Research Article Vol. 12, No. 6 / June 2024 / Photonics Research 1159

PHOTONICS Research

Diffractive neural networks with improved expressive power for gray-scale image classification

Minjia Zheng,[1] Wenzhe Liu,[2,5] Lei Shi,[1,2,3,4,6] and Jian Zi[1,2,3,4,7]

[1]State Key Laboratory of Surface Physics, Key Laboratory of Micro- and Nano-Photonic Structures (Ministry of Education) and Department of Physics, Fudan University, Shanghai 200433, China
[2]Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China
[3]Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China
[4]Shanghai Research Center for Quantum Sciences, Shanghai 210315, China
[5]e-mail: wliubh@connect.ust.hk
[6]e-mail: lshi@fudan.edu.cn
[7]e-mail: jzi@fudan.edu.cn

Received 29 November 2023; revised 14 January 2024; accepted 29 February 2024; posted 20 March 2024 (Doc. ID 513845); published 27 May 2024

In order to harness diffractive neural networks (DNNs) for tasks that better align with real-world computer vision requirements, the incorporation of gray scale is essential. Currently, DNNs are not powerful enough to accomplish gray-scale image processing tasks due to limitations in their expressive power. In our work, we elucidate the relationship between the improvement in the expressive power of DNNs and the increase in the number of phase modulation layers, as well as the optimization of the Fresnel number, which can describe the diffraction process. To demonstrate this point, we numerically trained a double-layer DNN, addressing the prerequisites for intensity-based gray-scale image processing. Furthermore, we experimentally constructed this double-layer DNN based on digital micromirror devices and spatial light modulators, achieving eight-level intensity-based gray-scale image classification for the MNIST and Fashion-MNIST data sets. This optical system achieved the maximum accuracies of 95.10% and 80.61%, respectively.

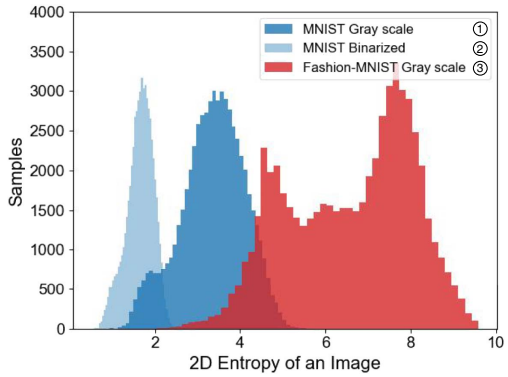
https://doi.org/10.1364/PRJ.513845

1. INTRODUCTION

As the revolution of deep learning is ongoing, it also revitalizes the field of computer vision (CV) [1]. CV is a field that bestows upon machines the ability to perceive and interpret the visual world, typically represented in gray scale, in the way humans do [2]. Some CV applications have become deeply integrated into our lives, including image classification [3,4], image segmentation [5,6], and target detection [7–11]. Algorithms for image processing require significant parallel computational resources [4,12–15]. Recently, to address the high parallelism and large-scale computational demands, optical neural networks (ONNs) have emerged [16–41]. An all-optical ONN framework, known as diffractive deep neural network (D^2NN), was introduced to leverage optical diffraction for computational operations with the potential of hundreds of billions of artificial neuron connections [23]. Its capabilities are also extended to encompass optical logical operations and image-processing tasks [38,42–46].

D^2NNs perform all-optical computations using free-space diffraction and optical parameter modulation. In D^2NNs, each diffractive neuron within the hidden layers modulates the phase/amplitude of the incoming light. The modulations between successive layers are connected by optical diffraction. The values of neurons are optimized via the error backpropagation algorithm. The passive hidden layers can be fabricated and assembled into the physical architecture of a DNN [23,43,47–49]. Alternatively, a DNN can also be realized by loading the phase values of neurons in the hidden layers onto a spatial light modulator (SLM) [50].

So far, there have been few studies that can achieve the classification capability for gray-scale images in terms of light intensity. Ozcan *et al.* encoded gray-scale information into the phase channel of light, achieving a numerical accuracy 81.13% on the Fashion-MNIST data set [23]. In general, gray scale serves as the initial step in CV for recognizing and comprehending the world. In order to emulate the operation of the human visual system within DNNs and to extend their applicability to a broader spectrum of practical CV scenarios, achieving enhanced complexity in DNNs and accomplishing their gray-scale processing capabilities are of paramount importance.



Benchmark

	Models	Data set ①	Data set ②	Data set ③
Digital	Linear NN[52]	88.0	-	78.2
Digital	CNN[52]	> 98.9	-	95.3
Optical		Numerical		
Optical	D^2NN[53]	98.59	91.75	91.06 (phase channel)
Optical	SL-DNN[50]	-	97.08	-
Optical	This work	97.90	-	86.02
Optical		Experimental		
Optical	D^2NN[23]	-	88 (#50)	90 (#50, phase channel)
Optical	SL-DNN[50]	-	92.70 (#1000)	-
Optical	This work	95.10 (#10000)	-	80.61 (#10000)

Fig. 1. 2D image entropy distribution of all gray-scale samples in the MNIST and Fashion-MNIST data sets [51] and binarized samples in the MNIST data set. The benchmark table showcases the performance of digital computer algorithms, which are the linear neural network (NN) and the convolutional neural network (CNN), in the first two rows [52]. In contrast, the performance of deep [23,53,54] and single-layer DNNs (SL-DNNs) in numerical simulations and experiments is also presented in the remaining table.

In CV, the difficulty of image-processing tasks is to some extent proportional to the amount of information contained within the image itself. Two-dimensional (2D) image entropy is a metric used to quantify the amount of information or uncertainty present in an image, and it also provides a measure of the image's complexity, randomness, or disorder. In Fig. 1, the 2D image entropy distribution of all training samples in the binarized/gray-scale MNIST and Fashion-MNIST data sets is shown. Excluding the influence of image noise, the mean of 2D image entropy of samples in the Fashion-MNIST data set is 6.58, which is higher than that of the gray-scale MNIST data set, which is 3.34. The average 2D image entropy of the binarized MNIST data set is minimal, measuring only 1.65. This result suggests that the samples in the Fashion-MNIST data set contain more information compared to those in the MNIST data set. Binarizing image samples leads to a loss of the original information contained in the gray-scale images, resulting in a decrease in their 2D image entropy. As is shown in the benchmark table in Fig. 1, there is a prominent difference in the image classification accuracy between the two data sets, with the accuracy being inversely proportional to the amount of image information. Moreover, due to the passive architecture design of DNNs at present, it is challenging to use intensity-based gray-scale images as inputs for image classification tasks during testing.

In this work, we introduce a novel architecture for a multilayer DNN based on digital micromirror devices (DMDs) and SLMs. We have achieved the task of eight-level intensity-based gray-scale image classification experimentally through a multilayer DNN at visible range. DNNs tasked with processing gray-scale images demand a more robust expressive capacity in comparison to their binary image processing counterparts. In our research, we harness the potential of a double-layer DNN that undergoes optimization concerning the Fresnel number, which yielded the highest accuracies with 97.90% for the intensity-based gray-scale MNIST data set and 86.02% for the Fashion-MNIST data set. Furthermore, our experiments mark a pioneering achievement by attaining a testing accuracy of 95.10% for the intensity-based gray-scale MNIST data set and 80.61% for the Fashion-MNIST data set when subjected to the assessment of the complete 10,000 gray-scale samples in the test set.

2. RESULTS AND DISCUSSION

A. Theoretical Analysis

A DNN constitutes a linear neural network, due to the fact that optical diffraction and phase/intensity modulation are all linear operations. Therefore, when vectorized, the input–output relationship of a DNN, $\mathbf{u}_{\text{intput}}$ and $\mathbf{u}_{\text{output}}$, can be linked through a diffraction matrix $\mathbf{M}$, which can be expressed as

$$\mathbf{u}_{\text{output}} = |\mathbf{M} \times \mathbf{u}_{\text{input}}|^2, \tag{1}$$

and here

$$\mathbf{M} = \mathbf{D} \times \prod_{i=L}^{1} [\text{diag}(\mathbf{p}_i) \times \mathbf{D}], \tag{2}$$

where $\mathbf{p}_i$ is the vectorized hidden layer, $\mathbf{D}$ can represent the free-space diffraction process, and L is the number of layers. The ability of the DNN to modulate the input light can be illustrated by analyzing the properties of $\mathbf{M}$.

The property of the diffraction matrix $\mathbf{M}$ plays a critical role in determining the performance of a DNN. When the count of phase modulation layers is zero, corresponding to a scenario akin to free-space diffraction, the amplitude of $\mathbf{M}$ in Fig. 2(a) implies that $\mathbf{M}$ is equal to $\mathbf{D}$, according to Eq. (2). In this context, optical diffraction alone does not have the capacity to modulate the input image. This limitation occurs because each complex-valued element within $\mathbf{M}$ is identical to three other symmetric elements along the two diagonals of the matrix. As such, the entire matrix possesses only 1 degree of freedom to control because the relationships between adjacent elements are also constrained by the diffraction-related parameter, which is the Fresnel number, defined as

$$F = \frac{a^2}{\lambda d}, \tag{3}$$

where a is the pixel size, λ is the working wavelength, and d is the diffraction distance.

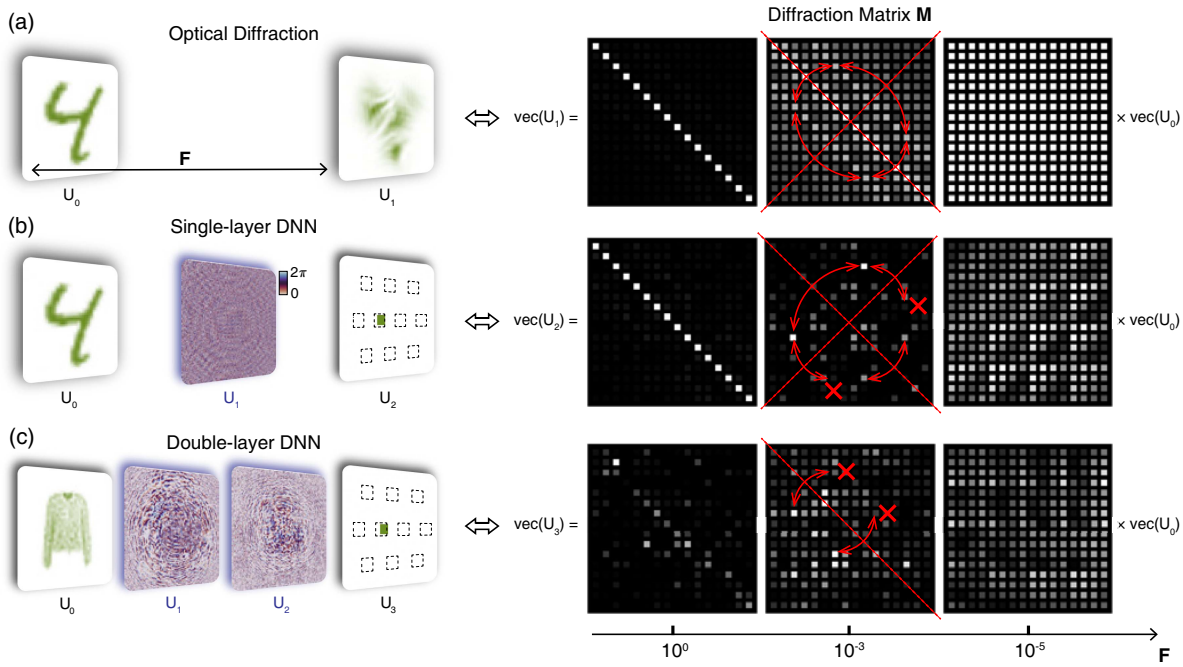


Fig. 2. Properties of the diffraction matrices **M** for optical diffraction and DNNs. (a) Schematic view of the free-space optical diffraction. Each element in **M** for optical diffraction is identical to the other three elements symmetrically positioned along the two diagonals (red dashed lines). (b) Schematic view of a single-layer DNN. One symmetric axis of matrix elements, which is the diagonal from the bottom-left to the top-right of **M**, is disrupted. (c) Schematic view of a double-layer DNN. The last symmetric axis of matrix elements, which is the diagonal from the top-left to the bottom-right of **M**, is disrupted. **M** with different Fresnel numbers F have different properties. When the optimal F is properly chosen, elements in **M** will be independent to take and DNN will have promising performance.

In Fig. 2(b), with the insertion of a single-phase modulation layer into the diffraction process, the symmetry in the amplitude of **M** along the diagonal from the bottom-left to the top-right is disrupted. This means the phase modulation layer provides an additional degree of freedom for light's modulation. Due to the constraints imposed by the remaining symmetrical axis on the values of the matrix elements, only half of elements in **M** or fewer can be used to modulate the incoming light. Moreover, the effectiveness of this enhancement also hinges on the Fresnel number F. A small F, less than roughly 10^{-5}, yields nearly identical elements in **M**. Consequently, irrespective of the input image's characteristics, the resulting light field remains largely consistent. On the other hand, a large F, more than approximately 10^{0}, gives only the elements at the diagonal line the ability to modulate the incoming light. In such a case, light diffracted from pixels in the previous layer cannot propagate to pixels in the subsequent layer, except at their corresponding positions. Our previous work has shown that with the optimal Fresnel number, a single-layer DNN can deliver promising performance when handling a binarized MNIST data set [50].

To enhance DNN's expressive power for processing intensity-based gray-scale images, increasing the number of DNN phase modulation layers is an effective approach. When two or more phase modulation layers are incorporated, the only symmetric axis of **M** is broken and every element is independent to some degree, because the correlation determined by optical diffraction between adjacent elements still persists. Consequently, this grants DNNs more degrees of modulation. In fact, the arbitrary elements of **M** provide almost optimal performance for the DNNs. For more complex and challenging data sets, deep DNNs typically should have better processing capabilities. In Fig. 2(c), without optimization of the Fresnel number, a double-layer DNN still struggles to achieve excellent expressive power. Therefore, even as the number of layers increases, we still need to adjust the diffraction-related parameters to a reasonable range. From the rightmost and leftmost columns of Fig. 2 together, it can be observed that for the same Fresnel number, increasing the number of layers can also reduce the correlation between elements in the matrix, thereby enhancing their independent adjustability. Therefore, increasing the number of layers in the DNN can improve its expressive power without the range of favorable Fresnel numbers. As the number of layers and diffractive neurons increases, the accumulation of errors also proportionally complicates the preparation of DNNs. Therefore, we consider that a double-layer DNN optimized with a proper Fresnel number can, to a great extent, maximize the arbitrariness of values between elements in **M** while limiting the possibility of inevitable error accumulation due to spatial complexity.

B. Experimental Design

Here, we employed a more expressive DNN consisting of two phase modulation layers to process gray-scale images. In Fig. 3, we introduce the architecture of a multilayer DNN based on a DMD and two SLMs. A DMD is used to display the intensity information of incoming light and diffract it by controlling the tilt of each tiny mirror. The well-trained phase values can be

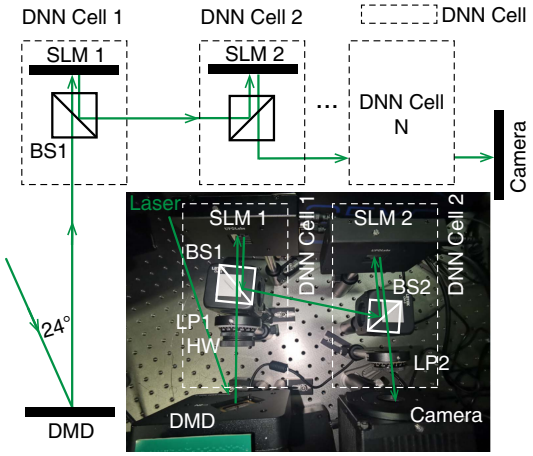


Fig. 3. Schematic and photo of the architecture of the multilayer DNN. It is a combination of a DMD, multiple DNN cells, each of which contains a phase-only SLM and NPBS, and a camera. An experimental setup for a double-layer DNN is shown. A linear polarizer, LP1, serves to adjust the polarization direction of the light to be parallel to the direction of the horizontal axis of the SLM panel. Another linear polarizer, LP2, serves as the analyzer. A half-wave plate, HW, is placed between LP1 and DMD to increase the component of the light with the same polarization direction as the desired direction.

encoded on the SLM, after its gamma correction is done. After the incident light illuminates the SLM, its wavefront is modulated. A 50:50 nonpolarized beam splitter (NPBS) reflects half of the modulated light. The combination of an SLM and an NPBS can be regarded as a cell of a deep DNN, whose primary duty is to modulate the incoming light and output it to another direction. Each DNN cell can be considered as a layer of a DNN. Cells can be connected by using the output of the previous cell as the input of the next cell. After the output of the last cell, light energy is received by a complementary metal-oxide semiconductor (CMOS) camera. We experimentally construct a double-layer DNN based on this optical architecture, which is also available for DNNs with any number of layers.

C. Simulation and Experimental Results

A double-layer DNN consists of three diffraction and two phase modulation processes. The first diffraction is from DMD to the SLM 1, the second diffraction is from SLM 1 to SLM 2, and the third diffraction is from the SLM 2 to the CMOS camera. In the experiments, these three distances d_i $(i = 1, 2, 3)$ must be priorly and precisely measured. Subsequently, we got $d_1 \approx 16.56$ cm, $d_2 \approx 24.99$ cm, and $d_3 \approx 15.45$ cm. The Fresnel number is approximately 6×10^{-4}, falling within the reasonable range. The measurement of three diffraction distances is necessary to reconstruct the forward propagation of the DNN model into a digital computer. The specific measurement methods are detailed in Section 3. The angular spectrum method is used to simulate a free-space optical diffraction process, which can be expressed as

$$\mathcal{F}(u_{i+1}) = \mathcal{F}(u_i) \circ H(d_i), \tag{4}$$

where u_i and u_{i+1} are the complex-valued light field of layer i and $i + 1$, H is the transfer function, and $\mathcal{F}(\cdot)$ is the Fourier transform. Zero padding is also needed to upscale the resolution of the Fourier plane, allowing for a more accurate simulation of the diffraction light-field distribution. The phase modulation process can be simply presented by a Hadamard product between the light field and the phase delay. The optimization of phase values is achieved using the error backpropagation algorithm. After all the diffraction and phase modulation processes, the light intensity at the output layer is used to match the ground truths manually set for every category of the data set. Our training employs both the softmax-cross-entropy (SCE) loss and the mean-squared error (MSE) loss as loss functions.

To demonstrate the excellent performance of the double-layer DNN, we initially choose the gray-scale MNIST handwritten digit data set for testing. After DNN is trained on a training set of 60,000 samples, it achieved its highest accuracy of 97.90% on a blind numerical test of 10,000 samples. All samples are converted to eight-level gray scale. The confusion matrix and energy distribution percentage of the simulation are shown in Fig. 4(b). We loaded the trained phase values onto two SLMs and experimentally tested a total of 10,000 test samples. In Fig. 4(a), the example testing sample of "2" is shown and is loaded onto the DMD. During the CMOS camera's exposure time, the DMD achieves eight-level gray-scale output through the flipping of micromirrors. The optical intensity of the output distribution is also shown. The target region with the maximum light intensities determines the classification result of the DNN for the input image. The positions of the selected regions are chosen compatible with the ground truths during the training process and fine-tuned based on the overall accuracy of the test set. The confusion matrix and energy distribution percentage of the experimental result of gray-scale MNIST handwritten digits classified by a double-layer DNN are shown in Fig. 4(c). We achieved a blind-testing accuracy of 95.10% on 10,000 samples in the test set. The decrease in experimental accuracy relative to simulated accuracy can be caused by several main factors. One factor is the phase and amplitude errors caused by the SLMs. The second factor is the measurement error of three diffraction distances. The third factor is the insufficient polarization purity, resulting in unexpected phase modulation. Nonetheless, this is still a promising performance on the gray-scale MNIST data set based on a DNN model.

Furthermore, we chose the Fashion-MNIST data set for testing. Instead of handwritten digits, Fashion-MNIST consists of a collection of gray-scale images of various fashion items and provides a more challenging problem compared to MNIST. We also trained a double-layer DNN of 60,000 samples in the training set, and it achieved its highest accuracy of 86.02% on a blind testing of 10,000 samples. The confusion matrix and energy distribution percentage of the simulation are shown in Fig. 5(b). In Fig. 5(a), the example testing sample of "Sandal" is shown and is loaded onto the DMD. The experimental gray-scale settings of DMD remain the same as what is set when testing the MNIST data set. The confusion matrix and energy distribution of the experimental result of Fashion-MNIST classified by a double-layer DNN are shown in Fig. 5(c). We achieved a blind-testing accuracy of 80.61% on 10,000 samples in the test set.

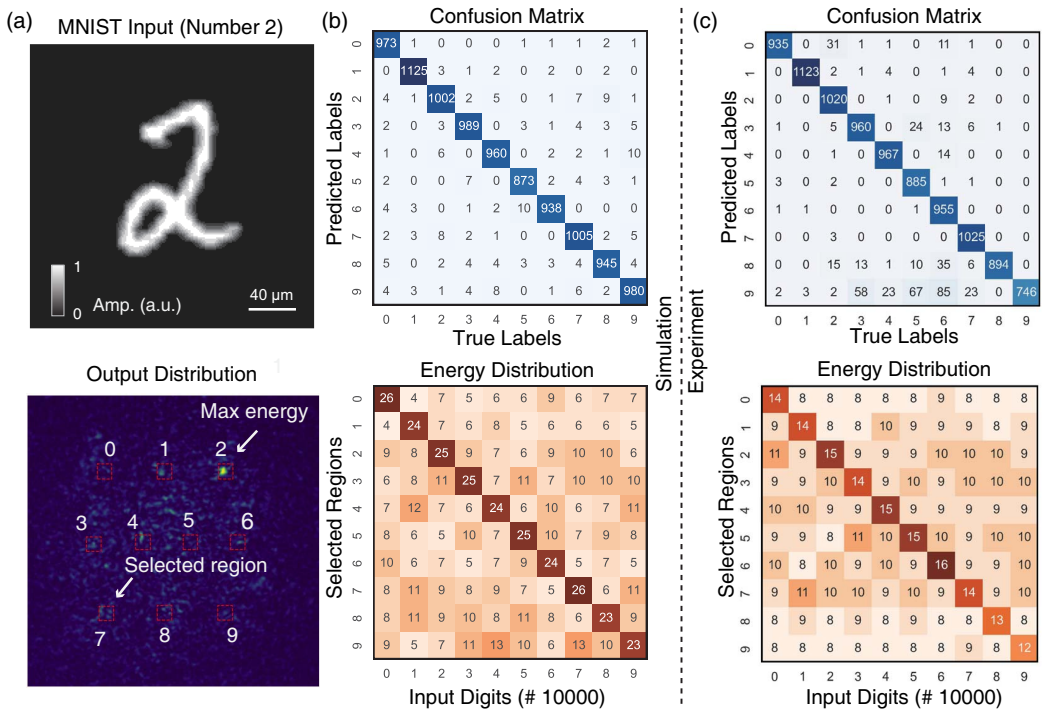


Fig. 4. Simulation and experimental result of gray-scale MNIST data set. (a) Images of MNIST handwritten digits are intensity-based eight-level gray scale. Ten light intensity regions are manually selected. The target region with the maximum intensity determines the classification result. (b) The confusion matrix and energy distribution percentage show numerical test results of blindly testing 10,000 samples, and it achieves the accuracy of 97.90%. (c) The confusion matrix and energy distribution percentage for the experimental results. All 10,000 samples in the test set are tested, and the double-layer DNN achieves the accuracy of 95.10%.

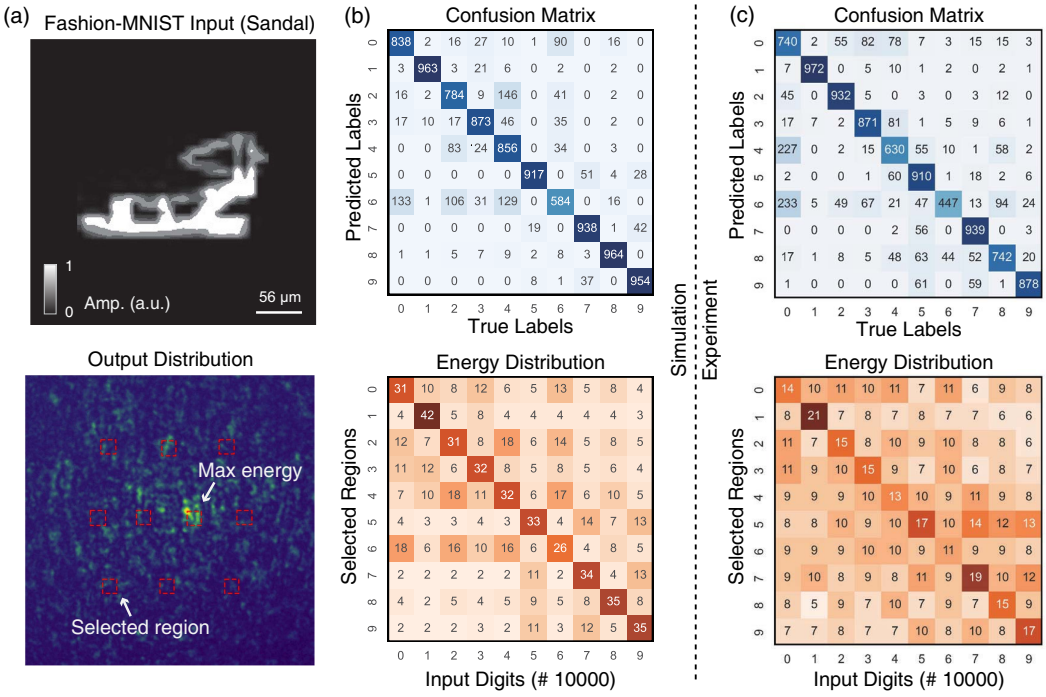


Fig. 5. Simulation and experimental result of Fashion-MNIST data set. (a) Images of Fashion-MNIST handwritten digits are intensity-based eight-level gray scale. Ten light intensity regions are manually selected. The target region with the maximum intensity determines the classification result. (b) The confusion matrix and energy distribution percentage show numerical test results of blindly testing 10,000 samples, and it achieves the accuracy of 86.02%. (c) The confusion matrix and energy distribution percentage for the experimental results. All 10,000 samples in the test set are tested, and the double-layer DNN achieves the accuracy of 80.61%.

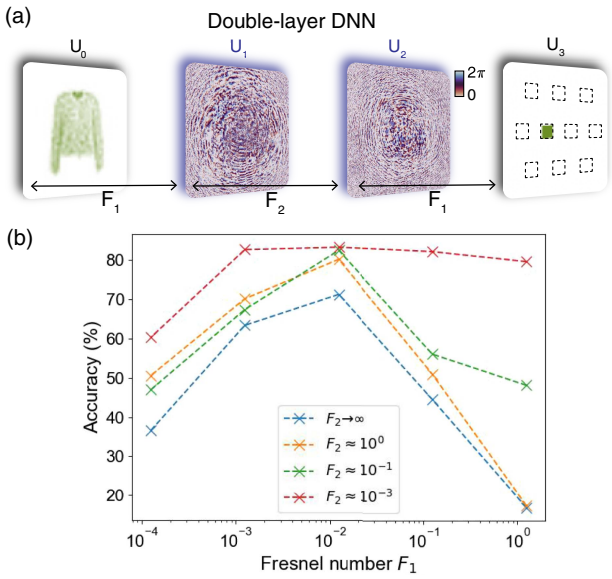


Fig. 6. Performance of a double-layer DNN with different Fresnel numbers. (a) In a double-layer DNN, there is three-segment free-space diffraction. We let the first and the last diffraction processes to be the same, where $F_1 = F_3$. The second diffaction process can be described by F_2. (b) Performance of the double-layer DNN with different combinations of F_1 and F_2.

The relationship of the Fresnel number F and the performance of a single-layer DNN has been sufficiently illustrated in our work recently [50]. The relationship between F and the double-layer DNN is also worthy of discussion. The process of three-segment free-space diffraction can be described by three Fresnel numbers: F_1, F_2, and F_3. Let $F_1 = F_3$; this is done to reduce the redundancy in numerical analysis. The diffraction between the input or output to the phase modulation layer and the diffraction between the two phase modulation layers are sufficient to capture the relationship between the performance of double-layer DNN with the Fresnel numbers. To illustrate this relationship, we conducted tests on double-layer DNN using the Fashion-MNIST data set. In Fig. 6(b), $F_2 \to \infty$ is the condition of adhering two phase modulation layers together, which is equivalent to the condition of a single-phase modulation layer. When F_1 ranges from 10^{-4} to 10^{-2}, DNN will have a decent performance. When F_2 becomes gradually smaller, the elements independence of diffraction matrix $\mathbf{M}$ increases (see Fig. 2), which provides the double-layer DNN with a stronger expressive power. When F_2 continues to get smaller, the correlation between elements of $\mathbf{M}$ increases again, which leads to a decrease in accuracy. But in any case, for the same Fresnel number, a double-layer DNN will always outperform a one-layer DNN as long as there is a practical diffraction process between the two phase modulation layers.

In summary, we theoretically analyzed the benefits of optimizing Fresnel number values and increasing the number of phase modulation layers to enhance the performance of DNNs. Based on this conclusion, we designed and developed an optical system using a DMD and multiple SLMs. In contrast to previous DNNs primarily used intensity binarization, we achieved testing on intensity-based gray-scale MNIST and Fashion-MNIST data sets, which contain more information. In simulations, we achieved accuracies as high as 97.90% and 86.02% on these two data sets. In experiments, we tested the complete test sets and achieved accuracies of 95.10% and 80.61%, respectively.

Successfully processing gray-scale images means that DNNs can now be applied not only to image classification tasks but also have practical potential for more complex CV objectives such as object recognition, saliency detection, and facial recognition. Image binarization is an image-processing technique that can be used for specific tasks, such as object detection and text recognition. However, in more practical and widespread applications, binarization leads to the loss of image details and gray-scale information. Choosing different threshold values can also result in decreased overall performance. Additionally, the process of image binarization requires electronic devices. So, implementing an all-optical DNN for gray-scale image processing is also meaningful. It is still worth discussing the performance of DNNs in processing either binary or gray-scale images in a more complicated data set like CIFAR-10. We believe that our work provides a theoretical and experimental foundation for such further validation and the application of more powerful DNNs in a broader range of scenarios.

3. METHODS

A. Experimental System

Our experimental optical system adopted commercially available optoelectronic devices as the blocks of a double-layer DNN. The coherent light source is generated from a continuous-wave diode-pumped laser (04-01 Series, Fandango, Cobolt) with a working wavelength of 515 nm. Following laser collimation, it is incident on the DMD (HDSLM756D65, UPOLabs) surface at an angle of 24 deg. The DMD consists of 1920 × 1080 micromirrors with a pitch of 7.56 μm. After encoding image information onto the DMD and reflection, we employed a half-wave plate (ZWP20H-520Q, JCOPTIX) and a linear polarizer (OPPF1-VIS, JCOPTIX) to modulate the polarization of the light. Two SLMs (HDSLM80R Plus, UPOLabs) with pixel sizes of 8 μm serve as the phase modulation layers. Two NPBSs (BS013, Thorlabs) are used to adjust directions of the reflected and transmitted light. The light intensity at the output layer was recorded using a CMOS camera (FL20BW, Tucsen).

B. Data Preprocessing

Both the MNIST and Fashion-MNIST data sets have 10 categories, with a total of 60,000 training samples and 10,000 testing samples. These images have a resolution of 28 × 28 pixels. For training and testing on the gray-scale MNIST data set, we upscaled the resolution to 200 × 200 pixels, and for training and testing on the Fashion-MNIST data set, we upscaled the resolution to 300 × 300 pixels. All images were set to eight-level intensity-based gray scale.

C. Diffraction Distance Measurement

To obtain accurate diffraction distances in experiments, using lens imaging is a simple and effective common method. One of

the primary functions of an SLM is to simulate the effect of a lens in an optical system by loading the phase distribution of a Fresnel lens. After encoding the phase distributions of Fresnel lens with three combinations of different focal lengths into the SLMs and recording the object plane using the CMOS camera, three independent equations can be listed to solve three unknown quantities: d_i $(i = 1, 2, 3)$. The three pairs of focal lengths are (f_1, ∞), (∞, f_2), and (f_1', f_2'). The focal length approaching ∞ means phase values of the SLM are set to be a constant value. The first two equations can be written as $1/d_1 + 1/(d_2 + d_3) = 1/f_1$ and $1/(d_1 + d_2) + 1/d_3 = 1/f_2$. The third equation can be written as $1/[d_2 - f_1' d_1/(d_1 - f_1')] + 1/d_3 = 1/f_2'$. In the experiment, we set up these three combinations of focal lengths to be (11.75 cm, ∞), (∞, 11.27 cm), and (20.00 cm, 13.71 cm) to achieve good imaging at the object plane. After solving the equations, we got $d_1 \approx 16.56$ cm, $d_2 \approx 24.99$ cm, and $d_3 \approx 15.45$ cm. Under the circumstances of DNNs with more layers, performing imaging experiments with any two SLMs using the method described above, with the phase values set to 0 for the remaining SLMs, three distances can be obtained in the first set. Then, by replacing the other two layers of SLMs and repeating the same procedure, another set of distances can be obtained. After multiple measurements, the diffraction distance for each segment can be measured. Owing to the limited resolution and the fill factor of SLMs, there may be a slight error between the simulated and the actual focal length when loading a lens phase distribution on SLMs. This error may influence the measurement of diffraction distances and cause the decrease in experimental accuracy.

Currently, this method requires precision in the alignment and diffraction distance among various optical components during practical experiments. Considerable effort is needed for calibration before DNN's implementation.

Funding. Major Program of National Natural Science Foundation of China (T2394481); Science and Technology Commission of Shanghai Municipality (2019SHZDZX01, 21DZ1101500, 22142200400, 23DZ2260100); National Key Research and Development Program of China (2022YFA1404800, 2023YFA1406900); National Natural Science Foundation of China (12234007, 12221004, 12321161645).

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
2. R. Szeliski, *Computer Vision: Algorithms and Applications* (Springer, 2010).
3. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, Vol. **25**, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds. (Curran Associates, Inc., 2012).
4. Y. LeCun, B. Boser, J. Denker, *et al.*, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems*, Vol. **2**, D. Touretzky, ed. (Morgan-Kaufmann, 1989).
5. R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," Comput. Vis. Graph. Image Process. **29**, 100–132 (1985).
6. S. Minaee, Y. Y. Boykov, F. Porikli, *et al.*, "Image segmentation using deep learning: a survey," IEEE Trans. Pattern Anal. Mach. Intell. **44**, 3523–3542 (2021).
7. A. Borji, M.-M. Cheng, H. Jiang, *et al.*, "Salient object detection: a benchmark," IEEE Trans. Image Process. **24**, 5706–5722 (2015).
8. H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," IEEE Trans. Image Process. **22**, 3766–3778 (2013).
9. W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," IEEE Trans. Image Process. **27**, 38–49 (2017).
10. A. Wang and M. Wang, "RGB-D salient object detection via minimum barrier distance transform and saliency fusion," IEEE Signal Process. Lett. **24**, 663–667 (2017).
11. A. Chaurasia and E. Culurciello, "Linknet: exploiting encoder representations for efficient semantic segmentation," in *IEEE Visual Communications and Image Processing (VCIP)* (IEEE, 2017), pp. 1–4.
12. O. Russakovsky, J. Deng, H. Su, *et al.*, "Imagenet large scale visual recognition challenge," Int. J. Comput. Vis. **115**, 211–252 (2015).
13. W. Zhang, K. Itoh, J. Tanida, *et al.*, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture," Appl. Opt. **29**, 4790–4797 (1990).
14. D. Powell and M. Duffy, "Neural networks and statistical models," in *Proceedings of the Nineteenth Annual SAS Users Group International Conference* (Citeseer, 1994), pp. 806–814.
15. R. Hamerly, L. Bernstein, A. Sludds, *et al.*, "Large-scale optical neural networks based on photoelectric multiplication," Phys. Rev. X **9**, 021032 (2019).
16. J. Bueno, S. Maktoobi, L. Froehly, *et al.*, "Reinforcement learning in a large-scale photonic recurrent neural network," Optica **5**, 756–760 (2018).
17. T. W. Hughes, M. Minkov, Y. Shi, *et al.*, "Training of photonic neural networks through *in situ* backpropagation and gradient measurement," Optica **5**, 864–871 (2018).
18. P. R. Prucnal, B. J. Shastri, and M. C. Teich, *Neuromorphic Photonics* (CRC Press, 2017).
19. D. Pérez, I. Gasulla, P. D. Mahapatra, *et al.*, "Principles, fundamentals, and applications of programmable integrated photonics," Adv. Opt. Photon. **12**, 709–786 (2020).
20. X. Xu, M. Tan, B. Corcoran, *et al.*, "11 tops photonic convolutional accelerator for optical neural networks," Nature **589**, 44–51 (2021).
21. B. J. Shastri, A. N. Tait, T. Ferreira de Lima, *et al.*, "Photonics for artificial intelligence and neuromorphic computing," Nat. Photonics **15**, 102–114 (2021).
22. J. Feldmann, N. Youngblood, M. Karpov, *et al.*, "Parallel convolutional processing using an integrated photonic tensor core," Nature **589**, 52–58 (2021).
23. X. Lin, Y. Rivenson, N. T. Yardimci, *et al.*, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004–1008 (2018).
24. Y. Shen, N. C. Harris, S. Skirlo, *et al.*, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**, 441–446 (2017).
25. A. N. Tait, T. F. De Lima, E. Zhou, *et al.*, "Neuromorphic photonic networks using silicon photonic weight banks," Sci. Rep. **7**, 7430 (2017).
26. M. Hermans, M. Burm, T. Van Vaerenbergh, *et al.*, "Trainable hardware for dynamical computing using error backpropagation through physical media," Nat. Commun. **6**, 6729 (2015).
27. D. Brunner, M. C. Soriano, C. R. Mirasso, *et al.*, "Parallel photonic information processing at gigabyte per second data rates using transient states," Nat. Commun. **4**, 1364 (2013).
28. M. M. P. Fard, I. A. D. Williamson, M. Edwards, *et al.*, "Experimental realization of arbitrary activation functions for optical neural networks," Opt. Express **28**, 12138–12148 (2020).
29. S. Pai, Z. Sun, T. W. Hughes, *et al.*, "Experimentally realized *in situ* backpropagation for deep learning in photonic neural networks," Science **380**, 398–404 (2023).

30. G. Wetzstein, A. Ozcan, S. Gigan, *et al.*, "Inference in artificial intelligence with deep optics and photonics," Nature **588**, 39–47 (2020).
31. I. Chakraborty, G. Saha, A. Sengupta, *et al.*, "Toward fast neural computing using all-photonic phase change spiking neurons," Sci. Rep. **8**, 12980 (2018).
32. J. Chang, V. Sitzmann, X. Dun, *et al.*, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," Sci. Rep. **8**, 12324 (2018).
33. L. Mennel, J. Symonowicz, S. Wachter, *et al.*, "Ultrafast machine vision with 2D material neural network image sensors," Nature **579**, 62–66 (2020).
34. Y. Zuo, B. Li, Y. Zhao, *et al.*, "All-optical neural network with nonlinear activation functions," Optica **6**, 1132–1137 (2019).
35. X. Luo, Y. Hu, X. Ou, *et al.*, "Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible," Light Sci. Appl. **11**, 158 (2022).
36. F. Ashtiani, A. J. Geers, and F. Aflatouni, "An on-chip photonic deep neural network for image classification," Nature **606**, 501–506 (2022).
37. T. W. Hughes, I. A. Williamson, M. Minkov, *et al.*, "Wave physics as an analog recurrent neural network," Sci. Adv. **5**, eaay6946 (2019).
38. H. Dou, Y. Deng, T. Yan, *et al.*, "Residual D^2NN: training diffractive deep neural networks via learnable light shortcuts," Opt. Lett. **45**, 2688–2691 (2020).
39. J. Li, Y.-C. Hung, O. Kulce, *et al.*, "Polarization multiplexed diffractive computing: all-optical implementation of a group of linear transformations through a polarization-encoded diffractive network," Light Sci. Appl. **11**, 153 (2022).
40. M. S. S. Rahman, X. Yang, J. Li, *et al.*, "Universal linear intensity transformations using spatially-incoherent diffractive processors," arXiv, arXiv:2303.13037 (2023).
41. B. Bai, Y. Li, Y. Luo, *et al.*, "All-optical image classification through unknown random diffusers using a single-pixel diffractive network," Light Sci. Appl. **12**, 69 (2023).
42. C. Qian, X. Lin, X. Lin, *et al.*, "Performing optical logic operations by a diffractive neural network," Light Sci. Appl. **9**, 59 (2020).
43. S. Jiao, J. Feng, Y. Gao, *et al.*, "Optical machine learning with incoherent light and a single-pixel detector," Opt. Lett. **44**, 5186–5189 (2019).
44. Z. Wu, M. Zhou, E. Khoram, *et al.*, "Neuromorphic metasurface," Photon. Res. **8**, 46–50 (2020).
45. Z. Wu and Z. Yu, "Small object recognition with trainable lens," APL Photon. **6**, 071301 (2021).
46. T. Zhou, X. Lin, J. Wu, *et al.*, "Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit," Nat. Photonics **15**, 367–373 (2021).
47. H. Chen, J. Feng, M. Jiang, *et al.*, "Diffractive deep neural networks at visible wavelengths," Engineering **7**, 1483–1491 (2021).
48. Y. Hu, X. Luo, Y. Chen, *et al.*, "3D-integrated metasurfaces for full-colour holography," Light Sci. Appl. **8**, 86 (2019).
49. Y. Chen, Z. Shu, S. Zhang, *et al.*, "Sub-10 nm fabrication: methods and applications," Int. J. Extreme Manuf. **3**, 032002 (2021).
50. M. Zheng, L. Shi, and J. Zi, "Optimize performance of a diffractive neural network by controlling the Fresnel number," Photon. Res. **10**, 2667–2676 (2022).
51. H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," arXiv, arXiv:1708.07747 (2017).
52. Y. LeCun, L. Bottou, Y. Bengio, *et al.*, "Gradient-based learning applied to document recognition," Proc. IEEE **86**, 2278–2324 (1998).
53. J. Li, D. Mengu, Y. Luo, *et al.*, "Class-specific differential detection in diffractive optical neural networks improves inference accuracy," Adv. Photon. **1**, 046001 (2019).
54. D. Mengu, Y. Luo, Y. Rivenson, *et al.*, "Analysis of diffractive optical neural networks and their integration with electronic neural networks," IEEE J. Sel. Top. Quantum Electron. **26**, 3700114 (2019).