

DOI: [10.29026/oea.2024.230005](https://doi.org/10.29026/oea.2024.230005)

Pluggable multitask diffractive neural networks based on cascaded metasurfaces

Cong He¹, Dan Zhao², Fei Fan², Hongqiang Zhou^{1,3}, Xin Li¹, Yao Li⁴, Junjie Li⁴, Fei Dong⁵, Yin-Xiao Miao⁵, Yongtian Wang^{1*} and Lingling Huang^{1*}

Optical neural networks have significant advantages in terms of power consumption, parallelism, and high computing speed, which has intrigued extensive attention in both academic and engineering communities. It has been considered as one of the powerful tools in promoting the fields of imaging processing and object recognition. However, the existing optical system architecture cannot be reconstructed to the realization of multi-functional artificial intelligence systems simultaneously. To push the development of this issue, we propose the pluggable diffractive neural networks (P-DNN), a general paradigm resorting to the cascaded metasurfaces, which can be applied to recognize various tasks by switching internal plug-ins. As the proof-of-principle, the recognition functions of six types of handwritten digits and six types of fashions are numerical simulated and experimental demonstrated at near-infrared regimes. Encouragingly, the proposed paradigm not only improves the flexibility of the optical neural networks but paves the new route for achieving high-speed, low-power and versatile artificial intelligence systems.

Keywords: optical neural networks; diffractive deep neural networks; cascaded metasurfaces

He C, Zhao D, Fan F et al. Pluggable multitask diffractive neural networks based on cascaded metasurfaces. *Opto-Electron Adv* 7, 230005 (2024).

Introduction

Deep learning is a form of machine learning that attempts to imitate the principles of the human brain for data interpretation. It has been applied to many specific tasks, including image classification^{1,2}, image encryption³, intelligent photonic devices^{4,5}, speech recognition^{6,7}, and language translation⁸. However, deep learning is a data-driven algorithm that requires frequent

reading and writing of large amounts of data using existing electronic computers. The calculation abilities of electronic computers are limited by the von Neumann architecture, which stores data and programs separately⁹. The bottleneck of computing performance caused by the mismatch between data reading and processing speed and the huge energy consumption caused by frequent data reading and writing have become blocks to the further development of artificial intelligence.

¹Beijing Engineering Research Center of Mixed Reality and Advanced Display, Key Laboratory of Photoelectronic Imaging Technology and System of Ministry of Education of China, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; ²Institute of Modern Optics, Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Nankai University, Tianjin 300350, China; ³Department of Physics and Optoelectronics, Faculty of Science, Beijing University of Technology, Beijing 100124, China; ⁴Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100191, China; ⁵Beijing Aerospace Institute for Metrology and Measurement Technology, Beijing 100076, China.

*Correspondence: YT Wang, E-mail: wyt@bit.edu.cn; Huang LL, E-mail: huanglingling@bit.edu.cn

Received: 14 January 2023; Accepted: 24 April 2023; Published online: 26 July 2023



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024. Published by Institute of Optics and Electronics, Chinese Academy of Sciences.

In the last several decades, optical neural networks (ONNs)^{10–18} have provided a solution by exploiting the unique properties of light and possess the advantages of high speed, high parallelism, and low energy consumption. Recently, deep neural networks have been implemented in optical systems using diffractive optical elements and validated in imaging processing and object recognition. The terahertz diffractive deep neural networks (D²NN) fabricated by 3D printing is the landmark example for ONNs¹⁹. Each unit structure of D²NN is regarded as a neuron, and the interconnections of neurons between layers are realized by diffraction of light. Metasurfaces, composed of subwavelength elements, are a novel type of two-dimensional planar structures, which provide promising platform for ultrathin flat optics compared to traditional diffractive optical elements^{20,21}. The amplitude and phase of light can be controlled simultaneously by changing the size, arrangement and shape of meta-atoms inside the metasurface. The realization of D²NN with metasurfaces is helpful to realize miniaturized and multifunctional intelligent integrated devices²².

Currently, D²NN has been widely studied for various tasks, including broadband pulse shaping²³, optical logic gate operation²⁴, and OAM beam multiplexing and demultiplexing^{25,26} among others^{27–29}. In addition, some D²NNs with improved structures have been applied to wider fields, such as the multi-view D²NN array scheme for 3D object recognition³⁰, where each D²NN corresponds to a 3D object view, ensemble learning of D²NN³¹, which introduces passive filters in space or Fourier space to preprocess input information and achieve higher-precision image classification, and Fourier-space D²NN³², which places the diffractive layer on the Fourier plane of the optical system to achieve all-optical saliency detection. However, once these architectures are trained and fabricated, they cannot be changed. If extra tasks need to be achieved, the parameters of the entire network have to be retrained. This process consumes a lot of computing resources to optimize the network parameters. Although there have been some studies on the reconfigurability problem of D²NN, such as programmable electromagnetic metasurface³³, optoelectronic fusion computing architecture³⁴, hardware-software co-design architecture³⁵, on-chip polarization multiplexed diffractive neural networks. These methods introduced additional energy consumption and required complex experimental setups. Recently, the on-chip polarization multiplexing neural network has been proposed to achieve multitasking in

the visible light band³⁶. However, this approach has limited performance with a single-layer metasurface.

Here, we proposed the pluggable diffractive neural networks (P-DNN), which can realize the switching of various recognition tasks such as handwritten digits and fashions by switching the pluggable components in the network. It improves the flexibility of network design while effectively reducing the consumption of computing resources and training time. We designed two-layer cascaded metasurfaces³⁷ to demonstrate the capabilities of P-DNN by using handwritten digits and fashions as input, respectively. Considering the experimental feasibility, the phase-only metasurfaces were used for verification. The experimental classification accuracies of handwritten digits and fashions classification tasks exceed 91.3% and 90.0%, while similar classification accuracies (91.8% and 90.2%) are obtained in P-DNN experimental verification. P-DNN is a general model for various classification tasks and provides an alternative for reconfigurability problems of D²NN. In addition, P-DNN can be used as an integrated component of artificial intelligence systems with different functions to provide low-energy, high-speed computing for specific tasks in the future, such as microscopy imaging, and autonomous driving assistance.

Results and discussion

The framework of the proposed P-DNN can be used to recognize various types of datasets such as handwritten digits and fashions, as shown in Fig. 1. We took the object to be classified as input, phase-encoding neurons as hidden layers and discretized detection plane as output. Our P-DNN can be divided into two parts: the first layer is a common layer for preprocessing input information, and the second layer is an alternative task-specific classification layer. We divide the detection output plane that corresponds to handwritten digits into six discrete regions, representing digits 0–5. When illuminating plane wave to the object (a mask of a specific shape), the amplitude-encoded incident information of equal phase was obtained. The diffractive light was focused on the corresponding horizontally arranged detection area through a two-layer P-DNN. Then, the fashion classification was implemented by replacing the second plugin layer. The detection plane was divided into six vertically arranged discrete regions representing six fashions (T-shirts, trousers, coats, sneakers, bags, and ankle boots). Our chosen training dataset is a subset of two classic

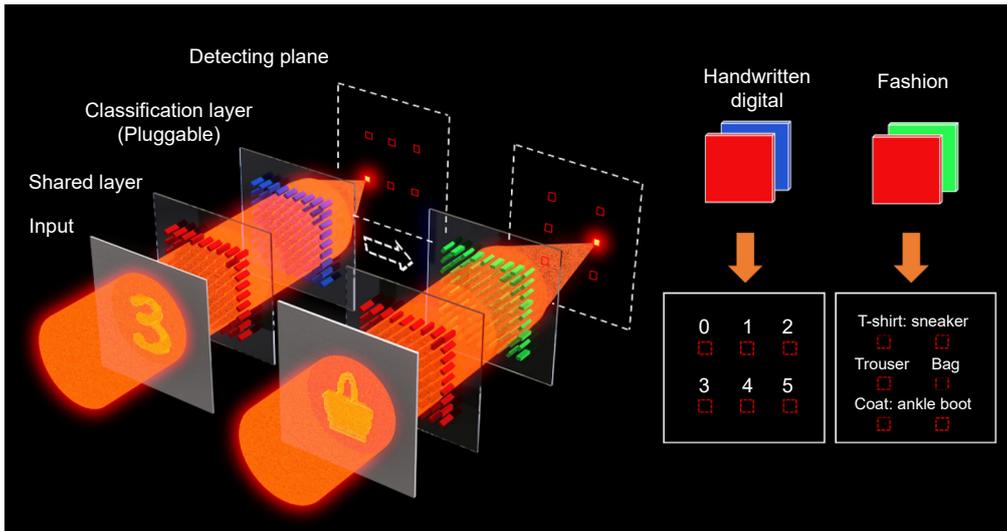


Fig. 1 | Concept of pluggable diffractive neural networks (P-DNN) for multiple tasks. P-DNN is composed of common layers (marked in red) and classification layers (marked in blue and green, respectively). The recognition of handwritten digits and fashion datasets can be achieved by switching plugins of classification layers. When parallel light is encoded as a specific input and passed through a two-layer pluggable D²NN, the light can be focused on a specified region of the detection plane to achieve classification.

machine learning datasets, that is, Modified national institute of standards and technology (MNIST)³⁸ and Fashion-MNIST³⁹.

The P-DNN system is trained based on the optical diffraction theory. According to the Huygens-Fresnel principle⁴⁰, every point of the wavefront can be regarded as a secondary spherical wave source. Each meta-atoms within the metasurface can be treated as an optical neuron, connecting to the neurons in the next layer by diffraction.

According to the Rayleigh-Sommerfeld diffraction theory⁴¹, the complex field $U(\mathbf{r}^{l+1})$ from l th to $(l + 1)$ th layer can be expressed as:

$$U(\mathbf{r}^{l+1}) = t(\mathbf{r}^l) \iint_{\mathcal{S}} U(\mathbf{r}^l) \cdot h(\mathbf{r}^{l+1} - \mathbf{r}^l) dx dy, \quad (1)$$

for the first hidden layer with $l = 1$, $U(\mathbf{r}^l)$ is the transmitted light encoded by the amplitude of the input layer. The complex field is modulated by the spatially varying complex transmittance $t(\mathbf{r}^l) = a^l e^{i\phi^l}$, where a^l and ϕ^l are the amplitude and the phase of $t(\mathbf{r}^l)$. The impulse response $h(\mathbf{r}^{l+1} - \mathbf{r}^l)$ can be defined as :

$$h(\mathbf{r}^{l+1} - \mathbf{r}^l) = \frac{z^{l+1} - z^l}{R^2} \left(\frac{1}{2\pi R} - \frac{1}{j\lambda} \right) \exp\left(j \frac{2\pi R}{\lambda}\right), \quad (2)$$

where λ is the illumination wavelength, $R = \sqrt{(x^{l+1} - x^l)^2 + (y^{l+1} - y^l)^2 + (z^{l+1} - z^l)^2}$, and $j = \sqrt{-1}$.

We explored an optimization algorithm flow based on transfer learning. The training criterion is to maximize each normalized signal corresponding to the detection

region, while minimizing the total signal outside the detection region. Considering the feasibility of the experiment, a phase-only modulation method was adopted and the sigmoid function was used to restrict the phase parameter in the range of $0-2\pi$. During optimization, we used a mean squared error (MSE) loss function to evaluate the difference between the output and truth by light intensity of different detector regions. According to the loss function, the phase parameters are randomly optimized by using stochastic gradient descent and error back-propagation algorithm. More detailed model training and derivation are described in the Methods section and Supplementary information.

More specifically, the P-DNN training process can be divided into two steps (Fig. 2). Firstly, the common layer (MS1) and classification layer (MS2) were simultaneously trained using handwritten digital images from the MNIST datasets for the task of handwritten digits recognition. The handwritten digit inputs were trained to map to six longitudinally distributed regions representing 0–5 handwritten digits. Next, the common layer network parameters were fixed, and the parameters of the fashion classification layer (MS3) were trained using images of fashions from the Fashion-MNIST datasets. The fashion inputs were trained to map to six distributed regions representing different categories of fashions (T-shirts, trousers, coats, sneakers, bags, and ankle boots).

According to the multi-task P-DNN design process, three phases can be optimized by using handwritten

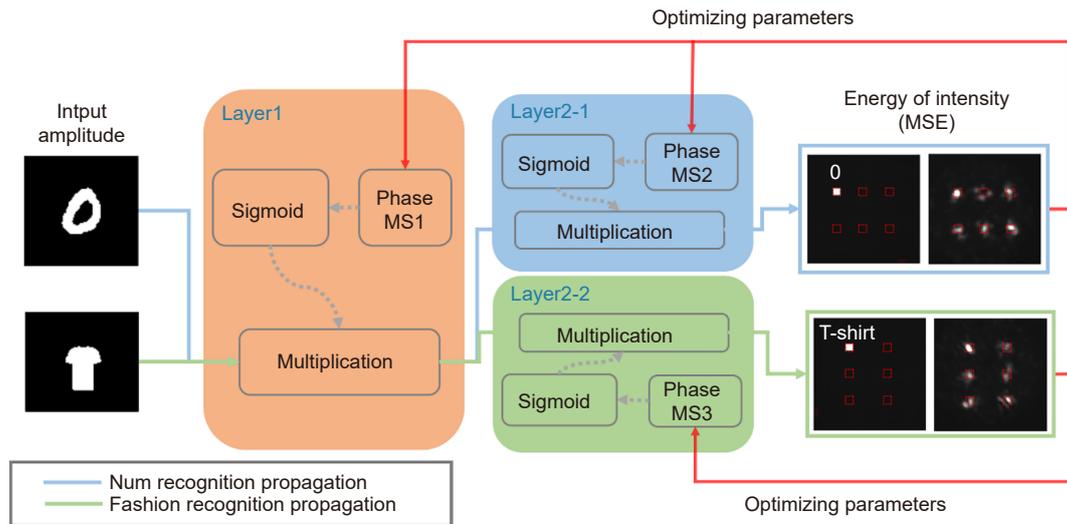


Fig. 2 | Flowchart of multi-task P-DNN design. The information of the input object is encoded into the amplitude channel, which propagates in free space. The propagated complex field is multiplied by the phase at each layer before being passed to the next layer. The network parameters are optimized according to the mean square error (MSE) of the output field energy. The sigmoid function is used to constrain the phase of each neuron. In the first training, the parameters of the common layer and the classification layer need to be trained simultaneously. In the subsequent training of other tasks, only the parameters of the classification layer need to be optimized. MS: metasurface

digital datasets and fashion datasets, corresponding to one sharing layer and two classification layers respectively. As a proof-of-concept, the cascaded metasurfaces are designed to realize corresponding phase modulation. The metasurfaces are composed of rectangular amorphous silicon nanofin deliberately designed on a glass substrate (Fig. 3(a)). The Berry phase modulation mechanism provides a dispersionless azimuthal angle dependent full phase control for circularly polarized light when it converts to its opposite helicity⁴².

The rigorous coupled wave analysis (RCWA) simulation method was used to design and optimize amorphous silicon nanofins. The electromagnetic response of each periodic array of nanostructures can be calculated. The height of the nanofins is fixed at 600 nm and the period is 500 nm in the x and y directions. The incident wavelength is 800 nm. Then, we obtained the simulated magnitudes of the circularly cross- and co-circularly polarized transmission coefficients of the nanofins by sweeping the length and width of the nanofins in 5 nm steps from 70 nm to 300 nm (Fig. 3(b)). Finally, we choose a nanofin with length L of 210 nm and width W of 135 nm (marked with white dots) to construct metasurfaces. The angle between the long axis of the nanofin and the x -axis of the substrate is φ as shown in Fig. 3(c). When the incident light is set as circularly polarized light, the phase modulation value obtained by rotating the nanofin is explored under cross circularly polarized

state. The relationship between the additional phase and amplitude is related to the rotation angle of the nanofin, shown in Fig. 3(d). The additional phase can cover the range from $0-2\pi$.

We build experimental system to demonstrate the performance of metasurface-based multi-task P-DNN (Fig. 4(a)). Considering the characteristics of the geometric phase modulation mechanism, a linear polarizer and a $1/4$ wave plate were used to generate the polarization state. The beam carrying the target information was generated by digital micromirror devices (DMD). In order to avoid the diffraction of the input beam during transmission, the $4f$ system was used to image the encoded target images to the front of the metasurface with distance of 4 mm as set in design. The pattern of the detection plane was collected and amplified by the microscope objective and the polarization state was filtered by the second set of quarter-wave plates and a linear polarizer. Finally, the camera was used to detect the images of the output plane.

We used two 3D displacement platforms to help align the cascaded metasurfaces. The distance between the two metasurfaces is set to 500 μm . Furthermore, a large pixel consisting of a 10×10 array of nanofins was intentionally created by repeatedly placing nanofins with the same azimuthal angle. The size of the metasurface is $500\times 500 \mu\text{m}^2$ and contains 100×100 pixels. The side length of each super-pixel is 5 μm . Figure 4(b) exhibits the scanning

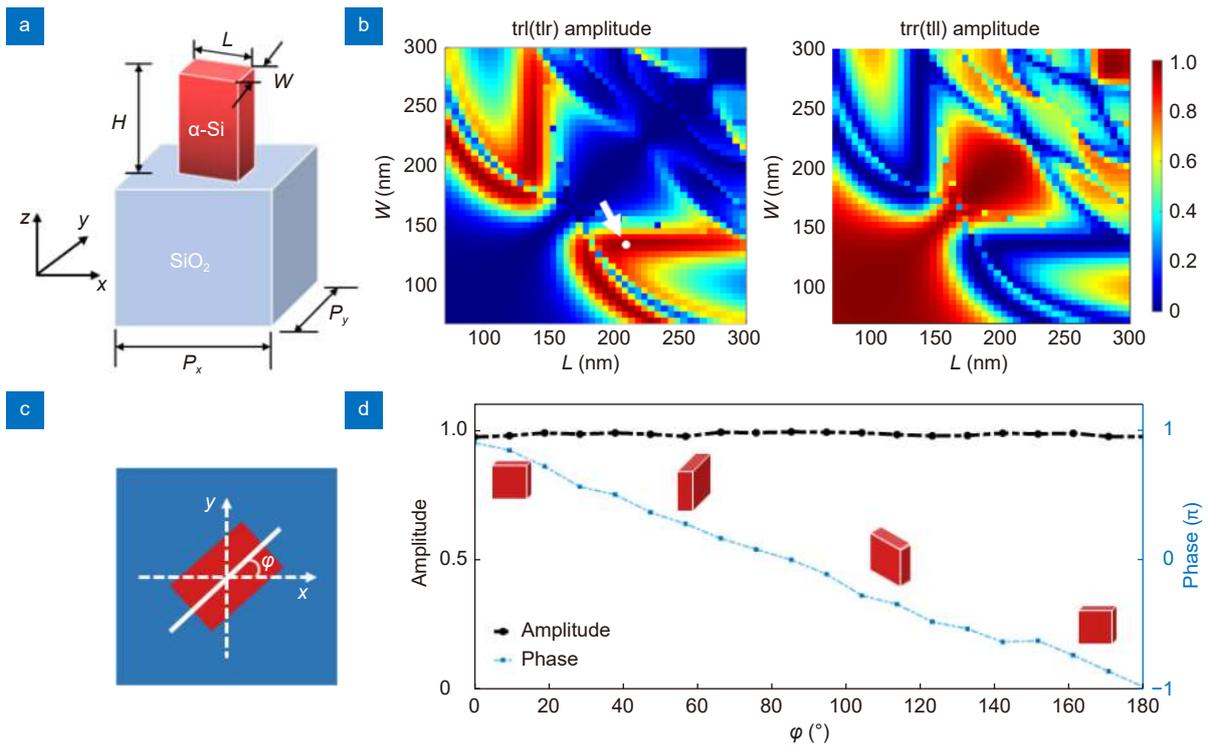


Fig. 3 | Design of the nanostructure based on geometric phase principle. (a) Schematic of an amorphous silicon nanorod fabricated on a glass substrate, where P_x and P_y are periods in the x and y directions, H is the height, and L and W are the length and width, respectively. (b) Amplitude map of the circular transmission coefficient in cross- and co-polarization for different geometry sizes. (c) Schematic diagram of the deflection angle of the nanofin. (d) Relationship between the rotation angle of the nanofin and the additional phase.

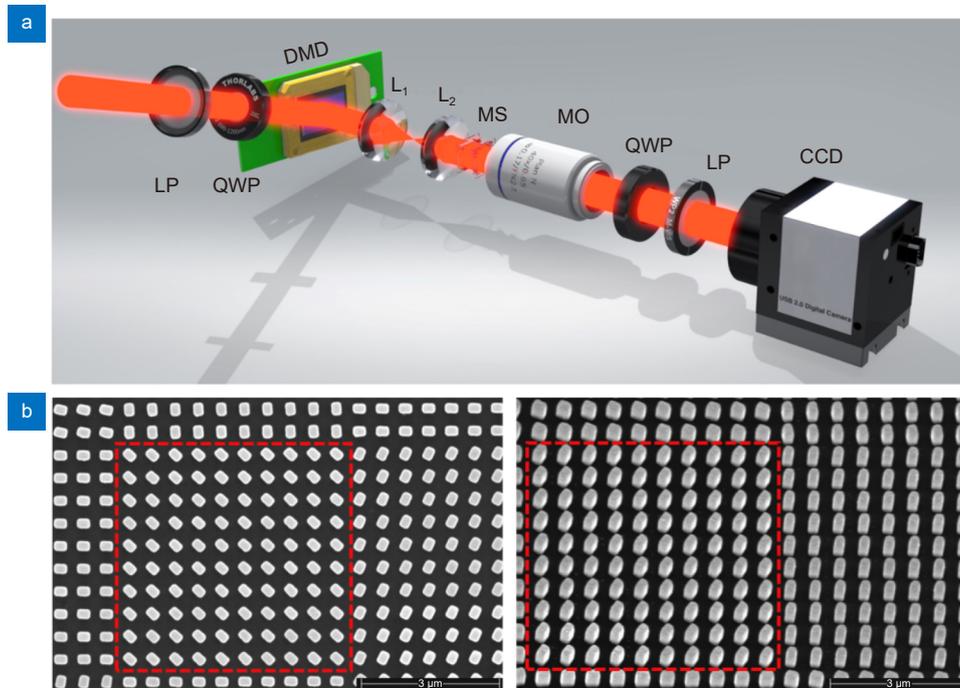


Fig. 4 | Experimental setup and the SEM images of the metasurface. (a) Schematic of the experimental setup for observing the object classification. P: linear polarizer, QWP: quarter waveplate, MS: metasurface, MO: microscope objective. (b) The SEM images of the metasurface in the top and side view, respectively. A large pixel consisting of a 10×10 array of nanofins is marked by red dotted lines.

electron microscope (SEM) images of the top and side views of the sample.

In digit classification simulation, the P-DNN handwritten digits classification component has been iteratively trained for 10 periods, and the simulation test accuracy has reached 91.8% (Fig. 5(a)). We used 6000 handwritten digit images as the test datasets, and randomly selected 300 images of datasets as the experimental validation datasets (50 images per classification category). The test results are presented using the confusion matrix, showing test details for correctly identified and misidentified instances of simulation and simulation (Fig. 5(b, c)). According to the experimental statistical results, the test accuracy reaches 90%. The experimental results are consistent well with the simulated res-

ults, indicating that the design theory is effective. The handwritten digital image was encoded into the amplitude channel as input (Fig. 5(d)). The light intensity distribution of the detection plane obtained in the simulation and experiment is shown in Fig. 5(e, f). By analyzing the energy distribution of the detection (Fig. 5(g)), The results show that the system can rightly recognize handwritten digits according to the energy distribution. Due to the influence of experimental error, the maximum energy distribution of experimental results is a bit lower than simulation results. In order to clearly display the difference between the percentage of maximum and second maximum energy, we have introduced $\Delta E = E_{\max}(\text{maximum energy}) - E_{\text{smax}}(\text{second maximum energy})$ as an indicator of the energy distribution difference,

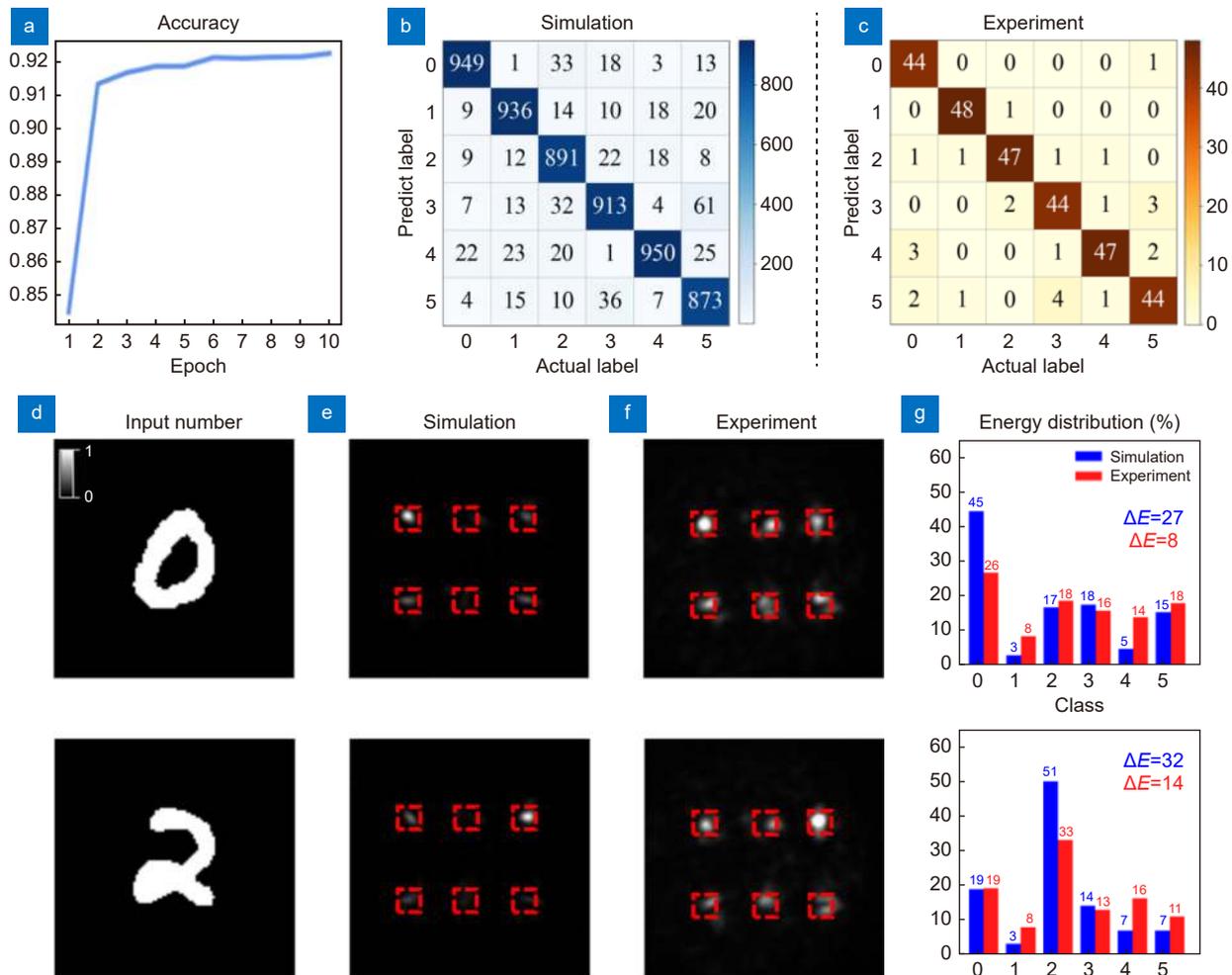


Fig. 5 | Simulation and experiment result of num-P-DNN. (a) Training accuracy is 92% after training 10 epochs. (b, c) Simulation and experimental results of confusion matrix for handwritten digits. The Simulation and experimental results test accuracy is 91.8% and 91.3% respectively, which is obtained by dividing the sum of elements on the main diagonal of confusion matrix by the sum of all elements. (d) Handwritten digital input images were encoded into amplitude channel. (e, f) Output energy distribution maps of handwritten digits in simulations and experiments. (g) The energy distribution of handwritten digits experimental results and simulation results. ΔE represents the difference between the percentage of maximum and second maximum energy.

where blue numbers represent simulation data and red numbers represent experimental data. It can be observed that the maximum energy is at least 71% higher than the second maximum energy in simulations, and at least 34% higher in experiments. The results indicate that P-DNN is capable of achieving high recognition accuracy and is not easily affected by errors.

Furthermore, we test the classification performance of P-DNN on more complex image datasets, by replacing the classification layer plug-in for fashion recognition. The datasets consist of six different fashion products (T-shirts, pants, jumpers, sneakers, bags, and ankle boots). It is worth noting that the training of the fashion plugin has achieved high classification accuracy after only five

training epochs. Compared to the training without transfer learning, the training time and parameters are reduced by half. We used 6000 fashion images as the test dataset, the numerical test has an accuracy rate of 90.2% (Fig. 6(a)). As above, we used the confusion matrix to display the statistical results, and selected 300 images from the test datasets for experimental verification (Fig. 6(b, c)). It can be seen that the experimental test accuracy reached 90%. Figure 6(d) shows the fashion images which were encoded to the amplitude channel. The system successfully realizes the classification of fashion according to the energy distribution of the output plane (Fig. 6(e, f)). It turned out that P-DNN only needs a small amount of retraining based on the original para-

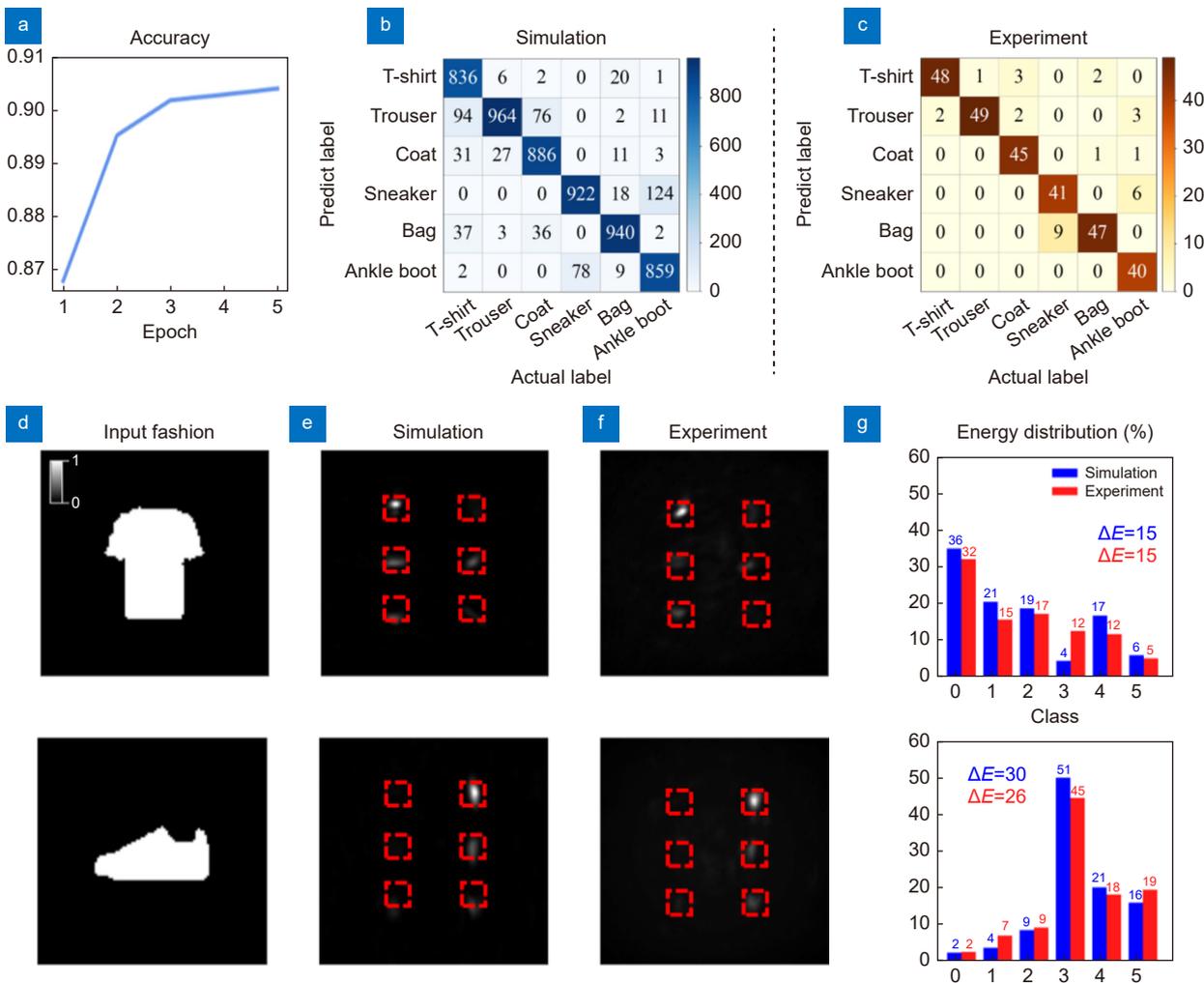


Fig. 6 | Simulation and experiment result of fashion-P-DNN. (a) Training accuracy is 91% after training 5 epochs. (b, c) Simulation and experimental results of confusion matrix for fashions classification. The Simulation and experimental results test accuracy is 90.2% and 90% respectively, which is obtained by dividing the sum of elements on the main diagonal of confusion matrix by the sum of all elements. In the simulation and experimental results, the test accuracy reaches 90.2% and 90%, respectively. (d) Fashion input images were encoded into amplitude channel. (e, f) Output plane energy distribution maps of fashions in simulations and experiments. (g) Energy distribution percentage of experimental and simulated results of fashions. ΔE represents the difference between the percentage of maximum and second maximum energy.

meters, and other more difficult tasks can also achieve good recognition performance.

Conclusions

Generally, the D²NN can only achieve a single task after the design is completed and the physically manufactured network cannot be modified. If other tasks need to be achieved, D²NN needs to be completely retrained and manufactured. Here we demonstrate that P-DNN can be applied to recognize various tasks in the near-infrared band by switching internal plug-ins. To verify the feasibility, we design cascaded metasurfaces for handwritten digits and fashion recognition. It can work similarly to the pluggable components in optical communication. According to user requirements, classification layer (pluggable layer) plug-ins can be customized to be used in combination with common layer plug-ins to achieve a variety of specific tasks (not limited to classification tasks). The experimental test accuracies were 91.3% and 90.0%, respectively, which were in good agreement with the numerical simulation results. Such plug-ins have good reconfigurability and can easily complete more tasks through plugging and unplugging. It has the characteristics of ultra-low energy consumption and light-speed calculation, and may provide extra flexibility for neural networks. Our proposed P-DNN can achieve a wide range of applications, such as intelligent optical filtering in microscopy imaging and real-time object detection in autonomous driving systems.

Material and method

Training of the P-DNN

Our P-DNN is trained using Python version 3.7.0. and TensorFlow framework version 2.4.1 (Google Inc.) on a desktop computer (Intel(R) Core(TM) i5-10500 CPU @3.10 GHz with 32 GB RAM, running the Windows 10 operating system (Microsoft)). In the training process, mean square error loss is selected as the loss function, which is usually used for target classification of machine learning, and the Adam optimizer is used to update the phase value of each layer in the network. We used 36000 handwritten digital images and fashion images as training datasets with a training batch size of 8 and a learning rate of 0.01.

Fabrication

The dielectric metasurface of amorphous silicon (α -Si) nanofins was fabricated on SiO₂ substrate. Firstly,

amorphous silicon films with a thickness of 600 nm were prepared by plasma-enhanced chemical vapor deposition. Subsequently, the polymethyl methacrylate resist layer was spin-coated. The pattern is then patterned using standard electron beam lithography. After development, a 30 nm thick chromium layer is plated on the surface of the sample. Finally, we performed the lift-off process in hot acetone and employed inductively coupled plasma reactive ion etching to transfer the desired structure from chromium to silicon.

References

1. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* **60**, 84–90 (2017).
2. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016); <http://doi.org/10.1109/CVPR.2016.90>.
3. Zhou HQ, Wang YT, Li X, Xu ZT, Li XW et al. A deep learning approach for trustworthy high-fidelity computational holographic orbital angular momentum communication. *Appl Phys Lett* **119**, 044104 (2021).
4. Guo YM, Zhong LB, Min L, Wang JY, Wu Y et al. Adaptive optics based on machine learning: a review. *Opto-Electron Adv* **5**, 200082 (2022).
5. Krasikov S, Tranter A, Bogdanov A, Kivshar Y. Intelligent meta-photonics empowered by machine learning. *Opto-Electron Adv* **5**, 210147 (2022).
6. Sainath TN, Mohamed AR, Kingsbury B, Ramabhadran B. Deep convolutional neural networks for LVCSR. In *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 8614–8618 (IEEE, 2013); <http://doi.org/10.1109/ICASSP.2013.6639347>.
7. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* **29**, 82–97 (2012).
8. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K et al. Natural language processing (almost) from scratch. *J Mach Learn Res* **12**, 2493–2537 (2011).
9. Markram H. The blue brain project. *Nat Rev Neurosci* **7**, 153–160 (2006).
10. Shen YC, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* **11**, 441–446 (2017).
11. Feldmann J, Youngblood N, Karpov M, Gehring H, Li X et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
12. Xu XY, Tan MX, Corcoran B, Wu JY, Boes A et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
13. Goi E, Chen X, Zhang QM, Cumming BP, Schoenhardt S et al. Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip. *Light Sci Appl* **10**, 40 (2021).
14. Ashtiani F, Geers AJ, Aflatouni F. An on-chip photonic deep

- neural network for image classification. *Nature* **606**, 501–506 (2022).
15. Zarei S, Marzban MR, Khavasi A. Integrated photonic neural network based on silicon metalines. *Opt Express* **28**, 36668–36684 (2020).
 16. Chen H, Feng JN, Jiang MW, Wang YQ, Lin J et al. Diffractive deep neural networks at visible wavelengths. *Engineering* **7**, 1483–1491 (2021).
 17. Liu J, Wu QH, Sui XB, Chen Q, Gu GH et al. Research progress in optical neural networks: theory, applications and developments. *Photonix* **2**, 5 (2021).
 18. Zhang X, Huang LL, Zhao RZ, Zhou HQ, Li X et al. Basis function approach for diffractive pattern generation with Dammann vortex metasurfaces. *Sci Adv* **8**, eabp8073 (2022).
 19. Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
 20. Zhao RZ, Huang LL, Wang YT. Recent advances in multi-dimensional metasurfaces holographic technologies. *Photonix* **1**, 20 (2020).
 21. Zhang YX, Pu MB, Jin JJ, Lu XJ, Guo YH et al. Crosstalk-free achromatic full Stokes imaging polarimetry metasurface enabled by polarization-dependent phase optimization. *Opto-Electron Adv* **5**, 220058 (2022).
 22. Badloe T, Lee S, Rho J. Computation at the speed of light: metamaterials for all-optical calculations and neural networks. *Adv Photon* **4**, 064002 (2022).
 23. Veli M, Mengü D, Yardimci NT, Luo Y, Li JX et al. Terahertz pulse shaping using diffractive surfaces. *Nat Commun* **12**, 37 (2021).
 24. Qian C, Lin X, Lin XB, Xu J, Sun Y et al. Performing optical logic operations by a diffractive neural network. *Light Sci Appl* **9**, 59 (2020).
 25. Wang PP, Xiong WJ, Huang ZB, He YL, Xie ZQ et al. Orbital angular momentum mode logical operation using optical diffractive neural network. *Photon Res* **9**, 2116–2124 (2021).
 26. Huang ZB, He YL, Wang PP, Xiong WJ, Wu HS et al. Orbital angular momentum deep multiplexing holography via an optical diffractive neural network. *Opt Express* **30**, 5569–5584 (2022).
 27. Rahman SS, Ozcan A. Computer-free, all-optical reconstruction of holograms using diffractive networks. *ACS Photonics* **8**, 3375–3384 (2021).
 28. Mengü D, Ozcan A. All-optical phase recovery: diffractive computing for quantitative phase imaging. *Adv Opt Mater* **10**, 2200281 (2022).
 29. Li JX, Hung YC, Kulce O, Mengü D, Ozcan A. 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).
 30. Shi JS, Zhou L, Liu TG, Hu C, Liu KW et al. Multiple-view D²NNs array: realizing robust 3D object recognition. *Opt Lett* **46**, 3388–3391 (2021).
 31. Rahman SS, Li JX, Mengü D, Rivenson Y, Ozcan A. Ensemble learning of diffractive optical networks. *Light Sci Appl* **10**, 14 (2021).
 32. Yan T, Wu JM, Zhou TK, Xie H, Xu F et al. Fourier-space diffractive deep neural network. *Phys Rev Lett* **123**, 023901 (2019).
 33. Liu C, Ma Q, Luo ZJ, Hong QR, Xiao Q et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat Electron* **5**, 113–122 (2022).
 34. Zhou TK, Lin X, Wu JM, Chen YT, Xie H et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat Photonics* **15**, 367–373 (2021).
 35. Li YJ, Chen RY, Sensale-Rodriguez B, Gao WL, Yu CX. Real-time multi-task diffractive deep neural networks via hardware-software co-design. *Sci Rep* **11**, 11013 (2021).
 36. Luo XH, Hu YQ, Ou XN, Li X, Lai JJ et al. Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible. *Light Sci Appl* **11**, 158 (2022).
 37. Georgi P, Wei QS, Sain B, Schlickriede C, Wang YT et al. Optical secret sharing with cascaded metasurface holography. *Sci Adv* **7**, eabf9718 (2021).
 38. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* **86**, 2278–2324 (1998).
 39. Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. arXiv: 1708.07747 (2017). <https://doi.org/10.48550/arXiv.1708.07747>
 40. Goodman JW. *Introduction to Fourier Optics and Holography* 3rd ed (Roberts and Company, Englewood, 2005).
 41. Mandel L, Wolf E. Some properties of coherent light. *J Opt Soc Am* **51**, 815–819 (1961).
 42. Marrucci L, Manzo C, Paparo D. Optical spin-to-orbital angular momentum conversion in inhomogeneous anisotropic media. *Phys Rev Lett* **96**, 163905 (2006).

Acknowledgements

The authors thank Dr. H. Wang for discussing the simulation algorithm, Dr. X. Zhang participated in the construction of the experimental optical path. The authors acknowledge the funding provided by the National Key R&D Program of China (2021YFA1401200), Beijing Outstanding Young Scientist Program (BJJWZYJH01201910007022), National Natural Science Foundation of China (No. U21A20140, No. 92050117, No. 62005017) program, Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (No. Z211100004821009). This work was supported by the Synergetic Extreme Condition User Facility (SECUF). We also acknowledge the fabrication and measurement service in the Analysis & Testing Center, Beijing Institute of Technology.

Author contributions

L. L. Huang proposed the idea, C. He, H. Q. Zhou, and X. Li developed the optimization algorithm, C. He and L. L. Huang conducted the designs and numerical simulations, Y. Li and J. J. Li fabricated the samples, C. He, D. Zhao and F. Fei performed experiment the measurements, C. He, L. L. Huang, F. Dong and Y.-X. Miao analyzed the data, C. He, L. L. Huang, and H. Q. Zhou prepared the manuscript. L. L. Huang and Y. T. Wang supervised the overall projects. All the authors analyzed and discussed the results.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at <https://doi.org/10.29026/oea.2024.230005>