IOE | Intelligent Opto-Electronics

Polarization-multiplexed meta-neural networks for simultaneous imaging and all-optical classification

Jialuo Cheng, Xu Li, Wenjun Zhu, Mi Zhou, Zihan Geng, Wenzhao Sun and Mu Ku Chen

Citation: Cheng JL, Li X, Zhu WJ, et al. Polarization-multiplexed meta-neural networks for simultaneous imaging and all-optical classification. *Intelligent Opto-Electronics* **2**, 250017 (2026).

Received: 24 December 2025; Accepted: 13 February 2026; Published online: 25 March 2026

IOE | Intelligent Opto-Electronics

Article

March 2026, Vol. 2, No. 1

Open Access

Polarization-multiplexed meta-neural networks for simultaneous imaging and all-optical classification

Jialuo Cheng[1†], Xu Li[2†], Wenjun Zhu[2], Mi Zhou[3], Zihan Geng[3], Wenzhao Sun[1,2,5] and Mu Ku Chen[1,4,5*]

Abstract: Deep learning has transformed perception and inference but remains constrained by memory–compute bottlenecks, latency, and energy costs. All-optical diffractive deep neural networks (D^{2}NNs) alleviate these limitations by computing with light, yet most implementations trade image formation for direct classification, limiting downstream processing. Here we introduce a polarization-multiplexed meta-neural network (PMNN) that unifies imaging and classification within a single, static optical platform. The PMNN employs cascaded metasurfaces whose meta-atoms jointly harness geometric (Pancharatnam–Berry) and propagation phases to engineer distinct phase profiles for left- and right-circularly polarized (LCP and RCP) channels. This polarization contrast enables dual-channel functionality. Under LCP illumination, the system performs lens-like imaging, whereas under RCP illumination, it executes all-optical classification via diffractive routing to predefined detection regions. Built on a differentiable angular-spectrum forward model and trained end-to-end, the PMNN achieves 96.51% accuracy on handwritten-digit recognition while delivering an imaging mean squared error of 5.38×10^{-3}, a peak signal-to-noise ratio of 22.70 dB, and a structural similarity index measure of 0.90. By coupling perception with inference without mechanical switching or electronic post-processing, the proposed approach enhances utility, reduces computational load, and offers a practical path toward compact, scalable, and energy-efficient optical intelligent systems.

Keywords: meta-device; metasurface; diffractive deep neural networks

DOI:

Citation: Cheng JL, Li X, Zhu WJ et al. Polarization-multiplexed meta-neural networks for simultaneous imaging and all-optical classification. *Intell Opto-Electron* **2**, 250017 (2026).

1 Introduction

Deep learning, an artificial neural network-based and brain-inspired paradigm for data-driven representation learning, has achieved remarkable successes across diverse tasks, including image classification[1], intelligent photonic devices[2], speech recognition[3], and machine translation[4]. Nevertheless, its dependence on large-scale datasets and iterative training, together with frequent memory access under von Neumann architectures, induces substantial latency, bandwidth pressure, and energy consumption[5]. These memory–compute bottlenecks slow inference and training, inflate operational costs, and limit scalability, thereby constraining the continued development and broader deployment of deep learning systems. All-optical diffractive deep neural networks (D^{2}NNs), a class of optical neural networks, exploit the diffraction of light to perform computation, thereby delivering ultrahigh throughput, massive spatial–spectral parallelism, intrinsically high bandwidth, and markedly reduced energy consumption relative to electronic accelerators[6–12]. Implemented with cascaded diffractive layers in which subwavelength-scale features act as neurons and free-space diffraction mediates interlayer connectivity, D^{2}NNs have been experimentally validated in imaging tasks and handwritten digit classification. Beyond these benchmarks, D^{2}NN architectures have realized diverse optical functions, including logic gate operations[13], broadband pulse shaping[14], and orbital angular momentum (OAM) beam multiplexing/demultiplexing[15,16].

Received: 24 December 2025

Accepted: 13 February 2026

Published online: 25 March 2026

[1]Department of Electrical Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR 999077, China; [2]Department of Electrical Engineering, City University of Hong Kong (Dongguan), Dongguan 523808, China; [3]Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; [4]The State Key Laboratory of Terahertz and Millimeter Waves, City University of Hong Kong, Kowloon, Hong Kong SAR 999077, China; [5]City University of Hong Kong Shenzhen Research Institute, Shenzhen 518057, China.

[†]These authors contributed equally to this work.

*Correspondence: MK Chen, E-mail: mkchen@cityu.edu.hk

Metasurfaces, comprising subwavelength artificial nanostructures (meta-atoms), enable sophisticated wavefront manipulation and have driven advances in beam steering[17], depth sensing [18], light-field imaging [19,20], augmented reality[21], and holography[22]. Unlike traditional 3D-printed diffractive elements that rely solely on phase accumulation during propagation, metasurfaces offer a higher degree of phase control that can be intrinsically coupled to the incident light's polarization state. Phase engineering in these platforms is typically realized via two complementary mechanisms: the geometric (Pancharatnam–Berry) phase [23,24], and the propagation phase [25,26]. The geometric phase is imparted by rotating anisotropic meta-atoms, whereas the propagation phase is tuned by varying their geometric dimensions. Notably, the geometric phase acquires opposite modulation for left- and right-circularly polarized (LCP and RCP) light, while the propagation phase is largely polarization-insensitive. This contrast enables polarization multiplexing, allowing multiple functionalities to be executed simultaneously within a metasurface, for example, bi-focal imaging[27,28], polarization-multiplexed metasurface holography[29], and dual-task optical computing[30].

In machine vision, image acquisition (i.e., forming an image of the object) is the most fundamental task, upon which all subsequent data processing depends. Handwritten digit recognition is likewise a foundational benchmark. In conventional pipelines, the digits are first captured and then classified algorithmically. By contrast, all-optical D^{2}NNs can perform direct digit classification without explicit image formation, inferring labels from the intensity distribution at the sensor plane. However, this approach loses access to the original image data, which is detrimental to downstream processing. In this work, a polarization-multiplexed meta-neural network (PMNN) is constructed to simultaneously perform imaging and classification. By jointly engineering the phase profiles of the LCP and RCP channels, the system simultaneously performs imaging and classification within a single optical platform. In experiments, the classification branch attains an accuracy of 96.51%, while the imaging branch achieves a mean squared error (MSE) of 5.38×10^{-3}, a peak signal-to-noise ratio (PSNR) of 22.70 dB, and a structural similarity index measure (SSIM) of 0.90. This dual-function paradigm enhances system utility while conserving computational resources, and it provides a practical path toward compact, scalable, and integrable optical intelligent systems that tightly couple perception with inference.

2 Results and discussions

The framework of the proposed PMNN supports heterogeneous tasks such as imaging and handwritten-digit classification within a single optical platform, as shown in Fig. 1. The target object (to be imaged or classified) serves as the input. Cascaded phase-encoding metasurfaces act as the hidden layers, and a sensor plane provides the output. Crucially, the PMNN exhibits markedly different phase responses to LCP and RCP illumination, thereby enabling polarization multiplexing and dual-channel functionality. Under plane-wave illumination, a shaped object (mask) produces an amplitude-encoded field with uniform phase that propagates into the PMNN. For the LCP channel, the engineered phase profile emulates a lensing operation, refocusing the transmitted light onto the camera plane to form an image of the object. In contrast, for the RCP channel, a two-layer PMNN directs the diffracted light into ten predefined detection regions, each corresponding to a digit class (0–9). The relative intensities within these regions yield the classification decision directly at the sensor, obviating any

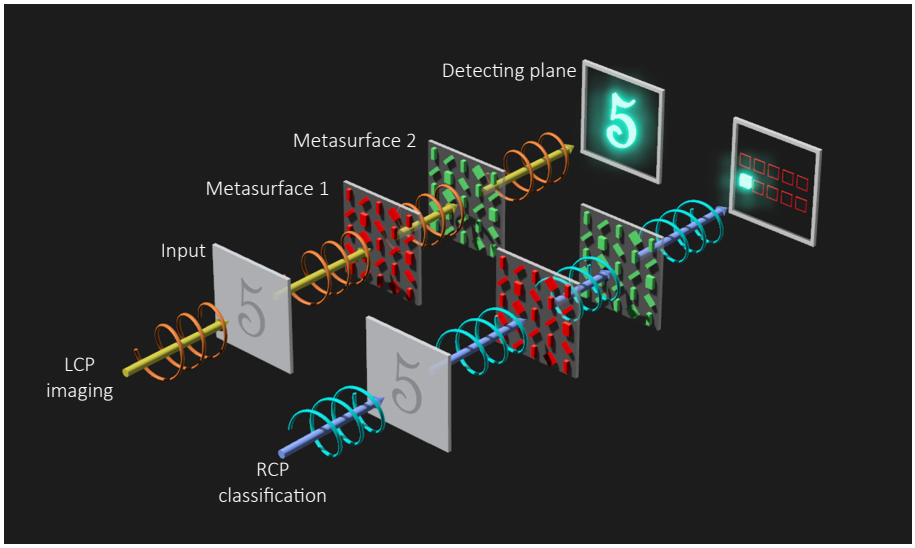

Fig. 1 | Schematic of the polarization-multiplexed meta-neural network (PMNN). Imaging and classification are integrated within a single meta-neural architecture. Under LCP illumination, the input digit image is reimaged onto the sensor plane. Under RCP illumination, the input digit is optically routed and focused into predefined detection regions corresponding to class labels for direct all-optical classification.

electronic post-processing or additional inference stages. This dual-function, polarization-multiplexed architecture tightly couples perception (imaging) with inference (classification).

The PMNN is trained within the framework of scalar diffraction theory. Under the angular-spectrum formulation, each point on a wavefront acts as a secondary spherical-wave source, and free-space propagation mixes these contributions through diffraction. In this representation, each meta-atom on a metasurface functions as an optical neuron whose complex transmission modulates the local field, while interlayer connectivity is naturally mediated by free-space propagation. Consequently, a diffractive neural network comprises two fundamental operations per layer transition: phase modulation by the metasurface and propagation through free space. The input complex field $\boldsymbol{x}_{l+1}$ for $(l+1)$th layer can be expressed as

$$\boldsymbol{x}_{l+1} = \boldsymbol{A}_l(\boldsymbol{t}_l \odot \boldsymbol{x}_l), \tag{1}$$

for the first hidden layer with $l = 1$, $\boldsymbol{x}_l$ is the transmitted light encoded by the amplitude of the input layer. The complex field is modulated by the lth meta-layer $\boldsymbol{t}_l = \boldsymbol{a}_l \odot e^{i\boldsymbol{\phi}_l}$, where $\boldsymbol{a}_l$ and $\boldsymbol{\phi}_l$ are the amplitude and the phase of $\boldsymbol{t}_l$. $\boldsymbol{A}$ denotes the linear measurement operator, which is calculated based on the angular spectrum propagation and can be further expressed as $\boldsymbol{A} = \boldsymbol{F}^{-1}\boldsymbol{H}_d\boldsymbol{F}$, where $\boldsymbol{F}$ and $\boldsymbol{F}^{-1}$ denote the Fourier transform and inverse Fourier transform, respectively. The transfer function at a given distance d is given by

$$\boldsymbol{H}_d(f_x, f_y) = \begin{cases} \exp\left(j\frac{2\pi}{\lambda}d\sqrt{1-\lambda^2 f_x^2 - \lambda^2 f_y^2}\right) & f_x^2 + f_y^2 \le \frac{1}{\lambda^2} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where λ is the illumination wavelength, f_x and f_y denote the spatial frequency coordinates. The light manipulation by the meta-atoms can be characterized by the Jones matrix[31]:

$$\boldsymbol{J} = \boldsymbol{R}(\theta)\begin{bmatrix} a_x \exp(j\varphi_x) & 0 \\ 0 & a_y \exp(j\varphi_y) \end{bmatrix}\boldsymbol{R}(-\theta), \tag{3}$$

where $\varphi_x - \varphi_y = \pi$, for PB phase, and $a_x = a_y = 1$, for all phase modulation. Due to the perfect orthogonality between LCP and RCP, the optimization process can be simplified by treating them as two entirely independent scalar diffraction channels during training. Upon completing the training, the required phase profiles for both LCP and RCP channels are obtained. The metasurface is then designed to convert the incident LCP light into RCP light with an encoded phase of φ_{LCP}, and vice versa, which can be expressed as

$$\begin{cases} \frac{1}{\sqrt{2}}\boldsymbol{J}\begin{bmatrix}1\\ i\end{bmatrix} = \frac{1}{\sqrt{2}}\exp(i\varphi_{\mathrm{LCP}})\begin{bmatrix}1\\ -j\end{bmatrix} \\ \frac{1}{\sqrt{2}}\boldsymbol{J}\begin{bmatrix}1\\ -i\end{bmatrix} = \frac{1}{\sqrt{2}}\exp(i\varphi_{\mathrm{RCP}})\begin{bmatrix}1\\ j\end{bmatrix} \end{cases}. \tag{4}$$

By solving these equations, the phase requirements can be directly decoupled into the propagation phase and the orientation angle (related to the PB phase).

$$\begin{cases} \varphi_x = \frac{\varphi_{\mathrm{LCP}} + \varphi_{\mathrm{RCP}}}{2} \\ \theta = \frac{\varphi_{\mathrm{LCP}} - \varphi_{\mathrm{RCP}}}{4} \end{cases} \tag{5}$$

With the forward optical model, we embed the PMNN into a PyTorch-based deep learning framework to enable end-to-end training and adopt the adam optimizer for stable and efficient convergence. More detailed model training and derivation are described in Section 1 of the Supplementary Information.

Task-specific loss functions are employed for the two polarization channels. For the LCP imaging branch, the MSE between the reconstructed image at the sensor plane and the ground-truth target is minimized. For the RCP classification branch, the cross-entropy loss is used to promote accurate digit discrimination. To realize all-optical readout, the detection plane is partitioned into a set of discrete subregions that serve as the network's physical output channels, with each region corresponding to a digit class, and the relative intensities in these regions directly yield the predicted label without electronic post-processing. Training and evaluation are conducted on the modified national institute of standards and technology database (MNIST) of handwritten digits.

To realize dual-function control, we simultaneously employ both geometric and propagation phases. Gallium nitride nanopillars are designed and optimized using the finite-difference time-domain (FDTD) simulation method. This method allows the calculation of the electromagnetic response of the nanostructure, with the nanopillar height fixed at 900 nm and the periods in the x- and y-directions both set to 280 nm, and an incident wavelength of 532 nm. The metasurfaces are composed of rectangular gallium nitride (GaN) designed on a sapphire substrate, as shown in Fig. 2(a). By scanning the length and width of the nanopillar from 100 nm to 280 nm in 5 nm steps, the simulation results indicate that the phase modulation achieved by varying the length and width covers the range of 0 to 2π while maintaining high transmittance across this parameter range, as shown in Fig. 2(b,c). Since both LCP and RCP are insensitive to transmission phase modulation, the resulting phase profiles and transmission maps remain identical. The comprehensive characterization and full sets of images are provided in Section 4 of the Supplementary Information.

Simultaneously, the amplitude values of the cross-circular polarization co-transmission coefficient and the co-circular polarization co-transmission coefficient of the nanopillars are obtained from the simulations, as shown in Fig. 3(a,b), respectively. A nanopillar ($L = 215$ nm, $W = 100$ nm) is selected as a representative unit to demonstrate the geometric phase mechanism. As shown in Fig. 3(c), this configuration functions as a local half-wave

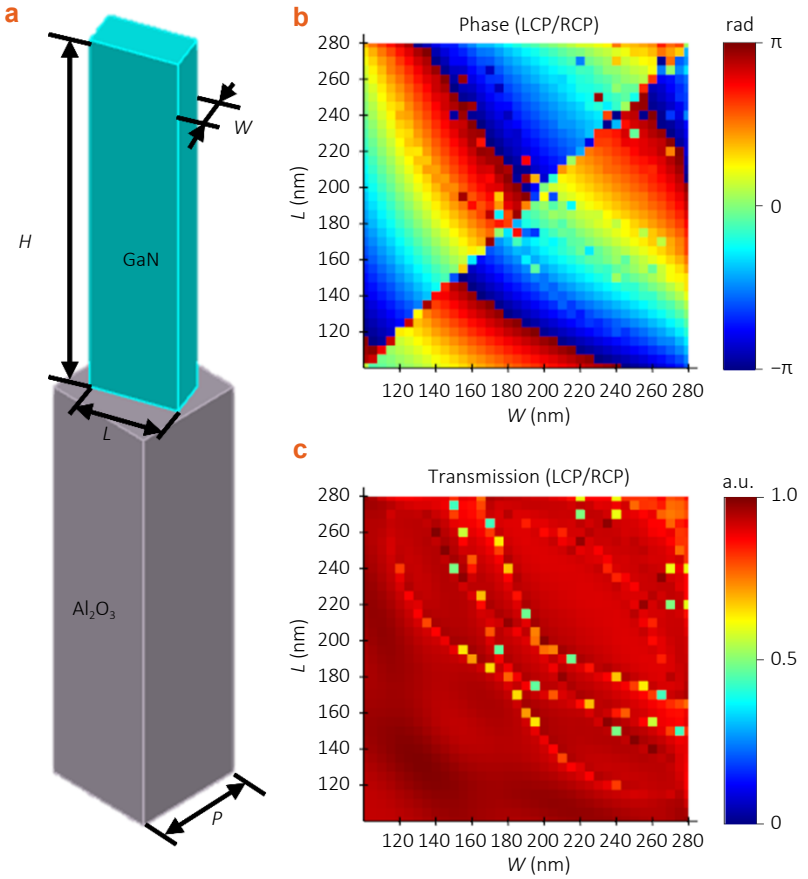

Fig. 2 | Design methodology of the metasurface. (a) Schematic of gallium nitride (GaN) nanopillars fabricated on a sapphire substrate. (b) The simulated phase responses under LCP and RCP illumination. (c) The simulated amplitude responses under LCP and RCP illumination. The simulations are performed at an incident wavelength of 532 nm, with simultaneous sweeps of nanopillar length and width over the range of 100–280 nm.

plate (HWP) satisfying the condition $\phi_x - \phi_y = \pi$, which ensures near-unity polarization conversion efficiency. Under circularly polarized incidence, the rotation of the nanopillar provides a complete 0–2π phase shift while maintaining a stable amplitude. To achieve multi-tasking, a library of 50 distinct nanopillar geometries is optimized. Each geometry satisfies the HWP condition but offers a different propagation phase based on its lateral dimensions. By synergistically mapping the required phase profiles to these dimensions (propagation phase) and their orientation angles (geometric phase), the PMNN gains the necessary degrees of freedom to independently modulate imaging and classification channels within a single metasurface neural network. To substantiate the physical realism of the design, a rigorous sensitivity analysis is performed regarding fabrication tolerances. The impact of dimensional and rotational deviations is modeled as stochastic phase and amplitude perturbations sampled from Gaussian distributions. The results indicate that the PMNN possesses exceptional robustness; specifically, the classification accuracy maintains its high performance even under an amplitude standard deviation of 0.3, while the imaging task preserves its structural integrity (SSIM > 0.90) despite significant amplitude fluctuations. This resilience stems from the neural network's ability to prioritize semantic features and relative contrast, effectively decoupling the system's functional output from minor pixel-wise intensity variations inherent in practical fabrication. The comprehensive information and images are provided in Section 6 of the Supplementary Information.

To enable simultaneous imaging and classification, polarimetric cameras capture images in both LCP and RCP channels. Since commercial polarization cameras are typically restricted to sensing linear polarization orientations (0°, 45°, 90°, and 135°), a quarter-wave plate (QWP) is integrated into the optical path. The QWP functions to transform LCP and RCP components into distinct linear polarization states, rendering them detectable by the camera's micro-polarizer array. By illuminating the system with horizontally polarized light, which can be decomposed into LCP and RCP components. The first metasurface swaps its handedness (LCP → RCP, RCP → LCP). After propagation through the second metasurface, the handedness is swapped back, restoring LCP and RCP. After the light passes through the quarter-wave plate, whose fast axis is aligned with the *x*-axis, the

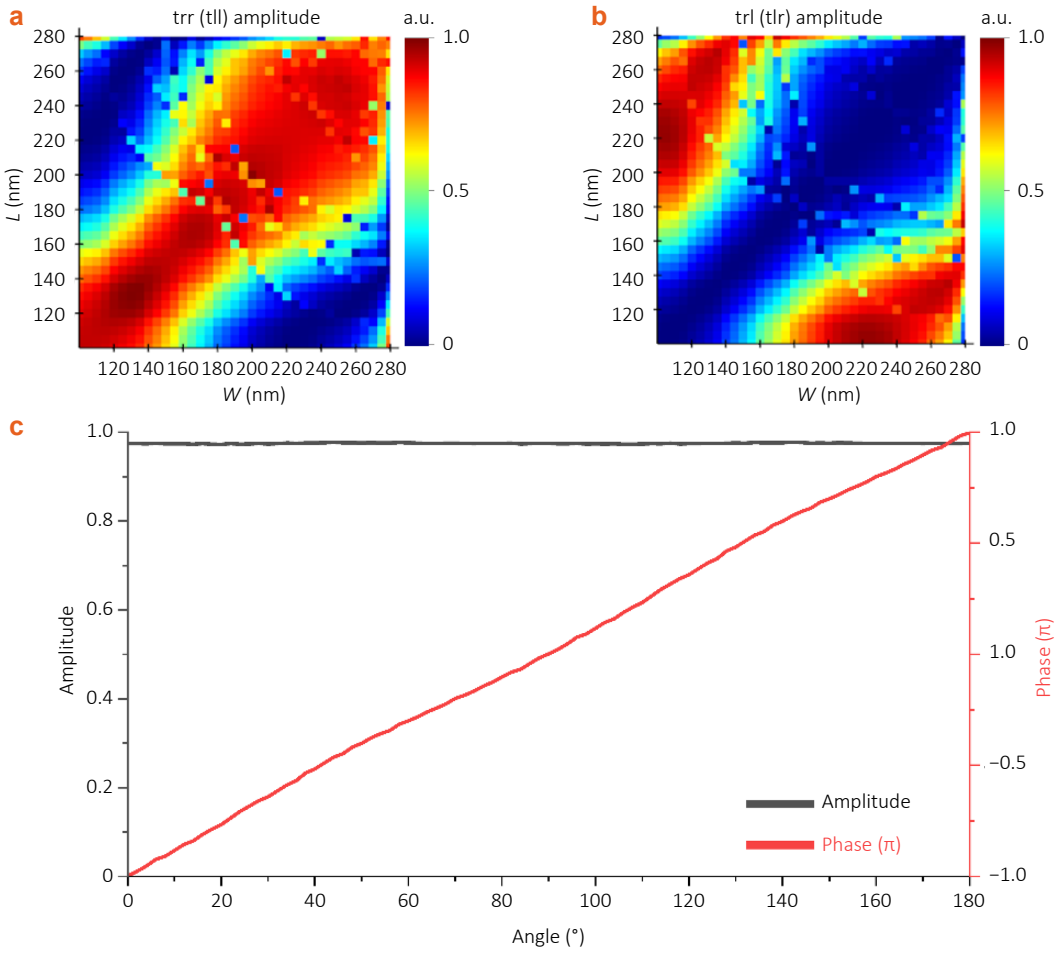

Fig. 3 | Simulation result for geometric phase. (a) Amplitudes of the cross-circular transmission coefficients for varying structural parameters. (b) Amplitudes of the co-circular transmission coefficients for varying structural parameters. (c) Relationship between nanopillar rotation angle and the induced phase and amplitude.

LCP will be changed to −45° LP, and the RCP will be changed to 135° LP.

$$\text{LCP changed to 135° LP}: \frac{e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}1 & 0\\ 0 & j\end{bmatrix}\begin{bmatrix}1\\ j\end{bmatrix} = \frac{e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}1\\ -1\end{bmatrix}$$

$$\text{RCP changed to 45°LP}: \frac{e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}1 & 0\\ 0 & j\end{bmatrix}\begin{bmatrix}1\\ -j\end{bmatrix} = \frac{e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}1\\ 1\end{bmatrix} \quad (6)$$

When these two LP lights are captured by a polarized camera, they pass through four linear polarizations, and the corresponding responses are

$$0°: \frac{e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix} + \frac{e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = \frac{e^{i2\varphi_{LCP}} + e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}1\\ 0\end{bmatrix}, \quad (7)$$

$$45°: \frac{e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix} + \frac{e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = \frac{e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}1\\ 1\end{bmatrix}, \quad (8)$$

$$90°: \frac{e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix} + \frac{e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = \frac{e^{i2\varphi_{RCP}} - e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}0\\ 1\end{bmatrix}, \quad (9)$$

$$135°: \frac{e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix} + \frac{e^{i2\varphi_{RCP}}}{\sqrt{2}}\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = \frac{e^{i2\varphi_{LCP}}}{\sqrt{2}}\begin{bmatrix}1\\ -1\end{bmatrix}. \quad (10)$$

The images measured in the 45° and 135° channels exhibit the identical intensity distributions as those obtained under

RCP and LCP illumination, corresponding to the classification and imaging functions, respectively.

In the imaging-task simulation, the PMNN is trained for 80 epochs on handwritten digits, yielding a test-set imaging MSE of 5.38×10^{-3}, a PSNR of 22.70 dB, and a SSIM of 0.90. We use 60,000 images for training and 10,000 images for testing (MNIST). The meta-atom period is set to 280 nm, and each metasurface comprises 2000 × 2000 meta-atoms. The illumination wavelength is 532 nm. The propagation distances are d_1 from the object plane to layer 1, d_2 from layer 1 to layer 2, and d_3 from layer 2 to the detector, with $d_1 = d_2 = d_3 = 0.1$ mm. For training convenience and physical realism, the dataset is preprocessed by resizing each image to 2000 × 2000 pixels and binarizing it to 0 and 1. During optimization, the input image serves as the ground truth for the imaging branch. The MSE between the sensor-plane reconstruction and the target is used as the loss, and gradient descent is applied to update the phase profiles of the two metasurfaces. The trained phase distributions for layer 1 and layer 2 are shown in Fig. 4(a). Representative inputs and their reconstructed images are presented in Fig. 4(b–e), with additional results provided in Section 2 of the Supplementary Information. We note that the reconstructed image quality is not as high as in some prior demonstrations[6], where the interlayer propagation distances used for imaging are much shorter than those typically required for classification. Shorter spacings can help preserve higher spatial frequencies. Further quantitative assessments regarding the impact of propagation distance on both imaging and classification tasks, along with the distance-dependent reconstruction fidelity and the system's generalization to unseen patterns, are provided in Section 5 of the Supplementary Information. In our PMNN, however, imaging and classification must be realized within the same optical stack without mechanical adjustment, using only a change in the incident polarization. This constraint necessitates a compromise, leading to a modest reduction in imaging fidelity in exchange for dual-function, polarization-multiplexed operation.

In the classification-task simulation, the PMNN is trained for 80 epochs on handwritten digits and achieves a test accuracy of 96.51%. The same metasurface parameters, propagation distances, and image preprocessing pipeline are used as in the imaging task. The trained phase distributions for layer 1 and layer 2 are shown in Fig. 5(a). Test performance is summarized by a confusion matrix, which details correct and incorrect predictions; the horizontal axis denotes the predicted labels, and the vertical axis denotes the ground-truth labels, showing that the vast majority of samples are correctly classified, as in Fig. 5(b). Representative inputs and the corresponding light-intensity distributions on the detection plane are presented in Fig. 5(c,d), respectively. By analyzing the energy distribution across the predefined detection regions, the system accurately recognizes digits

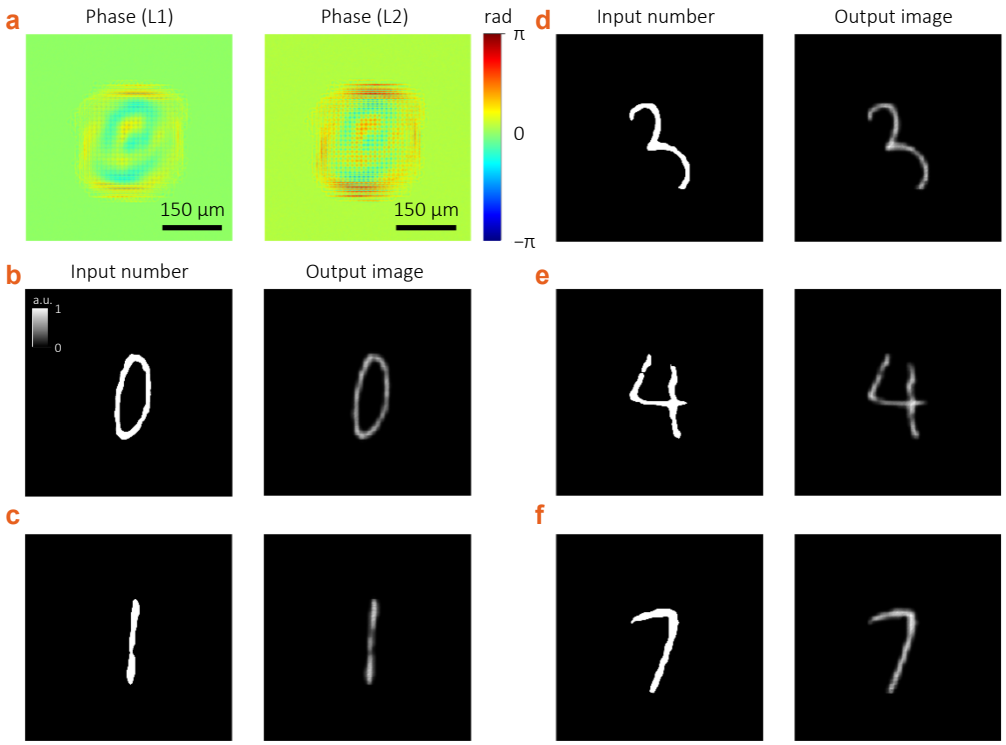

Fig. 4 | Imaging results under LCP illumination. (a) Trained phase distributions for metasurface layer 1 and layer 2. (b–e) Input digit patterns (ground truth) and the corresponding reconstructed images at the sensor plane.

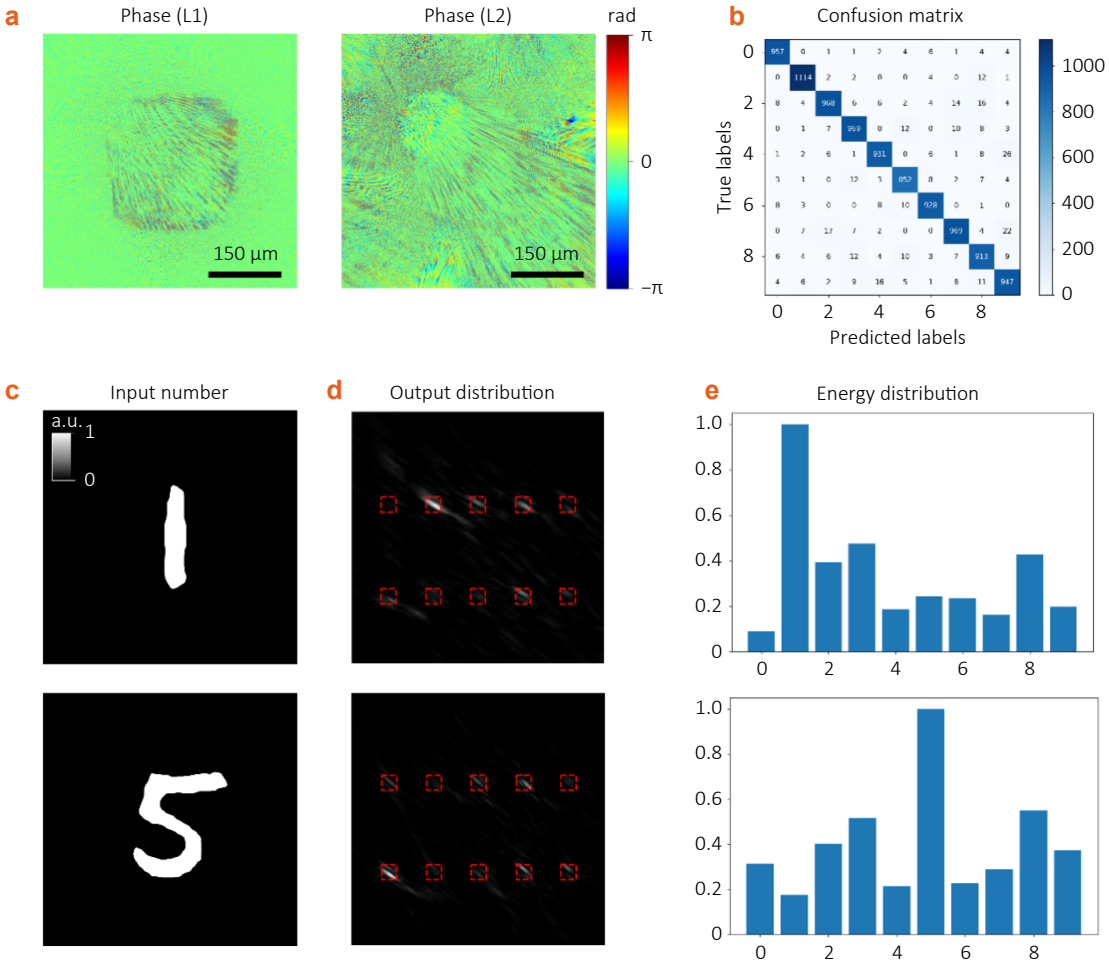

Fig. 5 | Classification results under RCP illumination. (a) Trained phase distributions for metasurface layer 1 and layer 2. (b) Confusion matrix for handwritten-digit classification. (c) Input digit patterns. (d) Output energy-distribution maps on the detection plane. (e) Energy distributions across the predefined detection regions.

purely optically, as illustrated in Fig. 5(e). For example, for the digits "1" and "5", the optical energy is significantly higher in their respective regions than elsewhere, enabling straightforward discrimination without any electronic post-processing. The additional results are provided in Section 3 of the Supplementary Information. A detailed comparison between this work and existing polarization-multiplexing-based optical neural networks is provided in Section 7 of the Supplementary Information.

3 Conclusions

Conventional machine vision pipelines perform recognition by first acquiring an image and then running electronic inference, which incurs latency and energy consumption. All-optical D^2NNs, by contrast, can perform all-optical computation and directly output class labels with low power and high speed. However, they do not recover the original image, limiting downstream processing. To address this trade-off, we design the PMNN that leverages both geometric and propagation phases. In conjunction with a polarization camera, the PMNN enables simultaneous imaging (LCP channel) and all-optical classification (RCP channel) within a single, static optical stack, realizing perception and inference without mechanical switching. The phase-design space and transmission efficiency are validated through FDTD simulations, demonstrating that the combined use of geometric and propagation phases provides sufficient degrees of freedom for polarization-selective wavefront engineering while maintaining high throughput. An end-to-end training framework is further developed, grounded in scalar diffraction theory and implemented in PyTorch, optimizing the metasurface phase profiles for both tasks. The classification branch attains an accuracy of 96.51%, and the imaging branch achieves an MSE of 5.38×10^{-3}, a PSNR of 22.70 dB and an SSIM of 0.90 on MNIST. This dual-function, polarization-multiplexed paradigm enhances system utility, reduces computational load, and supports low-latency, low-power operation. More broadly, it charts a practical route toward compact, scalable, and integrable optical intelligent systems that tightly couple image

formation with inference, and it provides a versatile platform for multi-task optical computing using a single, co-designed metasurface stack. Potential extensions include multi-channel multiplexing (e.g., wavelength or spatial-mode), task-conditioned outputs, and hardware implementations.

4 Materials and Methods

The PMNN is designed and trained in Python (v3.10.19) using PyTorch (v2.9.0+cu128) on Ubuntu 24.04.3 LTS. Experiments are conducted on a workstation equipped with an 11th Gen Intel Core i9-11900 CPU (2.50 GHz), 128 GB RAM, and an NVIDIA GeForce RTX 3090 GPU. The MNIST dataset is employed for both image processing and classification, with a batch size of 32, 40 training epochs, and the Adam optimizer with a learning rate of 1×10^{-3}. For the LCP imaging branch, the mean squared error between the sensor-plane reconstruction and the ground-truth target is minimized. For the RCP classification branch, the cross-entropy loss is the loss function.

References

1. He KM, Zhang XY, Ren SQ et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* **37**, 1904–1916 (2015).
2. Guo YM et al. Adaptive optics based on machine learning: a review. *Opto-Electron Adv* **5**, 200082 (2022).
3. Dahl GE, Sainath TN, Hinton GE. Improving deep neural networks for LVCSR using rectified linear units and dropout. *In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 8609–8613 (IEEE, 2013). http://doi.org/10.1109/ICASSP.2013.6639346.
4. Collobert R, Weston J, Bottou L et al. Natural language processing (almost) from scratch. *J Mach Learn Res* **12**, 2493–2537 (2011).
5. Markram H. The blue brain project. *Nat Rev Neurosci* **7**, 153–160 (2006).
6. Lin X, Rivenson Y, Yardimci NT et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
7. He C, Zhao D, Fan F et al. Pluggable multi-task diffractive neural networks based on cascaded metasurfaces. *Opto-Electron Adv* **7**, 230005 (2024).
8. Shen YC, Harris NC, Skirlo S et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* **11**, 441–446 (2017).
9. Xu XY, Tan MX, Corcoran B et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
10. Luo XG. Digital optics and optical intelligent agent. *Appl Phys Lett* **127**, 010501 (2025).
11. Guo YH, Lei YS, Pu MB et al. Vectorial digitelligent optics for high-resolution non-line-of-sight imaging. *Engineering* **45**, 70–78 (2025).
12. Tian X, Li RZ, Peng T et al. Multi-prior physics-enhanced neural network enables pixel super-resolution and twin-image-free phase retrieval from single-shot hologram. *Opto-Electron Adv* **7**, 240060 (2024).
13. Qian C, Lin X, Lin XB et al. Performing optical logic operations by a diffractive neural network. *Light Sci Appl* **9**, 59 (2020).
14. Veli M, Mengu D, Yardimci NT et al. Terahertz pulse shaping using diffractive surfaces. *Nat Commun* **12**, 37 (2021).
15. Wang PP, Xiong WJ, Huang ZB et al. Orbital angular momentum mode logical operation using optical diffractive neural network. *Photonics Res* **9**, 2116–2124 (2021).
16. Huang ZB, He YL, Wang PP et al. Orbital angular momentum deep multiplexing holography via an optical diffractive neural network. *Opt Express* **30**, 5569–5584 (2022).
17. Li WL, Chen BW, Hu XY et al. Modulo-addition operation enables terahertz programmable metasurface for high-resolution two-dimensional beam steering. *Sci Adv* **9**, eadi7565 (2023).
18. Chen MK, Liu XY, Wu YF et al. A meta - device for intelligent depth perception. *Adv Mater* **35**, 2107465 (2023).
19. Lin RJ, Su VC, Wang SM et al. Achromatic metalens array for full-colour light-field imaging. *Nat Nanotechnol* **14**, 227–231 (2019).
20. Fan ZB, Qiu HY, Zhang HL et al. A broadband achromatic metalens array for integral imaging in the visible. *Light Sci Appl* **8**, 67 (2019).
21. Li Y, Chen SY, Liang HW et al. Ultracompact multifunctional metalens visor for augmented reality displays. *PhotoniX* **3**, 29 (2022).
22. Jiang Q, Jin GF, Cao LC. When metasurface meets hologram: principle and advances. *Adv Opt Photonics* **11**, 518–576 (2019).
23. Chen XZ, Huang LL, Mühlenbernd H et al. Dual-polarity plasmonic metalens for visible light. *Nat Commun* **3**, 1198 (2012).
24. Huang K, Dong ZG, Mei ST et al. Silicon multi-meta-holograms for the broadband visible light. *Laser Photonics Rev* **10**, 500–509 (2016).
25. Zhan AL, Colburn S, Trivedi R et al. Low-contrast dielectric metasurface optics. *ACS Photonics* **3**, 209–214 (2016).
26. Fan ZB, Shao ZK, Xie MY et al. Silicon nitride metalenses for close-to-one numerical aperture and wide-angle visible imaging. *Phys Rev Appl* **10**, 014005 (2018).
27. Zhou JX, Tian FL, Hu J et al. Eagle-eye inspired meta-device for phase imaging. *Adv Mater* **36**, 2402751 (2024).
28. Cheng JL, Geng ZH, Zhou Y et al. Tunable meta-device for large depth of field quantitative phase imaging. *Nanophotonics* **14**, 1249–1256 (2025).
29. Zhao RZ, Sain B, Wei QS et al. Multichannel vectorial holographic display and encryption. *Light Sci Appl* **7**, 95 (2018).
30. Wang YZ, Yu AX, Cheng YY et al. Matrix diffractive deep neural networks merging polarization into meta-devices. *Laser Photonics Rev* **18**, 2300903 (2024).
31. Fan QB, Xu WZ, Hu XM et al. Trilobite-inspired neural nanophotonic light-field camera with extreme depth-of-field. *Nat Commun* **13**, 2130 (2022).

Acknowledgements

This work is financed by the Guangdong Basic and Applied Basic Research Foundation [2025A1515011846], the National Science Foundation of China (NSFC) [62405254]; the Fundamental and Applied Fundamental Research Foundation Project of Guangdong Province [No. 2023A1515140108]; the University Grants Committee/Research Grants Council of the Hong Kong Special Administrative Region, China [CRF Project: C5031-22G; and GRF Project: CityU11310522; CityU11300123]; City University of Hong Kong [Project No. 9610628 and No. 7020142]; Guangdong and Hong Kong Universities '1+1+1' Joint Research Collaboration Scheme.

Author contributions

M.K.C. and J.C. conceived the idea for this work. M.K.C. supervised the research. J.C. is responsible for developing algorithms and implementing methods. M.Z. and Z.G. conceived the development of light field computing. X.L., J.C., and W.Z. built FDTD simulations. W.S. and Z. G. provided the high-speed computing hardware and guidance for the simulation work. J.C., X.L., and W.Z.

perform data processing and analysis. All authors discussed the results and provided comments on the manuscript.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at https://doi.org/10.29026/ioe.2026.250017