Optical Compressive Encryption via Deep Learning

Yi Qin, Yuhong Wan, Shujia Wan, Chao Liu, and Wei Liu

Abstract—**The compression of the ciphertext of a cryptosystem is desirable considering the dramatic increase in secure data transfer via Internet. In this paper, we propose a simple and universal scheme to compress and decompress the ciphertext of an optical cryptosystem by the aid of deep learning (DL). For compression, the ciphertext is first resized to a relatively small dimension by bilinear interpolation and thereafter condensed by the JPEG2000 standard. For decompression, a well-trained deep neural network (DNN) can be employed to perfectly recover the original ciphertext, in spite of the severe information loss suffered by the compressed file. In contrast with JPEG2000 and JPEG, our proposal can achieve a far smaller size of the compressed file (SCF) while offering comparable decompression quality. In addition, the SCF can be further reduced by compromising the quality of the recovered plaintext. It is also shown that the compression procedure can provide an additional security level, and this may offer new insight into the compressive encryption in optical cryptosystems. Both simulation and experimental results are presented to demonstrate the proposal.**

Index Terms—**Optical security, ciphertext compression, deep learning.**

I. Introduction

OPTICAL security methods have received increasing interest since the pioneering work of double random phase encoding was reported by Javidi and Refregier in 1995 [1]. In contrast to conventional methods, optical cryptosystems possess the advantages of high-speed parallel-processing and multiple freedoms [2]. Therefore, a variety of optical techniques, including holography [3], joint transform correlator [4], interference [5], diffraction [6], and single-pixel imaging [7], are exploited to create optical cryptosystems. Among the existing optical encryption schemes, the ones that encrypt the primary image into phase-only masks (POMs) are quite attractive, as the decoded image can be directly captured by the intensity sensitive devices in such schemes [5], [8], [9]. In particular, a cascaded double phase encryption (CDPE) scheme is worthy of noting due to its high security [9], and some variants of it were also reported and studied recently [10], [11].

In addition to developing new optical cryptosystems, people also make efforts to compress the ciphertext to facilitate its transmission and storage [12]. This is especially true considering the dramatic increase in data transfer via Internet. Multiplexing and classical data compressing technique are two major ways for ciphertext compression in optical cryptosystems. Multiplexing is a powerful character that distinguishes optical encryption, as it allows for merging multiple massages into one single package. Till now, various multiplexing strategies have been developed. For instance, Situ and Zhang proposed wavelength multiplexing [13] and distance multiplexing [14] in the double random phase encoding scheme, Rueda *et. al.* showed the feasibility of lateral shift multiplexing [15] and key rotation multiplexing [16] in the joint transform correlator encrypting architecture, Barrera *et. al.* demonstrated the potential of multiplexing in securing movies [17]. In these multiplexing strategies, multiple ciphertexts associating with different plaintexts are directly superposed to yield the synthetic ciphertext. Although confused, the synthetic ciphertext still permits the approximately recovery of each plaintext with its unique decryption keys. Besides, by combining spectral multiplexing with some other means (e. g. quantization), Alfalou and co-authors developed a series of approaches for simultaneous compression and encryption of multiple images [18]–[20].

In contrast to multiplexing, direct processing of the ciphertext with classical compression techniques seems to be more straightforward. However, literatures on such topic are relatively rare, since the ciphertext always has a noisy appearance and thus makes the lossless compression method no longer effective [12]. Of course, lossy compression technique, such as quantization, can effectively reduce the volume of the ciphertext; however, this will result in obvious degradation in the reconstructed image [21]. In the traditional sense, the lossy compression always means the permanent loss of information. Fortunately, the emergence of deep learning (DL) brings the chance for coping with such issues. For example, Dong *et. al.* showed that the compression artifact of a natural image could be well repaired by a deep convolutional network [22], Jiao *et. al.* compressed a phase-only hologram with JPEG standard and recovered it via DL [23]. Also by DL, Shimobaba *et. al.* presented a dynamic-range compression scheme for phase-only holograms and attained a satisfied compression ratio [24]. Zhang *et. al.* reported a new scheme for image compression and encryption, and they demonstrated that the degraded image caused by ghost-imaging transmission can be substantially improved by DL [25]. In particular, because the optical ciphertext is of more

Manuscript received June 3, 2021; revised June 26, 2021; accepted July 6, 2021. Date of publication July 8, 2021; date of current version July 29, 2021. This work was supported in part by National Natural Science Foundation of China under Grants 61575009 and 61505091, in part by the Natural Science Foundation of Beijing Municipality under Grant 4182016, and in part by Foundation of Henan Educational Committee under Grants 2017GGJS130, 2019GGJS182. *(Corresponding author: Yuhong Wan.)*

Yi Qin is with the Faculty of Science, Beijing University of Technology, Beijing 100124, China, and also with the College of Mechanical and Electrical Engineering, Nanyang Normal University, Nanyang 473061, China (e-mail: 641858757@qq.com).

Yuhong Wan and Chao Liu are with the Faculty of Science, Beijing University of Technology, Beijing 100124, China (e-mail: yhongw@bjut.edu.cn; 1038478317@qq.com).

Shujia Wan and Wei Liu are with the College of Mechanical and Electrical Engineering, Nanyang Normal University, Nanyang 473061, China (e-mail: 412414140@qq.com; 331734554@qq.com).

Digital Object Identifier 10.1109/JPHOT.2021.3095712

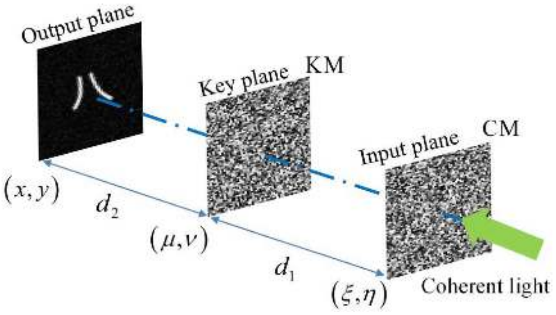

Fig. 1. Schematic of the decryption setup for CDPE.

randomness than natural image or hologram, the compression of it becomes extremely difficult. To our best knowledge, the only literature on such topic is contributed by Li *et. al.* [26], who compressed the ciphertext image into the one-dimensional signal and reconstructed it with DL. However, their compression approach is based on the complex optical setup that includes expensive optical elements (e. g. DMD), and therefore may not be applicable in some scenarios. In this paper, we present herein a novel compression and decompression scheme for optical ciphertext that employs a deep neural network (DNN). In our scheme, the bilinear interpolation and JPEG2000 work collaboratively to compress the ciphertext, while a novel DNN is employed to perform lossless decompression. The proposed scheme is simple and universal, and we take the CDPE system as an example to demonstrate it. It is shown that the performance of the proposal surpasses the excellent compression standards such as JPEG2000 and JPEG by a large margin. Both simulations and experimental results are presented to support the proposal.

II. Principle

A. The Cascaded Double Phase Encryption (CDPE)

The schematic of the decryption setup for CDPE is depicted in Fig. 1. It consists mainly of two cascaded POMs: the key mask (KM) and the ciphertext mask (CM). The values of the KM are fixed and are distributed randomly within $[0, 2\pi]$. The CM is generated by using an iterative algorithm detailed below and it varies with the plaintext. When the CM (i. e. input plane) is illuminated by the parallel coherent light, the plaintext hidden in it will be reconstructed at the output plane. The recovered plaintext can be directly recorded by an intensity-sensitive device such as CCD camera. For the convenience of the subsequent discussion, (ξ,η), (μ,ν), and (x,y) are used to denote the coordinates of the input plane, key plane, and output plane, respectively.

In order to encrypt the plaintext into the two POMs, an iterative algorithm simulating the reciprocal propagation of the wavefront between the input plane and the output plane should be utilized. To begin with, a random-valued matrix $CM^{(1)}(\xi,\eta)$ is generated to initialize the CM. The $n\text{th}, n = 1, 2, 3\ldots$ iteration starts from propagating $CM^{(n)}(\xi,\eta)$ forward to the output plane, where a complex amplitude of $O^{(n)}(x,y)$ is obtained. Afterwards, a type of amplitude constraint is performed on $O^{(n)}(x,y)$ by substituting its amplitude with the target image (i. e. plaintext). The updated $O^{(n)}(x,y)$ is then propagated back to the input plane to yield the complex amplitude of $I^{(n)}(\xi,\eta)$. Then we impose another type of amplitude constraint on $I^{(n)}(\xi,\eta)$ (i. e. replacing its amplitude by a special matrix whose elements all equal 1). The renewed $I^{(n)}(\xi,\eta)$ has the pure phase form and it serves as $CM^{(n+1)}(\xi,\eta)$ in the next iteration. The block diagram for illustrating the iterative algorithm is shown in Fig. 2, where $\mathrm{FrT}[\cdot]$ and $\mathrm{FrT}^{-1}[\cdot]$denote respectively the forward and back Fresnel diffraction in the free space, the symbols $\otimes$ and $*$ represent respectively the multiplication and conjugation operation. The iteration will be terminated when the correlation coefficient between the amplitude of $O^{(n)}(x,y)$ and the plaintext exceeds some predefined value. Assuming K iterations have proceeded before the termination, $CM^{(K)}(\xi,\eta)$ is saved as the final value of the CM.

B. The Proposal

The proposed scheme is outlined in Fig. 3. The compression of the CM involves two steps [Fig. 3(a)]. The first step reduces the size of the CM from $N \times N$ to $M \times M$ $(M < N)$ by bilinear interpolation, and the second step further compress it with the JPEG2000 standard. It is worth noting that JPEG2000 can substantially condense an image in volume without changing its dimensions. Fig. 3(b) shows the decompression principle: a well-trained DNN can be utilized to directly predict the original CM from the compressed one.

Fig. 4 shows the details of the network that we devised. It inherits the basic framework of the U-net [27]. The size of the network input/output is 64×64. The arrow with a certain color denotes a unique type of operation, as indicated by the legend. Each blue block represents a multi-channel feature map, and the number of channels and the x-y size are annotated at the top and lower edges of it, respectively. We use the residual block (See details in [28]) rather than the common convolution for extracting the feature throughout the network. The max pooling with 2×2 filters (2×2 Max pooling) and transposed convolution with 3×3 filters (3×3 Up-conv) work respectively as the layer downsampling and upsampling layer. The skip and concatenation bridge the contracting path and the expanding path for better convergence of the network. Every involved convolution operation has an identical kernel size of 3×3 pixels and is followed by the batch normalization (BN) and a rectified linear unit (Relu).

C. Data Preparation

The data generation is performed on the platform of Matlab R2011a. To yield the CMs, the Chinese characters with different fonts are used as the target images (i. e. plaintexts). The number of pixels of the target image and that of the CM are both set to 64×64 ($=N \times N$). The parameters for calculation are that the illumination wavelength is 532.8nm, the pixel pitch is 6.4um, and the axial distances d_1 and d_2 both take the value of 250mm. It should be stressed that all the subsequent numerical simulations also follow such settings of parameter. The generated CMs are further compressed by the proposed “Bilinear interpolation+JPEG2000” approach. The compressed CMs are with the size of 20×20 ($=M \times M$), and they will be resized to 64×64 to match the network input before training/predicting.

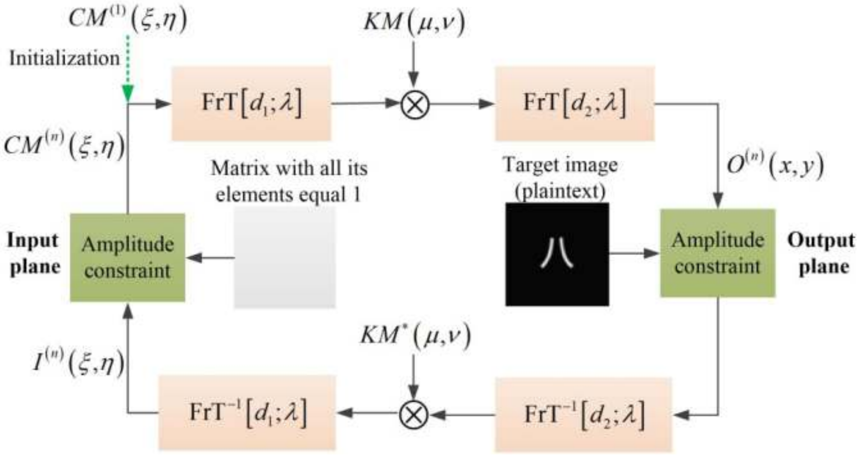

Fig. 2. The block diagram of the iterative algorithm used to yield the CM.

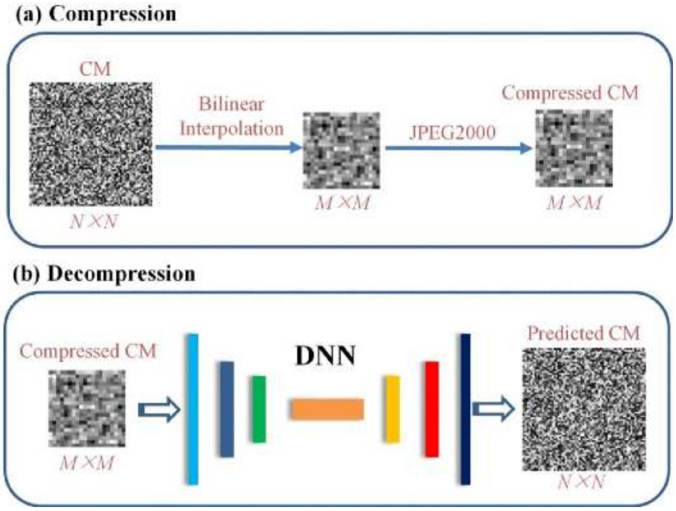

Fig. 3. Outline of the proposed scheme.

Several examples of the target images, as well as the corresponding CMs and compressed CMs, are depicted in Fig. 5. To train the DNN, we generate totally 100000 pairs of CMs and compressed CMs as the dataset, of which 95% are used as the training set and 5% as the testing set.

D. Training

Our DNN is constituted based on the Keras framework that uses Python 3.6. The training is carried out on a PC with the configuration of Intel core i7-8700 CPU and 32 GB RAM. The NVIDIA GeForce GTX 1070Ti GPU is enabled to accelerate the calculation. The mean absolute error (MAE) between the predicted CM and the primary CM is chosen as the loss function. In addition, we use optimizer Adam with a learning rate of 0.0002 and a batch size of 32. The training lasts for about 40 hours.

III. Results and Discussion

A. Experimental Arrangements

The experimental arrangements are prepared according to the schematic illustrated in Fig. 1. In principle, two phase modulation elements, such as spatial light modulator (SLM), should collaborate to complete the decryption. However, considering the difficulty in aligning them, only the propagation from the key plane to the output plane is optically realized, while that from the input plane to the key plane is calculated digitally. Moreover, as the currently available SLM is incapable of modulating the amplitude and phase simultaneously, we translate the complex amplitude behind the KM into a phase-only wavefront (POW) by discarding the amplitude information. This approximation is reasonable due to the importance of phase information [29], but it will inevitably cause degradation of the reconstruction. The proposed optical setup is shown in Fig. 6. The coherent light from the laser ($\lambda = 532.8$nm) is focused by the objective lens (OL), filtered by the spatial filter (SF), and then collimated by L1 ($f = 150$mm). The parallel wave passes through the polarizer (P) and the beam splitter and then illuminates the SLM. The reconstructed plaintext is imaged on the CCD through L2 ($f = 150$mm). The SLM (JD955B, Jasper Display Corp) used for displaying the POW is with 1920×1080 pixel count and 6.4um×6.4um pixel pitch, and the CCD camera (GigE Vision TL, Daheng imaging) is with 1920×1080 pixel count and 6.4um×6.4um pixel pitch. Besides, since the POW is also with a size of 64×64, it will be extended to 1920×1080 via zero-padding to fit the SLM.

B. Feasibility of the Proposal

We will introduce several indexes to ensure a comprehensive evaluation for each compression method. The first index we prefer is the correlation coefficient (CC), as it can evaluate the quality of a recovered image by comparing it with the primary one from the statistical perspective. The CC is defined as

$$\mathrm{CC} = \frac{E\left\{[|f - E(f)|]\,[|f_r - E(f_r)|]\right\}}{\sqrt{E\left\{[f - E(f)]^2\right\} E\left\{[f_r - E(f_r)]^2\right\}}}, \quad (1)$$

where $E[\,]$ stands for the mathematical expectation, f denotes the primary image, f_r denotes the recovered image with degraded quality. The value of CC ranges from 0 to 1, and the larger the value the better the quality of the recovered image. The second index we introduce is the peak signal to noise ratio

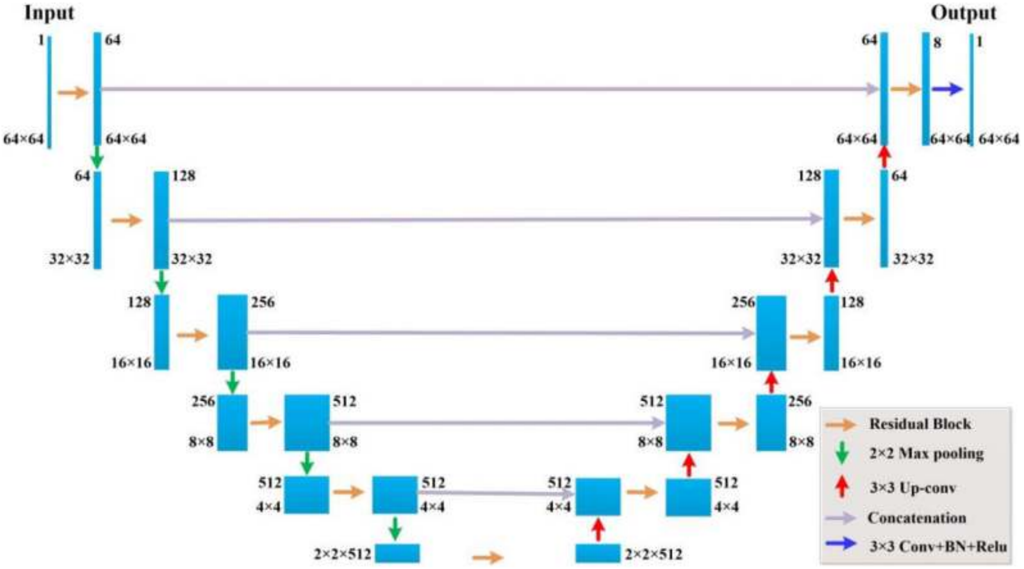

Fig. 4. The structure of the proposed DNN.

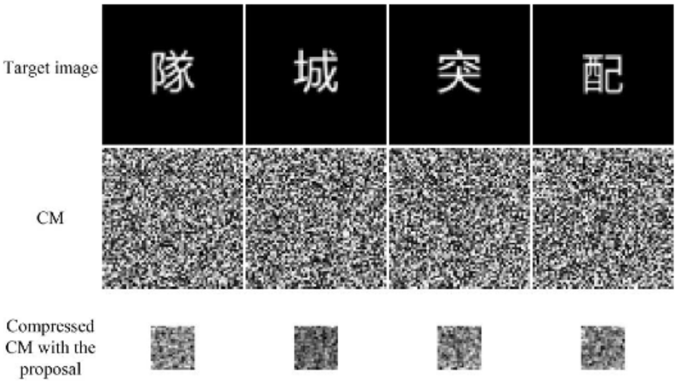

Fig. 5. Several examples of the target images (top row), the corresponding CMs (middle row), and the compressed CMs with our method (bottom row).

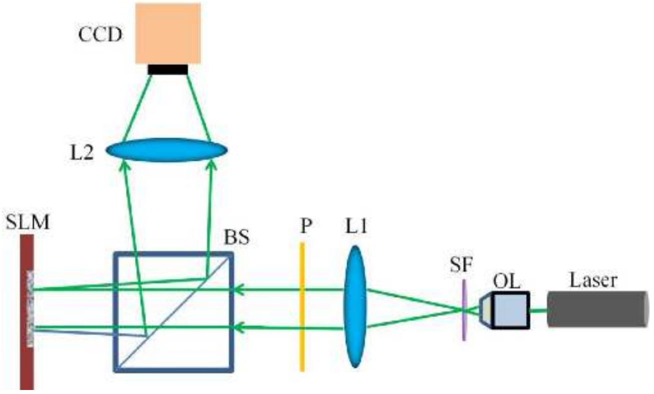

Fig. 6. The experimental setup for approximately reconstruction of the plaintext. OL, objective lens; SF, spatial filter; L1, L2: lens; P, polarizer; BS, beam splitter; SLM, spatial light modulator; CCD, charge-coupled device.

(PSNR). Assuming both f and f_r have the size of $J \times K$ and are of 8-bit depth, the PSNR for f_r can be mathematically expressed by

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{255^2}{\frac{1}{JK}\sum\sum \|f - f_r\|^2}\right), \tag{2}$$

where $\|\|$ denotes the L2 norm. Another index is the size of the compressed file (SCF), as it points out the absolute memory space the file occupies in the computer.

To verify the feasibility of our method, one CM is randomly selected from the validation set for test. Fig. 7(a) shows the original CM (Ground truth, GT), and its size in BMP format (5174 bytes) is regarded as the original file size. The decompressed CM from the BMP file is completely identical to the GT because BMP is a lossless coding format. The recovered plaintext from the original CM in simulation [Fig. 7(e)] or experiment [Fig. 7(i)] is also taken as the GT for the calculation of the CC/PSNR values. Fig. 7(b) shows the decompressed CM by use of our proposal. As can be seen, it highly resembles the GT (CC = 0.9859/ PSNR = 26.20dB); meanwhile, the SCF for it reaches to a surprisingly small value of 273 bytes. In the following discussions, we will compare the proposal with two outstanding modern compression standards: JPEG and JPEG2000. First of all, we give an evaluation of the three methods in the case that the decompressed CM with each of them is comparable in quality. To do this, we enforce the CC values of the decompressed CMs using JPEG2000 and JPEG to approach sufficiently to 0.9859 through adjusting the parameters of the IMWRITE function in Matlab (i. e. “CompressionRatio” for JPEG2000 and “quality” for JPEG), and we show the corresponding decompressed CMs in Figs. 7(c) and 7(d), respectively. The SCF for each method is annotated above the decompressed CM. It is seen that, yielding the CMs comparable in quality, the SCFs for JPEG2000 and JPEG are 1867 bytes and 2382 bytes, which are ~7 and ~9 times the volume of that for our proposal. Fig.s 7(e)-7(h)

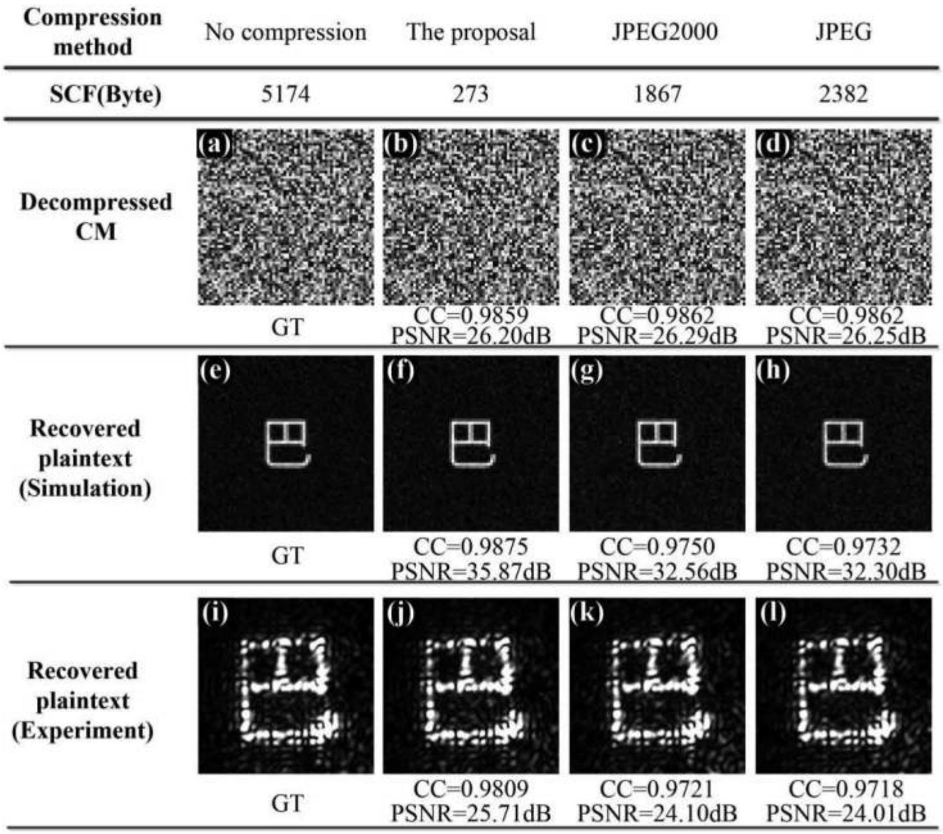

Fig. 7. The comparison of our method with JPEG and JPEG2000 when the decompressed CMs are comparable in quality.

show the numerically reconstructed plaintexts with these decompressed CMs. Interestingly, the reconstructed image with the proposal (CC = 0.9875/PSNR = 35.87dB) exceeds those with JPEG2000 (CC = 0.9750/PSNR = 32.56dB) and JPEG2 (CC = 0.9732/PSNR = 32.30dB). Similar conclusions can also be drawn from the experimental results shown in Figs. 7(i)-7(l).

Secondly, it is also significant to compare the three approaches in the case that the compressed CMs are identical in file size. Unfortunately, neither JPEG nor JPEG2000 allows continuous adjustment of the SCF. Hence, the SCF for either is set to the value which is closest to that for our method (i. e. 273 bytes). The SCF for JPEG is finally reduced to 856 bytes, because this is the minimum that JPEG can offer with the current CM (by setting the "quality" parameter to 0). With this SCF, the decompressed CM [Fig. 8(d)] differs from the GT obviously, and the recovered plaintexts obtained by both simulation [Fig. 8(h)] and experiment [Fig. 8(l)]] are extremely blurry. The eligible SCF for JPEG2000 is identified to be 315 bytes, and the corresponding decompressed CM [Fig. 8(c)] becomes totally different from the GT. The corresponding recovered plaintext with either simulation [Fig. 8(g)] or experiment [Fig. 8(k)] has a noise appearance and is thoroughly unrecognizable. It can be judged that the proposal has surpassed the compression limit that JPEG and JPEG2000 can achieve.

Fig. 9 shows in more detail the relationship between the SCF and the quality indexes of the decrypted image. It should be pointed out that, for a certain plaintext, the SCF obtained by using our method is fixed (∼270 bytes). This is because all the parameters for compression are constant during the generation of the dataset, which is employed to train the DNN. As a result, only when a compressed file is yielded with such parameters, can it be recognized by the DNN and successfully recovered. By comparison, both JPEG and JPEG2000 permit the adjustment of the SCF, and they produces gradually degraded decrypted results as the SCF declines. In this regard, JPEG and JPEG2000 can furnish the compression with more flexibility. It should be pointed out that the proposal is also potentially able to provide such flexibility, provided that variable compression parameters are adopted to produce the training data. However, this will give rise to dramatic expanding of the dataset; what is more, the network architecture, together with the relative parameters, may need to be modified to ensure the convergence.

TABLE I
THE ROLES THAT STEP 1 AND STEP 2 PLAY DURING THE COMPRESSION

Index	Original file	After step 1	After step 2	After step2 (without step1)
SCF(Byte)	5174	1440	273	2076

As mentioned in Section II-B, the proposed compression comprises of two steps. Table I shows the roles that they play during the compression. As can be seen, the first step compresses the ciphertext from its original size (i. e. 5174 bytes) to 1440

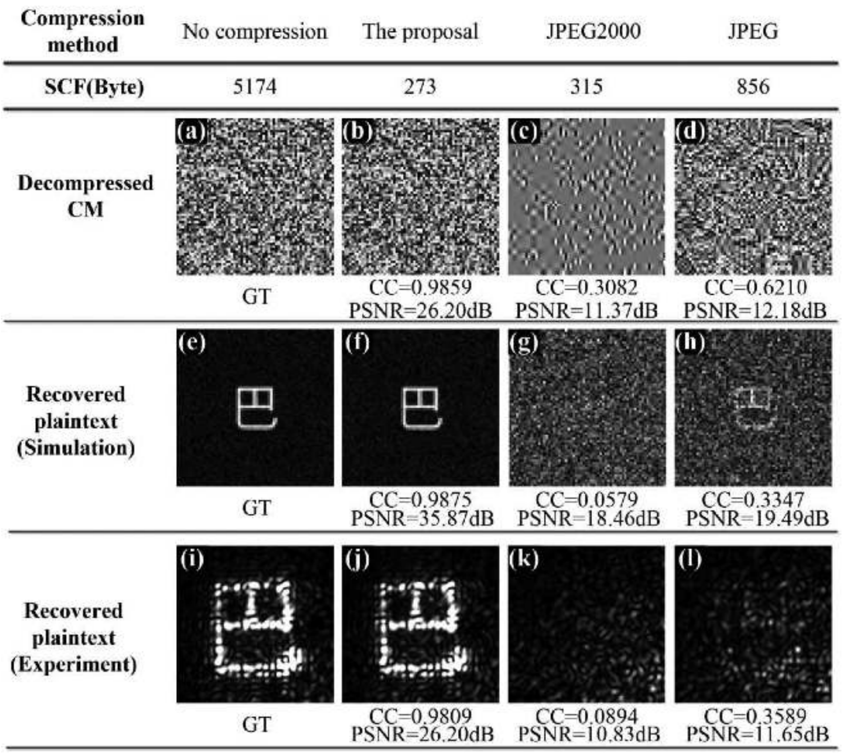

Fig. 8. The comparison of our method with JPEG and JPEG2000 when the compressed CMs are comparable in SCF.

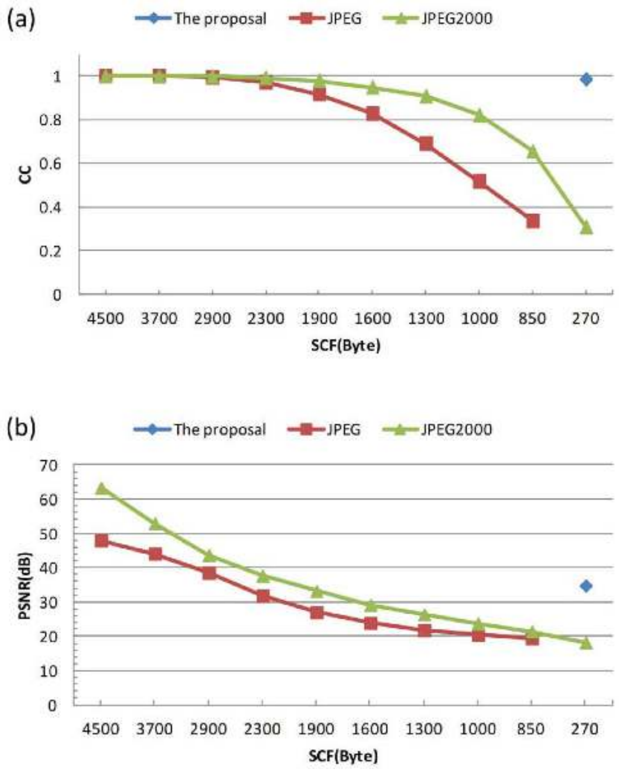

Fig. 9. The relationship between the SCF and the quality indexes of the decrypted image. (a) CC; (b) PSNR.

TABLE II
THE COMPARISON OF THE COMPRESSION AND DECOMPRESSION TIME OF THE THREE METHODS

Compression method	The proposal	JPEG2000	JPEG
Compression time(ms)	3.7	3.2	2.9
Decompression time(ms)	26.5	1.9	1.5

bytes, and the second step further reduces this value to 273 bytes. By comparison, if the first step is removed from the compression process (i. e. $M{=}N$), the SCF of the ciphertext will be finally 2076 bytes. Hence, both steps are indispensable for adequate compression.

We also compared the compression and decompression time of the three methods. It should be kept in mind that the DNN predicting is conducted on Python 3.6, while all the other related programs are carried out on MATLABR2011a. Approximately, the compression/decompression time of JPEG and JPEG2000 is acquired by calculating the running time of IMWRITE/IMREAD function. Because the proposed compression consists of bilinear interpolation and JPEG2000, the compression time of the proposal should be a sum of that to execute each of them. Likewise, the decompression time of the proposal comprises of those spent on JPEG2000 decompression and DNN predicting. The test results are summarized in Table II. It is seen that the compression time of our proposal is comparable

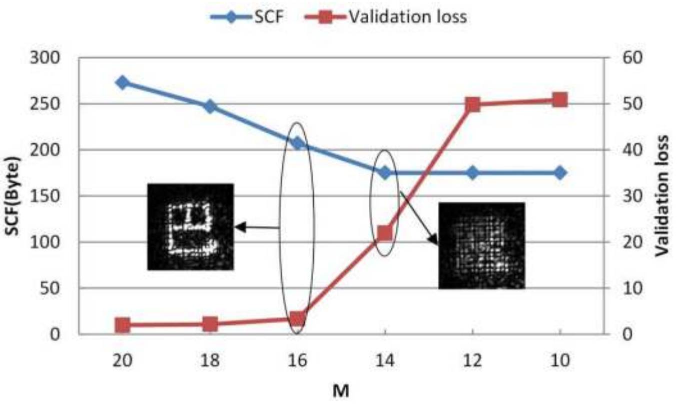

Fig. 10. The dependence of SCF and validation loss on the value of M.

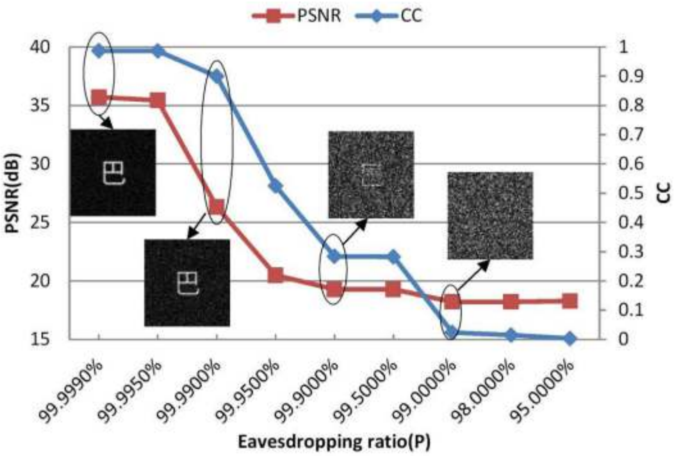

Fig. 11. The dependence of the decrypted image on the eavesdropping ratio.

with its counterparts; however, the decompression time of our method, of which the predicting time accounts for ∼90%, is much longer.

The current SCF of our approach has been shown to be far superior to that of JPEG or JPEG2000, while it is expected to be further reduced in some application scenarios. A reasonable way to achieve this goal is to decrease the value of M, as the file size of an image in any format inherently relates to its dimensions. However, an excessively small M will lead to severe information loss of the compressed CM and thus confuse its "feature". Consequently, the training of the DNN may become rather difficult or even failed. Thus, it is necessary to figure out how the value of M affects the SCF, the DNN model, as well as the decrypted plaintext. Reducing the value of M from 20 to 10 with an interval of 2 and using the same CM set as that in Section II-C, we produce six groups of compressed CMs that represent different degrees of compression. Six models are obtained by training the DNN with these data sets individually, and the training parameter settings follow those specified in Section II-D. For comparison, these models are used to predict the same CM shown in Fig. 7(a). The predicted CMs are further decrypted numerically and experimentally to recover the plaintexts. Partial results are illustrated in Fig. 10. As can be seen, the SCF dwindles with the decrease of M at first and achieves a minimum of 175KB at M = 14. In other words, the SCF no longer changes even if M takes more small values (i. e. 12, 10). This phenomenon is likely to be caused by the encoding rule of JPEG2000. Meanwhile, the validation loss increases as the value of M descends, indicating the gradually deterioration of the DNN model. It can be concluded that, to some extent, the SCF can be reduced by compromising the performance of the DNN model (i. e. the quality of the decompressed CM as well as the recovered plaintext), but there is a lower limit for it due to the encoding role of JPEG2000. Fig. 10 also shows that the DNN model with M = 16 ensures a good recovered image while that with M = 14 causes a totally unrecognizable one. Therefore, the minimum of the acceptable values of M is 16 in the current research, corresponding to a SCF of 207 bytes.

C. Security Analysis

The security of the CDPE scheme itself has been sufficiently investigated in [9]–[11], so we will focus on the security enhancement contributed by the compression procedure. According to the Kerckhoffs' assumption [30], the attacker has the total knowledge of the cryptosystem except the secret keys. To break the proposed scheme, the attacker should firstly transform the eavesdropped ciphertext $(M \times M)$ into its original size $(N \times N)$. However, such issue is essentially ill-posed, since the bilinear interpolation results in irreversible information loss. Consequently, the currently available cryptoanalysis, such as ciphertext-only attack and known-plaintext attack [30], will fail to crack the proposed scheme. In other words, the attacker has to utilize the brute-force attack to acquire the DNN. Let P represent the set of the parameters (weights and biases) that composes the DNN. We investigate the dependence of the decrypted image on the eavesdropping ratio (ER) of P and show it in Fig. 11. As is observed, the decrypted image renders no meaningful information of the plaintext even though the ER reaches to 99.9%; that is, a small amount of incorrect parameters of the DNN can lead to entire failure of the decryption. Furthermore, a perfect reconstruction of the original image requires the knowledge of more than 99.995% of the whole parameters. It can be concluded that the proposed scheme guarantees itself high security, as the DNN has a total of ∼50000000 parameters that will exhaust the unauthorized intruder.

IV. Conclusion

To summarize, we have reported a DL-based scheme for compression and decompression of the ciphertext of an optical cryptosystem. The compression is accomplished by the cooperation of bilinear interpolation and JPEG2000 standard, and the decompression is implemented by a DNN. Yielding the lossless decompression of the same ciphertext, our method attains a far smaller SCF than JPEG and JPEG2000. In particular, the proposal surpasses the compression limit that JPEG or JPEG2000 can achieve. Moreover, the SCF can be further reduced by compromising the quality of the recovered plaintext. Compared with JPEG and JPEG2000, our method takes comparable compression time but relatively longer decompression time. It is also demonstrated that the compression can immensely reinforce the security of the CDPE scheme, and this may offer new insight into the compressive encryption in optical cryptosystems.

References

[1] P. Refregier and B. Javidi, "Optical image encryption based on input plane and fourier plane random encoding," *Opt. Lett.*, vol. 20, pp. 767–769, 1995.

[2] W. Chen, B. Javidi, and X. Chen, "Advances in optical security systems," *Adv. Opt. Photon.*, vol. 6, pp. 120–155, 2014.

[3] E. Tajahuerce and B. Javidi, "Encrypting three-dimensional information with digital holography," *Appl. Opt.*, vol. 39, pp. 6595–6601, 2000.

[4] T. Nomura and B. Javidi, "Optical encryption using a joint transform correlator architecture," *Opt. Eng.*, vol. 39, pp. 2031–2035, 2000.

[5] Y. Zhang and B. Wang, "Optical image encryption based on interference," *Opt. Lett.*, 2008, vol. 33, pp. 2443–2445, 2008.

[6] W. Chen, X. Chen, and C. J. R. Sheppard, "Optical image encryption based on diffractive imaging," *Opt. Lett.*, vol. 35, pp. 3817–3819, 2010.

[7] S. Jiao, J. Feng, Y. Gao, T. Lei, and X. Yuan, "Visual cryptography in single-pixel imaging," *Opt. Exp.*, vol. 28, pp. 7301–7313, 2020.

[8] H. T. Chang, W. C. Lu, and C. J. Kuo, "Multiple-phase retrieval for optical security systems by use of random-phase encoding," *Appl. Opt.*, vol. 41, pp. 4825–4834, 2002.

[9] Y. Li, K. Kreske, and J. Rosen, "Security and encryption optical systems based on a correlator with significant output images," *Appl. Opt.*, vol. 39, pp. 5295–5301, 2000.

[10] L. Sui, X. Zhang, and A. Tian, "Multiple-image hiding based on cascaded free-space wave propagation using the structured phase mask for lensless optical security system," *IEEE Photon. J.*, vol. 9, pp. 1–14, Oct. 2017, Art. no. 7803414.

[11] L. Zhou, Y. Xiao, and W. Chen, "Learning-based attacks for detecting the vulnerability of computer-generated hologram based optical encryption," *Opt. Exp.*, vol. 28, pp. 2499–2510, 2020.

[12] A. Alfalou and C. Brosseau, "Optical image compression and encryption methods," *Adv. Opt. Photon.*, vol. 1, pp. 589–636, 2009.

[13] G. Situ and J. Zhang, "Multiple-image encryption by wavelength multiplexing," *Opt. Lett.*, vol. 30, pp. 1306–1308, 2005.

[14] G. Situ and J. Zhang, "Position multiplexing for multiple image encryption," *J. Opt. A*, vol. 8, pp. 391–397, 2006.

[15] E. Rueda, J. F. Barrera, and R. Henao, "Lateral shift multiplexing with a modified random mask in a joint transform correlator encrypting architecture," *Opt. Eng.*, vol. 48, 2009, Art. no. 027006.

[16] E. Rueda, C. Rios, J. F. Barrera, R. Henao, and R. Torroba, "Experimental multiplexing approach via code key rotations under a joint transform correlator scheme," *Opt. Commun.*, vol. 284, pp. 2500–2504, 2011.

[17] F. Mosso, J. F. Barrera, M. Tebaldi, N. Bolognini, and R. Torroba, "All-optical encrypted movie," *Opt. Exp.*, vol. 19, pp. 5706–5712, 2011.

[18] A. Alfalou, C. Brosseau, and N. Abdallah, "Simultaneous compression and encryption of color video images," *Opt. Commun.*, vol. 338, pp. 371–379, 2015.

[19] A. Alfalou, C. Brosseau, N. Abdallah, and M. Jridi, "Simultaneous fusion, compression, and encryption of multiple images," *Opt. Exp.*, vol. 19, pp. 24023–24029, 2011.

[20] A. Alfalou, C. Brosseau, N. Abdallah, and M. Jridi, "Assessing the performance of a method of simultaneous compression and encryption of multiple images and its resistance against various attacks," *Opt. Exp.*, vol. 21, pp. 8025–8043, 2013.

[21] T. J. Naughton and B. Javidi, "Compression of encrypted three-dimensional objects using digital holography," *Opt. Eng.*, vol. 43, pp. 2233–2238, 2004.

[22] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 576–584.

[23] S. Jiao, Z. Jin, C. Chang, C. Zhou, W. Zou, and X. Li, "Compression of phase-only holograms with JPEG standard and deep learning," *Appl. Sci.*, vol. 8, pp.1258, 2018.

[24] T. Shimobaba *et al.*, "Dynamic-range compression scheme for digital hologram using a deep neural network," *Opt. Lett.*, vol. 44, pp. 3038–3041, 2019.

[25] L. Zhang, R. Xiong, J. Chen, and D. Zhang, "Optical image compression and encryption transmission-based on deep learning and ghost imaging," *Appl. Phys. B*, vol. 126, pp. 1–10, 2020.

[26] Y. Li, J. Li, and J. Li, "High-quality object reconstruction from one-dimensional compressed encrypted signal based on multi-network mixed learning," *IEEE Access*, vol. 8, pp. 155224–155234, 2020.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[29] M. D. Pritt and D. C. Ghiglia, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*, Hoboken, NJ, USA: Wiley, 1998.

[30] W. Stallings, *Cryptography and Network Security*. Upper Saddle River, NJ, USA: Prentice Hall, 2004.