CHINESE OPTICS LETTERS

Vol. 24, No. 6 | June 2026

Physics-driven cryptographic framework for high security and rapid decryption

Jiaan Gan (甘家安)[1], Yong Yang (杨 勇)[2*], Siwei Zhu (朱思伟)[1], Shengjiang Chang (常胜江)[1**], and Xiaocong Yuan (袁小聪)[1,2,3***]

[1] Institute of Modern Optics and Institute of Translational Medicine, Nankai University, Tianjin 300350, China

[2] School of Science, Jiangnan University, Wuxi 214122, China

[3] Nanophotonics Research Center, Shenzhen Key Laboratory of Micro-Scale Optical Information Technology, Shenzhen University, Shenzhen 518060, China

*Corresponding author: yangyongjnu@gmail.com

**Corresponding author: sjchang@nankai.edu.cn

***Corresponding author: xcyuan@szu.edu.cn

Received November 12, 2025 | Accepted February 4, 2026 | Posted Online June 2, 2026

Speckle-based cryptography with deep learning shows strong potential for optical security. However, existing data-driven approaches rely on massive plaintext–ciphertext pairs. Due to intensity-only measurements and the absence of physical constraints, these methods require extensive ciphertext acquisition and measurement, leading to high experimental cost and limited generalization. We propose Phys-Encrypt-Net, a physics-driven optical encryption diffractive neural network, where a multilayer diffractive structure acts as a stacked scattering medium to encrypt the optical field. The phase distribution, pixel configuration, and interlayer spacing constitute the physical encryption key. Phys-Encrypt-Net enables direct optical-domain decryption of complex amplitudes without ciphertext measurements or digital latency, achieving 94.65% accuracy on unseen speckle data. This framework enables secure, energy-efficient, and fast optical decryption without large-scale ciphertext measurements, offering a scalable paradigm for secure optical communication.

Keywords: optical encryption; speckle-based cryptosystem; physics-driven design; high security; rapid decryption.

© The Authors. Published by Chinese Laser Press under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

DOI: 10.3788/COL202624.060501

1. Introduction

Information security has become a critical issue in modern society[1]. Optical cryptosystems, offering advantages of high security, high speed, and multidimensional capability, are considered highly promising solutions for information protection[2–8]. Among these, speckle-based optical encryption leverages random scattering media to achieve information scrambling. The inherent complexity and randomness of such media provide a high degree of security, making speckle encryption an important research direction in optical security[9–11].

In recent years, the integration of deep learning with scattering media has offered new insights into optical encryption and information processing[12–16]. These approaches typically employ scattering media such as frosted glass[13], random phase masks[17], or disordered metasurfaces[18] to transform plaintext into complex and random speckle fields[16–20]. By measuring the ciphertext intensity using detector, large scale plaintext-ciphertext pairs can be collected to serve as training datasets for neural network. Such methods do not require explicit knowledge of the optical transmission matrix. Instead, they rely on end-to-end training to optimize network weights and biases to learn the mapping between plaintext and ciphertext, thereby achieving speckle decryption. This has led to significant advances at the intersection of optical security and intelligent computation. Accordingly, acquiring sufficiently large training datasets to carefully train the network is crucial for these encryption systems.

Purely data-driven end-to-end decryption schemes may be limited by the availability of a sufficient number of real images for training. For example, speckle encryption schemes based on multimode fibers may require several days to acquire thousands of speckle samples[21], during which environmental perturbations, such as temperature fluctuations, vibrations, and fiber bending, can significantly alter the optical transmission matrix. It has been reported that bending of the optical fiber can render a

trained network ineffective[22], making it impossible to acquire enough real images for training. Moreover, because neural networks lack inherent physical constraints, purely data-driven schemes often require complex architectures and large-scale datasets to converge over extended training periods[23–25]. Additionally, during ciphertext measurement, detectors capture only intensity information, while phase information is lost, resulting in incomplete network input representations. This necessitates even larger datasets (intensity-only images) and longer training times to compensate for missing information. Repeated measurements to generate large-scale datasets are time-consuming and require substantial storage, incurring high acquisition costs. Consequently, in many practical scenarios, collecting sufficient training data is challenging, leading to network overfitting and limited generalization and thereby constraining applicability.

In existing approaches, neural networks are usually employed for decryption[16–20], requiring measured ciphertexts to be processed via GPU-based inference. This introduces latency and significant power consumption, deviating from the original intention of all-optical encryption, leading to the loss of the merits of parallel, high-speed, and low-power consumption properties to some extent. To date, the concept of employing neural networks to directly perform optical encryption and to realize decryption by exploiting the amplitude and phase of the ciphertext field has not yet been explored. Current techniques, relying solely on limited intensity-only measurements and networks lacking physical mechanisms, struggle to capture the physical mapping between plaintext and ciphertext effectively. Therefore, developing a framework in which neural networks perform encryption and can exploit the complete complex amplitude of the optical field while circumventing the need for large-scale data collection is of critical importance.

Here, we introduce a novel physics-driven encryption diffractive deep neural network (Phys-Encrypt-Net). Departing from conventional network-based decryption approaches, our system employs a multilayer diffractive deep neural network (D2NN)[26–30] as a stacked scattering medium to physically encrypt plaintext optical fields into ciphertext. The phase distribution, pixel count, pixel pitch size, and interlayer distances collectively constitute a physical key, providing a high level of security. During decryption, a complete physical model is employed. This allows the amplitude and phase of the ciphertext to be decrypted directly using a fixed phase plate as the physical decoder, operating entirely in the optical domain without digital processing delay. Optimization is achieved through the interaction between the network and the physical model. This process yields a feasible solution that satisfies the imposed physical constraints and avoids the phase loss inherent to intensity-only detection. By incorporating explicit physical constraints and the network, this approach eliminates the need for ciphertext acquisition and subsequent GPU-based numerical decryption. At the same time, it substantially enhances generalization performance and physical interpretability. Phys-Encrypt-Net achieves 94.65% decryption accuracy on speckle fields generated from 10,000 previously unseen MNIST images. Even when sufficient training data are unavailable, the system can perform encryption and decryption at the speed of light, with low power consumption and high speed. This provides a new paradigm for physically interpretable optical information security systems that do not require large-scale ciphertext acquisition.

2. Methods

2.1. Structure of Phys-Encrypt-Net cryptosystem

The optical design of the speckle-based encryption system built on the Phys-Encrypt-Net framework is illustrated in Fig. 1. A 633-nm-laser beam is expanded and projected onto a digital micromirror device (DMD)[31,32] to form the plaintext optical field, where handwritten digits from the MNIST dataset are displayed. The resulting plaintext field passes through a 4*f* system with a pinhole filter, serving as the input layer of the diffractive network. It then sequentially propagates through four diffractive layers, each consisting of 256 × 256 neurons with a pixel pitch of 8 μm, before forming a speckle ciphertext field at the output plane. Through the multi-layer diffraction process, the plaintext information becomes highly scrambled. Unlike conventional approaches that rely solely on intensity measurements for neural network-based digital decryption, Phys-Encrypt-Net utilizes both the amplitude and phase of the ciphertext optical field during decryption. A phase plate is designed as the physical decoder, and after propagating a certain distance, the ciphertext field is modulated by this phase plate and recorded by a camera on the detection plane. The decrypted result appears as the energy distribution across ten predefined regions on the detector where the index of the region with the maximum intensity corresponds to the recognized digit. Importantly, our system does not require ciphertext measurement followed by numerical decryption on a GPU. Instead, the optical setup itself serves as a physical decoder, enabling fully optical speckle decryption by exploiting

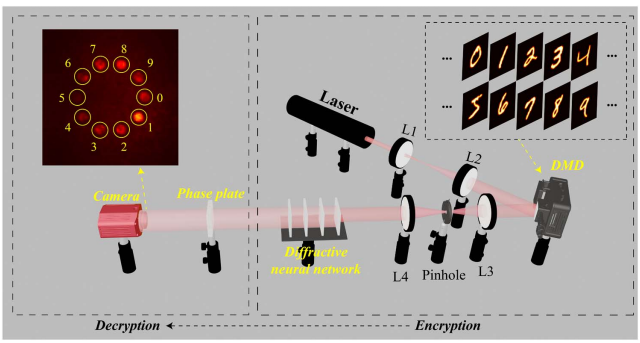

Fig. 1. Structure of the Phys-Encrypt-Net cryptosystem. A 633-nm-laser beam is expanded by a pair of lenses (L1 and L2). The plaintext image is displayed on a DMD. A 4*f* system formed by lenses L3 and L4 selects the first-order diffraction through a pinhole. The plaintext is then modulated by the diffractive neural network to generate the speckle ciphertext at the output plane. The ciphertext field is further modulated by a fixed phase plate, and the decrypted optical field is recorded by a camera.

the complete complex amplitude of the optical field. This design effectively eliminates the need for large-scale ciphertext data acquisition and avoids phase loss inherent to intensity-only detection.

2.2. Principle and concept

During the encryption process, as illustrated in Fig. 2(a), we construct a D2NN consisting of four diffractive layers, with each layer containing 256×256 pixels. These pixels act as artificial neurons capable of receiving, modulating, and propagating the optical field. Each neuron can be regarded as a secondary wave source, with its amplitude and phase determined by the product of the input field and the neuron's transmission coefficient. As shown in Fig. 2(a), the complex-valued input to the network is the plaintext optical field corresponding to different MNIST handwritten digits. The incident light sequentially passes through the modulation of each diffraction layer, where the system leverages the interaction between light and matter to achieve multi-layer joint modulation of the input field. The output layer produces the speckle ciphertext. More specifically, the neurons in each layer of Phys-Encrypt-Net are connected to the neurons in adjacent layers through free-space wave propagation, following the Rayleigh–Sommerfeld (RS) diffraction equation[26,33]:

$$w_i(x,y,z) = \frac{z-z_i}{r^2}\left(\frac{1}{2\pi r}+\frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right). \tag{1}$$

Here, i denotes the ith neuron at the position (x_i, y_i, z_i) in a given layer of the system, λ is the wavelength of the illumination source, $r = \sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}$, and $j = \sqrt{-1}$. The amplitude and relative phase of the secondary wave are determined by the product of the neuron's input field and its transmission coefficient t, both of which are complex-valued functions. Based on this, the output function $y_{\text{out}_i}(x,y,z)$ of the ith pixel located at (x_i, y_i, z_i) in a given layer of the diffractive neural network can be expressed as

$$y_{\text{out}_i}(x,y,z) = w_i(x,y,z)\cdot t_i(x_i,y_i,z_i)\cdot y_{\text{in}_i}(x_i,y_i,z_i), \tag{2}$$

where

$$y_{\text{in}_i}(x_i,y_i,z_i) = \sum_k (w_k(x,y,z)\cdot y_{0_k}(x_i,y_i,z_i)), \tag{3}$$

$$t_i(x_i,y_i,z_i) = a_i(x_i,y_i,z_i)e^{j\phi_i(x_i,y_i,z_i)}, \tag{4}$$

y_0 represents the complex amplitude distribution of the incident optical field, and a_i denotes the amplitude coefficient of each neuron.

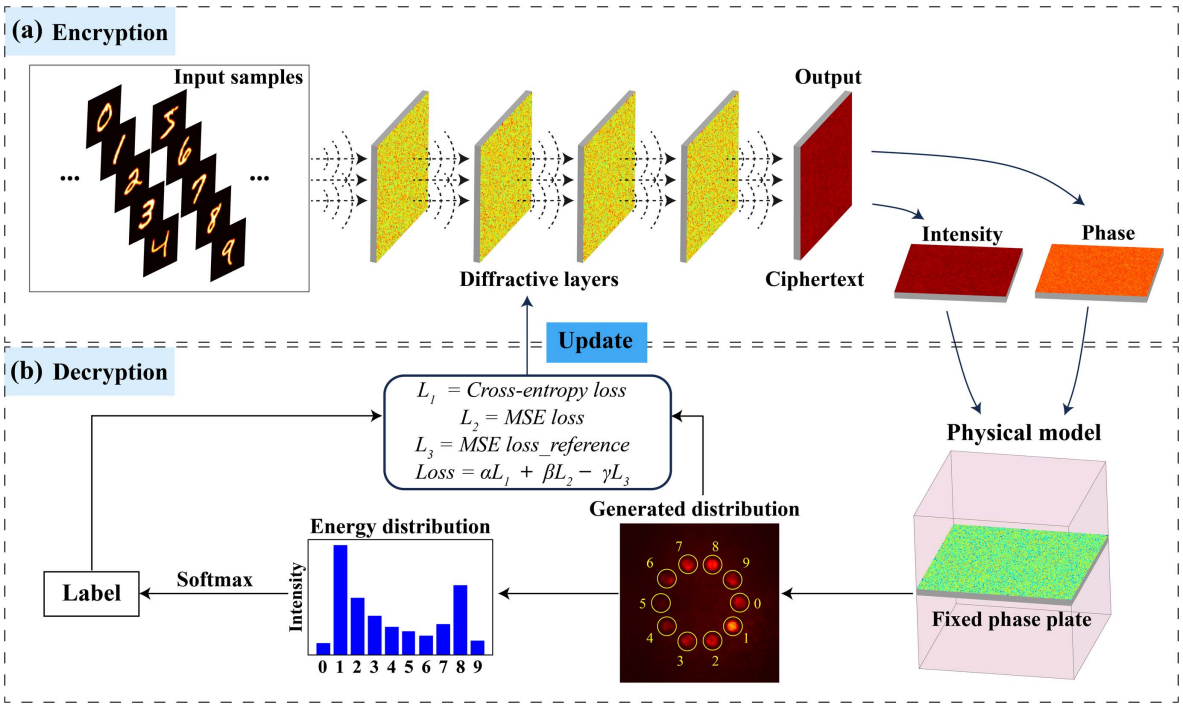

Fig. 2. Working principle of the Phys-Encrypt-Net cryptosystem. (a) The multilayer diffractive network modulates the plaintext input into a speckle ciphertext field, achieving optical encryption through the network. (b) During decryption, both the amplitude and phase information of the ciphertext field is utilized. The field is propagated through a fixed phase plate serving as the decoder according to the physical model to generate the distribution at the detection plane. Classification is determined by comparing the energy distributions across 10 detection regions, an error function composed of cross-entropy loss (L_1) and MSE loss (L_2) is used to update the network parameters, and an additional MSE loss (L_3) constrains the raw speckle field (without decoder modulation) to prevent information leakage at the detection plane.

The formula describes how the optical field's amplitude and phase evolve during propagation between diffractive layers. In the D2NN, this diffraction governs the spatial transformation of the input light across layers, enabling full optical connectivity and complex optical computations. In our implementation, using a phase-only D2NN architecture, the amplitude is assumed to be constant, ideally equal to 1. The transmission coefficient of each neuron is denoted as t_i, which incorporates both amplitude and phase terms, and ϕ represents the phase modulation imparted by each diffractive neuron.

In the proposed system, the operation is divided into two main stages. First, the diffractive neural network model, predicts and outputs the ciphertext optical field corresponding to different input fields. Second, the ciphertext field is propagated through a physical model that incorporates a specific decoder, generating the decrypted optical field at predetermined positions. Specifically, the physical constraints are directly embedded into the network optimization through the loss function. Although the network generates encrypted speckle fields, the loss is computed on the decrypted image obtained after modulation by the fixed decryption phase mask. As a result, the gradients backpropagated through the network inherently carry information about the decryption phase mask, effectively embedding the physical constraints into the optimization. This ensures that the learned speckle patterns are physically consistent with the fixed optical decryption mechanism, achieving true synergistic optimization between the neural network and the physical model.

As illustrated in Fig. 2(b), the complex amplitude of the ciphertext field is processed by the physical model to produce the generated optical field distribution. The detection plane captured by the camera is divided into ten circular subregions, each corresponding to a classification category in the D2NN. During inference, the system determines the decrypted class based on the intensity of each subregion. The region with the highest intensity indicates the predicted category, analogous to the softmax mechanism in electronic neural networks. The category corresponding to the peak in the energy distribution histogram is taken as the decrypted label.

During optimization, the weights of the diffractive neural network are determined by RS diffraction theory, while the phase distributions of the diffractive layers serve as learnable bias parameters. By iterating over the training samples, the phase distributions of the D2NN layers are updated through the interaction between the encryption-decryption process (the neural network and the physical model) ultimately yielding a feasible solution that satisfies the imposed physical constraints.

To optimize the network model, we trained Phys-Encrypt-Net using the 60,000-sample MNIST training set and evaluated its performance on 10,000 previously unseen test samples. Each 28×28 pixels MNIST image was first upsampled to 180×180 pixels using bilinear interpolation and then zero-padded to 256×256 pixels to match the network input size. When a plaintext image is provided to the input layer, the predicted ciphertext field is calculated according to RS diffraction theory and the previously described equations. The training objective is to ensure that the network-predicted ciphertext field satisfies the physical constraints imposed by the fixed phase plate in the decryption stage. Specifically, the output of the network should match the phase modulation of the decoder and, through the physical model, generate the decrypted optical field at the designated detection positions. To achieve this, we designed a three-term loss function. First, a cross-entropy loss (L_1) measures classification accuracy by determining the region with the maximum energy among the ten detection subregions. Second, a mean squared error (MSE) loss (L_2) constrains the intensity distribution of the decrypted field, encouraging the energy to be concentrated in the target detection region (set to 1) while suppressing the others (set to 0). This MSE term further guides the network optimization, maximizing energy utilization and classification stability. Third, an additional MSE loss (L_3) on the unregulated (reference) speckle field without passing through the decoder ensures that the speckle pattern itself does not leak class information without the decoder. The final loss is defined as $\text{Loss} = \alpha L_1 + \beta L_2 - \gamma L_3$, with $\alpha = 1$, $\beta = 0.1$, and $\gamma = 0.1$. We adopted an iterative training procedure, cycling through all training samples. The learnable phase ϕ of the diffractive layers was updated via backpropagation using the Adam optimizer[34]. The network was implemented in TensorFlow (Google Inc.) with Python 3.6.13 and trained on a system equipped with an NVIDIA GeForce RTX 3090 GPU and Intel Xeon Silver 4210R CPU @ 2.40 GHz with 32 GB RAM. The network input size was 256×256 pixels, with a batch size of 128 and a learning rate of 0.01. Training over the 60,000-sample MNIST dataset for 50 epochs required approximately 153 min.

3. Results

To visually illustrate the performance of the Phys-Encrypt-Net system, Fig. 3(a) presents 10 representative examples, including the plaintext, ciphertext, decrypted field, and the corresponding energy distribution histogram over 10 detection regions. It can be seen that the multilayer diffractive network scrambles the plaintext into random speckle ciphertext, rendering the original image completely unrecognizable and achieving effective optical encryption. The system optimizes the network parameters through training such that different ciphertext fields can match the fixed decoder during decryption, recovering the decrypted optical fields as shown. The decryption results are determined from the energy distribution over the 10 detection regions, with the region exhibiting the highest energy (highlighted in purple) corresponding to the decoded digit. To validate the effectiveness of the system, we evaluated the decryption accuracy of Phys-Encrypt-Net on 10,000 unseen test samples that were not included during training. The evolution of accuracy for both training and test sets over the course of training is shown in Fig. 3(b). As the number of epochs increases, the accuracy for both sets gradually improves and stabilizes after the 20th epoch. The final

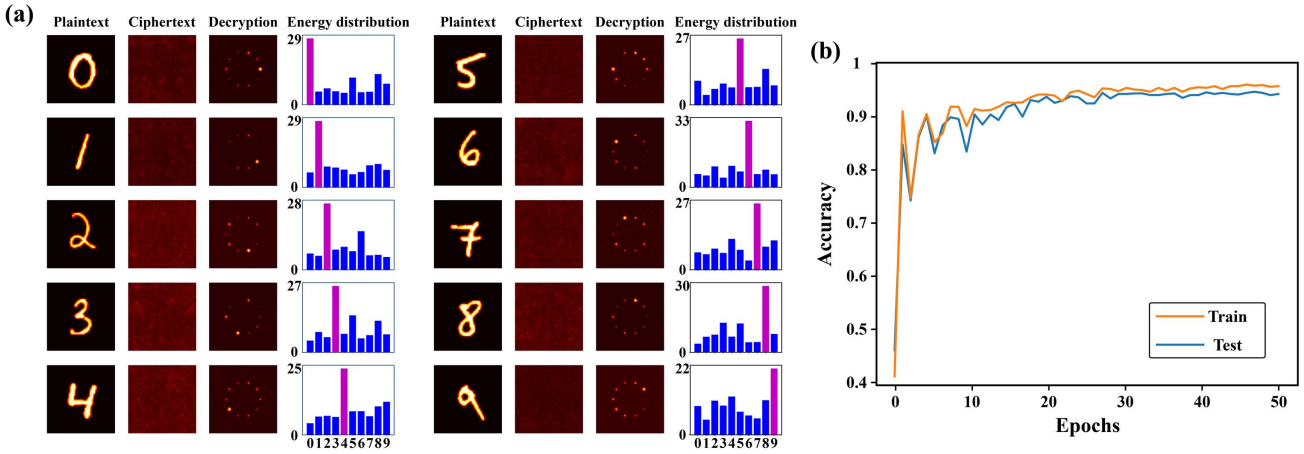

Fig. 3. Encryption and decryption performance of Phys-Encrypt-Net. (a) Ten handwritten digit examples showing the plaintext, ciphertext, decrypted image, and the corresponding energy distribution histogram over 10 detection regions. (b) Evolution of decryption accuracy on the training and test sets during the network training process.

training accuracy reaches approximately 95.5%, while the test accuracy is around 94.65%. These results demonstrate that the system achieves high decryption accuracy while maintaining strong generalization capability. Compared with conventional D2NN-based light-speed classification methods, Phys-Encrypt-Net preserves similar classification accuracy while simultaneously enabling secure information transmission and light-speed optical decryption, offering a new approach for optical encryption.

The system employs a D2NN composed of multiple diffractive layers as a physical encryption key, generating unique speckle patterns through optical diffraction. The trained model and its parameters, including phase distribution, number of layers, number of neurons, pixel pitch, and inter-layer distances, together constitute a non-clonable encryption key. During decryption, a fixed phase plate serves as the physical decoder, whose phase distribution and spatial positioning also form essential components of the secure key. Correct decryption is only possible when the encryption network and decoder are perfectly matched. As shown in Fig. 4(a), when all key components are correct, the system achieves a decryption accuracy of 94.65% on 10,000 unseen test samples, as reflected in the corresponding confusion matrix. Key sensitivity is a critical aspect of cryptosystems. As shown in Fig. 4(b), using incorrect network parameters leads to performance degradation. Specifically, we replaced the fourth diffractive layer of the network with a random phase distribution while keeping the remaining layers unchanged. The decryption accuracy dropped sharply to 6.46%. During training, the goal of the diffractive network is to generate ciphertext fields that accurately match the modulation of the decoder, thereby satisfying the decryption rules. Consequently, the phase distribution, number of pixels, and feature dimensions of the physical decoder are critical key parameters that directly affect decryption performance. As illustrated in Fig. 4(c), when an incorrect decoder is used and all other key parameters are correct, the confusion matrix shows that the accuracy decreases to 14.84%, making recovery of meaningful information impossible. These results underscore the critical influence of both the encrypted diffractive neural network and the physical decoder on decryption performance, highlighting their dual role as essential elements for accurate decryption and as secure keys for the system.

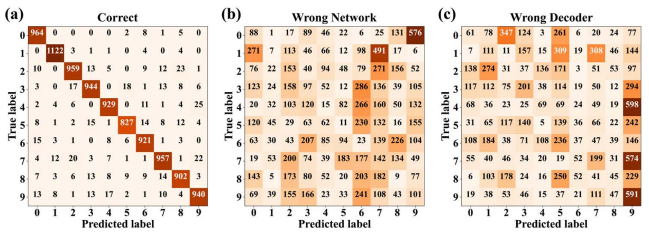

Fig. 4. Sensitivity analysis of the encryption and decryption keys. (a) Confusion matrix of decryption results for 10,000 unseen test samples when both the encryption and decryption physical keys are correct. (b) Confusion matrix of system decryption results when incorrect diffractive network parameters are used for encryption (incorrect encryption physical key). (c) Confusion matrix of system decryption results when an incorrect decoder is used for decryption (incorrect decryption physical key).

Furthermore, we investigated the impact of the camera's axial displacement on the decryption performance of the system. In general, axial shifts of the camera reduce the correlation of the speckle field and may degrade decryption quality. As shown in Fig. 5(a), the camera was moved along the optical axis in 5 mm increments over a range of ±15 mm relative to the original detection plane. It presents the decrypted intensity distributions and corresponding energy histograms for the plaintext digit "2" at seven different detection positions. It can be observed that the highest contrast in the decrypted image occurs at 0 mm displacement. As the displacement increases, the speckle field becomes more diffused, and the contrast among the 10 detection regions decreases. Within the ±5 mm range, the energy peak corresponding to digit "2" clearly remains higher than other categories, whereas beyond

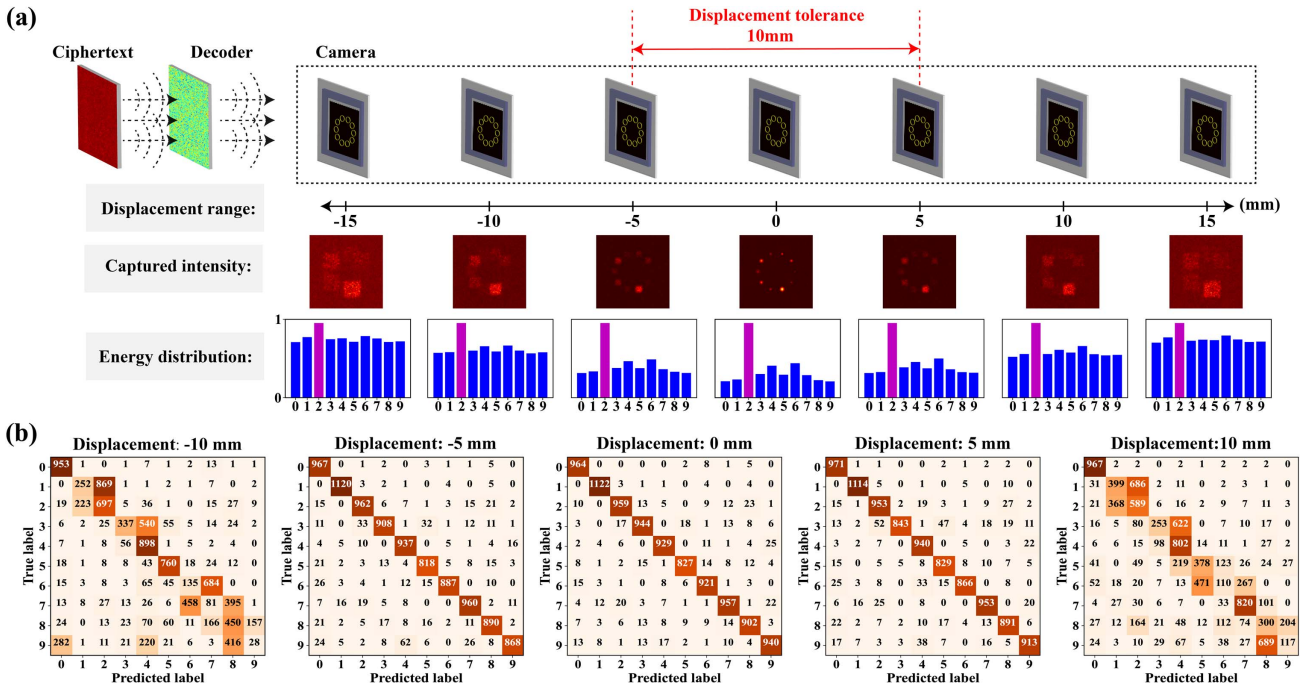

Fig. 5. Analysis of tolerance to displacement errors. (a) Decrypted images and corresponding energy distribution histograms obtained at seven different axial camera positions. (b) Confusion matrices of the system for all test samples when the camera axial displacement is ±10 mm, ±5 mm, and 0 mm.

this range, the energy differences across regions gradually diminish, consistent with the observed degradation in the decrypted images. To quantitatively assess the effect of axial displacement on decryption performance, Fig. 5(b) shows the confusion matrices for different displacements. The decryption accuracies at −10 mm, −5 mm, 0 mm, +5 mm, and +10 mm were 45.91%, 93.17%, 94.65%, 92.73%, and 47.35%, respectively. These results indicate that the system maintains high decryption performance within a ±5 mm range and exhibits good tolerance to positional errors up to 10 mm. When the displacement exceeds ±10 mm, the accuracy drops below 50%, and the confusion matrices reveal that many decrypted digits are misclassified as neighboring classes. Overall, these results demonstrate that Phys-Encrypt-Net maintains high decryption accuracy within a 10 mm axial range, exhibiting strong robustness and positional tolerance. This physics-driven speckle cryptosystem provides a practical pathway for overcoming challenges in optical encryption related to low positional tolerance and degraded decryption quality, offering a robust and reliable solution.

To evaluate the practical alignment sensitivity of the proposed Phys-Encrypt-Net, we further investigate its tolerance to angular misalignment. As shown in Fig. 6, a controlled angular tilt is applied to the phase modulation plane by introducing a linear phase ramp, and the corresponding decryption accuracy is measured. The results indicate that the system preserves stable performance for small angular deviations, while the decryption quality gradually degrades as the tilt angle increases. This behavior is consistent with the coherent and interference-based nature of the diffractive decoding process.

To further assess the generalization capability of the proposed Phys-Encrypt-Net, we tested the system with digit images of sizes different from the training samples. All scale generalization results are obtained in a zero-shot setting. After training the Phys-Encrypt-Net at the original input scale, scaled test images are directly evaluated using the same trained model without any retraining or fine-tuning. Specifically, 9 sets of test samples were constructed with digit sizes varying from 0.8 to 1.2 times that of the training data to systematically evaluate the network's performance under scale variations.

When the input digit sizes were 0.9, 1.0, and 1.1 times those of the training samples, the corresponding decryption results and confusion matrices are presented in Fig. 7(a), showing decryption accuracies of 92.62%, 94.65%, and 92.84%,

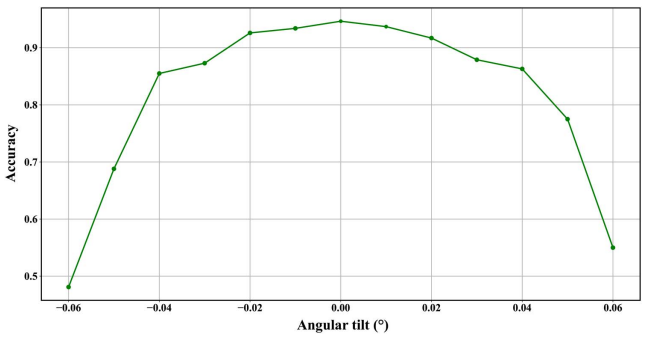

Fig. 6. Analysis of tolerance to angular misalignment. The angular misalignment is introduced by adding a linear phase ramp, with the tilt angle varied from −0.06° to +0.06°.

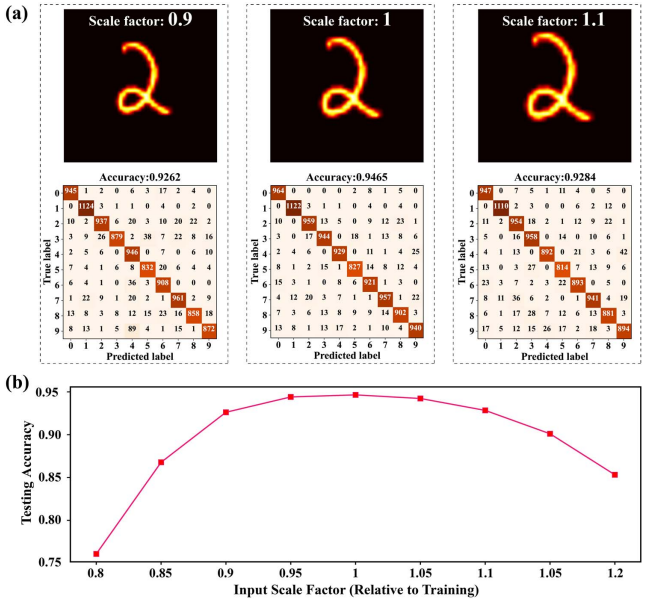

Fig. 7. Generalization performance under varying input scales. (a) Decryption results and corresponding confusion matrices for test sets with different digit sizes, where the digit size scale factors relative to the training data are 0.9, 1.0, and 1.1. (b) Variation of decryption accuracy as a function of the input size scale factor across the test sets.

respectively, indicating effective recognition of the corresponding speckle ciphertexts. Testing on 30,000 previously unseen samples further demonstrates that the system maintains over 92% accuracy within the 0.9–1.1 scale range, highlighting its stable generalization performance. The variation of decryption accuracy across the 9 sets of test samples with different input sizes is shown in Fig. 7(b). The results indicate that performance is optimal when the input size matches the training data, while slight decreases in accuracy are observed as the size factor increases or decreases. For scale factors of 0.8 and 1.2, the system achieves accuracies of 76.08% and 85.32%, respectively. Despite substantial geometric variations in the input digits, the system retains a significant level of generalization without sharp performance degradation. These results suggest that Phys-Encrypt-Net does not merely memorize statistical features of the training samples but rather learns a stable physical mapping among the input wavefront, speckle, and the embedded physical decoder. In contrast, conventional purely data-driven networks, which rely on large-scale plaintext-ciphertext pairs for pretraining and only capture intensity (amplitude) information of the ciphertexts, typically exhibit limited generalization when insufficient training data are available. By leveraging both amplitude and phase information of the ciphertext field and through the interaction between the diffractive network and the embedded physical model, Phys-Encrypt-Net avoids the high cost of acquiring large-scale ciphertext measurements and enables efficient optimization. The proposed system maintains stable performance under input perturbations, demonstrating strong generalization and practical potential for speckle encryption and decryption applications.

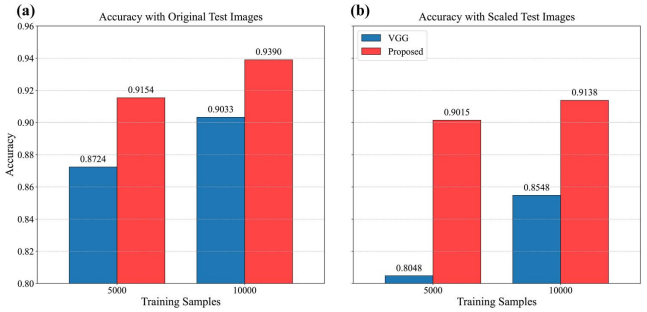

Fig. 8. Accuracy comparison of the proposed method against the baseline VGG model as a function of training sample size (5000 and 10,000). (a) Results on original test images. (b) Results on scaled test images.

To further evaluate the performance of our method, we conducted a comparative experiment with a standard Visual Geometry Group (VGG)[12] type convolutional neural network as a baseline. Both methods were trained on limited datasets containing 5000 and 10,000 images and evaluated on a completely unseen test set of 10,000 images. As shown in Fig. 8, the left panel (standard test) presents results on test images that have the same size as the training images but were not seen during training, reflecting the standard generalization capability. The right panel (scale-generalization test) shows results on test images scaled to 0.9× of their original size, simulating a distribution shift and assessing robustness to variations in input scale. The results clearly demonstrate that our method outperforms VGG even when trained on a small dataset. For instance, when trained with 5000 images, our method achieves higher accuracy than VGG trained with 5000 or even 10,000 images, in both standard and scale-generalization tests. This indicates that our physics-driven approach requires significantly less training data, reducing the cost and effort associated with collecting large-scale datasets. Furthermore, unlike convolutional neural network-based decryption, which relies on GPU-based forward inference, our method performs decryption directly in the optical domain, avoiding digital processing delays. Therefore, this comparison not only demonstrates superior performance under limited data but also highlights the unique advantage of light-speed optical decryption, which is beyond what conventional convolutional neural network-based approaches can achieve.

4. Conclusion

In this work, we propose a physics-driven optical cryptosystem, Phys-Encrypt-Net, which for the first time explores the use of diffractive neural network for encryption and embeds a physical model to perform decryption using the full complex amplitude of the ciphertext. This approach eliminates the need for ciphertext measurement, avoiding the high cost associated with acquiring large-scale plaintext-ciphertext pairs while offering high security, low power consumption, and rapid decryption. Phys-Encrypt-Net employs a network as a

multilayer scattering medium for encryption, and leverages the amplitude and phase of the ciphertext field to achieve optical-domain decryption through a fixed phase decoder. Unlike purely data-driven approaches, Phys-Encrypt-Net performs decryption directly in the optical domain, thereby avoiding the phase loss problem caused by intensity detection, as well as the decryption latency and power consumption of network inference on GPUs. This physics-driven design enables synergistic optimization between the physical mechanisms and the neural networks, yielding feasible solutions that rigorously satisfy optical constraints. Consequently, it overcomes the generalization limitations inherent in purely data-driven networks trained on limited intensity-only ciphertext images, which struggle to capture the underlying physical mapping between plaintexts and ciphertexts. The diffractive network and the decoder together serve as physical keys, forming a large key space that ensures system security. Phys-Encrypt-Net demonstrate a decryption accuracy exceeding 94.65%, comparable to the classification performance of conventional diffractive neural networks[26,35,36].

Importantly, our system not only enables light-speed information transmission and recognition but also realizes encryption and secure data transfer. Compared with purely data-driven methods, Phys-Encrypt-Net overcomes the limitations of intensity-only measurements and data scarcity, achieving high-security speckle-based cryptosystems at light speed and low power. The proposed framework relies on precise lateral alignment. Such sensitivity is a common characteristic of coherent diffractive and optical encryption systems, where precise spatial alignment is required for accurate wavefront reconstruction. From a security perspective, this lateral alignment sensitivity can be regarded as an additional physical constraint, as successful decryption requires not only the correct encryption parameters but also precise spatial positioning, which is difficult to achieve without full knowledge of the system configuration. Improving tolerance to large lateral misalignments is beyond the scope of the present work and will be investigated in future studies. This work establishes a new paradigm for physically interpretable optical information security systems that do not require large-scale ciphertext acquisition, offering significant implications for meeting modern communication demands for high-speed and secure data transmission.

Acknowledgements

This work was supported by the Guangdong Major Project of Basic Research (No. 2020B0301030009), the National Natural Science Foundation of China (Nos. 12174204, 12174203, 12074203, 62335012, and 62435010), the Natural Science Foundation of Guangdong Province (No. 2023A1515012888), the Science and Technology Innovation Commission of Shenzhen (Nos. JCYJ20220818101417039 and JCYJ20241202124428038), the Medical-Engineering Interdisciplinary Research Foundation of Shenzhen University (No. 86901/00000311), the Scientific Instrument Developing Project of Shenzhen University (No. 2023YQ001), the Shenzhen University 2035 Initiative (No. 2023B004), and the Key Research and Development Program of Zhejiang (No. 3000-3AA240100).

References

1. B. Javidi, A. Carnicer, M. Yamaguchi, *et al.*, "Roadmap on optical security," J. Opt. **18**, 083001 (2016).
2. P. Refregier and B. Javidi, "Optical image encryption based on input plane and Fourier plane random encoding," Opt. Lett. **20**, 767 (1995).
3. P. Caramazza, O. Moran, R. Murray-Smith, *et al.*, "Transmission of natural scene images through a multimode fibre," Nat. Commun. **10**, 2029 (2019).
4. I. Mehra and N. K. Nishchal, "Image fusion using wavelet transform and its application to asymmetric cryptosystem and hiding," Opt. Express **22**, 5474 (2014).
5. W. Chen, X. Chen, and C. J. Sheppard, "Optical image encryption based on diffractive imaging," Opt. Lett. **35**, 3817 (2010).
6. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," Opt. Express **24**, 13738 (2016).
7. Y. Zhang and B. Wang, "Optical image encryption based on interference," Opt. Lett. **33**, 2443 (2008).
8. G. Situ and J. Zhang, "Double random-phase encoding in the Fresnel domain," Opt. Lett. **29**, 1584 (2004).
9. L. Bian, X. Chang, S. Jiang, *et al.*, "Large-scale scattering-augmented optical encryption," Nat. Commun. **15**, 9807 (2024).
10. H. Ruan, J. Xu, and C. Yang, "Optical information transmission through complex scattering media with optical-channel-based intensity streaming," Nat. Commun. **12**, 2411 (2021).
11. L. Zhou, Y. Xiao, and W. Chen, "Learning complex scattering media for optical encryption," Opt. Lett. **45**, 5279 (2020).
12. N. Borhani, E. Kakkava, C. Moser, *et al.*, "Learning to see through multimode fibers," Optica **5**, 960 (2018).
13. S. Li, M. Deng, J. Lee, *et al.*, "Imaging through glass diffusers using densely connected convolutional networks," Optica **5**, 803 (2018).
14. B. Rahmani, D. Loterie, G. Konstantinou, *et al.*, "Multimode optical fiber transmission with a deep learning network," Light Sci. Appl. **7**, 69 (2018).
15. H. Li, Z. Yu, Q. Zhao, *et al.*, "Learning-based super-resolution interpolation for sub-Nyquist sampled laser speckles," Photon. Res. **11**, 631 (2023).
16. Q. Zhao, H. Li, Z. Yu, *et al.*, "Speckle-based optical cryptosystem and its application for human face recognition via deep learning," Adv. Sci. **9**, 2202407 (2022).
17. X. Wang, H. Wei, M. Jin, *et al.*, "Experimental optical encryption based on random mask encoding and deep learning," Opt. Express **30**, 11165 (2022).
18. Z. Yu, H. Li, W. Zhao, *et al.*, "High-security learning-based optical encryption assisted by disordered metasurface," Nat. Commun. **15**, 2607 (2024).
19. X. Wang, W. Wang, H. Wei, *et al.*, "Holographic and speckle encryption using deep learning," Opt. Lett. **46**, 5794 (2021).
20. L. Zhang, S. Lin, Q. Zhou, *et al.*, "Speckle-based optical encryption with complex-amplitude coding and deep learning," Opt. Express **31**, 35293 (2023).
21. R. Lan, Y. Li, H. Shen, *et al.*, "All-optical color image encryption using multimode fiber speckles and diffractive deep neural networks," Opt. Express **33**, 36133 (2025).
22. H. Yu, Z. Huang, S. Lamon, *et al.*, "All-optical image transportation through a multimode fibre using a miniaturized diffractive neural network on the distal facet," Nat. Photonics **19**, 486 (2025).
23. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, *et al.*, "Physics-informed machine learning," Nat. Rev. Phys. **3**, 422 (2021).
24. M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. **378**, 686 (2019).
25. K. Liu, J. Wu, Z. He, *et al.*, "4K-DMDNet: diffraction model-driven network for 4K computer-generated holography," Opto-Electron. Adv. **6**, 220135 (2023).
26. X. Lin, Y. Rivenson, N. T. Yardimci, *et al.*, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004 (2018).

27. T. Zhou, X. Lin, J. Wu, *et al.*, "Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit," Nat. Photonics **15**, 367 (2021).
28. H. Chen, S. Lou, Q. Wang, *et al.*, "Diffractive deep neural networks: theories, optimization, and applications," Appl. Phys. Rev. **11**, 021332 (2024).
29. J. Hu, D. Mengu, D. C. Tzarouchis, *et al.*, "Diffractive optical computing in free space," Nat. Commun. **15**, 1525 (2024).
30. E. Goi, S. Schoenhardt, and M. Gu, "Direct retrieval of Zernike-based pupil functions using integrated diffractive deep neural networks," Nat. Commun. **13**, 7531 (2022).
31. Y. Yang, A. Forbes, and L. Cao, "A review of liquid crystal spatial light modulators: devices and applications," Opto-Electron. Sci. **2**, 230026 (2023).
32. K. Y. Bliokh, E. Karimi, M. J. Padgett, *et al.*, "Roadmap on structured waves," J. Opt. **25**, 103001 (2023).
33. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).
34. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv, arXiv:1412.6980 (2014).
35. F. Léonard, A. S. Backer, E. J. Fuller, *et al.*, "Co-design of free-space metasurface optical neuromorphic classifiers for high performance," ACS Photonics **8**, 2103 (2021).
36. H. Chen, J. Feng, M. Jiang, *et al.*, "Diffractive deep neural networks at visible wavelengths," Engineering **7**, 1483 (2021).