Advanced all-optical classification using orbital-angular-momentum-encoded diffractive networks

Kuo Zhang[a,†] **Kun Liao**[b,†] **Haohang Cheng,**[a] **Shuai Feng,**[a,*] **and Xiaoyong Hu**[b,c,d,*]
[a]Minzu University of China, School of Science, Beijing, China
[b]Peking University, Collaborative Innovation Center of Quantum Matter, Nano-Optoelectronics Frontier Center of Ministry of Education,
  State Key Laboratory for Mesoscopic Physics, Department of Physics, Beijing, China
[c]Shanxi University, Collaborative Innovation Center of Extreme Optics, Taiyuan, China
[d]Peking University Yangtze Delta Institute of Optoelectronics, Nantong, China

Abstract. As a successful case of combining deep learning with photonics, the research on optical machine learning has recently undergone rapid development. Among various optical classification frameworks, diffractive networks have been shown to have unique advantages in all-optical reasoning. As an important property of light, the orbital angular momentum (OAM) of light shows orthogonality and mode-infinity, which can enhance the ability of parallel classification in information processing. However, there have been few all-optical diffractive networks under the OAM mode encoding. Here, we report a strategy of OAM-encoded diffractive deep neural network (OAM-encoded D^2NN) that encodes the spatial information of objects into the OAM spectrum of the diffracted light to perform all-optical object classification. We demonstrated three different OAM-encoded D^2NNs to realize (1) single detector OAM-encoded D^2NN for single task classification, (2) single detector OAM-encoded D^2NN for multitask classification, and (3) multidetector OAM-encoded D^2NN for repeatable multitask classification. We provide a feasible way to improve the performance of all-optical object classification and open up promising research directions for D^2NN by proposing OAM-encoded D^2NN.

Keywords: diffractive deep neural network; deep learning; orbital angular momentum multiplexing; optical classification.

1 Introduction

The exponential growth of information and data processing has led to bottlenecks in the continuous improvement of performance for traditional electronic hardware processors.[1] To address this problem, all-optical computing using photons as information carriers has become a promising solution.[2–6] Compared with traditional electronic hardware computing, optical computing offers several advantages, including ultrafast computing speed,[7,8] ultralow energy consumption,[9] and significant potential for parallel computing.[10,11] In recent years, with the rapid development of deep learning,[12] optical computing based on deep learning with different implementation schemes has been increasingly applied to various tasks,[13] such as vowel recognition,[9] image classification,[11,14–17] mathematical operations,[7] and matrix operations.[18–25]

A diffractive deep neural network (D^2NN) is a series of successive diffractive layers designed in a computer using error backpropagation and stochastic gradient descent methods.[11] Unlike machine vision systems that use conventional optics, the diffractive layer of D^2NN consists of a series of two-dimensional passive pixel arrays. Each pixel point on the diffractive layer is a parameter that can be learned by the computer and can be used for independent complex-valued tuning of the light field. Based on its capabilities in optical information processing, D^2NN has been applied to image recognition,[11,26–40] optical logic operations,[41–43] terahertz pulse shaping,[44] phase retrieval,[45] and image reconstruction,[15,46–48] etc.

Diffractive networks performed in passive optical elements have the advantages of fast processing speed and low energy consumption, while also enabling flexible utilization of various degrees of freedom of light in the network. For example, when using broadband light instead of monochromatic light to illuminate the diffractive networks, spectrally encoded machine vision applications,[15,38] parallel computing,[39] snapshot multispectral imaging,[48] and spatially controlled wavelength multiplexing/demultiplexing[49] can be accomplished. In addition, the linear transformation of polarization multiplexing can be achieved by using the polarization properties of light in diffractive networks instead of being based on birefringence or polarization-sensitive materials,[50] which fully demonstrates the classification and computational potential of diffractive networks in complex-valued matrix vector operations. So far, the phase, amplitude, polarization, and wavelength of light have been applied in different diffractive networks to perform the specific required computational tasks.

As another important property of light, the orbital angular momentum (OAM) modes carried by vortex beams (VBs) are widely used in various fields by virtue of the unique properties brought about by their wavefront structure.[51-60] In terms of the combination of the D²NN with OAM modes, multiplexing/demultiplexing of OAM modes,[61-63] optical logic gates,[42] holography,[64,65] and spectral detection[66] have been reported in recent years. These works show the great potential of D²NN in handling complex OAM modes. Since parallel object classification requires multiple independent channels as carriers for information processing, the orthogonality and near-infinite mode of the OAM can present significant pattern differentiation and recognition robustness during propagation, which is well suited for application in all-optical parallel classification. However, the near-infinite OAM mode has not yet been utilized in D²NN to achieve advanced all-optical classification.

Here, we report on the strategy of OAM-encoded diffractive deep neural networks (OAM-encoded D²NNs) which encodes the spatial information of objects into OAM modes of light by using deep-learning-trained diffractive layers to perform recognition and classification in vortex light multiplexed by different OAM modes. We use a VB that multiplexes 10 OAM modes with different topological charges while maintaining equal weights. And the beam is used to illuminate handwritten digits, which then pass through five diffractive layers of D²NN. The modulated vortex light is obtained at the output, and its OAM spectrum is analyzed. The normalized intensity distribution of each OAM in the OAM spectrum is assigned to a digit/class.

(1) First, we demonstrate a single detector OAM-encoded D²NN for a single task classification. We achieve a blind accuracy of 85.43% for the Mixed National Institute of Standards and Technology (MNIST) data set.[67] For comparison, the spectrally encoded single-pixel machine vision without image reconstruction achieved blind test accuracy of 84.02% for the same data sets.[15] (2) In addition, we show a single detector OAM-encoded D²NN for multitask classification. To evaluate the discriminative criteria for multi-object classification, we propose the self-defined MNIST array data set and MNIST repeatable array data set (see Sec. 4.4). Most of the previous multitask classification works were performed on several different data sets for parallel recognition.[16,39] However, their accuracies were calculated separately and independently for each data set; few of them were computed in parallel on the same data set.

The MNIST array data set and MNIST repeatable array data set will give some digits as a digit array for classification each time. When any one or more digits in the input are inferred incorrectly, we assume that the digit array is judged incorrectly. So, there are a large number of cases where the correct inference of just one digit in an array is attributed to misclassification. We achieved a blind test accuracy of 64.13% for the MNIST array data set. In fact, there are 45 inferred categories in the MNIST array data set, which is significantly larger than the 10 categories in the MNIST data set. (3) Moreover, we design a multidetector OAM-encoded D²NN for repeatable multitask classification. By measuring multiple OAM spectra of beams and comparing their intensities, we achieve parallel classification for two-digit, three-digit, and four-digit MNIST repeatable array data sets. Although using the MNIST array data set and the MNIST repeatable array data set instead of the MNIST data set undoubtedly increases the difficulty of judgment, the advantages of advanced parallel classification are highlighted by the process of promoting a single task into multiple tasks.

As shown in Table 1, this work achieves a breakthrough in parallel classification by utilizing the OAM degree of freedom compared to other existing D²NN designs. We believe that OAM-encoded D²NNs provide a powerful framework to further improve the capability of all-optical parallel classification and OAM-based machine vision tasks. In the near future, the development of OAM mode multiplexing/demultiplexing technology may enable the application of OAM combs consisting of hundreds of OAM modes.[60] The advancement will be possible to introduce more OAM modes into the OAM-encoded D²NN, and thus break through to a higher degree of parallelism for solving more complex multitasking parallel classifications.

2 Results

2.1 Design of OAM-Encoded D²NNs

In this paper, we demonstrate an approach to incorporate OAM into D²NN, which encode the spatial information of objects into the OAM modes of light. Our approach is based on the Fresnel scalar diffraction theory, and we propose three different variants of OAM-encoded D²NNs, as shown in Fig. 1. The schematic diagram illustrates the OAM-encoded D²NN structures and highlights the similarities and differences between the proposed OAM-encoded D²NNs. The similarity among the proposed OAM-encoded D²NNs is that they are all composed of five diffractive layers, with a constant spacing of 1.55 mm between the input layer and the diffractive layer, as well as between the diffractive layers and between the diffractive layer and the output layer. The distance is determined by the qualifying conditions of the Fresnel scalar diffraction theory. The number of diffractive units per layer is 200×200. These diffractive networks are trained to run independently without being coupled to other networks, although they have the same number of layers and neurons. At the input, an OAM mode is generated by using a Laguerre–Gaussian (LG) beam operating at 1550 nm, with a waist radius of 3λ. Ten OAM modes with $m \in [-5, +5]$ are selected, each corresponding to one of the 10 categories of handwritten digits in the MNIST data set. The $+1$ to $+5$ OAM modes represent digits 0 to 4, while the -1 to -5 OAM modes represent digits 5 to 9. A VB multiplexes 10 OAM modes with equal weights to illuminate handwritten digits. The equation we employed for multiplexing LG beams carrying different OAM modes can be expressed as follows:

Table 1 Comparison with other D^{2}NN using more than three degrees of freedom.

Reference	Degree of freedom	Footprint	Function	Performance	Parallel classification	Single detector
This work	OAM	164.3 μm × 164.3 μm	Image recognition	Accuracy: 85.49%	Yes	Yes
11	—	8 cm × 8 cm	Image recognition	Accuracy: 93.39%	No	No
15	Wavelength	8 cm × 8 cm	Image recognition	Accuracy: 91.29% (84.02%)[a]	No	Yes
38	Wavelength	6 cm × 6 cm	Image recognition	Accuracy: 87.74%	No	Yes
39	Wavelength	0.8 mm × 0.8 mm	Image recognition	Accuracies of four tasks are 92.8%, 83.0%, 81.0%, and 90.4%, respectively	Yes	No
48	Wavelength	88.2 μm × 88.2 μm	Multispectral imaging	Filter transmission efficiency: >79%	—	—
49	Wavelength	5 cm × 5 cm	Spectral filters	Process optical waves over a continuous, wide range of frequencies	—	—
16	Polarization	11.2 μm × 11.2 μm	Image recognition	Accuracy: 93.75%	Yes	No
50	Polarization	24λ × 24λ	Linear transformations	Perform multiple complex-valued, arbitrary linear transformations using polarization multiplexing	—	—
42	OAM	3 cm × 3 cm	Logic operation	Proposed an OAM logical operation	—	—
61	OAM	3 cm × 3 cm	Optical communication	The diffraction efficiency and mode conversion purity: >96%. The bit error rates: <10^{-4}	—	—
64	OAM	2.5 μm × 2.5 μm	Holography	10 multiplexed OAM modes among five spatial depths in deep multiplexing holography	—	—
66	OAM	100λ × 100λ	Spectral detection	Optical operations/electronic operations: ~10^3	—	—

[a]Accuracy without reconstructed image is shown in parentheses.

$$f_{\text{multiple OAM}}(r, \varphi, z) = f_{\text{OAM1}}(r, \varphi, z) + f_{\text{OAM2}}(r, \varphi, z) + \ldots$$
$$+ f_{\text{OAM}m}(r, \varphi, z), \qquad (1)$$

where $f_{\text{OAM}m}(r, \varphi, z)$ represents the input OAM beams and m represents the topological charge of the OAM beams. After irradiating the digits, due to the different transmission light distribution of different digits, each OAM mode will generate independent post-transmission complex amplitude information and encode the spatial position information of the digits into the OAM mode information.

The first scheme with the OAM-encoded D^{2}NN demonstrated encoding of a single digit using the OAM mode, then transmitting it through diffractive layers. The results showed that the OAM beam generated in the output plane corresponded to the handwritten digit input, as shown in Fig. 1(a). The OAM-encoded D^{2}NN was then used for parallel image recognition. As shown in Fig. 1(b), two different categories of digits were positioned in separate spatial locations, encoded with the OAM mode, and transmitted simultaneously through the diffractive networks. The result was an independent multiplexed OAM beam at the output, with the OAM modes corresponding to the two initial input digit categories. However, using a single detector for parallel detection resulted in an inability to distinguish between identical digits, as the single detector OAM-encoded D^{2}NN lacked the ability to detect sequential information.

To address this issue, a multidetector OAM-encoded D^{2}NN was used to discriminate repeating digits [see Fig. 1(c)]. Compared to the single detector OAM-encoded D^{2}NN, the ability of the multiple detectors to encode sequential information between repeating digits allows them to recognize the same digits while further increasing the parallel classification power of the diffractive network.

2.2 Single Detector OAM-Encoded D^{2}NN for Single Task Classification

Here is a demonstration of the recognition of an OAM encoded digit "1" using the single detector (not a single-pixel detector) OAM-encoded D^{2}NN. A multiplexed OAM beam is used to illuminate the MNIST handwritten digit "1" and then passes through the diffractive layers, resulting in a modulated OAM beam at the output receiver plane [see Fig. 2(b)]. The optical field distribution of the input OAM-encoded digit "1" in each layer after modulation by the trained diffractive networks is shown in Fig. 2(a). It can be seen that the input digit "1" exhibits a residual, which is caused by the uneven distribution of light intensity in the mixed OAM beam. By our comparison, this type of irradiation does not affect the accuracy of the blind test recognition. After the modulation of the diffractive layers, a second-order OAM beam is reconstructed at the output, which can show that our diffractive network is able to perform the given task
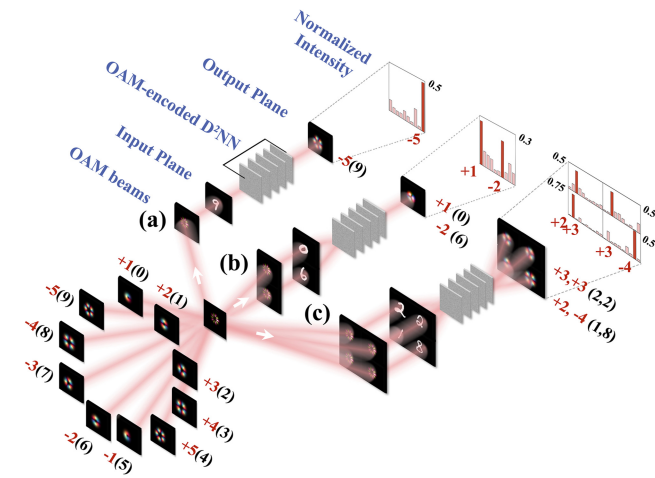
Fig. 1 Schematic diagrams of the three types of the OAM-encoded D²NN. The OAM beams illuminating the digits are multiplexed by 10 OAM modes ranging from –5 to +5 in equal proportions. The red numbers represent the topological charges of the OAM modes, while the black numbers in brackets correspond to the assumed digits associated with the OAM modes. The digit inputs are illuminated by the multiplexed OAM beams, and the predicted OAM beams are obtained in the output plane after modulation by the OAM-encoded D²NNs. The right side of the output plane shows the OAM spectra of the OAM beams. Three different configurations of OAM-encoded D²NNs have been described below: (a) single detector OAM-encoded D²NN for single-task classification, (b) single detector OAM-encoded D²NN for multitask classification, and (c) multidetector OAM-encoded D²NN for multitask classification.

relatively well. Although the output light contains non-single OAM modes due to modulation limitations and diffraction effects, the classification can still be inferred from the intensity distributions among different OAM modes. We obtained the normalized intensity distribution of each OAM mode by analyzing the OAM spectra of the OAM beams at the output (see Sec. 4.3). The category of the inferred digit is determined by the highest normalized intensity of the OAM mode. As shown in Fig. 2(c), the intensity of the OAM mode with $m = +2$, corresponding to the digit "1," is 79.37%, which is significantly higher than that of other OAM modes, demonstrating effective filtering of the vortex light with other OAM modes.

During the training process, the single-detector OAM-encoded D²NN reduces the loss value by continuously updating and adjusting the phase and amplitude distribution of the diffractive layers. The loss and accuracy functions for both the training and testing phases are shown in Fig. 2(d), where the dashed lines represent the results of each recognition, and the solid lines represent the average of the results of the three recognitions. From the mean curves, it can be seen that the single-detector OAM-encoded D²NN experiences a sharp drop in the loss function at the beginning of the iterative process and then stabilizes after a few iterations. In addition, the test accuracy is slightly higher than the training accuracy, and the loss function exhibits smoother fluctuations during the test phase. The blind accuracy of the single-detector OAM-encoded D²NN for the MNIST data set was found to be 85.49% [as shown in Fig. 2(e)]. The accuracy of D²NN using OAM encoding is essentially the

same as that of D²NN with wavelength encoding when compared to spectrally encoded single-pixel machine vision using diffractive networks that do not reconstruct images.[15] It should be noted that the single detector of OAM-encoded D²NN is not a single-pixel detector, but rather a single interferometer-like detector (see Sec. 4.3). This shows that this single-detector OAM-encoded D²NN design can efficiently perform a single-digit recognition task.

2.3 Single-detector OAM-Encoded D²NN for Multitask Classification

Following our demonstration of single image classification using OAM-encoded D²NN, we present a more challenging application of the same framework: single-detector OAM-encoded D²NN for multitask classification. In Fig. 3(b), by simultaneously irradiating two different digits, "7" and "0," with independent spatial distributions as an array to the input layer, OAM beams are generated at the center of the output layer, multiplexing the OAM modes with $m = 3$ and $m = +1$ corresponding to each of the two digits. The OAM-encoded D²NN multiplexes the spatial information of both digits into the same OAM beams, effectively utilizing the orthogonality of the OAM modes. However, if the two input digits have the same label, using the highest normalized intensity measure may lead to indistinguishable outcomes. For example, whether we input two digits "2" as an array or one digit "2" combining an array of other digits, a single-detector OAM-encoded D²NN cannot accurately determine how many digits "2" are present at the input because only the highest intensity is considered as the judgment criterion, which can lead to a large error in the network. To address this issue, we utilize a modified MNIST array data set that prevents the inclusion of digits with the same label in a single array (see Sec. 4.4). In Fig. 3(a), the two input digits are modulated by the diffractive layers to produce the optical field in the output plane with the expected OAM modes. By detecting the OAM spectra of the OAM beams at the output, the two OAM modes with the highest normalized intensity represent the classes of the presumed digits [see Fig. 3(c)]. Among them, the normalized intensity of the OAM mode with $m = 3$ corresponding to the digit "7" is 38.97%, and the normalized intensity of the OAM mode with $m = +1$ corresponding to the digit "0" is 35.57%, which far exceeds the other modes. Although OAM modes with the same proportional intensity distribution should theoretically be obtained at the output, the problem of different intensities between the two OAM modes is inevitable due to the limitations of the diffractive network modulation capability. However, this uneven distribution of intensities only slightly affects the accuracy of the inference (after testing, the accuracy error caused by this distribution does not exceed 1%).

After iterative training convergence, our single-detector OAM-encoded D²NN for multitask classification can achieve a blind measurement accuracy of 64.13% [see Fig. 3(d)]. The test results obtained indicate that the accuracy of the single-detector OAM-encoded D²NN, which performs parallel recognition of multiple digits, is lower compared to the previously reported D²NNs. In terms of accuracy requirements, the OAM-encoded D²NN must correctly recognize all digits in the input array. As can be seen from the confusion matrix, there are actually 45 categories to be recognized in the MNIST array data set, which is significantly larger than the 10 categories in the MNIST data set [see Fig. 3(e)]. It is the substantial increase
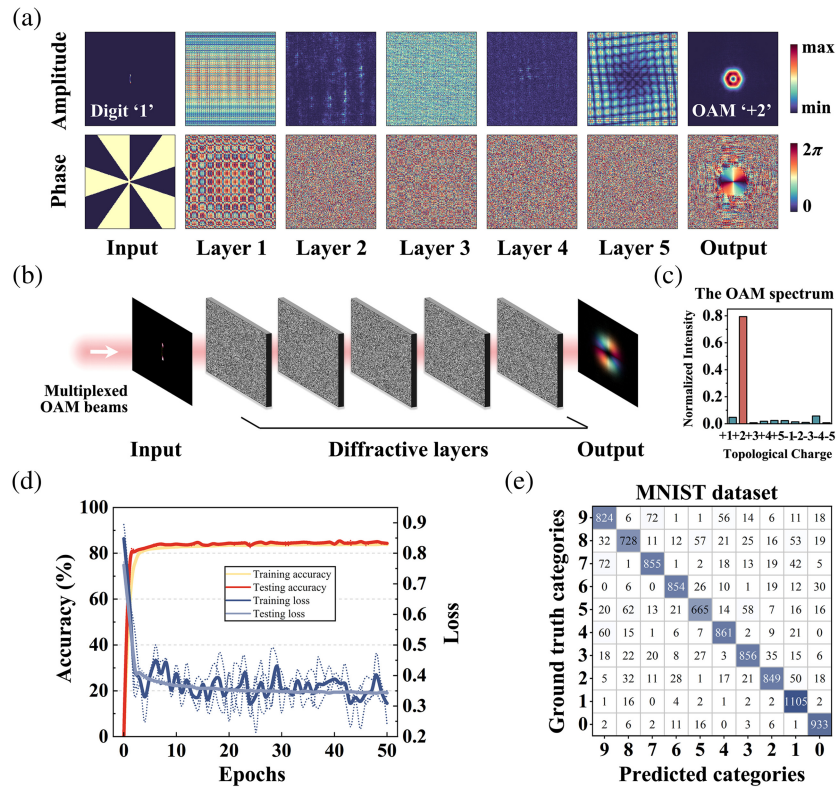
Fig. 2 (a) The amplitude and phase distributions of the OAM beams are shown for the input plane, the diffractive layers, and the output plane. The input image is a handwritten digit "1" encoded as an OAM beam with +2 mode. (b) Schematic of the modulation of the light field by the single-detector OAM-encoded D²NN. (c) The OAM spectrum of the output OAM beams. The red plot corresponding to the OAM mode with the highest normalized intensity indicates the inferred category of the input digit. (d) The loss and accuracy functions for both the training and test sets. Three simulations were conducted for each set, and the corresponding results are represented by the three dashed lines. The solid lines represent the average results of the three function curves depicted by the dashed lines. (e) A confusion matrix summarizes the numerical classification results in the test set. The matrix provides a comprehensive overview of the performance of the single-detector OAM-encoded D²NN in recognizing the handwritten digits from the MNIST data set.

in task complexity that causes the plummeting of our blind test accuracy for multitask classification compared to that for single-task classification.

2.4 Multidetector OAM-Encoded D²NN for Repeatable Multitask Classification

Next, when considering the ability of OAM-encoded D²NN to perform parallel recognition of large batches of images, it is necessary to load the sequence of digits into the light field. In addition, the reason we use multiple detectors is to simultaneously measure the OAM spectrum of multiplexed OAM beams at the output plane, which cannot be realized by using a single detector. If we separate the OAM beams at the output and utilize multiple detectors for OAM detection, we can enhance the capability of the OAM-encoded D²NN to process multiple images and introduce multiple digits at the input for multitask classification. In addition, we can use the positional information between different detectors to encode the sequential information of the same digits in an array and achieve parallel recognition of repeatable digit tasks.

Therefore, we propose a multidetector OAM-encoded D²NN for repeatable multitask classification that can encode repeatable numerical order using spatial information to enhance the parallel ability of the diffractive network to process more complex information. Unlike the first two schemes that generate a single multiplexed OAM beam at the central location, multiple OAM beams are generated at discrete spatial locations in the output plane. The number of generated OAM beams is equal to the number of digits in the input array, facilitating the use of multiple detectors for identification and classification. Figure 4(b) shows a schematic demonstration of the four-detector OAM-encoded D²NN. When the four digits are modulated by the diffractive layers, they will produce OAM beams with the corresponding OAM modes at the specified spatial locations in the output layer. Figure 4(a) shows the amplitude and phase of the input two, three, and four digits at different positions in the input layer, diffractive layers, and output layer, respectively. It can be seen that the intensities of different output OAM beams are not uniformly distributed, which is similar to the problem encountered in single-detector OAM-encoded D²NN for single-task classification, and is caused by the limitation of the diffractive
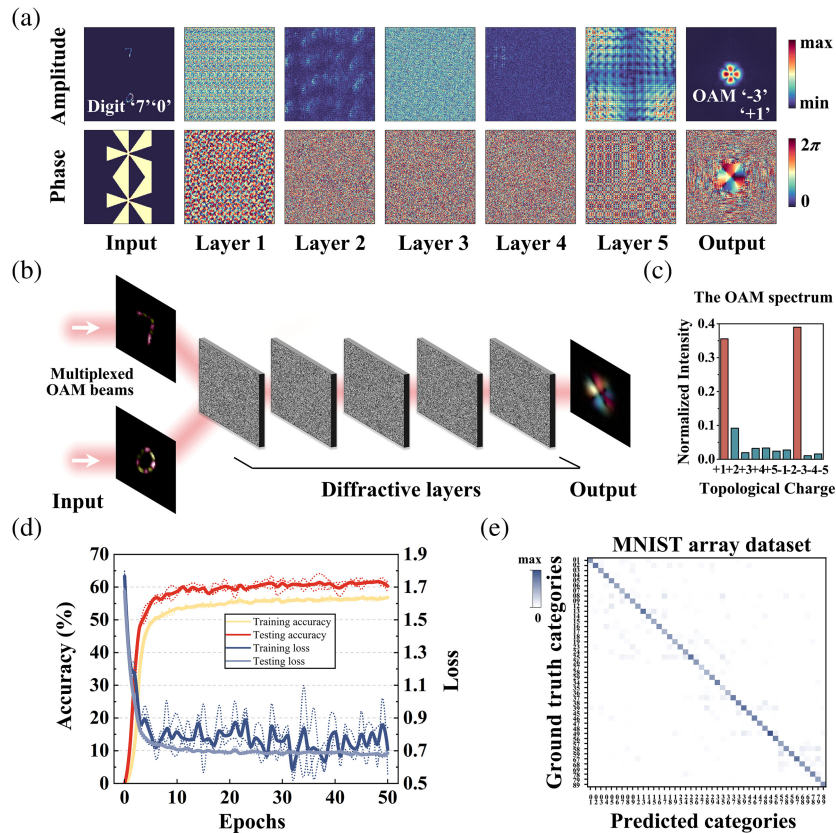
Fig. 3 (a) The amplitude and phase distribution of the OAM beams in the input plane, diffractive layers, and output plane. The input handwritten digits are "7" and "0," which correspond to the multiplexed OAM beams that produce "-3" and "+1" OAM modes. (b) Schematic of the light field modulation by single-detector OAM-encoded D^2NN for multitask classification. The OAM beam encodes two handwritten digits as the input. After undergoing OAM-encoded D^2NN modulation, it produces a new OAM beam corresponding to two modes at the same spatial location. (c) The OAM spectrum of the output OAM beams. The two OAM modes detected by the detector with the highest normalized intensity represent the assumed categories of the input digits, and their classes are indicated by the red bars. (d) Loss function and accuracy during training and testing. Solid lines indicate the average result of the three-function curve represented by the dashed line. (e) The confusion matrix summarizes the numerical classification result in the test set.

network's own modulation capability. In addition, it is shown in Fig. 4(a) that there is only a logical correspondence between our input and output layers for digital recognition, and no direct correspondence in the optical path propagation. When the digits "6" and "0" are entered, the intensity of the generated OAM mode $m = -2$ and $+1$ corresponding to their digit classification accounts for 46.55% and 69.77% of the OAM beam, respectively. When the arrays "6," "1," and "3" with repeatable digits are input, the normalized intensities of the corresponding OAM modes $m = -2$, $+2$, and $+4$ are 51.78%, 40.98%, and 45.20% of the output, respectively. And the OAM modes $m = +3$, $+2$, -3, and -4 corresponding to the array containing repeatable digits "2," "1," "7," and "3" account for 46.77%, 42.27%, 38.84%, and 34.73% of the total intensity, respectively. These proportions exceed the intensity accounted for by the other OAM modes [see Fig. 4(c)]. It can be seen that the multidetector OAM-encoded D^2NN can handle the parallel recognition task excellently when spatially separated OAM beams are generated at the output and jointly detected by the same number of detectors.

The accuracy curves obtained from successive iterative tests show that the multidetector OAM-encoded D^2NN achieves blind test accuracies of 70.94%, 52.41%, and 40.13% for two-digit, three-digit, and four-digit MNIST repeatable array data sets [see Fig. 5(a)]. Facing the same challenge as the single-detector OAM-encoded D^2NN for multitask classification, the rapid increase in the number of labels in the repeatable array data set further degrades the blind testing accuracy of the network. The two-digit, three-digit, and four-digit data sets have 100, 1000, and 10,000 labels, respectively. The difficulty is much higher than that of the original MNIST data set because it requires correctly classifying every digit in the array. In three-detector and four-detector OAM-encoded D^2NNs, there are too many labels consisting of different digits, and it is not feasible to display a pixel map of this size within the limited space for the inserted image. However, if we only capture a portion of the confusion matrix, we would sacrifice the comprehensiveness of all the data. Therefore, we choose a scaled-down version of the confusion matrix for the inserted image while employing a localized zoom approach [Fig. 5(b)]. In addition,
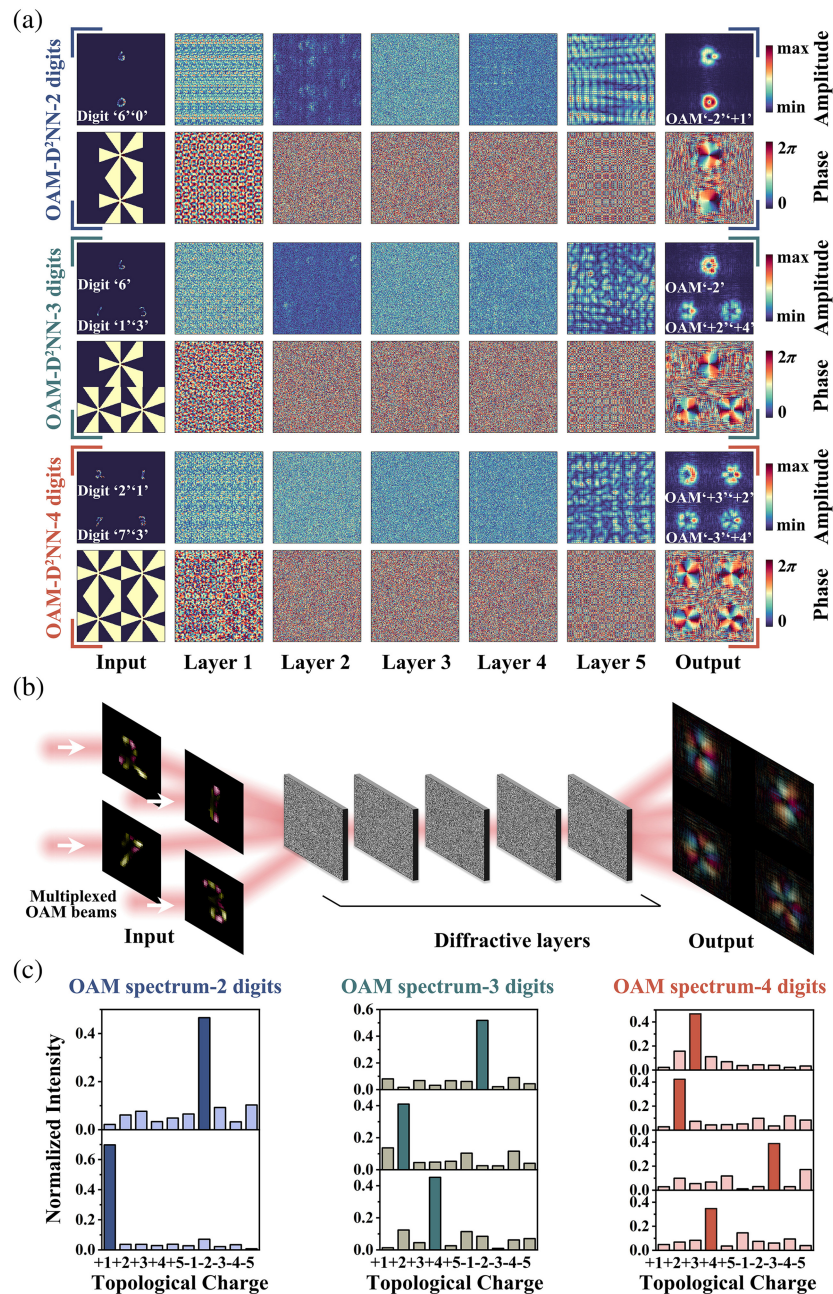
Fig. 4 (a) From top to bottom, the multidetector OAM-encoded D^2NN provides recognition for two digits, three digits, and four-digits, respectively. The amplitude and phase distribution of the OAM beams in the input plane, diffractive layers, and output plane. (b) Schematic of the light field modulation by four-detector OAM-encoded D^2NN for multitask classification. Each input OAM beam at different positions encodes only one digit and generates the corresponding OAM mode of that digit at the output, which is detected by a detector at a fixed position. (c) The OAM spectrum of the output OAM beams. The two blue OAM spectra correspond to the OAM beams generated by the two-detector OAM-encoded D^2NN, from top to bottom, respectively. The green OAM spectrum in the first row corresponds to the separate OAM beam in the first row of the three-detector OAM-encoded D^2NN, and the green OAM spectra in the second and third rows correspond to the two OAM beams from left to right in the second row, respectively. The four red OAM spectra are arranged in a sequential relationship from left to right and from top to bottom.

the results of the multidetector OAM-encoded D^2NN for repeatable multitask classification show that using more digits for parallel classification within the same array leads to a further decrease in classification accuracy. The ability of the OAM-encoded D^2NN to handle more digits can be improved by adopting certain approaches, such as increasing the size of the diffractive layer and expanding the number of neurons used for recognition.
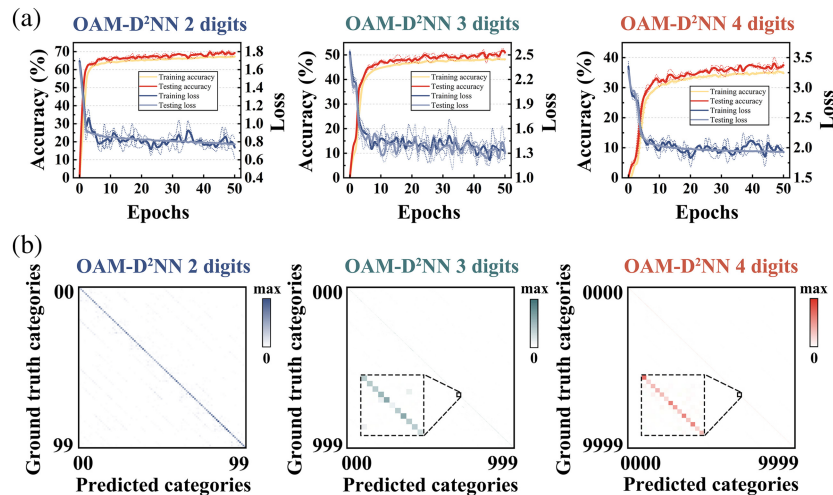
Fig. 5 (a) The loss function and accuracy function of the two-detector, three-detector, and four-detector OAM-encoded D^{2}NNs in training and testing are arranged from left to right. The solid line represents the average result of the function curves for the three simulations, which is represented by the dashed line. Their average accuracy in the test set is 70.94%, 52.41%, and 40.13%, respectively. (b) Confusion matrices of the three multidetector OAM-encoded D^{2}NNs, summarizing the numerical classification results of the test set. Due to the large number of pixel points in the confusion matrices of the three-detector and four-detector OAM-encoded D^{2}NNs, the confusion matrices are reduced and localized zoomed-in images are inserted.

3 Discussion and Conclusions

Experimental implementation of D^{2}NN typically uses a spatial light modulator to modulate the light source and 3D printing to fabricate metasurfaces designed by an electronic computer. Limited by the precision size of 3D printing, this fabrication method is typically only available for terahertz bands. There are two main challenges in building OAM encoded D^{2}NNs experimentally: sample fabrication and experimental measurement. Here, the OAM-encoded D^{2}NN operates at the wavelength of 1550 nm, which corresponds to pixel sizes of ~800 nm. The diffractive layer of the OAM-encoded D^{2}NN can be fabricated by micro/nanoprocessing technology compatible with CMOS technology, as the current state-of-the-art e-beam lithography technology has a fabrication resolution of only a few nanometers. However, there are still certain challenges left to be considered in the fabrication process due to the on-chip multilayer structures. These challenges may include issues related to overlay, alignment, and other aspects[68,69] that need to be solved with improved technology.

When detecting the spectrum of the output OAM beam, it can be analyzed using interferometric methods, diffractive methods, and other detection methods.[60,61,67] In terms of measuring the diffractive network, here we take the interferometric method as an example. This method can detect the OAM spectra of multiplexed OAM beams, not only the single OAM mode. The measurement details of the detector are outlined in Sec. 4.3. For the MNIST data set and the MNIST array data set, a single detector at the output plane of the diffractive network is sufficient for OAM spectrum analysis. However, for the MNIST repeatable array data set, we need to use multiple detectors to achieve simultaneous detection of different OAM modes corresponding to different categorized digits.

At the same time, the OAM-encoded D^{2}NNs require an interferometer detector with a high signal-to-noise ratio and high

sensitivity, considering reflections, material absorptions, scattering, and other loss issues; we can attempt to decrease the sensitivity and robustness requirements of the detector. One approach is to increase the intensity of the optical signal received by the detector, which can be achieved by reducing the number of layers to minimize absorption and reflection losses. Note that there is always a trade-off between classification accuracy and output efficiency. As we are dealing with an optical classification network, we only need the detected effective optical signal to meet the minimal requirements for classification. Despite the difficulties, we believe that there is great potential for realizing this scheme of OAM-encoded D^{2}NN as technology develops.

In summary, we have proposed and investigated an all-optical parallel classification using OAM mode-encoded diffractive networks, which can encode the spatial information of multiple objects as OAM modes of the VB. And then we utilize OAM spectra to analyze the OAM mode normalized intensity distribution for multitask optical classification. If the inference accuracy of the existing OAM-encoded D^{2}NN can be further improved, it can be extended from target recognition to other deep-learning tasks, such as multilabel classification and dynamic image recognition. We also envision introducing more OAM modes (this may require the use of a more advanced multimode OAM comb as a light source[60]) to solve more complex tasks. Finally, we expect that the OAM-encoded D^{2}NN can provide a new feasible pathway for all-optical parallel classification and OAM-based machine vision.

4 Appendix: Materials and Methods

4.1 Forward Propagation Model of the OAM-Encoded D^{2}NN

Traditional deep neural networks rely on forward propagation, backward propagation, and gradient descent algorithms for

brain-like electronic computation by continuously adjusting the weights of electronic neurons. The diffraction of light that occurs during propagation is very similar to the way neurons are connected in deep neural networks. Based on the Rayleigh–Sommerfeld diffraction,[70] each diffractive unit/neuron can be regarded as a coherent superposition of light propagating from every diffractive unit/neuron in the preceding diffractive layer. It can also be seen as the source of a secondary wave that is fully connected to the subsequent layer. The equation of light propagation between diffraction layers is given as

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

where $w_i^l(x, y, z)$ is the complex-valued field propagated to each diffractive unit located at (x, y, z) in layer $l + 1$'th by using the i'th diffractive unit located at (x_i, y_i, z_i) in layer l'th with a wavelength of λ as the wave source, $r = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$ and $j^2 = -1$. The light field function of the i'th neuron of the l'th layer u_i^l can be considered as

$$u_i^l(x_i, y_i, z_i) = \sum_{j \in N} u_j^{l-1}(x_j, y_j, z_j) \cdot t^l(x_i, y_i, z_i) \cdot w_i^{l-1}(x_i, y_i, z_i),$$

$$(3)$$

where N denotes all the pixels on the previous diffractive layer. $t^l(x_i, y_i, z_i)$ is the complex-valued modulation of the optical field by the l'th diffractive layer, which has the functional expression $t^l(x_i, y_i, z_i) = a^l(x_i, y_i, z_i) \cdot \exp[j\phi^l(x_i, y_i, z_i)]$, where a and ϕ denote the amplitude and phase coefficients, respectively, and both of which are trainable parameters in the diffractive networks, where a and ϕ are allowed in the range from 0 to 1 and 0 to 2π, respectively.

Due to the significant computational burden associated with solving the conventional D²NN model using the Rayleigh–Sommerfeld formula, the use of Fresnel scalar diffraction theory can effectively reduce the computational effort. This theory can replace the Rayleigh–Sommerfeld formula in the results under the conditions of the layer spacing we use. Here, we use the Fresnel scalar diffraction theory to construct the forward propagation model of OAM-encoded diffractive neural networks. The complex amplitude of the OAM beam of the i'th neuron of the l'th layer u_i^l can be considered as

$$u_i^l(x_i, y_i) = F^{-1}\{F[u_i^{l-1}(x_i, y_i) \cdot t^{l-1}(x_i, y_i)] \cdot H(f_x, f_y)\}, \quad (4)$$

$$H(f_x, f_y) = \exp[jk(z - z_i)] \cdot \exp[-j\lambda\pi(z - z_i)(f_x^2 + f_y^2)], \quad (5)$$

where F and F^{-1} denote the fast Fourier transform and reverse fast Fourier transform, respectively, which are functions that represent the transformation of the optical field between the spatial and frequency domains, where $H(f_x, f_y)$ is the transformation function in the frequency, which represents the propagation of the OAM beam in free space. $k = \frac{2\pi}{\lambda}$ represents the wave-number.

4.2 Error Analysis of OAM-encoded D²NN

In the main text, the OAM-encoded D²NN is based entirely on the ideal case with fixed parameters. When considering the experiments, different factors such as fabrication size errors, optical alignment errors, and material absorption may affect the performance of the diffractive network. Here, we present a systematic analysis of the various types of error problems that may be encountered by OAM-encoded D²NN.

4.2.1 Deviation analysis of the pixel size and the layer spacing

According to the Fresnel scalar diffraction theory, the spacing between layers of the diffractive network should be at least 10 times larger than the size of entire layer. Therefore, we grouped the pixel size and optical full-sized errors together for analysis. We assumed a deviation of ±20% in the manufacturing dimensions, which is much larger than the fabrication error of the CMOS machining process.[68,69] We considered an error range of 0.8 times the pixel size and the corresponding layer spacing, as well as a range of 1.2 times the pixel size and the corresponding layer spacing. As shown in Fig. 6(a), the accuracy of the OAM-encoded D²NN varies within 1% of this error range. Therefore, we believe that the errors in pixel size and layer spacing caused by processing and manufacturing do not affect the OAM-encoded D²NN.

4.2.2 Deviation analysis of the object misalignment

First, we consider the possible object misalignment error between the incident OAM beam and the digital mask. We introduced deviations of 2%, 4%, 6%, 8%, and 10% in both the horizontal and vertical directions of the object. For each of these object misalignment errors, we tested all five types of diffractive networks mentioned in our main text. As shown in Fig. 6(b), when the deviation of object misalignment is within 5% in both the horizontal and vertical directions, the accuracy of all OAM-encoded D²NNs, except for S-OAM-encoded D²NN-M (see Table 2 for the nomenclature), fluctuates within 1%. Therefore, our diffractive networks could ensure that the deviation of the incident beam from the digital mask does not exceed 5%, which is smaller than the range of fabrication error.[68,69]

In addition, we also observed an interesting phenomenon regarding the three-detector and four-detector OAM-encoded D²NN. Surprisingly, their accuracy seems to increase when the object misalignment error is around 5%. We hypothesize that this effect may be caused by misidentification of certain numbers when the incident beam deviates (e.g., when the OAM beam shifts horizontally to the right, it can cause the light intensity distribution of the number "8" to resemble that of the number "3" due to the nonuniform distribution of the light intensity of multiplexed OAM beams).

4.2.3 Deviation analysis of layer misalignment

Here, we selected two values for the misalignment error: 5% and 10%. This indicates that the layers would experience dislocations of 5% or 10% in random directions. As shown in Fig. 6(c), the horizontal coordinates represent the number of diffractive layers where the corresponding misalignment error occurred. It has been proven that the OAM-encoded D²NN is highly robust against layer alignment errors, with minimal impact on accuracy. In addition, to explore the limit of the OAM-encoded D²NN's sensitivity to layer alignment errors, we conducted
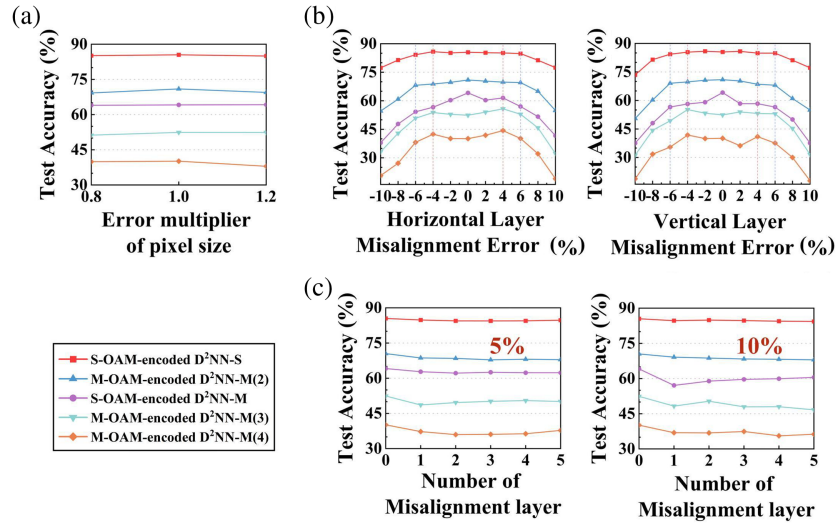
Fig. 6 The different colored curves represent different diffractive networks, as illustrated in the square diagram located in the lower left corner. (a) The deviation of the pixel size and the layer spacing. The horizontal coordinate represents the error range from 0.8 times the pixel size and the corresponding layer spacing to 1.2 times the pixel size and the corresponding layer spacing. (b) The analysis of the deviation of the object misalignment in horizontal and vertical directions. (c) The analysis of the deviation of the misalignment layer. The left image represents a random misalignment error of 5% for each layer, while the right image represents a random misalignment error of 10% for each layer.

Table 2 Various indices for single-detector OAM-encoded D²NN for single-task classification (S-OAM-encoded D²NN-S), single-detector OAM-encoded D²NN for multi-task classification (S-OAM-encoded D²NN-M), multidetector OAM-encoded D²NN for repeatable multitask classification (M-OAM-encoded D²NN-M).

	Training time (h)	Training loss	Training accuracy (%)	Test loss	Test accuracy (%)
S-OAM-encoded D²NN-S	12.74	0.402	84.30	0.343	85.43
S-OAM-encoded D²NN-M	5.69	0.708	57.42	0.667	64.13
M-OAM-encoded D²NN-M(2)	6.04	0.820	67.69	0.772	70.94
M-OAM-encoded D²NN-M(3)	4.09	1.345	48.94	1.238	52.41
M-OAM-encoded D²NN-M(4)	3.19	1.970	36.25	1.932	40.13

additional tests on the single-detector OAM-encoded D²NN for single-task classification with a 20% misalignment error (see Fig. 6). The accuracy of the OAM-encoded D²NN starts to exhibit a slight decline of 1% under these conditions. Consequently, we conclude that the performance of diffractive network can be reliably maintained as long as the alignment bit error between layers remains within 20% during sample processing and experimental testing.

4.2.4 Absorption error analysis of materials

As for the absorption effect, the material we used for the diffractive layer is silicon nitride, which corresponds to an extinction coefficient $k = 0$ in the wavelength of 1550 nm and does not have an absorption effect in the simulation. Considering that the fabricated silicon nitride material may have a small extinction coefficient during the experimental test, we assumed k to be 0.05 and incorporated it into the updated diffractive network for testing. After testing, the loss of D²NN is <1%. This may be due

to the thickness of the diffractive network is about 1 μm, which almost fails to produce any absorption.

4.2.5 Reflection error analysis of diffractive layers

The loss of the whole OAM-encoded D²NN system is mainly due to the reflection from the diffractive layers. When we assume that the beam enters the diffractive layer with positive incidence, the transmittance T can be calculated as

$$T = 1 - \frac{(n_2 - n_1)^2}{(n_2 + n_1)^2}, \tag{6}$$

where n_1 and n_2 are the refractive indices of the two media, respectively. In the wavelength range of 1550 nm, the refractive index of silicon nitride is approximated to be 2, while the refractive index of air is 1. Therefore, it can be calculated that the transmission of each diffractive layer is ~89%. So, the transmission efficiency of the entire diffractive network is estimated to

be around 56%. During the experimental test, the loss of the network will be higher than the theoretically calculated value. While we can attempt to reduce losses in the system, such as by reducing the number of layers in the diffractive network, thus minimizing absorption and reflection losses. Note that there is always a trade-off between classification accuracy and output efficiency. As we are dealing with an optical classification network, we only need to detect the effective optical signal against noise to meet the minimal requirements for classification. Despite the difficulties, we believe that there is great potential to realize this scheme of OAM-encoded D^2NN as technology develops.

4.3 OAM Spectrum Analysis

Multiple OAM states can appear in the same beam and are not limited to a single OAM mode. Similar to the spectrum that represents the intensity weights of different frequencies or wavelengths, the intensity weights of different OAM channels on the same beam are called the OAM spectrum. The spiral harmonic $\exp(jm\phi)$ is the eigenwave function of OAM, and the beam $E(r,\phi,z)$ can be represented by the spiral harmonic $\exp(jm\phi)$ in the column coordinates as

$$E(r,\phi,z) = \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{+\infty} a_l(r,z) \exp(jm\phi), \tag{7}$$

with the complex coefficient a_m

$$a_m(r,z) = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} E(r,\phi,z) \exp(-jm\phi)\mathrm{d}\phi, \tag{8}$$

where r represents the beam waist radius of the OAM beam, z represents the radial distance of the beam propagation, and m is the topological charge of the OAM. Thus, the intensity of the m'th order helical harmonic is

$$C_m = \int_0^{+\infty} |a_m(r,z)|^2 r\mathrm{d}r. \tag{9}$$

Since the value C_m is independent of the parameter z, the relative intensity of such a helical harmonic is

$$R_m = \frac{C_m}{\sum_{q=-\infty}^{+\infty} C_q}, \tag{10}$$

which is the OAM spectrum of $E(r,\phi,z)$. Among these considerations, detecting complex amplitude information in the output optical field is crucial. In simulations, acquiring the complex amplitude information of the output OAM beam is straightforward. However, in experimental detection, obtaining the complex amplitude information of the output OAM beam is not direct. Taking the interferometric method as an example, the phase information in the output optical field is obtained from the interference field between the beam to be measured and the probing Gaussian beam. Subsequently, when combined with the amplitude information detected by the CCD camera, we can obtain the complex amplitude information of the output beam. As long as the complex amplitude information of the output VB is obtained, we can further determine the corresponding OAM spectrum using the equations mentioned above. Therefore, we only need to obtain information on the complex amplitude of the

output OAM light in the simulation to obtain its corresponding OAM spectrum.

4.4 Preparation of Data Sets

The MNIST array data set and the MNIST repeatable array data set are used in the study to evaluate the discriminative criteria for multi-object classification in the proposed OAM-encoded D^2NN.

MNIST array data set: The digits in the MNIST data set are divided into 10 classes according to different labels, and the number of digits in each class is recorded. The labels of two random classes are arbitrarily selected using the shuffle function and combined into a label group containing two labels in no distinguishable order. Then, the data corresponding to the labels is selected separately from the data set, and the two selected data are stitched together into a new array. The generation of new arrays and label groups is performed in an iterative process until all digits in a given category have been selected. In addition, it is worth noting that the order of the digits also carries additional information. For example, the digits "0" and "1" result in a different light field distribution than the digits "1" and "0." The resulting MNIST array data set contains ~27,000 to 28,000 training sets and 4400 to 4500 test sets. The distribution of digits within each category in the MNIST data set is not uniform, which impacts the number of training and test sets. The MNIST array data set is regenerated after each round of the iterative process, and discarded data may be selected in subsequent rounds. As the number of training sessions increases, the probability of each digit appearing in the MNIST array data set gradually tends toward equality.

MNIST repeatable array data set: it builds on the MNIST array data set. Unlike the MNIST array data set, identical digits can be entered in the process of forming an array using random digits. The introduction of identical digits also requires encoding the order of combinations in the array. Due to the repeatability of the digits in the array within this data set, the MNIST repeatable array data set does not require rounding of digits.

4.5 Loss Function of OAM-Encoded D^2NN

We define the classical mean square error (MSE) loss function L_{MSE} to calculate the difference between the predicted output E and the ground truth target G, which can be expressed as

$$L_{\mathrm{MSE}} = \frac{1}{N} \sum_i^N |E_i - G_i|^2, \tag{11}$$

where N is the number of diffractive units in the output layer, which is set to 200×200 in the OAM-encoded D^2NNs.

In traditional D^2NN training, the softmax cross-entropy (SCE) loss function is often used in addition to the MSE loss function. The SCE loss function quantifies the degree of difference between two different probability distributions of the same random variable, which in diffractive networks is expressed as the difference between the true and predicted probability distributions. The smaller the value of the cross-entropy, the better the model prediction. The function L_{SCE} can be expressed as

$$E_i = \frac{y_i}{\sum_j y_j}, \tag{12}$$

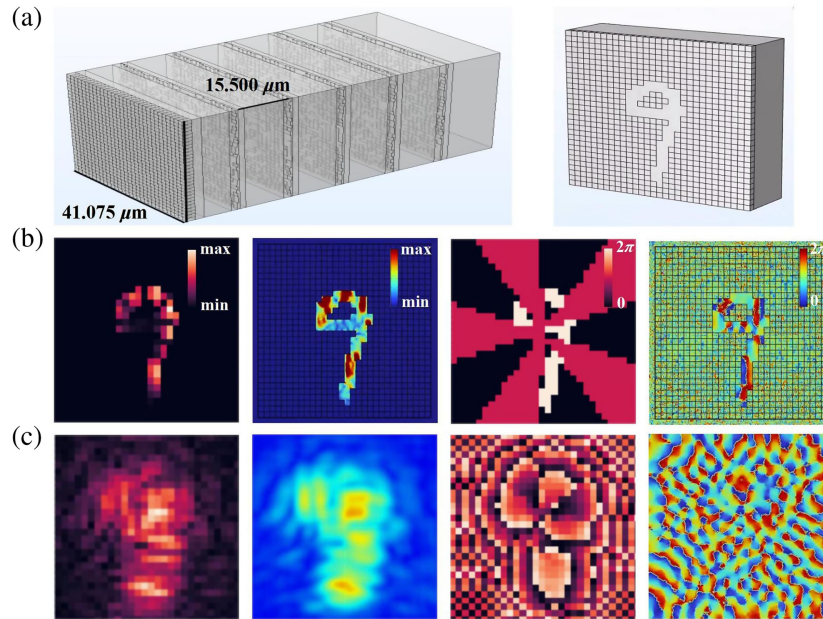$$L_{\mathrm{SCE}} = -\sum_i G_i \log E_i, \tag{13}$$

Fig. 7 (a) The left figure shows the geometrical model of the five layer D²NN with the pixel size of 50×50, and the right figure shows the mask model of the number "9" illuminated by the OAM beam. (b) The simulation of the incident OAM beam. (c) The simulation of the output plane by a one-layer D²NN with the pixel size of 30×30. (b), (c) The figures from left to right are amplitude distribution simulated with Python, amplitude distribution simulated with COMSOL Multiphysics software, phase distribution simulated with Python, and phase distribution simulated with COMSOL Multiphysics software.

where it is assumed that there is an array Y with a total of j numbers and y_i denotes the i'th element in Y with a softmax value of E_i. G represents the ground-truth target. The SCE loss function reduces the contrast of the output light in different spatial distributions, thereby effectively enhancing the inference accuracy of the classification. However, this performance improvement comes at the expense of the expected power efficiency of the network's output. In the case of OAM-encoded D²NNs, the output purity of the OAM beam is also a critical factor to consider. Therefore, pursuing higher accuracy at the expense of generating a loss function that compromises output purity is not a viable option. While the SCE loss function is useful in certain scenarios, it is not the optimal choice for OAM-encoded D²NNs, where both accuracy and output purity are important factors.

Table 2 shows the relevant performance parameters for our different network models. Our models were performed on a server [GeForce RTX 3080 Ti graphical processing unit (GPU, Nvidia Inc.), Intel(R) Core(TM) i9-10900K @3.70 GHz central processing unit (CPU, Intel Inc.) and 64 GB of RAM, running the Windows 10 operating system (Microsoft)] with Python (v3.9.13) and PyTorch (1.11.0+cu113) for simulation computations. All the models were trained with 50 epochs. All the models were optimized using the built-in Adam optimizer. The learning rate was set to 0.01.

4.6 Optical Demonstration of OAM-Encoded D²NN

The demonstration of optically simulating the entire model of the OAM-encoded D²NN is challenging to realize. Taking COMSOL Multiphysics software as an example, the size of the diffractive layer of OAM-encoded D²NN is $(200 \times 0.53 \times 1.55) = 164.3 \ \mu$m, and the total length of the model is $(1000 \times 1.55 \times 6) = 9300 \ \mu$m. The limit of the mesh delineation in COMSOL calculations ranges from one-quarter of a wavelength to one-sixth of a wavelength (i.e., between 0.2583 and 0.3875 μm). To simulate the full OAM-encoded D²NN, the required computer memory would be astronomical and unattainable. In order to show the consistency of our theoretical results in Python with the COMSOL Multiphysics software, we used COMSOL Multiphysics software to build a five-layer structure with 50 pixels $\times$ 50 pixels for model demonstration, as well as a single-layer structure with 30 pixels $\times$ 30 pixels for simulation. Figure 7(b) shows the light field distribution in the input side of the digit "9" when irradiated by a multiplexed OAM beam. Figure 7(c) shows the light-field distribution modulated by the diffractive layer at the output plane. It can be seen that the simulation results from the COMSOL Multiphysics software are highly consistent with the theoretical results obtained from Python. We believe that the simulation results can provide support and guidance for the experiments.

Disclosures

The authors declare no competing financial interests.

Code and 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. M. M. Waldrop, "The chips are down for Moore's law," *Nature* **530**, 144–147 (2016).
2. C. Sun et al., "Single-chip microprocessor that communicates directly using light," *Nature* **528**, 534–538 (2015).
3. D. Solli and B. Jalali, "Analog optical computing," *Nat. Photonics* **9**, 704–706 (2015).
4. K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with neuromorphic computing," *Nature* **575**, 607–617 (2019).
5. G. Wetzstein et al., "Inference in artificial intelligence with deep optics and photonics," *Nature* **588**, 39–47 (2020).
6. H. J. Caulfield and S. Dolev, "Why future supercomputing requires optics," *Nat. Photonics* **4**, 261–263 (2010).
7. K. Liao et al., "All-optical computing based on convolutional neural networks," *Opto-Electron. Adv.* **4**, 200060 (2021).
8. R. Salem, M. A. Foster, and A. L. Gaeta, "Application of space–time duality to ultrahigh-speed optical signal processing," *Adv. Opt. Photonics* **5**, 274–317 (2013).
9. Y. Shen et al., "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* **11**, 441–446 (2017).
10. J. Feldmann et al., "Parallel convolutional processing using an integrated photonic tensor core," *Nature* **589**, 52–58 (2021).
11. X. Lin et al., "All-optical machine learning using diffractive deep neural networks," *Science* **361**, 1004–1008 (2018).
12. G. Batra et al., "Artificial-intelligence hardware: new opportunities for semiconductor companies," McKinsey & Company, 2019, https://www.mckinsey.com/industries/semiconductors/our-insights/artificial-intelligence-hardware-new-opportunities-for-semiconductor-companies.
13. K. Liao et al., "Integrated photonic neural networks: opportunities and challenges," *ACS Photonics*, **10**(7), 2001–2010 (2023).
14. Y. Shi et al., "Nonlinear germanium-silicon photodiode for activation and monitoring in photonic neuromorphic networks," *Nat. Commun.* **13**, 6048 (2022).
15. J. Li et al., "Spectrally encoded single-pixel machine vision using diffractive networks," *Sci. Adv.* **7**, eabd7690 (2021).
16. X. Luo et al., "Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible," *Light Sci. Appl.* **11**, 158 (2022).
17. E. Goi et al., "Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip," *Light Sci. Appl.* **10**, 40 (2021).
18. K. Liao et al., "Matrix eigenvalue solver based on reconfigurable photonic neural network," *Nanophotonics* **11**, 4089–4099 (2022).
19. S. Li et al., "Programmable unitary operations for orbital angular momentum encoded states," *Natl. Sci. Open* **1**(3), 2097–1168 (2022).
20. J. Bueno et al., "Reinforcement learning in a large-scale photonic recurrent neural network," *Optica* **5**, 756–760 (2018).
21. T. Zhou et al., "Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit," *Nat. Photonics* **15**, 367–373 (2021).
22. H. Zhou et al., "Photonic matrix multiplication lights up photonic accelerator and beyond," *Light Sci. Appl.* **11**, 30 (2022).
23. R. Tang et al., "Ten-port unitary optical processor on a silicon photonic chip," *ACS Photonics* **8**, 2074–2080 (2021).
24. A. Annoni et al., "Unscrambling light—automatically undoing strong mixing between modes," *Light Sci. Appl.* **6**, e17110 (2017).
25. D. A. B. Miller, "Analyzing and generating multimode optical fields using self-configuring networks," *Optica* **7**, 794–801 (2020).
26. D. Mengu et al., "Analysis of diffractive optical neural networks and their integration with electronic neural networks," *IEEE J. Sel. Top. Quantum Electron.* **26**, 3700114 (2020).
27. J. Li et al., "Class-specific differential detection in diffractive optical neural networks improves inference accuracy," *Adv. Photonics* **1**, 046001 (2019).
28. T. Yan et al., "Fourier-space diffractive deep neural network," *Phys. Rev. Lett.* **123**, 023901 (2019).
29. D. Mengu, Y. Rivenson, and A. Ozcan "Scale-, shift-, and rotation-invariant diffractive optical networks," *ACS Photonics* **8**, 1, 324–334 (2020).
30. M. S. S. Rahman et al., "Ensemble learning of diffractive optical networks," *Light Sci. Appl.* **10**, 14 (2021).
31. O. Kulce et al., "All-optical information-processing capacity of diffractive surfaces," *Light Sci. Appl.* **10**, 25 (2021).
32. C. Liu et al., "A programmable diffractive deep neural network based on a digital-coding metasurface array," *Nat. Electron.* **5**, 113–122 (2022).
33. M. Zheng, L. Shi, and J. Zi, "Optimize performance of a diffractive neural network by controlling the Fresnel number," *Photonics Res.* **10**(11):2667 (2022).
34. T. Zhou et al., "*In situ* optical backpropagation training of diffractive optical neural networks," *Photonics Res.* **8**(8), 1323 (2020).
35. G. Qu et al., "All-dielectric metasurface empowered optical-electronic hybrid neural networks," *Laser Photonics Rev.* **16**, 2100732 (2022).
36. C. Qian et al., "Dynamic recognition and mirage using neuro-metamaterials," *Nat. Commun.* **13**, 2694 (2022).
37. T. Fu et al., "Photonic machine learning with on-chip diffractive optics," *Nat. Commun.* **14**, 70 (2023).
38. B. Bai et al., "All-optical image classification through unknown random diffusers using a single-pixel diffractive network," *Light Sci. Appl.* **12**, 69 (2023).
39. Z. Duan, H. Chen, and X. Lin, "Optical multi-task learning using multi-wavelength diffractive deep neural networks," *Nanophotonics*, **12**(5), 893–903 (2023).
40. T. Yan et al., "All-optical graph representation learning using integrated diffractive photonic computing units," *Sci. Adv.* **8**, eabn7630 (2022).
41. C. Qian et al., "Performing optical logic operations by a diffractive neural network," *Light Sci. Appl.* **9**, 59 (2020).
42. P. Wang et al., "Orbital angular momentum mode logical operation using optical diffractive neural network," *Photonics Res.* **9**, 2116–2124 (2021).
43. Y. Luo, D. Mengu, and A. Ozcan, "Cascadable all-optical NAND gates using diffractive networks," *Sci. Rep.* **12**, 7121 (2022).
44. M. Veli et al., "Terahertz pulse shaping using diffractive surfaces," *Nat. Commun.* **12**, 37 (2021).
45. 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).
46. Y. Luo et al., "Computational imaging without a computer: seeing through random diffusers at the speed of light," *eLight* **2**, 4 (2022).
47. Y. Chen et al., "Photonic unsupervised learning variational autoencoder for high-throughput and low-latency image transmission," *Sci. Adv.* **9**, eadf8437 (2023).
48. D. Mengu et al., "Snapshot multispectral imaging using a diffractive optical network," *Light Sci. Appl.* **12**, 86 (2023).
49. Y. Luo et al., "Design of task-specific optical systems using broadband diffractive neural networks," *Light Sci. Appl.* **8**, 112 (2019).
50. J. Li 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).

51. N. Bozinovic et al., "Terabit-scale orbital angular momentum mode division multiplexing in fibers," *Science* **340**, 1545–1548 (2013).
52. J. Wang et al., "Terabit free-space data transmission employing orbital angular momentum multiplexing," *Nat. Photonics* **6**, 488–496 (2012).
53. M. P. J. Lavery et al., "Detection of a spinning object using light's orbital angular momentum," *Science* **341**, 537–540 (2013).
54. G. Ruffato et al., "Diffractive optics for combined spatial- and mode-division demultiplexing of optical vortices: design, fabrication and optical characterization," *Sci. Rep.* **6**, 24760 (2016).
55. C. Kai et al., "The performances of different OAM encoding systems," *Opt. Commun.* **430**, 151–157 (2019).
56. H. Zhou et al., "Polarization-encrypted orbital angular momentum multiplexed metasurface holography," *ACS Nano* **14**(5), 5553–5559 (2020).
57. X. Fang, H. Ren, and M. Gu, "Orbital angular momentum holography for high-security encryption," *Nat. Photonics* **14**, 102–108 (2020).
58. Y. Shen et al., "Optical vortices 30 years on: OAM manipulation from topological charge to multiple singularities," *Light Sci. Appl.* **8**, 90 (2019).
59. S. Fu et al., "Universal orbital angular momentum spectrum analyzer for beams," *PhotoniX* **1**, 19 (2020).
60. S. Fu et al., "Orbital angular momentum comb generation from azimuthal binary phases," *Adv. Photonics Nexus* **1**(1), 016003 (2023).
61. Z. Huang et al., "All-optical signal processing of vortex beams with diffractive deep neural networks," *Phys. Rev. Appl.* **15**, 014037 (2021).
62. J. Zhang et al., "Polarized deep diffractive neural network for sorting, generation, multiplexing, and de-multiplexing of orbital angular momentum modes," *Opt. Express* **30**, 26728–26741 (2022).
63. P. Wang et al., "Diffractive deep neural network for optical orbital angular momentum multiplexing and demultiplexing," *IEEE J. Sel. Top. Quantum Electron* **28**(4), 7500111 (2022).
64. Z. Huang et al., "Orbital angular momentum deep multiplexing holography via an optical diffractive neural network," *Opt. Express* **30**, 5569–5584 (2022).
65. J. Guo et al., "Spatially structured-mode multiplexing holography for high-capacity security encryption," *ACS Photonics* **10**(3), 757–763 (2023).
66. H. Wang et al., "Intelligent optoelectronic processor for orbital angular momentum spectrum measurement," *PhotoniX* **4**, 9 (2023).
67. Y. Lecun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**, 2278–2324 (1998).
68. J. Mulkens et al. "Holistic approach for overlay and edge placement error to meet the 5 nm technology node requirements," *Proc. SPIE* **10585**, 105851L (2018).
69. W. H. Arnold, "Toward 3 nm overlay and critical dimension uniformity: an integrated error budget for double patterning lithography," *Proc. SPIE* **6924**, 692404 (2008).
70. L. Mandel and E. Wolf, "Some properties of coherent light," *J. Opt. Soc. Am.* **51**, 815–819 (1961).

Biographies of the authors are not available.