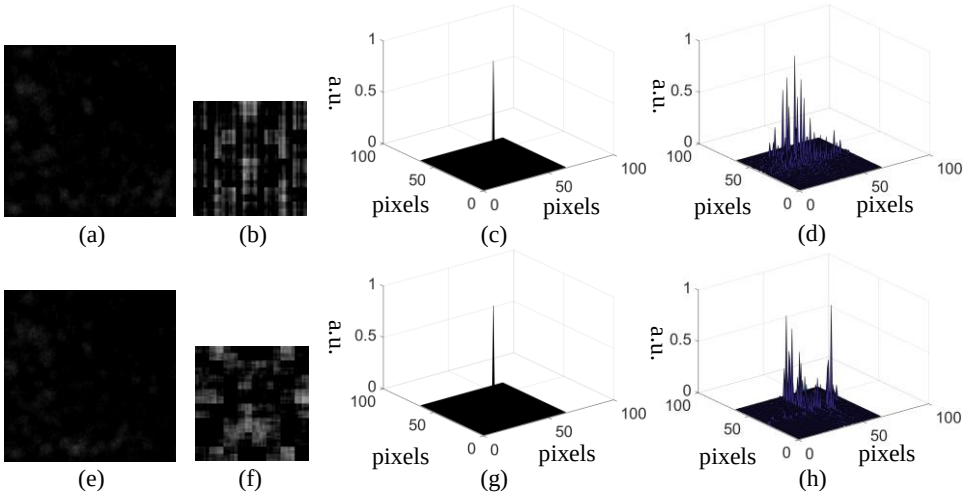


Fig. 13. Experimental results obtained by the proposed learning-based optical authentication using the optical setup with virtual phase-only masks and sampling rate of 5%. (a) and (e) The generated amplitude-only patterns just before virtual mask V2 when the MINST database and the fashion MNIST database are respectively used, (b) and (f) the retrieved plaintexts obtained by using the correspondingly trained learning model when correct security keys are used to retrieve the plaintext, (c) and (g) the generated nonlinear correlation maps respectively corresponding to (b) and (f), and (d) and (h) the generated nonlinear correlation maps obtained when wrong security keys, i.e., wrong weights and biases, are used to retrieve the plaintexts.

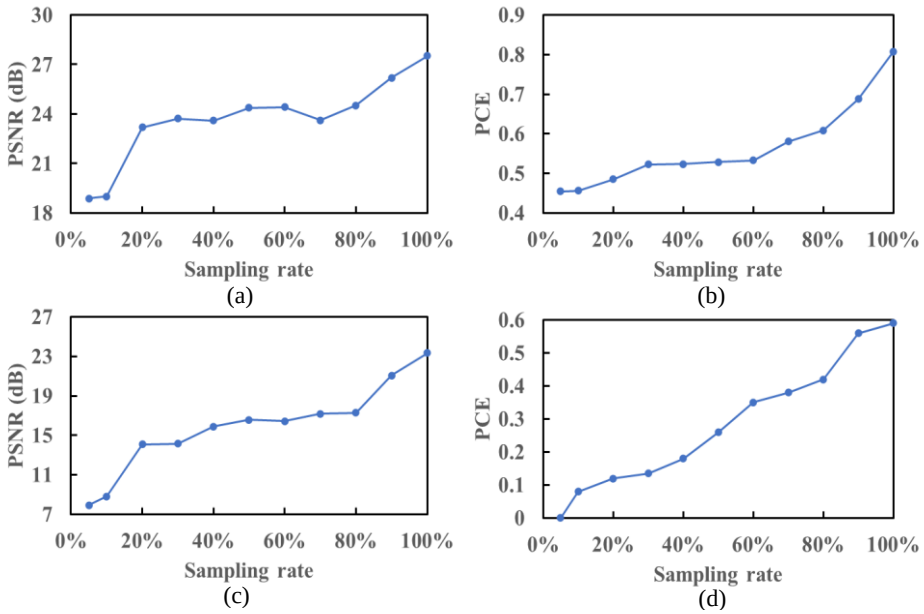


Fig. 14. The PSNRs of the retrieved plaintexts and PCEs of the generated nonlinear correlation distributions with respect to different sampling rates of 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. (a) and (c) PSNR curves respectively for the MNIST database and fashion MNIST database, and (b) and (d) PCE curves respectively for the MNIST database and fashion MNIST database.

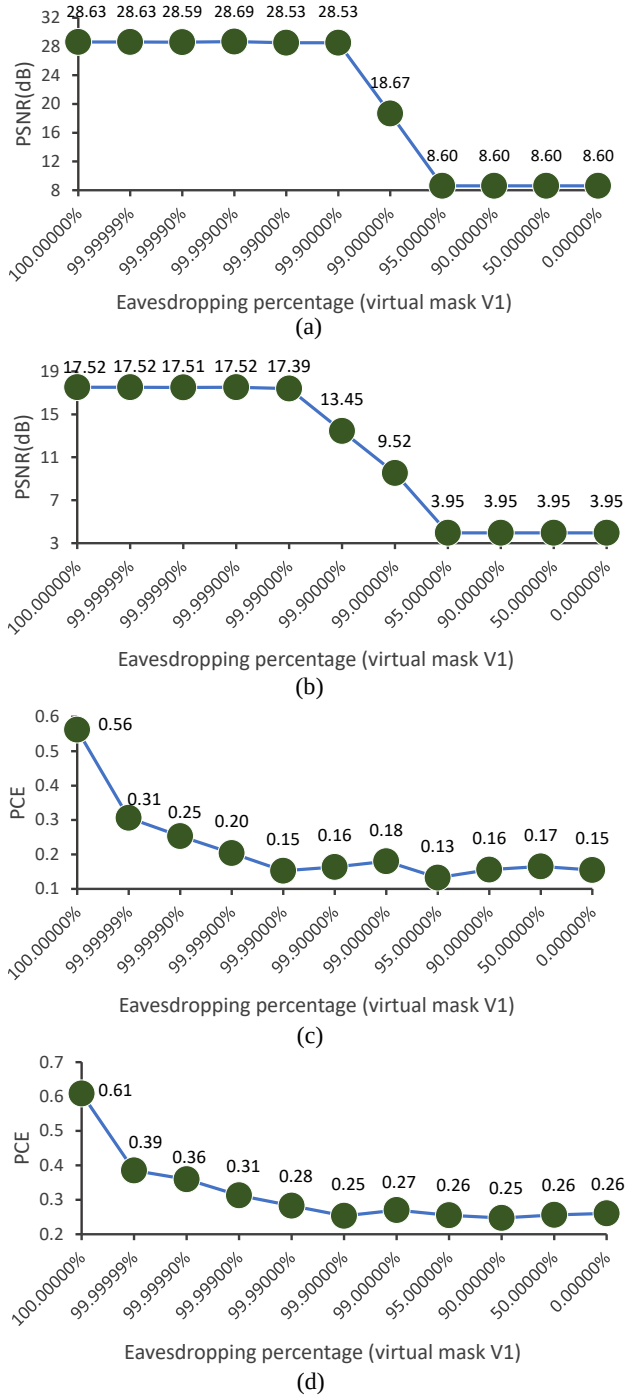


Fig. 15. PSNRs of the retrieved plaintexts and PCEs of the generated nonlinear correlation distributions respectively with respect to different eavesdropping percentages

of virtual phase-only mask V1. (a) and (b) PSNR curves respectively for the MNIST database and fashion MNIST database, and (c) and (d) PCE curves respectively for the MNIST database and fashion MNIST database. The sampling rate, i.e., only 5% of Hadamard spectrum coefficients, is fixed here and used to generate plaintexts P_R.

Eavesdropping analysis is also conducted to vet the security of the proposed learning-based optical authentication with virtual phase-only masks. The performance of virtual phase-only mask V1 is studied here, and other security keys are assumed to be correct. For the MNIST database, Figs. 15(a) and 15(c) show the performance of virtual phase-only mask V1 with different eavesdropping percentages using PSNR and PCE values to evaluate the quality. For the fashion MNIST database, Figs. 15(b) and 15(d) show the performance of virtual phase-only mask V1 with different eavesdropping percentages using PSNR and PCE values to evaluate the quality. Although different PSNR values are obtained with respect to different eavesdropping percentages of virtual phase-only mask V1 in Figs. 15(a) and 15(b), all the retrieved plaintexts cannot visually render recognizable information since only 5% of Hadamard spectrum coefficients are used to generate plaintexts P_R. As can be seen in Figs. 15(c) and 15(d), when the eavesdropping percentage of virtual phase-only mask V1 is lower than 99.99999%, PCE values of the generated nonlinear correlation distributions are low.

The experimental results and analyses aforementioned have validated the proposed learning-based optical authentication in complex scattering media. Instead of directly using parameters in the optical setup as security keys, the proposed method makes use of training parameters with the trained learning model as principal security keys. Moreover, virtual phase-only masks can be flexibly applied to enlarge key space. Therefore, conditions for the attacking algorithms [16–22] become invalid, and withstanding the attacks can be realized. When correct security keys are used, the retrieved plaintexts can be effectively authenticated instead of directly viewing the plaintext information and only one sharp peak is obtained in the generated nonlinear correlation distributions. The proposed learning-based optical authentication approach in complex scattering media can provide a promising strategy to enhance the security and enrich optical cryptosystems.

4. Conclusions

Learning-based optical authentication in complex scattering media has been proposed and experimentally verified. Ciphertexts are generated by using an optical setup in complex scattering media, and the trained learning model with its training parameters is used as security keys rather than parameters in the optical setup. Moreover, other parameters, e.g., virtual phase-only masks, can be flexibly designed and integrated to provide additional security keys. The retrieved plaintexts do not visually render recognizable information, and are further authenticated by using nonlinear correlation to establish an additional security layer. Eavesdropping analyses have also been conducted to vet the security of the proposed method. Advantages of the proposed method are briefly described as follows: 1) The training parameters (i.e., security keys) can be readily updated by new iterations. 2) Key space can be further enlarged by adding more neurons and more layers to the designed learning model. 3) The retrieved plaintexts obtained from the trained learning model cannot be visually recognized, but can be effectively authenticated by nonlinear correlation algorithm leading to the enhanced security. 4) Extra parameters, such as virtual phase-only masks, can be flexibly designed and used to further enlarge key space. It is worth noting that the proposed method needs some time for training the learning model. It is experimentally demonstrated that the proposed learning-based optical authentication in complex scattering media is feasible and effective. The proposed method provides a promising strategy for optical security, and this new scheme can be applied and integrated into various optical cryptosystems to enhance the security and withstand the attacks.

Acknowledgements

This work was supported by National Natural Science Foundation of China (NSFC) (61605165), Shenzhen Science and Technology Innovation Commission (JCYJ20160531184426473), and Hong Kong Research Grants Council (25201416, C5011-19G).

Disclosures

The authors declare that there are no conflicts of interest.

References

[1] R. C. Merkle, Secure communications over insecure channels, Commun. ACM 21(4), 294–299 (1978).
[2] B. Javidi, Securing information with optical technologies, Phys. Today 50(3), 27–32 (1997).
[3] W. Chen, B. Javidi, and X. Chen, Advances in optical security systems, Adv. Opt. Photon. 6, 120–155 (2014).
[4] O. Matoba, T. Nomura, E. Perez-Cabre, M. S. Millan, and B. Javidi, Optical techniques for information security, Proc. IEEE 97, 1128–1148 (2009).
[5] P. Refregier and B. Javidi, Optical image encryption based on input plane and Fourier plane random encoding, Opt. Lett. 20, 767–769 (1995).
[6] S. Liu and J. T. Sheridan, Optical encryption by combining image scrambling techniques in fractional Fourier domains, Opt. Commun. 287, 73–80 (2013).
[7] P. L. Yadav and H. Singh, Optical double image hiding in the fractional Hartley transform using structured phase filter and Arnold transform, 3D Research, 9(2), 20 (2018).
[8] M. R. Abuturab, Securing color information using Arnold transform in gyrator transform domain, Opt. Lasers Eng. 50(5), 772–779 (2012).
[9] 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).
[10] X. Chai, J. Bi, Z. Gan, X. Liu, Y. Zhang and Y. Chen, Color image compression and encryption scheme based on compressive sensing and double random encryption strategy, Sig. Process. 176, 107684 (2020).
[11] X. Chai, H. Wu, Z. Gan, Y. Zhang, Y. Chen and K. W. Nixon, An efficient visually meaningful image compression and encryption scheme based on compressive sensing and dynamic LSB embedding, Opt. Lasers Eng. 124, 105837 (2020).
[12] S. Xi, X. Wang, L. Song, Z. Zhu, B. Zhu, S. Huang, N. Yu, and H. Wang, Experimental study on optical image encryption with asymmetric double random phase and computer-generated hologram, Opt. Express 25(7), 8212–8222 (2017).
[13] 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).
[14] W. Chen, X. Chen, and C. J. R. Sheppard, Optical image encryption based on diffractive imaging, Opt. Lett. 35(22), 3817–3819 (2010).
[15] E. Pérez-Cabré, M. Cho, and B. Javidi, Information authentication using photon-counting double-random-phase encrypted images, Opt. Lett. 36(1), 22–24 (2011).
[16] 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).
[17] W. Qin, X. Peng, X. Meng, and B. Z. Gao, Vulnerability to chosen-plaintext attack of optoelectronic information encryption with phase-shifting interferometry, Opt. Eng. 50(6), 065601 (2011).
[18] 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).
[19] 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).
[20] 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).
[21] X. Wang and D. Zhao, A special attack on the asymmetric cryptosystem based on phase-truncated Fourier transforms, Opt. Commun. 285(6), 1078–1081 (2012).
[22] L. Zhou, Y. Xiao, and W. Chen, Learning-based attacks for detecting the vulnerability of computer-generated hologram based optical encryption, Opt. Express 28(2), 2499–2510 (2020).
[23] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature, 521(7553), 436–444 (2015).
[24] C. Dong, C. C. Loy, and X. Tang, Accelerating the super-resolution convolutional neural network, European conference on computer vision. Springer, Cham, 391–407 (2016).
[25] http://yann.lecun.com/exdb/mnist/

[26] H. Xiao, K. Rasul, and R. Vollgraf, Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, arXiv preprint arXiv:1708.07747 (2017).
[27] W. Chen, X. Chen, A. Stern, and B. Javidi, Phase-modulated optical system with sparse representation for information encoding and authentication, IEEE Photon. J. 5, 6900113 (2013).
[28] Y. Xiao, L. Zhou, and W. Chen, Single-Pixel Imaging Authentication Using Sparse Hadamard Spectrum Coefficients, IEEE Photon. Technol. Lett. 31(24), 1975–1978 (2019).
[29] W. K. Pratt, J. Kane, and H. C. Andrews, Hadamard transform image coding, Proc. IEEE 57(1), 58–68 (1969).
[30] 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(3), 1139–1147 (2013).
[31] L. Zhou, Y. Xiao, and W. Chen, Learning complex scattering media for optical encryption, Opt. Lett. 45, 5279–5282 (2020).