

Complex-valued universal linear transformations and image encryption using spatially incoherent diffractive networks

Xilin Yang^{a,b,c,†} Md Sadman Sakib Rahman,^{a,b,c,†} Bijie Bai,^{a,b,c} Jingxi Li^{a,b,c} and Aydogan Ozcan^{a,b,c,*}

^aUniversity of California, Los Angeles, Electrical and Computer Engineering Department, Los Angeles, California, United States

^bUniversity of California, Los Angeles, Bioengineering Department, Los Angeles, California, United States

^cUniversity of California, Los Angeles, California NanoSystems Institute (CNSI), Los Angeles, California, United States

Abstract. As an optical processor, a diffractive deep neural network (D²NN) utilizes engineered diffractive surfaces designed through machine learning to perform all-optical information processing, completing its tasks at the speed of light propagation through thin optical layers. With sufficient degrees of freedom, D²NNs can perform arbitrary complex-valued linear transformations using spatially coherent light. Similarly, D²NNs can also perform arbitrary linear intensity transformations with spatially incoherent illumination; however, under spatially incoherent light, these transformations are nonnegative, acting on diffraction-limited optical intensity patterns at the input field of view. Here, we expand the use of spatially incoherent D²NNs to complex-valued information processing for executing arbitrary complex-valued linear transformations using spatially incoherent light. Through simulations, we show that as the number of optimized diffractive features increases beyond a threshold dictated by the multiplication of the input and output space-bandwidth products, a spatially incoherent diffractive visual processor can approximate any complex-valued linear transformation and be used for all-optical image encryption using incoherent illumination. The findings are important for the all-optical processing of information under natural light using various forms of diffractive surface-based optical processors.

Keywords: optical computing; optical networks; machine learning; diffractive optical networks; diffractive neural networks; image encryption.

Received Oct. 19, 2023; revised manuscript received Dec. 3, 2023; accepted for publication Jan. 3, 2024; published online Jan. 19, 2024.

© The Authors. Published by SPIE and CLP 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: [10.1117/1.APN.3.1.016010](https://doi.org/10.1117/1.APN.3.1.016010)]

1 Introduction

The recent resurgence of analog optical information processing has been spurred by advancements in artificial intelligence (AI), especially deep-learning-based inference methods.^{1–9} These advances in data-driven learning methods have also benefited optical hardware engineering, giving rise to new computing architectures such as diffractive deep neural networks (D²NN), which exploit the passive interaction of light with spatially engineered surfaces to perform visual information processing. D²NNs, also referred to as diffractive optical networks, diffractive networks, or diffractive processors, have emerged as

powerful all-optical processors^{9,10} capable of completing various visual computing tasks at the speed of light propagation through thin passive optical devices; examples of such tasks include image classification,^{11–13} information encryption,^{14–17} and quantitative phase imaging (QPI),^{18,19} among others.^{20–24} Diffractive optical networks comprise a set of spatially engineered surfaces, the transmission (and/or reflection) profiles of which are optimized using machine-learning techniques. After their digital optimization (a one-time effort), these diffractive surfaces are fabricated and assembled in 3D to form an all-optical visual processor, which axially extends at most a few hundred wavelengths (λ).

Our earlier work^{10,25} demonstrated that a spatially coherent D²NN can perform arbitrary complex-valued linear transformations between a pair of arbitrary input and output apertures if its

*Address all correspondence to Aydogan Ozcan, ozcan@ucla.edu

[†]These authors contributed equally to this work.

design has a sufficient number (N) of diffractive features that are optimized, i.e., $N \geq N_i N_o$, where N_i and N_o represent the space-bandwidth product of the input and output apertures, respectively. In other words, N_i and N_o represent the size of the desired complex-valued linear transformation $\mathbf{A} \in \mathbb{C}^{N_o \times N_i}$ that can be all-optically performed by an optimized D²NN. For a *phase-only* diffractive network, i.e., only the phase profile of each diffractive layer is trainable, the sufficient condition becomes $N \geq 2N_i N_o$ due to the reduced degrees of freedom within the diffractive volume. Similar conclusions can be reached for a diffractive network that operates under spatially incoherent illumination: Rahman et al.²⁶ demonstrated that a diffractive network can be optimized to perform an arbitrary nonnegative linear transformation of optical intensity through phase-only diffractive processors with $N \geq 2N_i N_o$. However, encoding information with spatially incoherent light inherently confines both the input and output to nonnegative values, as they are represented by intensity patterns at the input and output apertures of a D²NN. To process complex-valued data with spatially incoherent light, other optical approaches were also developed,^{1,27–29} however, these earlier systems are limited to one-dimensional (1D) optical inputs and do not cover arbitrary input and output apertures, limiting their functionality and processing throughput. An extension of these earlier 1D input approaches introduced the processing of 2D incoherent source arrays using relatively bulky and demanding optical projection systems that are hard to operate at the diffraction limit of light.^{30,31}

Here, we demonstrate the processing of complex-valued data with compact diffractive optical networks under spatially incoherent illumination. We show that a spatially incoherent diffractive network that axially spans $<100 \times \lambda$ can perform any arbitrary complex-valued linear transformation on complex-valued input data with negligible error if the number of optimizable diffractive features is above a threshold dictated by the multiplication of the input and output space-bandwidth products, determined by both the spatial extent and the pixel size of the input and output apertures. To represent complex-valued spatial information using spatially incoherent illumination, we preprocessed the input information by mapping complex-valued data to a real and nonnegative, optical intensity-based representation at the input field of view (FOV) of the diffractive network. We term this mapping the “mosaicking” operation, indicating the utilization of multiple intensity pixels at the input FOV to represent one complex-valued input data point. Similarly, we used a postprocessing step, which involved mapping the output FOV intensity patterns back to the complex number domain, which we termed the “demosaicking” operation. Through these mosaicking/demosaicking operations, we show that a spatially incoherent D²NN can be optimized to perform an arbitrary complex-valued linear transformation between its input and output apertures while providing optical information encryption. The presented spatially incoherent visual information processor, with its universality and thin form factor ($<100 \times \lambda$), shows significant promise for image encryption and computational imaging applications under natural light.

2 Results

Figure 1(a) outlines a spatially incoherent D²NN architecture to synthesize an arbitrary complex-valued linear transformation (\mathbf{A}) such that $\mathbf{o} = \mathbf{A}\mathbf{i}$, where the input is $\mathbf{i} \in \mathbb{C}^{N_i}$, the target is $\mathbf{o} \in \mathbb{C}^{N_o}$ and $\mathbf{A} \in \mathbb{C}^{N_o \times N_i}$. The mosaicking process involves

finding the nonnegative (optical intensity-based) representation of each complex-valued element of \mathbf{i} using E nonnegative values; here, E bases, $e_k, k = 0, \dots, E - 1$ [see Fig. 1(c)], are used for representing the intensity-based encoding of complex numbers. Based on this representation, the 2D input aperture of a spatially incoherent D²NN will have EN_i nonnegative (optical intensity) values, denoted as $\mathbf{i}_r \in \mathbb{R}_+^{EN_i}$, representing the input information under spatially incoherent illumination. The output intensity distribution, denoted with $\mathbf{o}_r \in \mathbb{R}_+^{EN_o}$, undergoes a demosaicking process where a complex number is synthesized from the intensity values of E output pixels, yielding the complex output vector $\hat{\mathbf{o}} \in \mathbb{C}^{N_o}$ such that $\hat{\mathbf{o}} \approx \mathbf{A}\mathbf{i}$.

In our analyses, we used $E = 3$, except in Fig. S5 in the [Supplementary Material](#), where $E = 4$ results are shown for comparison. We chose the basis complex numbers as $e_k = \exp(jk \frac{2\pi}{E})$, $k = 0, \dots, E - 1$ such that the set of bases \mathcal{S} is closed under multiplication, and the product of any two of the bases in the set is also a basis; for example, for $E = 3$ we have $e_k e_l = e_{(k+l \bmod 3)}$. Based on this representation of information, with $E = 3$ and e_0, e_1, e_2 , we can decompose any arbitrarily selected complex-valued transformation matrix \mathbf{A} into $E = 3$ matrices ($\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$) with real nonnegative entries such that

$$\mathbf{A} = e_0 \mathbf{A}_0 + e_1 \mathbf{A}_1 + e_2 \mathbf{A}_2. \quad (1)$$

For a given complex-valued input $\mathbf{i} = e_0 \mathbf{i}_0 + e_1 \mathbf{i}_1 + e_2 \mathbf{i}_2$, where $\mathbf{i}_k \in \mathbb{R}_+$, the corresponding target output vector can be written as

$$\mathbf{o} = \mathbf{A}\mathbf{i} = (e_0 \mathbf{A}_0 + e_1 \mathbf{A}_1 + e_2 \mathbf{A}_2)(e_0 \mathbf{i}_0 + e_1 \mathbf{i}_1 + e_2 \mathbf{i}_2), \quad (2)$$

$$\begin{aligned} \mathbf{o} = & e_0 (\mathbf{A}_0 \mathbf{i}_0 + \mathbf{A}_2 \mathbf{i}_1 + \mathbf{A}_1 \mathbf{i}_2) + e_1 (\mathbf{A}_1 \mathbf{i}_0 + \mathbf{A}_0 \mathbf{i}_1 + \mathbf{A}_2 \mathbf{i}_2) \\ & + e_2 (\mathbf{A}_2 \mathbf{i}_0 + \mathbf{A}_1 \mathbf{i}_1 + \mathbf{A}_0 \mathbf{i}_2), \end{aligned} \quad (3)$$

i.e., we have

$$\mathbf{o}_r = \begin{bmatrix} o_0 \\ o_1 \\ o_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_2 & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 & \mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 \end{bmatrix} \begin{bmatrix} \mathbf{i}_0 \\ \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix} = \mathbf{A}_r \mathbf{i}_r, \quad (4)$$

with a nonnegative real-valued matrix \mathbf{A}_r

$$\mathbf{A}_r = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_2 & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 & \mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 \end{bmatrix}. \quad (5)$$

For $E = 4$, where $e_k e_l = e_{(k+l \bmod 4)}$ and $\mathbf{A} = e_0 \mathbf{A}_0 + e_1 \mathbf{A}_1 + e_2 \mathbf{A}_2 + e_3 \mathbf{A}_3$, a similar analysis yields

$$\mathbf{A}_r = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_2 & \mathbf{A}_3 & \mathbf{A}_1 \\ \mathbf{A}_2 & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_3 \\ \mathbf{A}_1 & \mathbf{A}_3 & \mathbf{A}_0 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_0 \end{bmatrix}. \quad (6)$$

Based on these equations, one can conclude that to all-optically implement an arbitrary complex-valued transformation $\mathbf{o} = \mathbf{A}\mathbf{i}$ using a spatially incoherent D²NN, the layers of the D²NN need to be optimized to perform an intensity linear

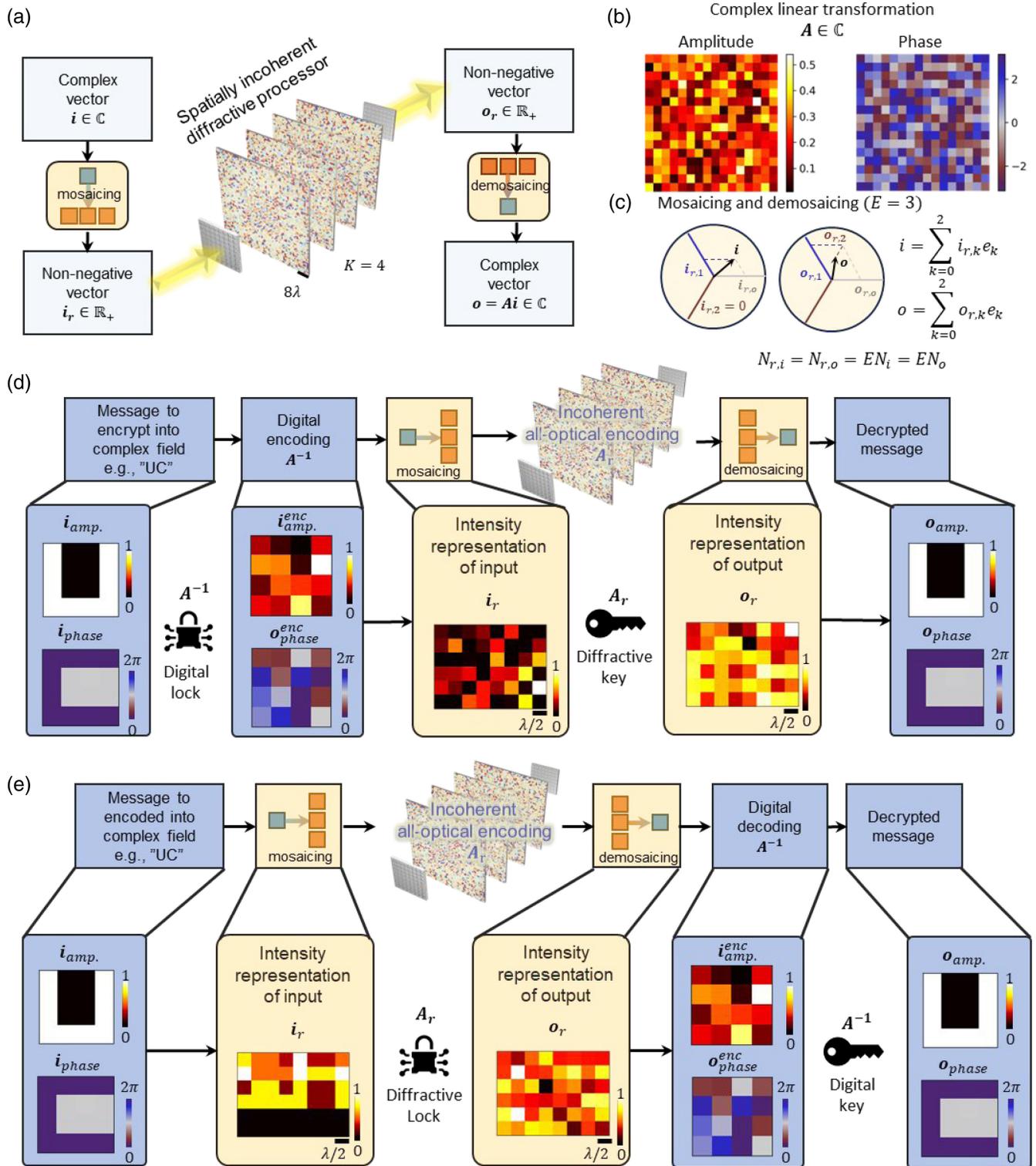


Fig. 1 (a) Complex-valued universal linear transformations using spatially incoherent diffractive optical networks. (b) Amplitude and phase of the target complex-valued linear transformation. (c) Mosaicking and demosaicking processes. (d)–(e) Image encryption. (d) Complex-valued images are digitally encrypted (A^{-1}), and subsequently decrypted using the diffractive system that performs A (diffractive key). (e) The encryption is performed through the spatially incoherent diffractive network (diffractive lock), and the decryption is performed digitally (digital key).

transformation $\mathbf{A}_r \in \mathbb{R}_+^{E^2 N_i N_o}$ such that $\mathbf{o}_r = \mathbf{A}_r \mathbf{i}_r$. The entire system, upon convergence, performs the predefined complex-valued linear transformation \mathbf{A} on any given input data using spatially incoherent light, based on Eqs. (2) and (4). In the following sections, we numerically explore the number of optimizable diffractive features (N) needed for accurate approximation of \mathbf{A} using a spatially incoherent D²NN.

2.1 Complex-Valued Linear Transformations through Spatially Incoherent Diffractive Networks

We numerically demonstrated the capabilities of diffractive optical processors to universally perform any arbitrarily chosen complex-valued linear transformation with spatially incoherent light. Throughout the paper, we used $N_i = N_o = 16$. To visually represent the data, we rearranged the 16-element vectors into 4×4 arrays of complex numbers, hereafter referred to as the “complex image.” We arbitrarily selected a desired complex-valued transformation, $\mathbf{A} \in \mathbb{C}^{16 \times 16}$, as shown in Fig. 1(b).

To explore the number of diffractive features needed, we trained nine models with varying values of N and evaluated the mean-squared-error (MSE) between the numerically measured ($\hat{\mathbf{A}}_r$) and the target all-optical linear transformation, \mathbf{A}_r (see Fig. 2). Our results, summarized in Fig. 2, highlight that with a sufficient number of optimizable diffractive features, i.e., $N \geq 2E^2 N_i N_o = 2N_{i,r} N_{o,r}$, our system achieves a negligible approximation error with respect to the target $\mathbf{A}_r \in \mathbb{R}_+^{48 \times 48}$. In Fig. 2(c), we also visualize the resulting all-optical intensity transformation $\hat{\mathbf{A}}_r$ compared to the ground truth \mathbf{A}_r . In essence, this comparison reveals the spatially varying incoherent point spread functions (PSFs) of our diffractive system optimized using deep learning; a negligible MSE between $\hat{\mathbf{A}}_r$ and \mathbf{A}_r shows that the resulting spatially varying incoherent PSFs match the target set of PSFs dictated by \mathbf{A}_r .

We also evaluated the numerical accuracy of our complex-valued transformation in an end-to-end manner, as illustrated in Fig. 2(d). For this numerical test, we sequentially set each entry of \mathbf{i} to e_0 , evaluated the corresponding complex output $\hat{\mathbf{o}}$, and stacked them to form $\hat{\mathbf{A}}_0$, where the subscript represents that the measurement was evaluated using the complex impulse along the basis e_0 as input. Then, we repeated this process for the other two bases to obtain $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_2$, and stacked these matrices as a block matrix $[\hat{\mathbf{A}}_0 | \hat{\mathbf{A}}_1 | \hat{\mathbf{A}}_2]$, shown in Fig. 2(d). Each row of the images $\text{amp}(\hat{\mathbf{o}})$ and $\text{phase}(\hat{\mathbf{o}})$ in Fig. 2(d) represents one of these complex output vectors, while the corresponding target vectors are presented in the same figure through $\text{amp}(\mathbf{o})$ and $\text{phase}(\mathbf{o})$. The small magnitude of the error $\boldsymbol{\varepsilon} = |\hat{\mathbf{o}} - \mathbf{o}|^2$ shown in Fig. 2(d) illustrates the success of this spatially incoherent D²NN model in accurately approximating the complex-valued linear transformation $\mathbf{o} = \mathbf{A}\mathbf{i}$, implemented for an arbitrarily selected \mathbf{A} .

2.2 Complex Number-based Image Encryption Using Spatially Incoherent Diffractive Networks

In this section, we demonstrate a complex number-based image encryption–decryption scheme using a spatially incoherent D²NN. In the first scheme, shown in Fig. 1(d), the message is encoded into a complex image, employing either amplitude and phase encoding or real and imaginary part encoding. Then, a digital lock encrypts the image by applying a linear transformation (\mathbf{A}^{-1}) to conceal the original message within the image.

At the optical receiver, the encrypted message is deciphered by an optimized incoherent D²NN that all-optically implements the inverse transformation \mathbf{A} . In an alternative scheme, as depicted in Fig. 1(e), the key and lock are switched, i.e., the spatially incoherent D²NN is used to encrypt the message with a complex-valued \mathbf{A} while the decryption step involves the digital inversion using \mathbf{A}^{-1} .

For our analysis, we used the letters “U,” “C,” “L,” and “A” as sample messages. “U” and “C” are used in amplitude-phase-based encoding (Fig. 3), whereas “L” and “A” are used for real-imaginary-based encoding of information (Fig. S1 in the [Supplementary Material](#)), forming complex-number-based images. To accurately model the spatially incoherent propagation²⁶ of light through the D²NN, we averaged the output intensities over a large number of $N_\varphi = 20,000$ of randomly generated 2D phase profiles at the input (see Sec. 4 for details).

In Fig. 3(a), we show the results corresponding to digital encryption and optical diffractive decryption, i.e., the system shown in Fig. 1(d). The digitally encrypted complex information $\mathbf{i} = \mathbf{A}^{-1}\mathbf{o}$, and its intensity representation \mathbf{i}_r , are shown in Fig. 3(a). The optically decrypted output $\hat{\mathbf{o}}$ (through the spatially incoherent D²NN) and its intensity-based representation $\hat{\mathbf{o}}_r$ are shown in the same Fig. 3(a), together with the resulting error maps, i.e., $|\hat{\mathbf{o}} - \mathbf{o}|^2$ and $|\hat{\mathbf{o}}_r - \mathbf{o}_r|^2$, which reveal a very small degree of error. This agreement of the recovered and the ground-truth messages in both the intensity and complex-valued domains confirms the accuracy of the diffractive decryption process through an optimized spatially incoherent D²NN. Figure 3(b) shows the successful performance of the sister scheme [Fig. 1(e)], which involves diffractive encryption through a spatially incoherent D²NN and digital decryption, also revealing a negligible amount of error in both $|\mathbf{A}^{-1}\hat{\mathbf{o}} - \mathbf{i}|^2$ and $|\hat{\mathbf{o}}_r - \mathbf{o}_r|^2$. As reported in Fig. S1 in the [Supplementary Material](#), we also conducted a numerical experiment using the letters “L” and “A,” encoded using the real and imaginary parts of the message. The visualizations are arranged the same way as in Fig. 3, where for both schemes depicted in Figs. 1(d) and 1(e), the degree of error between the recovered and the original messages is negligible, affirming the success of using the real and imaginary part-based encoding method. For the assessment of the approximation errors when the number of diffractive features is smaller, we compared the decryption performance of three models with different numbers of diffractive features/neurons, i.e., $N = (0.5, 0.7, 2) \times 2E^2 N_i N_o$, for the same setup outlined in Fig. S1(a) in the [Supplementary Material](#). The results are summarized in Fig. S2 in the [Supplementary Material](#): for models with $N < 2E^2 N_i N_o$, the decryption quality is compromised, exhibiting a pixel absolute error of >0.1 . However, this error reduces to <0.05 for $N = 4E^2 N_i N_o$ where the decrypted images display significantly enhanced contrast and reduced noise levels.

To further evaluate the efficacy of our encryption method, we analyzed the complex image entropy, examining both the real and imaginary components separately (refer to the Sec. 4 for details). The original image \mathbf{i} , D²NN encrypted output $\hat{\mathbf{o}}$, and the digitally encrypted output $\mathbf{A}\mathbf{i}$, along with the corresponding image entropies, are shown in Fig. S3(a) in the [Supplementary Material](#) for two complex image examples. We repeated this analysis for a set of 1000 complex images with the resulting entropy distributions reported in Fig. S3(b) in the [Supplementary Material](#). These results demonstrate that the entropy of the encrypted images is statistically higher than that of the original

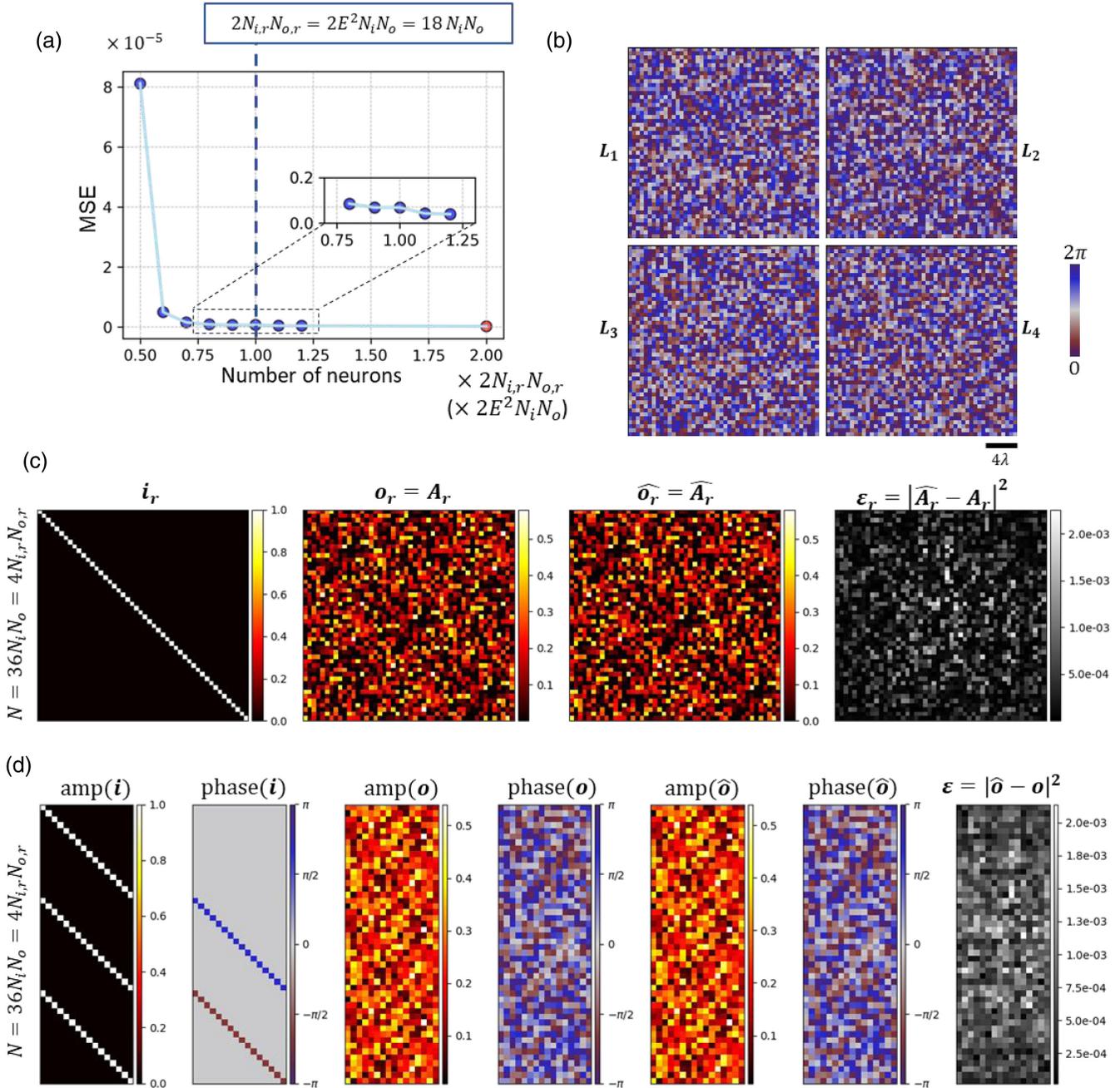


Fig. 2 Performance of spatially incoherent diffractive networks on arbitrary complex-valued linear transformations. (a) The all-optical linear transformation error as a function of the number of diffractive features (N). The red dot represents the design corresponding to the results shown in (b)–(d). (b) The phase profiles of the $K = 4$ diffractive layers of the optimized model ($N = 2 \times 2N_{i,r}N_{o,r}$). (c) Evaluation of the resulting all-optical intensity transformation, i.e., the spatially varying PSFs. (d) The complex linear transformation evaluation. For ϵ_r and ϵ , $|\cdot|^2$ represents an element-wise operation.

images. This increase in entropy signifies a heightened level of randomness within the encrypted images, thereby validating the effectiveness of our encryption process. In addition, the entropy distributions of the D²NN encrypted images show excellent agreement with the digitally encrypted corresponding images, further demonstrating the success of our spatially incoherent optical encryption scheme.

2.3 Different Mosaicking and Demosaicking Schemes in a Spatially Incoherent D²NN

How we assign each element in the vector i_r and o_r to the pixels at the input and output FOVs of the diffractive network does not affect the final accuracy of the image/message reconstruction. For example, we can arrange the FOVs in such a manner that

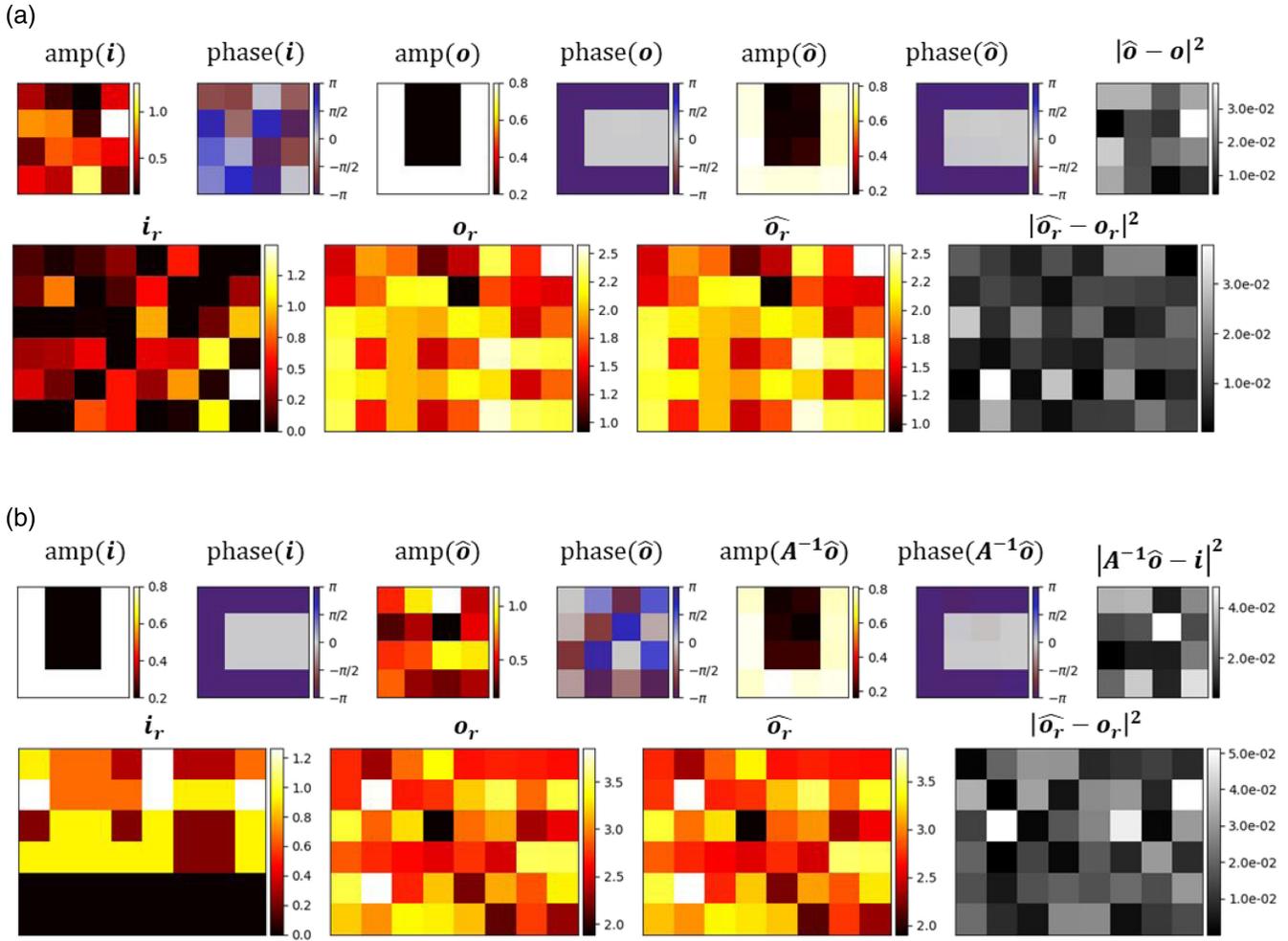


Fig. 3 Image encryption with the letters “U” and “C” encoded into amplitude and phase, respectively, of the complex-valued image. (a) The input, target, output, and the approximation error, both in complex and real nonnegative (intensity) domains. The original information is represented by \mathbf{o} , while \mathbf{i} is obtained by digital encrypting \mathbf{o} following Fig. 1(d). (b) The input, output (resulting from optical encryption), and digitally decrypted output and the error between the input and the decrypted output. The result of digital decryption matches the input information. The second row shows the corresponding input, target, and output intensities and the approximation error. $|\cdot|^2$ represents an element-wise operation.

the components $i_{r,k}$ corresponding to a basis e_k are assigned to the neighboring pixels, in two adjacent rows, as shown in Fig. S4(a) in the [Supplementary Material](#); in an alternative implementation, the assignment/mapping can be completely arbitrary, which is equivalent to applying a random permutation operation on the input and output vectors (see Sec. 4). When compared to each other, these two approaches of mosaicking and demosaicking schemes show negligible differences in the error of the final reconstruction of the letters “U,” “C,” “L,” and “A” as shown in Fig. S4(b) in the [Supplementary Material](#). These results underscore that the specific arrangement of the mosaicking/demosaicking schemes at the input and output FOVs does not impact the performance of the incoherent D²NN system.

3 Discussion and Conclusion

In this article, we employed a data-free PSF-based D²NN optimization method (see Sec. 4),²⁶ since we can determine

the nonnegative intensity transformation A_r from the target complex-valued transformation A based on the mosaicking and demosaicking schemes; the columns of A_r represent the desired spatially varying PSFs of the D²NN. The advantage of this data-free learning-based D²NN optimization approach is that computationally demanding simulation of wave propagation with large N_φ is not required during the training. Coherent propagation is appropriate for simulating the spatially varying PSFs, point by point, since a point emitter at the input aperture coherently interferes with itself during optical diffraction within a D²NN; this approach makes the training time much shorter. On the other hand, this approach necessitates prior knowledge of A_r , which might not always be available, e.g., for tasks such as data classification. An alternative to this data-free PSF-based optimization approach is to train the diffractive network in an end-to-end manner, using a data-driven direct training approach.²⁶ This strategy advances by minimizing the differences between the outputs and the targets on a large number of randomly

generated examples, thereby learning the spatially varying PSFs *implicitly* from numerous input-target intensity patterns corresponding to the desired task – instead of learning from an explicitly predetermined A_r . This direct approach, however, requires a longer training time, necessitating the simulation of incoherent propagation for each training sample on a large data set.

In our presented approach, the choice of E is not restricted to $E = 3$, as we have used throughout the main text. As another example of encoding, we show the image encryption results with $E = 4$ in Fig. S5 in the [Supplementary Material](#), where the four bases are $\exp(j\frac{\pi}{2}k)$ ($k = 0, 1, 2, 3$). The reconstructed “U,” “C,” “L,” and “A” letters are also reported in the same figure, confirming that given sufficient degrees of freedom (with $N \geq 2E^2N_iN_o$), the linear transformation performances are similar to each other. However, compared to $E = 3$, this choice of $E = 4$ necessitates 4/3 times more pixels on both the diffractive network input and output FOVs—reducing the throughput (or spatial density) of complex-valued linear transformations that can be performed using a spatially incoherent D²NN. Accordingly, more diffractive features and a larger number of independent degrees of freedom (by 16/9-fold) are required within the D²NN volume to achieve an output performance level that is comparable to a design with $E = 3$. Note that while $E \geq 3$ is sufficient to reconstruct the original complex-valued images regardless of the image complexity, the redundancy provided by larger E values might offer increased resilience against noise at the cost of reducing the image-processing throughput (per input aperture area) with larger E .

Our framework offers several flexibilities in implementation, which could be useful for different applications. First, the flexibility to arbitrarily permute the input and the output pixels following different mosaicking and demosaicking schemes (as introduced earlier in Sec. 2) could enhance the security of optical information transmission. A user would not be able to either spam or hack valuable information that is transferred optically without specific knowledge of the mosaicking *and* demosaicking schemes, thus ensuring the security of this scheme. Note that this enhancement in security is achieved without adding complexity to the system by just permuting the assignment of data elements to the pixels of the input and output devices, e.g., spatial light modulators (SLMs) and complementary metal-oxide-semiconductor (CMOS) detector-arrays. Second, the flexibility in choosing E , as discussed above, could be useful in adding an extra layer of security against unauthorized access, albeit with a trade-off in system throughput that comes with larger E . Furthermore, we can use different sets of bases for mosaicking and demosaicking operations by applying offset phase angles θ_i and θ_o , respectively, to the original bases $e_k = \exp(jk\frac{2\pi}{E})$, $k = 0, \dots, E - 1$. This will result in a set of modified/encrypted bases: $e_{k,i} = \exp[j(k\frac{2\pi}{E} + \theta_i)]$ for mosaicking and $e_{k,o} = \exp[j(k\frac{2\pi}{E} + \theta_o)]$ for demosaicking. This powerful flexibility in representation further enhances the security of the system.

Regarding image encryption-related applications, we demonstrated two approaches [Figs. 1(d) and 1(e)] to utilize D²NNs for encryption or decryption. However, it is also possible to deploy a pair of diffractive systems in tandem, with one undertaking the matrix operation A for encryption and the other undertaking the inverse operation A^{-1} for decryption. Furthermore, potential extensions of our work could explore a harmonized integration of

polarization state controls³² and wavelength multiplexing³³ to build a multifaceted, fortified encryption platform. In addition to increasing the data throughput, these additional degrees of freedom enabled by different illumination wavelengths and polarization states would further enhance the security of a diffractive processor-based system.

In this work, we focused on the numerical analysis of the presented concept. However, we should note that various D²NNs designed using deep-learning-based approaches have been experimentally validated over different parts of the electromagnetic spectrum, e.g., from terahertz (THz)^{9,14} to near-infrared (NIR)¹⁵ and visible wavelengths,²⁴ showing a good agreement between the numerical and experimental results. To address some of the experimental challenges associated with fabrication errors and mechanical misalignments, a “vaccination” strategy^{34,35} has been devised. This approach enhances the robustness of the diffractive optical designs by incorporating such aberrations/imperfections as random variables during the training phase, thereby preparing the system to better withstand and adapt to the uncertainties inherent in real-world experimental conditions.

Although spatially coherent light is more suitable for complex-valued information processing in laboratory settings, the use of spatially incoherent light offers various practical advantages. For example, speckle noise, which is inevitable in coherent systems, can be suppressed by using partially or fully incoherent illumination. An additional benefit of spatially incoherent designs is the range of viable illumination sources that can be used: instead of using specialized coherent sources, a spatially incoherent system can work with standard light-emitting diodes (LEDs), or even under natural light, which is important for some applications of diffractive information processing.

To conclude, we demonstrated the capability of spatially incoherent diffractive networks to perform arbitrary complex-valued linear transformations. By incorporating various forms of mosaicking and demosaicking operations, we paved the way for a wider array of applications by leveraging incoherent D²NNs for complex-valued data processing. We also showcased potential applications of these spatially incoherent D²NNs for complex number-based image encryption or decryption, highlighting the security benefits arising from the system’s flexibility. Our exploration marks a significant stride toward enhanced versatility and robustness in optical information processing with spatially incoherent diffractive systems that can work under natural light.

4 Appendix: Methods

4.1 Linear Transformation Matrix

In this paper, we use $N_i = N_o = 16$ so that $A \in \mathbb{C}^{16 \times 16}$; see Fig. 1(b). To generate A , we randomly sample the amplitude of each element from the uniform distribution $Uniform(0, 1)$ and the phases from $Uniform(0, 2\pi)$. For the encryption application, to ensure that the result of inversion is not sensitive to small errors, we performed QR-factorization on A to obtain a condition number of one.³⁶

4.2 Real-Valued Nonnegative Representation of Complex Numbers

Following Eq. (4), the complex-valued input and target vectors $\mathbf{i} \in \mathbb{C}^{N_i}$ and $\mathbf{o} \in \mathbb{C}^{N_o}$ are represented by the corresponding real and nonnegative intensity vectors $\mathbf{i}_r = [i_0^r \ \dots \ i_{E-1}^r]^T \in \mathbb{R}_+^{EN_i}$

and $\mathbf{o}_r = [\mathbf{o}_0^T \cdots \mathbf{o}_{E-1}^T]^T \in \mathbb{R}_+^{EN_o}$, where $\mathbf{i} = \sum_{k=0}^{E-1} e_k \mathbf{i}_k$ and $\mathbf{o} = \sum_{k=0}^{E-1} e_k \mathbf{o}_k$. The desired all-optical intensity transformation \mathbf{A}_r between \mathbf{i}_r and \mathbf{o}_r is derived from the target complex-valued linear transformation \mathbf{A} following Eqs. (1) and (5). We should note that deriving \mathbf{A}_r from \mathbf{A} requires mapping each complex element a to its real and nonnegative representation (a_0, \dots, a_{E-1}) based on the $E \geq 3$ complex bases e_k such that $a = \sum_{k=0}^{E-1} e_k a_k$. To define a unique mapping, we follow an algorithm²⁹ by imposing additional constraints: $a_k = 0$ if $\frac{2\pi}{E} \leq \text{phase}(e_k a^*) \leq 2\pi - \frac{2\pi}{E}$, i.e., $a_k = 0$ if the angle between a and e_k is greater than $\frac{2\pi}{E}$; here a^* represents the complex conjugate of a . The same constraints were also used while mapping the complex input vectors \mathbf{i} to the real and nonnegative intensity vectors \mathbf{i}_r .

4.3 Mosaicking and Demosaicking Schemes

For mosaicking (demosaicking) assignment of each element of \mathbf{i}_r (\mathbf{o}_r) to one of the $N_{i,r} = EN_i$ ($N_{o,r} = EN_o$) pixels of the 2D input (output), the arrangement of the FOV can be *regular*, e.g., in a row-major order as shown in Fig. S4(a) in the [Supplementary Material](#), “Regular mosaicking.” Alternatively, the pixel assignment on the input (output) FOV can follow any arbitrary mapping which can be defined by a permutation matrix \mathbf{P}_i (\mathbf{P}_o) operating on the input (output) vector; see Fig. S4(a) in the [Supplementary Material](#), “Arbitrary mosaicking.” For such cases, when ordered in a row-major format, intensities on the input (output) FOVs \mathbf{i}_r (\mathbf{o}_r) can be written as $\mathbf{i}_r = \mathbf{P}_i [\mathbf{i}_0^T \cdots \mathbf{i}_{E-1}^T]^T$ ($\mathbf{o}_r = \mathbf{P}_o [\mathbf{o}_0^T \cdots \mathbf{o}_{E-1}^T]^T$). Accordingly, such an arbitrary arrangement of pixels was accounted for by redefining the all-optical intensity transformation as $\mathbf{P}_o \mathbf{A}_r \mathbf{P}_i^T$.

4.4 Spatially Incoherent Light Propagation through a D²NN

The 1D vector \mathbf{i}_r is rearranged as a 2D distribution of intensity $I(x, y)$ at the input FOV of the D²NN. To numerically model the spatially incoherent propagation of the input intensity distribution $I(x, y)$ through the D²NN, we coherently propagated the optical field $\sqrt{I} \exp(j\varphi)$ through the trainable diffractive surfaces to the output plane, where φ is a random 2D phase distribution, i.e., $\varphi(x, y) \sim \text{Uniform}(0, 2\pi)$ for all (x, y) . If we denote the coherent field propagation operator as $\mathfrak{D}\{\cdot\}$ (see Sec. 4.5), then the instantaneous output intensity is $|\mathfrak{D}\{\sqrt{I(x, y)} \exp[j\varphi(x, y)]\}|^2$ and the time-averaged output intensity $O(x, y)$ for spatially incoherent light can be written as

$$O(x, y) = \langle |\mathfrak{D}\{\sqrt{I(x, y)} \exp[j\varphi(x, y)]\}|^2 \rangle. \quad (7)$$

The average output intensity can be approximately calculated by repeating the coherent wave propagation $\mathfrak{D}\{\cdot\}$ N_φ -times, each time with a different random phase distribution $\varphi_r(x, y)$, and averaging the resulting N_φ output intensities,

$$O(x, y) = \lim_{N_\varphi \rightarrow \infty} \frac{1}{N_\varphi} \sum_{r=1}^{N_\varphi} |\mathfrak{D}\{\sqrt{I(x, y)} \exp[j\varphi_r(x, y)]\}|^2. \quad (8)$$

We used $N_\varphi = 20,000$ for estimating the incoherent output intensity $O(x, y)$ corresponding to any arbitrary input intensity $I(x, y)$. Note that when only one pixel at the input aperture is

activated, with all other input pixels being inactive with zero intensity, as is the case while evaluating spatially varying PSFs, the application of Eq. (8) becomes redundant, although one could still use it. In this scenario, all the light diffracted from a single point source is mutually coherent. Consequently, for the purposes of evaluating the spatially varying PSFs of the system, as elaborated later in Sec. 4.7, employing a coherent propagation model for each point emitter at the input aperture is accurate and provides a faster solution.

4.5 Coherent Propagation of Optical Fields: $\mathfrak{D}\{\cdot\}$

The propagation of spatially coherent light patterns through a diffractive processor, denoted by $\mathfrak{D}\{\cdot\}$, involves a series of interactions with consecutive diffractive surfaces, interleaved by wave propagation through the free space separating these surfaces. We assume that these modulations are introduced by phase-only diffractive surfaces, i.e., the field amplitude remains unchanged during the light-matter interaction. Specifically, we assume that a diffractive surface alters the incident optical field, symbolized as $u(x, y)$, in a localized manner according to the optimized phase values $\phi_M(x, y)$ of the diffractive features, resulting in the phase-modulated field $u(x, y) \exp[j\phi_M(x, y)]$. The diffractive surfaces are coupled by free-space propagation, allowing the light to travel from one surface to the next. We used the angular spectrum method to simulate the free-space propagation,³⁷

$$u(x, y; z = z_0 + d) = \mathcal{F}^{-1} \{ \mathcal{F} \{ u(x, y; z = z_0) \} \times H(f_x, f_y; d) \}, \quad (9)$$

where $\mathcal{F}\{\cdot\}$ is the 2D Fourier transform and $\mathcal{F}^{-1}\{\cdot\}$ is its inverse operation. $H(f_x, f_y; d)$ is the free-space transfer function corresponding to propagation distance d . For wavelength λ ,

$$H(f_x, f_y; d) = \begin{cases} \exp \left[j \frac{2\pi}{\lambda} d \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \right], & f_x^2 + f_y^2 < 1/\lambda^2 \\ 0, & \text{otherwise} \end{cases}. \quad (10)$$

The fields were discretized with a lateral sampling interval of $\delta \approx 0.53\lambda$ to accommodate all the propagating modes and sufficiently zero-padded to remove aliasing artifacts.³⁸

4.6 Diffractive Network Architecture

We modeled the diffractive surfaces by their laterally discretized heights h , which correspond to phase delays $\varphi_M = \frac{2\pi}{\lambda} (n-1)h$, where n is the refractive index of the material. The connectivity between consecutive diffractive layers⁹ was kept equal across the diffractive designs with varying N by setting the separation between the layers as $d = \frac{W\delta}{\lambda}$, where the width of each diffractive layer is $W = \sqrt{\frac{N}{K}}\delta$. Here, K is the number of diffractive layers; we used $K = 4$ throughout the paper.

4.7 Training and Evaluation of Spatially Incoherent Diffractive Processors

For performing an arbitrary complex-valued linear transformation with a diffractive processor, we used the PSF-based

data-free design approach, where the diffractive features were optimized so that the all-optical intensity transformation of the diffractive processor achieves $\hat{\mathbf{A}}_r \approx \mathbf{A}_r$. To evaluate $\hat{\mathbf{A}}_r$, we used EN_i intensity vectors $\{\mathbf{i}_{r,t}\}_{t=1}^{EN_i}$, where $\mathbf{i}_{r,t}[l] = 1$ if $l = t$ and 0 otherwise. In other words, $\{\mathbf{i}_{r,t}\}_{t=1}^{EN_i}$ are unit impulses, located at different input pixels. We simulated the all-optical output intensity vectors $\{\mathbf{o}'_{r,t}\}_{t=1}^{EN_i}$ corresponding to these unit impulses and stacked them, i.e.,

$$\hat{\mathbf{A}}'_r = [\mathbf{o}'_{r,1} | \mathbf{o}'_{r,2} | \cdots | \mathbf{o}'_{r,EN_i}]. \quad (11)$$

Finally, we compensated for the optical diffraction efficiency-related scale mismatch through multiplication by a scalar, i.e.,

$$\hat{\mathbf{A}}_r = \sigma \hat{\mathbf{A}}'_r, \quad (12)$$

where σ was defined as

$$\sigma = \frac{\sum_{n=1}^{EN_i} \sum_{m=1}^{EN_o} \mathbf{A}_r[m, n] \hat{\mathbf{A}}'_r[m, n]}{\sum_{n=1}^{EN_i} \sum_{m=1}^{EN_o} (\hat{\mathbf{A}}'_r[m, n])^2}. \quad (13)$$

The MSE loss function to be minimized was defined as

$$\mathcal{L}_{\text{PSF}} = \frac{1}{N_i N_o} \sum_{n=1}^{EN_i} \sum_{m=1}^{EN_o} (\mathbf{A}_r[m, n] - \hat{\mathbf{A}}_r[m, n])^2. \quad (14)$$

The height h of the diffractive features at each layer was constrained between zero and a maximum value h_{max} by employing a latent variable h_{latent} . The relationship between the constrained height h and the latent variable h_{latent} was defined as $h = \frac{h_{\text{max}}}{2} \times [\sin(h_{\text{latent}}) + 1]$, where we chose $h_{\text{max}} \approx \frac{\lambda}{n-1}$, which corresponds to a differential phase modulation of 2π . The latent variables were initialized randomly from the standard normal distribution $\mathcal{N}(0,1)$.

The optimization of the diffractive layers was carried out using the AdamW optimizer³⁹ for 12,000 iterations, with an initial learning rate of 10^{-3} . The model state corresponding to the minimum of the MSEs evaluated after every 400 iterations was selected for the final evaluation. The D²NN models were implemented and trained using PyTorch (v1.12.1)⁴⁰ with Compute Unified Device Architecture (CUDA) version 12.2. Training and testing were done on GeForce RTX 3090 graphics processing units (GPUs) in workstations with 256 GB of random-access memory (RAM) and Intel Core i9 central processing unit (CPU). The training time of the models varied with the size of the models. For example, the model used in Figs. 2(b) and 2(c) took around 1 h for 12,000 iterations. Inference for each input vector with $N_\varphi = 20,000$ takes around 30 s.

To visualize the all-optical transformation error in Fig. 2, we used the error matrix $\boldsymbol{\varepsilon}_r = (\mathbf{A}_r - \hat{\mathbf{A}}_r)^2$; here $(\cdot)^2$ denotes an element-wise operation. To evaluate the error $\boldsymbol{\varepsilon}$ at complex linear transformation, we applied demosaïcking to the columns of $\hat{\mathbf{A}}_r$ to form the block matrix $[\hat{\mathbf{A}}_0 | \cdots | \hat{\mathbf{A}}_{E-1}] \in \mathbb{C}^{N_o \times EN_i}$. Here, the subscript k represents that $\hat{\mathbf{A}}_k$ is measured by applying the columns of $\mathbf{e}_k \mathbf{I}$ as input and stacking the corresponding demosaïcked (complex-valued) output vectors. Accordingly, we have

$$\boldsymbol{\varepsilon} = |[e_0 \mathbf{A} - \hat{\mathbf{A}}_0 | \cdots | e_{E-1} \mathbf{A} - \hat{\mathbf{A}}_{E-1}]|^2. \quad (15)$$

Here, $|\cdot|^2$ represents an element-wise operation.

4.8 Entropy Evaluation

For the evaluation of the image encryption strength, we computed the entropy separately for the real and imaginary parts of a complex image as follows:

$$H^{\text{Re/Im}}(\mathbf{x}) = -\sum_i p_i^{\text{Re/Im}} \cdot \log(p_i^{\text{Re/Im}}), \quad (16)$$

where we calculated the distribution of either the real or imaginary part (denoted by the superscript) over the pixels of \mathbf{x} ; here \mathbf{x} denotes the complex image. p_i is the probability (normalized histogram count) for a certain pixel value i .

For the histograms presented in Fig. S3(b) in the [Supplementary Material](#), the data set is adapted from the Extended MNIST (EMNIST).⁴¹ For the creation of the input complex images, we randomly selected two distinct images from the EMNIST data set, using one as the real part and the other as the imaginary part of the complex image. To ensure compatibility with the input dimensionality, these images were bilinearly downsampled to a resolution of 4×4 pixels. We randomly formed a set of 1000 such complex images to compile the histograms presented in Fig. S3(b) in the [Supplementary Material](#).

Code and Data Availability

The data and methods required to assess the conclusions drawn in this study are included within the main text and supplementary information files. The optimization of machine-learning models used in this research was conducted using the publicly available PyTorch library. Additional data can be requested from the corresponding author.

Acknowledgments

The Ozcan Research Group at UCLA acknowledges the support of the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under Award # DE-SC0023088.

References

1. J. W. Goodman, A. R. Dias, and L. M. Woody, "Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms," *Opt. Lett.* **2**(1), 1–3 (1978).
2. H. Kwon et al., "Nonlocal metasurfaces for optical signal processing," *Phys. Rev. Lett.* **121**(17), 173004 (2018).
3. R. Hamerly et al., "Large-scale optical neural networks based on photoelectric multiplication," *Phys. Rev. X* **9**(2), 021032 (2019).
4. A. Silva et al., "Performing mathematical operations with metamaterials," *Science* **343**(6167), 160–163 (2014).
5. B. J. Shastri et al., "Photonics for artificial intelligence and neuromorphic computing," *Nat. Photonics* **15**(2), 102–114 (2021).
6. G. Wetzstein et al., "Inference in artificial intelligence with deep optics and photonics," *Nature* **588**(7836), 39–47 (2020).
7. X. Xu et al., "11 TOPS photonic convolutional accelerator for optical neural networks," *Nature* **589**(7840), 44–51 (2021).
8. S. M. Kamali et al., "Angle-multiplexed metasurfaces: encoding independent wavefronts in a single metasurface under different illumination angles," *Phys. Rev. X* **7**(4), 041056 (2017).
9. X. Lin et al., "All-optical machine learning using diffractive deep neural networks," *Science* **361**(6406), 1004–1008 (2018).
10. O. Kulce et al., "All-optical synthesis of an arbitrary linear transformation using diffractive surfaces," *Light Sci. Appl.* **10**(1), 196 (2021).

11. M. S. S. Rahman et al., "Ensemble learning of diffractive optical networks," *Light Sci. Appl.* **10**(1), 14 (2021).
12. J. Li et al., "Class-specific differential detection in diffractive optical neural networks improves inference accuracy," *Adv. Photonics* **1**(4), 046001 (2019).
13. M. S. S. Rahman and A. Ozcan, "Time-lapse image classification using a diffractive neural network," arXiv:2208.10802 (2022).
14. B. Bai et al., "To image, or not to image: class-specific diffractive cameras with all-optical erasure of undesired objects," *eLight* **2**(1), 14 (2022).
15. B. Bai et al., "Data-class-specific all-optical transformations and encryption," *Adv. Mater.* **35**(31), 2212091 (2023).
16. Y. Gao et al., "Multiple-image encryption and hiding with an optical diffractive neural network," *Opt. Commun.* **463**, 125476 (2020).
17. Y. Su et al., "Optical image conversion and encryption based on structured light illumination and a diffractive neural network," *Appl. Opt.* **62**(23), 6131–6139 (2023).
18. D. Mengü and A. Ozcan, "All-optical phase recovery: diffractive computing for quantitative phase imaging," *Adv. Opt. Mater.* **10**(15), 2200281 (2022).
19. C.-Y. Shen et al., "Multispectral quantitative phase imaging using a diffractive optical network," *Adv. Intell. Syst.* **5**(11), 2300300 (2023).
20. T. Yan et al., "Fourier-space diffractive deep neural network," *Phys. Rev. Lett.* **123**(2), 023901 (2019).
21. B. Bai et al., "Pyramid diffractive optical networks for unidirectional magnification and demagnification," arXiv:2308.15019 (2023).
22. E. Goi, S. Schoenhardt, and M. Gu, "Direct retrieval of Zernike-based pupil functions using integrated diffractive deep neural networks," *Nat. Commun.* **13**(1), 7531 (2022).
23. Z. Huang et al., "All-optical signal processing of vortex beams with diffractive deep neural networks," *Phys. Rev. Appl.* **15**(1), 014037 (2021).
24. X. Luo et al., "Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible," *Light Sci. Appl.* **11**(1), 158 (2022).
25. O. Kulce et al., "All-optical information-processing capacity of diffractive surfaces," *Light Sci. Appl.* **10**(1), 25 (2021).
26. M. S. S. Rahman et al., "Universal linear intensity transformations using spatially incoherent diffractive processors," *Light Sci. Appl.* **12**(1), 195 (2023).
27. B. H. Soffer et al., "Programmable real-time incoherent matrix multiplier for optical processing," *Appl. Opt.* **25**(14), 2295–2305 (1986).
28. W. Swindell, "A noncoherent optical analog image processor," *Appl. Opt.* **9**(11), 2459–2469 (1970).
29. J. W. Goodman and L. M. Woody, "Method for performing complex-valued linear operations on complex-valued data using incoherent light," *Appl. Opt.* **16**(10), 2611–2612 (1977).
30. W. Schneider and W. Fink, "Incoherent optical matrix multiplication," *Opt. Acta Int. J. Opt.* **22**(11), 879–889 (1975).
31. A. R. Dias, "Incoherent optical matrix-matrix multiplier," NASA. Langley Research Center Opt. Inform. Process. for Aerospace Appl. (1981).
32. Y. Li et al., "Universal polarization transformations: spatial programming of polarization scattering matrices using a deep learning-designed diffractive polarization transformer," *Adv. Mater.* **35**(51), 2303395 (2023).
33. J. Li et al., "Massively parallel universal linear transformations using a wavelength-multiplexed diffractive optical network," *Adv. Photonics* **5**(1), 016003 (2023).
34. D. Mengü et al., "Misalignment resilient diffractive optical networks," *Nanophotonics* **9**(13), 4207–4219 (2020).
35. D. Mengü, Y. Rivenson, and A. Ozcan, "Scale-, shift-, and rotation-invariant diffractive optical networks," *ACS Photonics* **8**(1), 324–334 (2021).
36. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press (2012).
37. J. W. Goodman, *Introduction to Fourier Optics*, W. H. Freeman (2005).
38. T. Kozacki and K. Falaggis, "Angular spectrum-based wave-propagation method with compact space bandwidth for large propagation distances," *Opt. Lett.* **40**(14), 3420–3423 (2015).
39. I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," arXiv:1711.05101 (2019).
40. A. Paszke et al., "PyTorch: an imperative style, high-performance deep learning library," in *Adv. Neural Inf. Process. Syst.*, Curran Associates, Inc., Vol. 32 (2019).
41. G. Cohen et al., "EMNIST: extending MNIST to handwritten letters," in *Int. Joint Conf. Neural Netw. (IJCNN)*, IEEE, pp. 2921–2926 (2017).

Biographies of the authors are not available.