
 Research
 Optical Neural Networks—Article

Diffractive Deep Neural Networks at Visible Wavelengths

 Hang Chen^a, Jianan Feng^a, Minwei Jiang^a, Yiqun Wang^b, Jie Lin^{a,c,*}, Jiubin Tan^a, Peng Jin^{a,c,*}

^a Center of Ultra-precision Optoelectronic Instrument, Harbin Institute of Technology, Harbin 150001, China

^b Nanofabrication Facility, Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences, Suzhou 215123, China

^c Key Laboratory of Micro-Systems and Micro-Structures Manufacturing, Ministry of Education, Harbin Institute of Technology, Harbin 150001, China

ARTICLE INFO

Article history:

Received 10 February 2020

Revised 19 May 2020

Accepted 20 July 2020

Available online 13 February 2021

Keywords:

Optical computation

Optical neural networks

Deep learning

Optical machine learning

Diffractive deep neural networks

ABSTRACT

Optical deep learning based on diffractive optical elements offers unique advantages for parallel processing, computational speed, and power efficiency. One landmark method is the diffractive deep neural network (D²NN) based on three-dimensional printing technology operated in the terahertz spectral range. Since the terahertz bandwidth involves limited interparticle coupling and material losses, this paper extends D²NN to visible wavelengths. A general theory including a revised formula is proposed to solve any contradictions between wavelength, neuron size, and fabrication limitations. A novel visible light D²NN classifier is used to recognize unchanged targets (handwritten digits ranging from 0 to 9) and targets that have been changed (i.e., targets that have been covered or altered) at a visible wavelength of 632.8 nm. The obtained experimental classification accuracy (84%) and numerical classification accuracy (91.57%) quantify the match between the theoretical design and fabricated system performance. The presented framework can be used to apply a D²NN to various practical applications and design other new applications.

© 2021 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Deep learning is a type of machine learning method that can be used to achieve data representation, abstraction, and advanced tasks by simulating a multi-layer artificial neural network in a computer [1]. Recent advances in deep learning have attracted much attention. In some fields, deep learning performance has been shown to be superior to that of human experts. Deep learning has revolutionized the fields of artificial intelligence (AI) and computer science, and great advances have been made in these fields. Deep learning has been widely applied to computer vision [2], voice/image recognition [3–5], robotics [6], and other applications [7–9]. However, electronic and active deep learning is strongly limited by the von Neumann architecture in terms of processing time and energy consumption [10,11]. In the last several decades, optical information processing, which implements the operations of convolution, correlation, and Fourier transformation in an optical system, has been found to exhibit unique advantages for parallel processing and has been widely investigated [12–17]. Computer-based deep learning has been achieved in optical systems by use of diffractive optical elements, and optical deep

learning based on diffractive optical elements has been validated using image classification [18–21]. The terahertz diffractive deep neural network (D²NN) based on three-dimensional (3D) printing technology is the landmark method for optical deep learning [18].

Based on deep learning-based design and error back-propagation methods, D²NN is trained using a computer. During the training period, each pixel in the diffractive optical element layer is a neuron. The transmission coefficients (or complex amplitude distribution) for these pixels are optimized by the computer to control diffraction light from the input plane to the output plane to perform the required task. When this training phase is complete, these passive diffractive optical element layers can be physically fabricated. By stacking these layers together to form an all-optical network, these diffractive layers can execute the trained function without the use of energy, except for the coherent illumination light that is used to encode the input object's information and the output detectors. Various nonlinear functions, such as Fourier-space D²NN, where the optical nonlinearity is introduced using ferroelectric thin films [17], and linear functions, such as handwritten digits and a fashion product classifier [18–20] without the use of nonlinear optoelectronic materials, have been performed using a D²NN [18–21].

A 3D-printed terahertz D²NN has been proven to be a great success. The promising performance and validity of D²NN for terahertz

* Corresponding authors.

E-mail addresses: linjie@hit.edu.cn (J. Lin), p.jin@hit.edu.cn (P. Jin).

platforms has been verified in many studies [18–20]. Despite the huge advantages, terahertz itself suffers from some well-known limitations in practical applications, such as material losses [22] and limited interparticle coupling [22]. Therefore, there is a growing requirement to revise the terahertz scheme for use with visible or near-infrared [23] wavelengths. Extension of the working wavelength from the terahertz bandwidth range to visible light has the potential to offer more novel perspectives [24–29]. However, some restrictions and shortcomings must be addressed to adapt a D²NN from a long wavelength to a short wavelength. Some contradictions exist between the working wavelength, neuron size, and fabrication limitations. A shorter wavelength has a smaller neuron size, which makes the process more difficult, and traditional maximal half-cone diffraction angle theory cannot be used to overcome this contradiction [18–20].

In this work, a new general formula containing detailed analysis of the D²NN framework, including the different parameters required for the design space, is proposed to overcome the above-mentioned contradiction. This new formula is proposed by introducing a formula that includes related variables based on the traditional maximum half-cone diffraction angle theory. The proposed general formula is a quantitative analysis on how to expand D²NN for visible light sources such as a helium (He)–neon (Ne) laser source. A series of simulation analyses are designed to verify the proposed formula. As an example, a D²NN classifier is used with a subset of handwritten digits from the Mixed National Institute of Standards and Technology (MNIST) training dataset to reduce the training time and training complexity. The digital classifier is used to recognize handwritten digits from 0 to 9. In this situation, a phase-only, five-layer D²NN is used and a total of 32 000 neurons are used to prevent overfitting. The numerically blind testing accuracy for different cases with a new test dataset is used to verify the proposed formula.

Based on the proposed general formula, a novel visible light D²NN classifier was designed and further experimentally verified. The working light was a He–Ne laser with a wavelength of 632.8 nm. In contrast with existing D²NN classifiers [18–20], the proposed visible light D²NN classifier can classify handwritten digits from 0 to 9 even for a changing target, such as the digits being covered or altered, which are two common cases. The visible light D²NN's identification capability can be improved by extending the training dataset period. For this test, a total of 55 000 handwritten digits in the MNIST training dataset were extended to a new training dataset with 80 000 images of handwritten digits. The incremental 25 000 handwritten digits were used as the deformed MNIST training dataset, which were covered or altered. By using the five-layer, phase-only D²NN with five million neurons, a numerically blind testing accuracy of 91.57% was achieved in the new test dataset of 11 000 images. The visible light D²NN was fabricated using a multi-step photolithography-etching process on a silicon dioxide (SiO₂) substrate. The inputs for the experiment were 50 transmissive digital objects that were fabricated using micro-fabrication technology. A blind testing accuracy of 84% was achieved. The experimental classification accuracy (84%) and the numerical classification accuracy (91.57%) quantify the match between theoretical design and the fabricated system performance. The relatively small reduction in the performance of the experimental network compared to the numerical testing proves the validity of the design theory for the visible light D²NN.

By using these systematic advances for designing a D²NN, the reported method and the improvement set the state of the art for visible light D²NNs. Deep neural networks are sometimes called black boxes [30], as the hidden layers can be difficult to extract to explain how data is processed. The proposed D²NN may provide

some insights into this issue. Additionally, understanding the interaction between biological neurons in the human brain is of fundamental interest when building deep neural networks [31]. The proposed D²NN may provide insight into the current understanding of the interactions between biological neurons. The presented framework and theory can be used to apply D²NNs to various practical applications for human–computer interaction equipment and AI interaction devices and can promote progress in biology and computational science.

2. Material and methods

2.1. Deep learning and D²NN architecture

The proposed visible light D²NN follows optical diffraction theory and a deep learning structure [32–35]. Compared with traditional deep learning systems, the D²NN system has some differences: ① The D²NN system obeys optical diffractive theories, such as Huygens' principle and Rayleigh–Sommerfeld diffraction [36], for the forward-propagation model; ② the training loss function used in this system (i.e., the softmax-cross-entropy (SCE) loss function), is based on the light power incident on different detector regions in the output plane. Therefore, based on the calculated error with respect to the target output and according to the desired loss function, the network structure and its neuron phase values can be optimized using an error back-propagation algorithm.

Generally, in deep learning, the phase values for the neurons in each layer are iteratively adjusted (or trained) to perform a specific function by feeding training data to the input layer followed by computing the network's output through optical diffraction. Before a specific analysis of the forward-propagation model within a D²NN, some definitions are provided for the *l*th layer as follows (where *l* represents the layer number of the network):

- t^l is the *l*th layer's complex transmission function. The complex transmission coefficient of the *i*th neuron, where *i* represents the neuron located at (x_i, y_i, z_i) of layer *l* (where (x_i, y_i, z_i) is the spatial coordinate of the *i*th neuron along the *x*-, *y*-, and *z*-axis, respectively), is composed of amplitude and phase terms. It can be further defined as $t_i^l(x_i, y_i, z_i) = a_i^l(x_i, y_i, z_i) \exp(j\Phi_i^l(x_i, y_i, z_i))$, where $a_i^l(x_i, y_i, z_i)$ is the amplitude term, $\Phi_i^l(x_i, y_i, z_i)$ is the neuron phase term, *j* is an imaginary number and $j = \sqrt{-1}$. For a phase-only D²NN architecture, the amplitude $a_i^l(x_i, y_i, z_i)$ is assumed to be a constant value equal to 1.
- W^{l+1} is the secondary diffractive wave just after the *l*th layer. The secondary diffractive wave after the *i*th neuron located at (x_i, y_i, z_i) for layer *l* is W_i^{l+1} . The amplitude and phase of W^{l+1} are determined by the product of the input wave to layer *l* and its complex transmission coefficient t^l , which are complex-valued functions.
- Y^{l+1} is the output function when the secondary diffractive wave function W^{l+1} propagates across the distance between the diffractive layers *l* and *l* + 1. Y^{l+1} is also the input function for the next layer *l* + 1. When the input function Y^l is vertically irradiated at the *l*th diffraction layer, the complex amplitude distribution of W^{l+1} is the product of the transmission function of t^l and the input function Y^l :

$$W^{l+1} = Y^l \circ t^l \quad (1)$$

where \circ is the Hadamard product operation.

After the secondary diffractive wave W^{l+1} propagates from the diffractive layer *l* to the diffractive layer *l* + 1, a phase shift is introduced with a corresponding phase factor as follows:

$$\exp(jkd_l \cos \gamma) = \exp \left[jkd_l \sqrt{1 - (\cos^2 \alpha + \cos^2 \beta)} \right] \quad (2)$$

where d_l is the distance between the diffractive layer l and $l + 1$; α , β , and γ are the angles for the propagation direction of the secondary diffractive wave W^{l+1} along the x -, y -, and z -axis, respectively; k is the wave number and $k = 2\pi/\lambda$, λ is the wavelength.

Assuming that $u = \frac{\cos \alpha}{\lambda}$ and $v = \frac{\cos \beta}{\lambda}$, where u and v are spatial frequency, the corresponding phase factor after propagation of the secondary diffractive wave W^{l+1} from the diffractive layer l to the diffractive layer $l + 1$ can be expressed as follows:

$$\exp(jkd_l \cos \gamma) = \exp \left[jkd_l \sqrt{1 - \lambda^2(u^2 + v^2)} \right] \quad (3)$$

We then obtain

$$Y^{l+1} = W^{l+1} \cdot \exp \left[jkd_l \sqrt{1 - \lambda^2(u^2 + v^2)} \right] \quad (4)$$

Therefore, the overall analysis of the forward-propagation model within a D²NN is given as follows:

$$Y^{l+1} = (Y^l \circ t^l) \cdot \exp \left[jkd_l \sqrt{1 - \lambda^2(u^2 + v^2)} \right] \quad (5)$$

2.2. TensorFlow-based design for a D²NN and processing flow

The proposed visible light D²NN was realized using Python (v3.7) and the TensorFlow (v1.12.0, Google Inc., USA) framework. The proposed D²NN system was trained for 20 epochs using a desktop computer with a GeForce GTX 1080 Ti graphical processing unit (GPU) and an Intel®Core™ E5-2650 central processing unit (CPU) @2.00 GHz and 128 GB of random access memory (RAM), running the Windows 10 operating system (Microsoft Corporation, USA).

The trainable parameters in a diffractive neural network are the modulation values for each layer, which, here, were optimized using the back-propagation method of the adaptive moment estimation (Adam) optimizer with a learning rate of 10^{-3} . Furthermore, the number of network layers and the axial distance between these layers are also design parameters. The training time for a five-layer visible light D²NN to classify both unchanged and changed (covered or altered) handwritten digits was approximately 20 h.

The input digit objects were encoded based on the input amplitude into the D²NN and were fabricated by laser direct writing (LDW). The target objects were fabricated on a soda glass substrate. The glass substrate was first cleaned using acetone and isopropyl alcohol. The clear substrate was coated with a layer of chromium (Cr) with a thickness of a few hundred nanometres using electron beam evaporation. After spin-coating positive photoresist and a prebake process, the handwritten digit patterns were exposed using LDW technology. The exposed resist was stripped using a developer and the uncovered Cr was stripped using chrome mordant. Finally, any remnant resist was also cleaned using acetone and isopropyl alcohol.

To enhance the fabrication of the visible wavelength D²NN, a sigmoid function was applied to limit the phase value of each neuron to $0-\pi$, which enabled the neurons to be easily fabricated using a traditional multi-step photolithography-etching method. Before processing, the neuron phase values Φ need to be converted into a relative height map Δh ($\Delta h = \lambda\Phi/(2\pi\Delta n)$, where Δn is the refractive index difference between the fabricated substrate and air). The D²NN layers were fabricated onto a SiO₂ substrate using a similar cleaning process as above. After cleaning, equipment pre-treated with hexamethyldisilazane was used to change the surface activity of the SiO₂ substrate to enhance the adhesion

between the photoresist and the substrate. After spin-coating photoresist and exposure, the exposed resist was stripped using a developer. Then, after an oxygen plasma sizing treatment, magnetic neutral loop discharge etching was used. This process was repeated until the D²NN layer structure was achieved. More processing details can be found in Figs. S1 and S2 in Appendix A.

3. Results

By using a multi-step photolithography-etching process on a SiO₂ substrate, a D²NN was fabricated as five layers of diffractive optical elements, which were mounted as shown in Fig. 1. The identification capability of the visible light D²NN was improved by extending the target dataset for the training period. For a changing target, for example, a target being covered or altered, existing deep learning systems will falsely identify the target [18–20], even with sufficient object recognition accuracy improvements [20]. Therefore, our D²NN was trained as a digit classifier to perform automated classification of handwritten digits. The designed D²NN can classify unchanged number targets from 0 to 9 as well as targets that have been changed (by being covered or altered), as shown in Fig. 1(a). For these tasks, a phase-only transmission five-layer D²NN was designed by training 80 000 images, comprising 55 000 unchanged handwritten digits that were obtained from the MNIST training dataset and 25 000 changed handwritten digits (i.e., covered or altered digits) which were derived from the deformed MNIST training dataset. The input digits were encoded into the D²NN based on the input amplitude. The diffractive neural network was trained to map the input digits to eleven detector regions, which were marked by different numbers, as shown in Fig. 1(a). The unchanged input digits from 0 to 9 were mapped to the No. 0 to No. 9 detector regions, respectively. The changed input digits were all mapped to the No. X detector region. These detectors are also shown in Fig. 1(a). The classification criterion was used to find the detector with the maximum optical signal.

Once the training was completed, the improved D²NN digit classifier was numerically tested using 11 000 additional images, which were not used as training image sets and comprised 10 000 unchanged handwritten digits that were obtained from the MNIST test dataset and 1 000 deformed handwritten digits (i.e., covered or altered digits) that were derived from the deformed MNIST test dataset. The improved system achieved a blind classification accuracy of 91.57%.

Using the numerical phase values for the neurons in each layer, as shown in Fig. 1(b), the designed five-layer D²NN was processed. The phase of the neurons in each layer was physically encoded based on the relative thickness of each layer point. Therefore, the diffractive optical element for the D²NN was processed using a multi-step photolithography-etching process on a SiO₂ substrate. The experimental schematic for the whole five-layer D²NN is shown in Fig. 1(c). The experimental setup based on a He–Ne laser (25-STP-912-230, Melles Griot, USA) is shown in Figs. 1(d) and (e). In the experiment, a He–Ne laser beam was collimated by lenses 1 and 2, and a pinhole was used as a filter. The collimated He–Ne laser beam was used to illuminate the target objects in the input plane. Using the five-layer D²NN, the diffractive field for the output plane was detected by a charge coupled device (CCD) (Beijing Daheng Imaging Vision Co., Ltd., China). Since the training set contains a large quantity of samples, each layer contains one million neurons (1000×1000), and there are five million neurons in total in the five-layer D²NN. The wavelength of the applied He–Ne laser was 632.8 nm. The power of the applied He–Ne laser was 5 mW. Each neuron had a size of approximately $4 \mu\text{m}$ and each layer had an area of $4 \text{ mm} \times 4 \text{ mm}$. The distance between two adjacent diffractive layers is approximately 5 cm. The details for the alignment of the five-layer D²NN can be found in Fig. S3 in Appendix A.

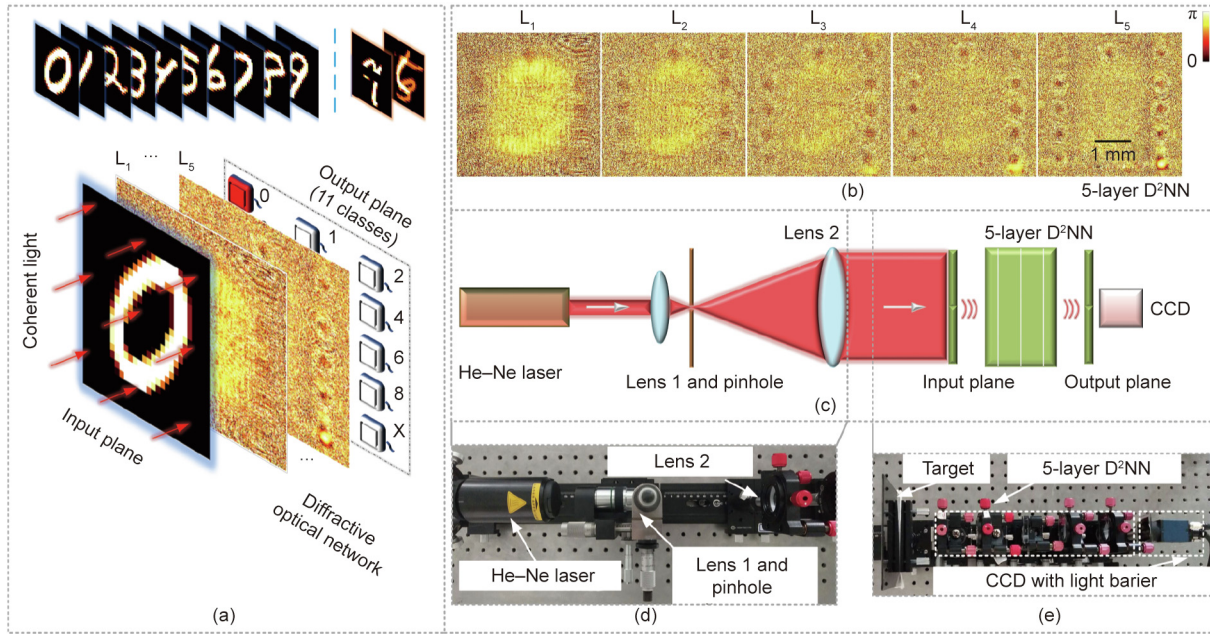


Fig. 1. Schematic diagram and experimental setup of the visible light D²NN. (a) Schematic diagram of the classifier used for the unchanged handwritten digit targets from 0 to 9 and the changed handwritten digit goals, such as the covered digit 7 and the altered digit 5. The spatial distribution of the detectors is also shown in this figure. (b) Numerical phase values for the neurons of the five layers L₁, L₂, L₃, L₄, and L₅. (c) The experimental schematic. (d, e) The experimental setup. CCD: charge coupled device.

For the numerical testing of the 11 000 test images, the classification accuracy of the designed five-layer D²NN was determined to be 91.57%. The confusion matrix is given in Fig. 2(a) and shows the details and distribution of the correctly identified examples and the incorrectly identified examples. For the 50 digital objects fabricated by LDW, the experimental blind testing accuracy was found to be 84%. The relatively small reduction in the performance of the experimental network compared to the numerical testing indicates that the design theory is correct. The confusion matrix in Fig. 2(b) shows the experimental details for examples of correct and incorrect identification. A CCD with a specially designed light

barrier was applied to each illuminated input object to obtain the D²NN output. The transmission regions for the light barrier correspond to the detector positions from No. 0 to No. X, respectively. The remaining regions are opaque. The first step in this test was to assess the recognition capability of the unchanged and changed handwritten test numbers. A handwritten 3, an altered handwritten 3, a handwritten 4, and a covered handwritten 4 were chosen as the input objects, as shown in Fig. 3(a). The simulated results and the experimental results in Fig. 3(b) indicate that the visible light D²NN can be used to easily classify the deformed object inputs. As shown in Fig. 3(c), the energy distribution shows that

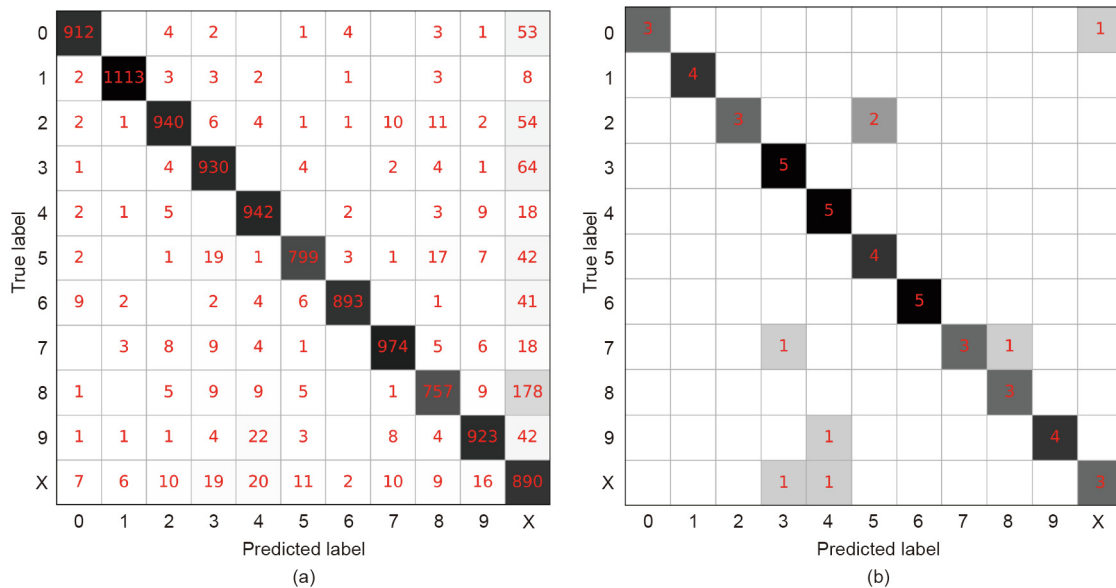


Fig. 2. Confusion matrix for the simulated and experimental results. (a) Confusion matrix for the simulated results. Numerical testing of the five-layer D²NN design achieves a classification accuracy of 91.57% over around 11 000 different test images. (b) Confusion matrix for the experimental results obtained for 50 different handwritten digits prepared by LDW. The classification accuracy is approximately 84%.

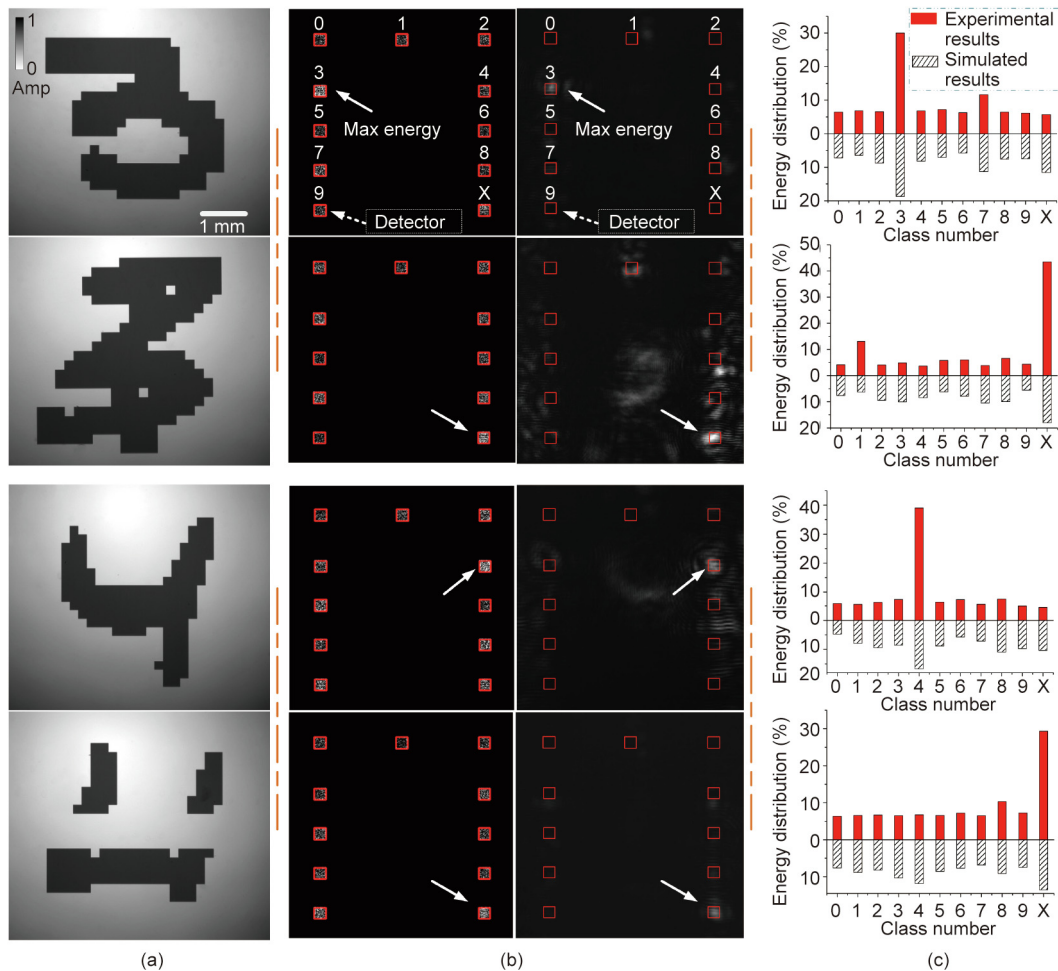


Fig. 3. Handwritten digit classifier for a visible light D²NN. (a) Objects under an optical microscope, including a handwritten 3, an altered handwritten 3, a handwritten 4, and a covered handwritten 4. Amp: amplitude. (b) Simulated results and experimental results showing that the handwritten digit classifier D²NN successfully classifies handwritten input digits based on 11 different detector regions at the output plane of the network, each corresponding to one digit. As an example, the output of the handwritten input of 3 and 4 are focused onto the No. 3 and No. 4 detectors, as indicated by the white arrows. The altered and covered handwritten input of 3 and 4 are all indicated by the No. X detector. Max: maximal. (c) Energy distribution percentage for our experimental results and simulated results, which demonstrates the success of the fabricated diffractive neural network and its inference capability.

the system can identify the maximum optical signal for the correct detector. The second step in this test was to use four different forms of the handwritten digit 6 as input objects, as shown in Fig. 4. The simulated results and the experimental results in Fig. 4(c) demonstrate that the fabricated diffractive neural network and the inference capability are valid. The average intensity distribution at the output plane of the visible light D²NN can converge the maximum input energy to the corresponding detector assigned to that digit.

In summary, the proposed D²NN illuminated by a He–Ne laser was demonstrated to successfully recognize unchanged targets (from 0 to 9) and changed targets (i.e., targets that are covered or altered) at a visible wavelength of 632.8 nm. Additionally, the proposed visible D²NN system was shown to have a transfer learning ability, as shown in Fig. 5. When the laser is passed directly into the D²NN system without passing through any handwritten digits, the existing D²NN system [18–20] will still diffract light to a digital detector. This indicates that incident light is misjudged to be a number. When the laser is directly incident onto the proposed D²NN system, as shown in Fig. 5(a), the proposed visible light D²NN system focuses the incident light into the No. X detector, which indicates that an incorrect number has been identified, which is not part of the classification set. The experimental results

shown in Fig. 5(b) show strong agreement with the simulated results.

The demonstrated D²NN can be used to address the contradictions that occur when adapting from a long wavelength to a visible light source. The quantitative analysis performed here demonstrates the building of a visible light D²NN and addresses the existing contradictions between wavelength, neuron size, and processing difficulty.

Connectivity between layers is a dominant factor that directly influences the diffraction of neurons. Therefore, the information transfer and the inference performance of the D²NN were determined. A fully connected D²NN can achieve sufficient information transfer and optical interconnection between neurons. A fully connected network requires that the diffraction angle of all neurons should be large enough to optically cover the diffractive optical element in the next layer. The maximal half-cone diffraction angle of a neuron (φ_{\max}) governed by wavelength and the neuron size can be qualitatively described for a fully connected structure as follows:

$$\varphi_{\max} = \sin^{-1}[\lambda/(2d_f)] \quad (6)$$

where d_f is the neuron size.

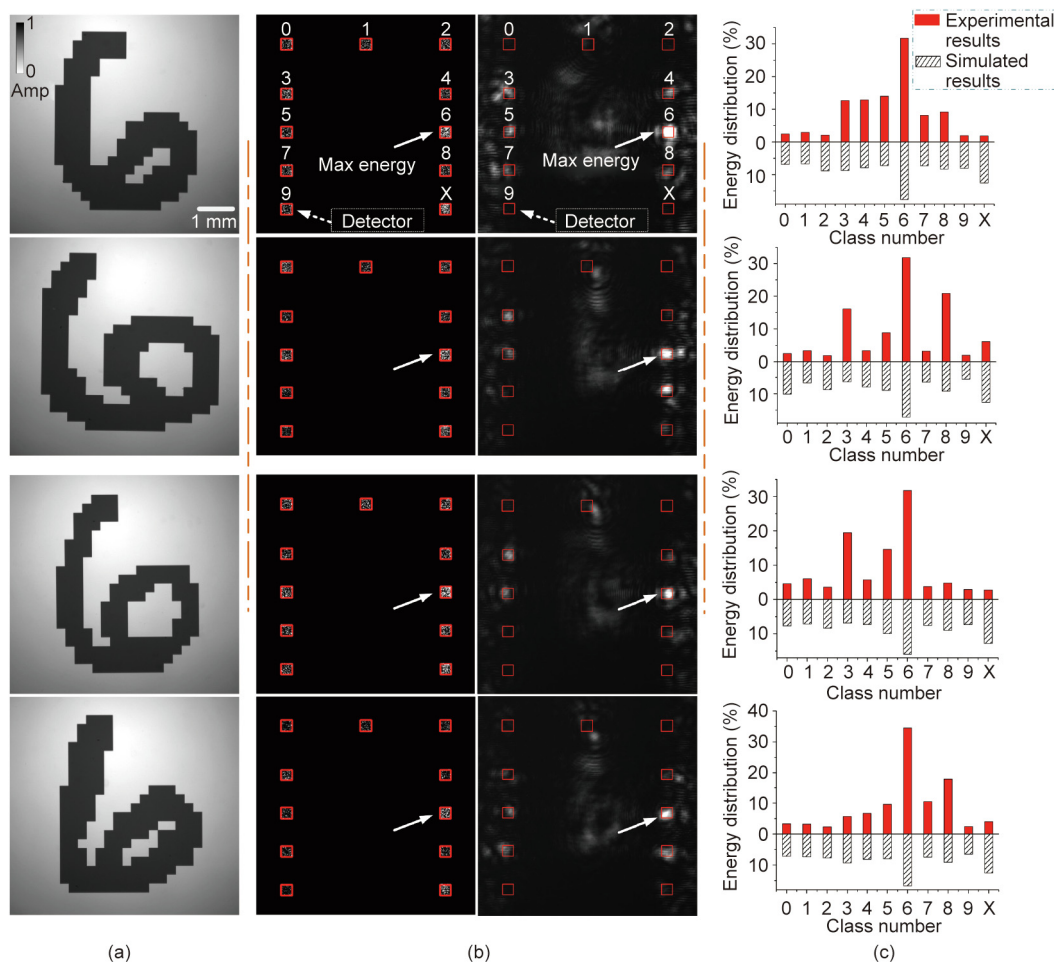


Fig. 4. Handwritten digit classifier for a visible light D^2NN . (a) Objects under an optical microscope, including four different forms of the handwritten 6. (b) Simulated results and experimental results showing that the handwritten digit classifier D^2NN successfully classifies different types of handwritten input digits. Four different forms of 6 were all focused onto the No. 6 detector, as indicated by the white arrows. (c) Energy distribution percentage for our experimental results and simulated results, which demonstrate the success of the fabricated diffractive neural network and its inference capability.

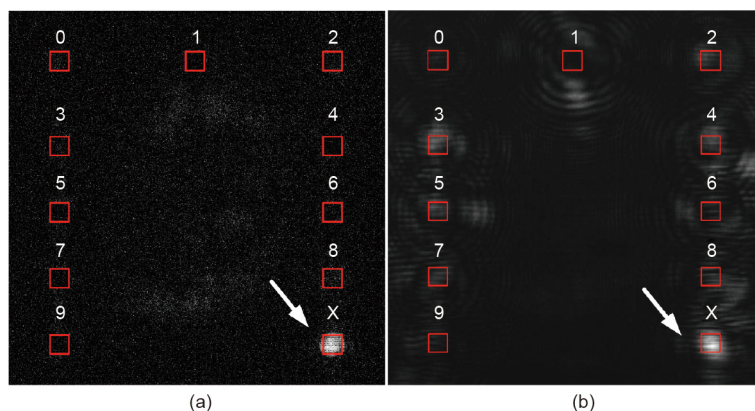


Fig. 5. Verification of the transfer learning ability of the proposed D^2NN . (a) Simulated results for the light-field distribution in the output plane when a plane wave passes directly into our system without passing through any handwritten digits. Most of the light is concentrated in the No. X region, which indicates that an incorrect number or incorrect case has been identified. (b) Experimental results for the light-field distribution in the output plane. The experimental results are in good agreement with the simulated results.

To obtain a large diffraction angle, it is necessary to have a small neuron size and a long wavelength. In previous work [18], a terahertz source wavelength of 0.75 mm, neuron size of approximately

0.4 mm, and maximal half-cone diffraction angle of approximately 70° were used. However, for visible light, the wavelength of the He–Ne laser used here is 632.8 nm, which is approximately 1200

times smaller than the terahertz wavelength. To obtain a 70° half-cone diffraction angle, the neuron size should be less than 330 nm, which is also 1200 times smaller than the neuron size for the terahertz bandwidth. The maximal unit size of 330 nm requires a complicated fabrication technique, which can result in a contradiction between wavelength, neuron size, and processing difficulty. Therefore, a general method should be applied when designing visible light D²NNs.

For a propagation distance D between two adjacent diffractive layers, the radius R of the diffraction spot of each neuron can be expressed as follows:

$$R = D \times \tan \varphi_{\max} \quad (7)$$

If d_f is the neuron size and N is the number of neurons in each diffractive layer, the side length w of each diffractive layer is

$$w = \sqrt{N \times d_f^2} \quad (8)$$

where the diffraction layer is assumed to be a square [18–20]. To obtain a fully connected D²NN, the condition of $R \geq w$ needs to be met. Therefore, based on the traditional maximal half-cone diffraction angle theory, a new formula is proposed as follows:

$$\begin{cases} \varphi_{\max} = \sin^{-1}[\lambda/(2d_f)] \\ D \geq \sqrt{N} \cdot d_f \cdot \sqrt{(4d_f^2/\lambda^2) - 1} \end{cases} \quad (9)$$

The improved formula can be used to quantitatively analyse the D²NN connectivity. A fully connected D²NN has better inference performance when the parameters satisfy Eq. (9). This formula indicates that the connectivity of a D²NN is affected by wavelength, neuron size, number of neurons, and the distance between layers. The contradiction between wavelength, neuron size, and processing difficulty can be alleviated by adjusting the number of neurons and spacing but using a longer wavelength. This formula provides a general case for the application of a D²NN to any wavelength.

The experimental parameters in Figs. 2–5 were chosen using Eq. (9) and the accuracy of this new formula was confirmed by the experimental results. To further verify the proposed formula, a series of simulation analyses were performed. To reduce the training time and complexity, a phase-only, five-layer D²NN was trained as a digital classifier to recognize only handwritten digits from 0 to 9 using a subset of the MNIST training image dataset, as shown in Fig. 6 (a). The training set contained 10 000 handwritten digit images (from 0 to 9); there were approximately 1000 images of each type of handwritten digit, which were randomly selected. These 10 000 input digits were encoded to the amplitude of the input field into the D²NN. The diffractive network was trained to map the input digits to ten detector regions, with one region for each digit. The classification criterion sought to find the detector with the maximum optical signal. After training, the D²NN digit classifier design was numerically tested using 500 images, which were also randomly selected from the MNIST test dataset and not contained within the training or validation image sets. The blind testing accuracy for the test set was used to verify the new proposed theory.

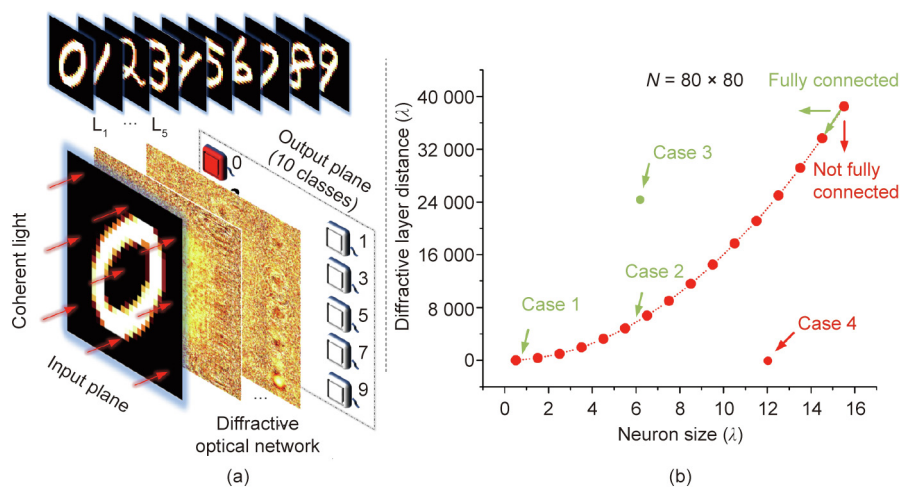
A quantitative analysis of the D²NN connectivity was performed, and the fitting curve for Eq. (9) is given in Fig. 6(b). For example, in order to prevent overfitting, the number of neurons in each layer was assumed to be 6400 (80 × 80) based on previous experience. The connectivity space for the D²NN was divided using Fig. 6(b), taking the relationship between the wavelength, distance, and neuron size into consideration. Once the parameters are within or above the fitting curve, as indicated by the green arrow,

the D²NN will realize full connectivity and perfect inference is guaranteed. For the case marked by the red arrow, the D²NN cannot achieve full connectivity. In Fig. 6(b), cases 1 and 2 are within the fitting curve, case 3 is above the fitting curve, and case 4 is below the fitting curve. The blind testing accuracies for cases 1 to 3 are all above 90%, while the accuracy for case 4 is approximately 0.1%. The confusion matrix for cases 1–4 are shown in detail in Figs. 6(d)–(g), respectively. These results prove the accuracy of the improved theory for the connectivity. The improved theory can offer a quantitative analysis for building a D²NN and demonstrates the performance advantages of a fully connected D²NN. The simulated results shown in Fig. 6(c) are consistent with previous studies [18–20]. By comparing cases 1 and 2, it can be seen that the proposed fully connected theory can overcome the contradiction between neuron size and processing difficulty by adjustment of the distance. The D²NN performance over a long distance D ($5.7 \times 10^3 \lambda$) and a large neuron size d_f (6λ) is consistent with that for a short distance D (15λ) and a small neuron size d_f (0.53λ), since the neuron size d_f is adjusted based on the distance D , which reduces the processing difficulty. We also analysed the influence of alignment errors between diffractive layers and the phase depth error for the diffractive layer. Further details can be found in Figs. S4 and S5 in Appendix A.

4. Discussion and conclusions

In this work, a general model for a D²NN at visible wavelengths was proposed. A visible wavelength D²NN can be used to overcome some of the drawbacks of the terahertz bandwidth and has many potential practical applications [25–29]. However, there are some restrictions and shortcomings, which make it challenging to change the bandwidth from terahertz to the visible light region. The first difficulty is the contradiction between the working wavelength, neuron size, and fabrication limitations. Shorter wavelengths require smaller neuron sizes, which make the processing more complex. A general theory that includes a revised formula was proposed to overcome these contradictions. A series of simulation analyses were designed that were able to successfully verify the proposed formula. Based on this theory, a novel visible light D²NN classifier was used to recognize unchanged targets (handwritten images of digits ranging from 0 to 9) as well as changed targets (i.e., covered or altered targets) at a visible wavelength of 632.8 nm. A numerical classification accuracy of 91.57% was obtained and is highly matched with an experimental classification accuracy of 84%, proving that both the theoretical analysis and the designed system can be successfully used.

Although there has been some recent success implementing deep neural networks for optical platforms [18–22,24], an all-optical design has not yet been fully demonstrated and realized. For example, computers are still required for the training process and the advantages of low energy consumption and high speed offered by optical information processing have not yet been realized. Additionally, applications for optical deep learning techniques are still emerging and many early attempts [18–20] use standard machine learning models, which may not be the best choice for an optical deep learning design. Other learning paradigms, such as unsupervised learning [37], generative adversarial networks [38], and reinforcement learning [39,40], should also be integrated into an optical neural network. It is expected that faster and more accurate optical deep learning frameworks will be proposed in the future, which may be able to offer capabilities that go beyond even current human knowledge.



Case	Neuron size (λ)	Diffractive layer distance (λ)	Blind testing accuracy (%)
Case 1	0.53	15	90.0
Case 2	6.00	$\sim 5.7 \times 10^3$	91.0
Case 3	6.00	$\sim 2.4 \times 10^3$	91.2
Case 4	12.00	15	~ 0.1

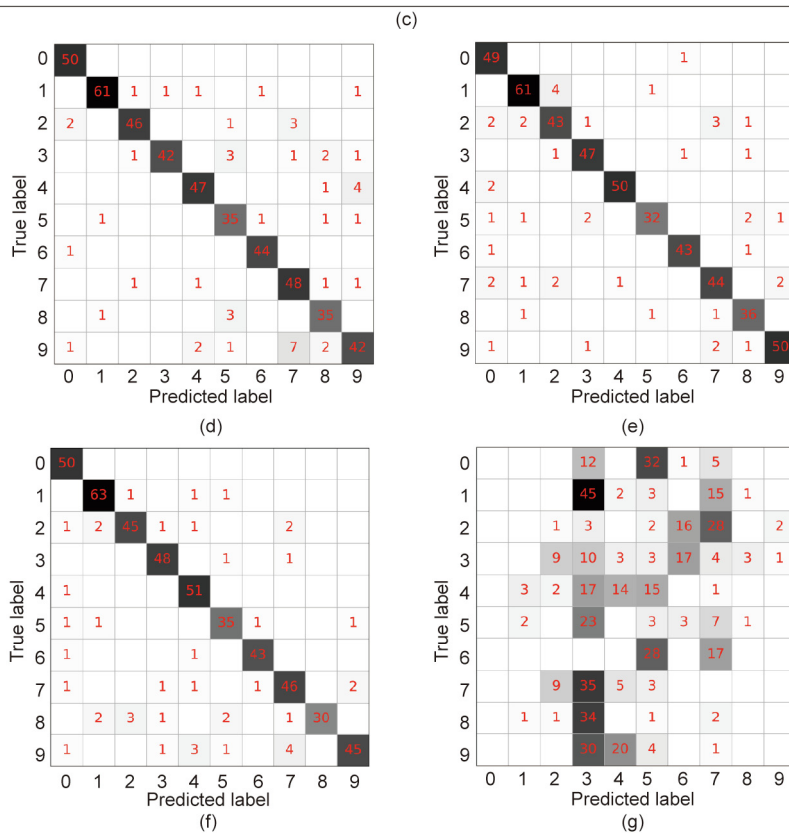


Fig. 6. Schematic diagram of the D²NN classifier used to verify the improved Eq. (9). (a) Schematic diagram of the D²NN classifier used to reduce the training time for the unchanged handwritten digit targets from 0 to 9. The location of each detector is displayed. (b) The fitting curve. (c) Four cases. The blind testing accuracy for different parameters in (c) were investigated and the confusion matrix is shown in (d) case 1, (e) case 2, (f) case 3, and (g) case 4, which demonstrate the success of the revised theory.

Acknowledgements

This research was supported in part by National Natural Science Foundation of China (61675056 and 61875048).

Compliance with ethics guidelines

Hang Chen, Jianan Feng, Minwei Jiang, Yiqun Wang, Jie Lin, Jiubin Tan, and Peng Jin declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2020.07.032>.

References

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [2] Bohannon J. Helping robots see the big picture. *Science* 2014;346(6206):186–7.
- [3] Hinton G, Deng L, Yu D, Dahl G, Mohamed AR, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 2012;29(6):82–97.
- [4] Schofield D, Nagrani A, Zisserman A, Hayashi M, Matsuzawa T, Biro D, et al. Chimpanzee face recognition from videos in the wild using deep learning. *Sci Adv* 2019;5(9):eaaw0736.
- [5] Barbastathis G, Ozcan A, Situ G. On the use of deep learning for computational imaging. *Optica* 2019;6(8):921–43.
- [6] Giusti A, Guzzi J, Ciresan DC, He FL, Rodriguez JP, Fontana F, et al. A machine learning approach to visual perception of forest trails for mobile robots. *IEEE Robot Autom Lett* 2016;1(2):661–7.
- [7] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016;529(7587):484–9.
- [8] Topol EJ. A decade of digital medicine innovation. *Sci Transl Med* 2019;11(498):eaaw7610.
- [9] Akhtar N, Mian A. Threat of adversarial attacks on deep learning in computer vision: a survey. *IEEE Access* 2018;6:14410–30.
- [10] Markram H. The blue brain project. *Nat Rev Neurosci* 2006;7(2):153–60.
- [11] Hines ML, Carnevale NT. The NEURON simulation environment. *Neural Comput* 1997;9(6):1179–209.
- [12] Deng R, Liu X. Optical multiplexing: tunable lifetime nanocrystals. *Nat Photonics* 2014;8(1):10–2.
- [13] Zijlstra P, Chon JWM, Gu M. Five-dimensional optical recording mediated by surface plasmons in gold nanorods. *Nature* 2009;459(7245):410–3.
- [14] Li X, Ren H, Chen X, Liu J, Li Q, Li C, et al. Athermally photoreduced graphene oxides for three-dimensional holographic images. *Nat Commun* 2015;6(1):6984.
- [15] Ren H, Li X, Zhang Q, Gu M. On-chip noninterference angular momentum multiplexing of broadband light. *Science* 2016;352(6287):805–9.
- [16] Deng R, Qin F, Chen R, Huang W, Hong M, Liu X. Temporal full-colour tuning through non-steady-state upconversion. *Nat Nanotechnol* 2015;10(3):237–42.
- [17] Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J, et al. Fourier-space diffractive deep neural network. *Phys Rev Lett* 2019;123(2):023901.
- [18] Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Yi, Jarrahi M, et al. All-optical machine learning using diffractive deep neural networks. *Science* 2018;361(6406):1004–8.
- [19] Mengu D, Luo Y, Rivenson Y, Ozcan A. Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J Sel Top Quantum Electron* 2019;26(1):3700114.
- [20] Li J, Mengu D, Luo Y, Rivenson Y, Ozcan A. Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *Adv Photonics* 2019;1(4):46001.
- [21] Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, et al. Nanophotonic media for artificial neural inference. *Photonics Res* 2019;7(8):823–7.
- [22] Yao K, Unni R, Zheng Y. Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale. *Nanophotonics* 2019;8(3):339–66.
- [23] Goi E, Gu M. Laser printing of a nano-imager to perform full optical machine learning [presentation]. In: Conference on Lasers and Electro-Optics/Europe; 2019 Jun 23–27; Munich, Germany; 2019.
- [24] Gu M, Fang X, Ren H, Goi E. Optically digitalized holography: a perspective for all-optical machine learning. *Engineering* 2019;5(3):363–5.
- [25] Yamazato T, Takai I, Okada H, Fujii T, Yendo T, Arai S, et al. Image-sensor-based visible light communication for automotive applications. *IEEE Commun Mag* 2014;52(7):88–97.
- [26] Khan LU. Visible light communication: applications, architecture, standardization and research challenges. *Digit Commun Netw* 2017;3(2):78–88.
- [27] Uysal M, Ghassemlooy Z, Bekkali A, Kadri A, Menouar H. Visible light communication for vehicular networking: performance study of a V2V system using a measured headlamp beam pattern model. *IEEE Veh Technol Mag* 2015;10(4):45–53.
- [28] Arsalan M, Hong HG, Naqvi RA, Lee MB, Kim MC, Kim DS, et al. Deep learning-based iris segmentation for iris recognition in visible light environment. *Symmetry* 2017;9(11):263.
- [29] Komine T, Nakagawa M. Fundamental analysis for visible-light communication system using LED lights. *IEEE Trans Consum Electron* 2004;50(1):100–7.
- [30] Benitez JM, Castro JL, Requena I. Are artificial neural networks black boxes? *IEEE Trans Neural Netw* 1997;8(5):1156–64.
- [31] Zhang Q, Yu H, Barbiero M, Wang B, Gu M. Artificial neural networks enabled by nanophotonics. *Light Sci Appl* 2019;8:42.
- [32] Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A. Deep learning microscopy. *Optica* 2017;4(11):1437–43.
- [33] Li Y, Xue Y, Tian L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* 2018;5(10):1181–90.
- [34] Lyu M, Wang W, Wang H, Wang H, Li G, Chen Ni, et al. Deep-learning-based ghost imaging. *Sci Rep* 2017;7:17865.
- [35] Nehme E, Weiss LE, Michaeli T, Shechtman Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* 2018;5(4):458–64.
- [36] Goodman JW. Introduction to Fourier optics. Colorado: Roberts & Company Publishers; 2005.
- [37] Barlow HB. Unsupervised learning. *Neural Comput* 1989;1(3):295–311.
- [38] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems; December 8–13 2014; Montreal, QC, Canada; 2014. p. 2672–80.
- [39] Bueno J, Maktoobi S, Froehly L, Fischer I, Jacquot M, Larger L, et al. Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* 2018;5(6):756–60.
- [40] Naruse M, Terashima Y, Uchida A, Kim SJ. Ultrafast photonic reinforcement learning based on laser chaos. *Sci Rep* 2017;7:8772.