Supplementary Material for: "Achieving Superior Accuracy in Photonic Neural Networks with Physical Multi-synapses"

Zhuonan Jia[1], Haopeng Tao[1], Guang-Bin Huang[2,3]*& Ting Mei[1]*

[1]Key Laboratory of Light Field Manipulation and Information Acquisition, Ministry of Industry and Information Technology, School of Physical Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710129, China

[2] School of Automation, Southeast University, Nanjing, China

[3] Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing, China

*E-mail: gbhuang@ieee.org; ting.mei@ieee.org

Table of Contents

Supplementary Note 1: Calculation details of RNT-based neural network

As the input in the image classification task, binary images in MNIST and Fashion-MNIST were encoded with {0, 1}, and color images in CIFAR-10 were converted to grayscale images. Fig. S1(a) illustrates a random number template (544×544) used in RNT-based neural network generated using the 'rand' function in MATLAB within a range of -1 to 1. To ensure consistent random projections and the matrix size of hidden neurons under different duplicate array formats, an input **x** and its duplicate array formats were set to a matrix of size 544×544 by zero padding. The matrix was convolved with the random number template using 'conv2' function in MATLAB to obtain a matrix of size 544×544, and then a square function was used for non-linear mapping. Subsequently, down sampling and flattening operations were performed to obtain hidden neurons **h''** after normalization of **h**. The hidden layer matrix **H** was constructed using **h''** employing all training samples for training. Inference testing was performed on the testing samples in the trained network to obtain the test accuracy. Fig. S1(b)-(d) show test accuracies for the three datasets, and illustrate the impact of different duplicate array formats (1×1, 5×5, 9×9) on the test accuracies for CIFAR-10. The final network's trained weights, matrix **β**, were determined by the regularization coefficient *C* selected in Table.S1. Fig. S1(e)-(i) show the confusion matrices for the testing samples with 10,000 hidden neurons. Test accuracies of 97.24% and 88.19% were achieved for the MNIST and Fashion-MNIST datasets, respectively. For CIFAR-10, test accuracies are consistently close to each other with different duplicate array formats, indicating that multi-synaptic connections are not effective for RNT-based networks. Fig. S2 presents RNT-based network parameters, including input, hidden layer feature matrix, and trained weight matrix, using an airplane image from CIFAR-10(grayscale) as example.

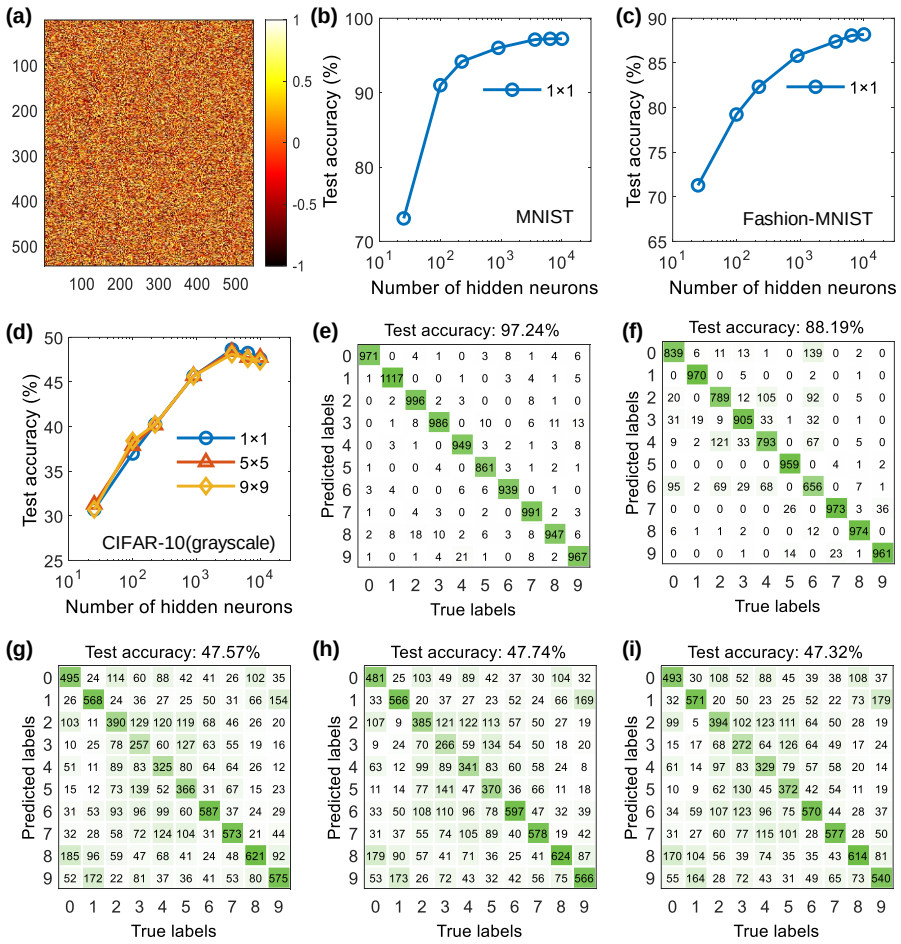

Fig. S1: Results of image classification by RNT-based neural network. (a) Random number template. (b)-(d), Test accuracy versus numbers of hidden neurons for MNIST, Fashion-MNIST and CIFAR-10(grayscale) with different duplicate array formats. (e)-(g), Confusion matrices for the three datasets in the 1×1 format with 10,000 hidden neurons. Confusion matrices for CIFAR-10(grayscale) in the 5×5 (h) and 9×9 (i) format with 10,000 hidden neurons.

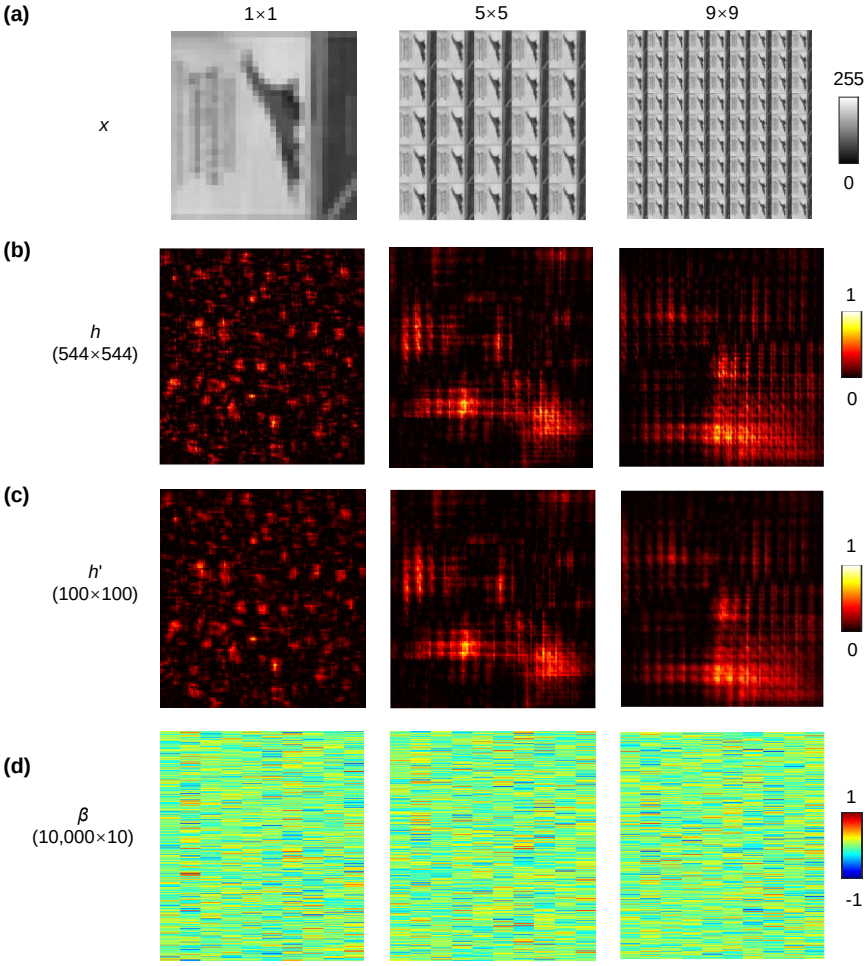

Fig. S2: Network parameters for RNT- based neural network. (a) Input image examples from CIFAR-10(grayscale) dataset in gray-encoded form. (b) Hidden layer feature matrix **h**. (c) Hidden layer feature matrix **h'** after down sampling. (d) Trained weight matrix **β** with 10,000 hidden neurons.

Supplementary Note 2: Calculation details of optical model-based neural network

For optical model-based neural networks, the sample image first needs to undergo zero-padding and encoding operations to obtain the incident field. The incident field corresponds to the input **x** in the general ELM neural network. We provide the discrete numerical expression from the sample image to the incident field distribution for different duplicate array formats. Methods Eq. 11 is for 1×1. For 5×5 and 9×9,

$$F_0(N_u, N_v) = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \mathbf{F}_{i11}[p,q] & \cdots & \mathbf{F}_{i1c}[p,q] & \vdots \\ 0 & \vdots & \ddots & \vdots & 0 \\ \vdots & \mathbf{F}_{ic1}[p,q] & \cdots & \mathbf{F}_{icc}[p,q] & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}, \quad (1)$$

where $\mathbf{F}_{icc}[p,q]$ represents $\mathbf{F}_i[p,q]$ in the *c*-th row and *c*-th column of the duplicated array formats. Here, binary images in MNIST and Fashion-MNIST were phase-encoded with $\{0, \pi\}$, grayscale images in CIFAR-10 were linearly mapped within a phase range of $[0, 1.8\pi]$. We applied phase encoding to the three datasets, where the incident field is $e^{jF_0(N_u,N_v)}$.

After obtaining the diffraction filed intensity matrix using Methods Eq. 16 and Eq. 8, we cropped the matrix from the center. To ensure consistency in the area size of 4.512 mm×4.512 mm between the diffraction intensity matrix in optical model-based neural network and the ROI region in photonic neural network, considering a pixel interval of 8.3 μm in optical model, we cropped a matrix size of 544×544 as the final diffraction intensity, representing the hidden neurons **h**. After normalizing the diffraction intensity of each sample, we performed down sampling and flattening operations. Using **h''** from all training samples, we constructed the hidden layer matrix **H** for training. Fig. S3(a)-(c) show test accuracies of the three datasets under different duplicated array formats, demonstrating the performance enhancement of multi-synaptic connections in optical model-based neural network. The final network's trained weights matrix **β** are determined by the regularization coefficients *C* selected in Table.S2. Fig. S3(d)-(l) show the confusion matrices for the testing samples with 10,000 hidden neurons. With

multi-synaptic connections, the optical model-based network achieves improved accuracy, especially notably on the CIFAR-10 dataset, where accuracy increased from 43.54% (1×1) to 53.49% (5×5) with 10,000 hidden neurons, surpassing the RNT-based network. This enhancement demonstrates strengthened feature extraction capabilities of the optical model after incorporating multi-synaptic connections. Fig. S4 shows the optical model-based network parameters, including the input encoding type, the hidden layer feature matrix, and the trained weight matrix.

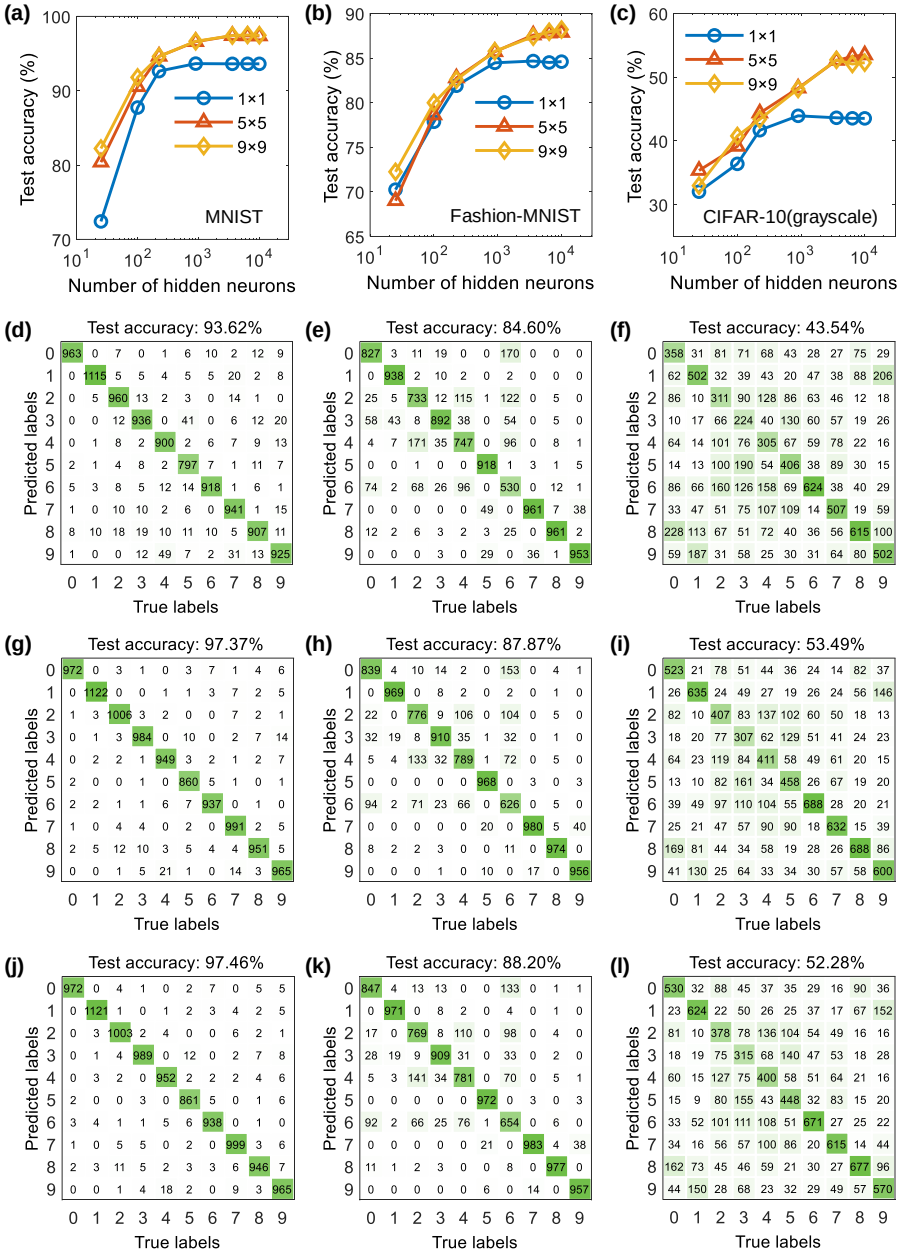

Fig. S3: Calculation results of image classification for optical model-based neural network. (a)-(c) Test accuracies under different numbers of hidden neurons and duplicate array formats for MNIST, Fashion-MNIST and CIFAR-10(grayscale). Confusion matrices for the three datasets in the 1×1 ((d)-(f)), 5×5 ((g)-(i)) and 9×9 ((j)-(l)) format with 10,000 hidden neurons.

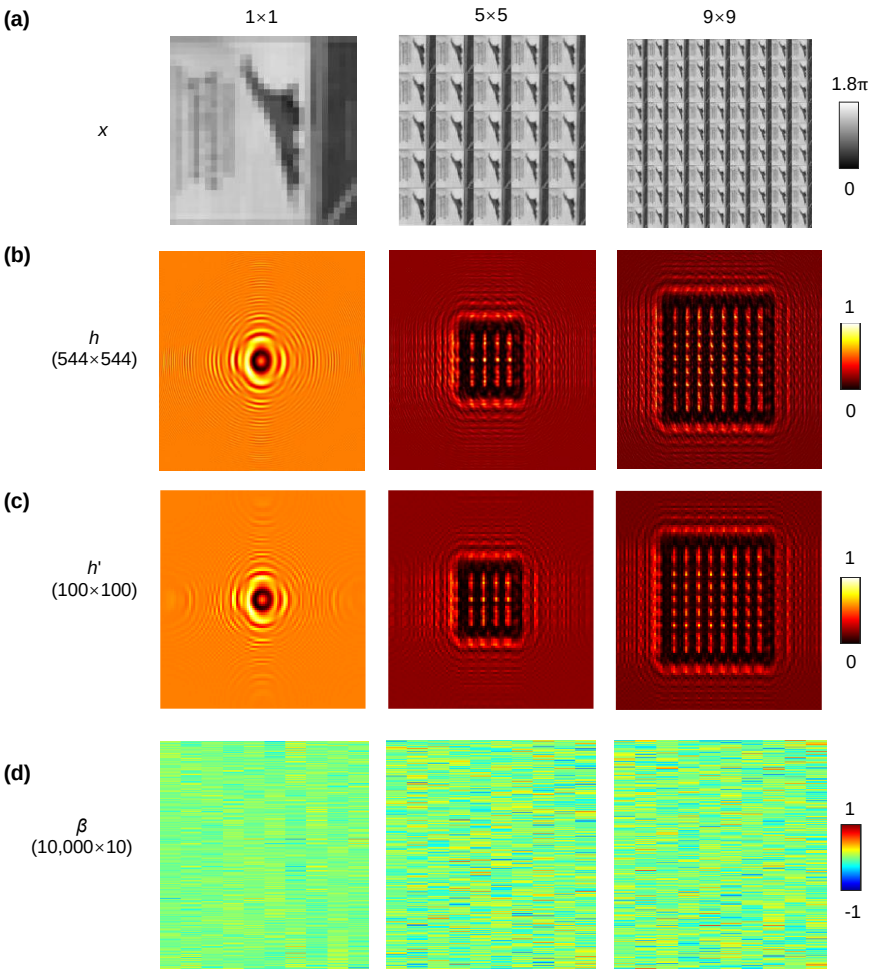

Fig. S4: Network parameters for optical model-based neural network. (a) Input Image examples from CIFAR-10(grayscale) dataset in phase-encoded form. (b) Hidden layer feature matrix **h**. (c) Hidden layer feature matrix **h'** after down sampling. (d) Trained weight matrix **β** with 10,000 hidden neurons.

Supplementary Note 3: Experimental details of photonic neural network

Fig. S5 shows the experimental workflow of photonic neural network, where the duplicate arrays of an input image are phase-encoded on SLM, all the rest pixels are set to a zero-phase level. For CIFAR-10, in addition to grayscale images, we also conducted classification tests on the color images of CIFAR-10 using the network structure shown in Fig. 5(e). For color images, each channel (RGB) is encoded separately, consistent with the encoding method for grayscale images.

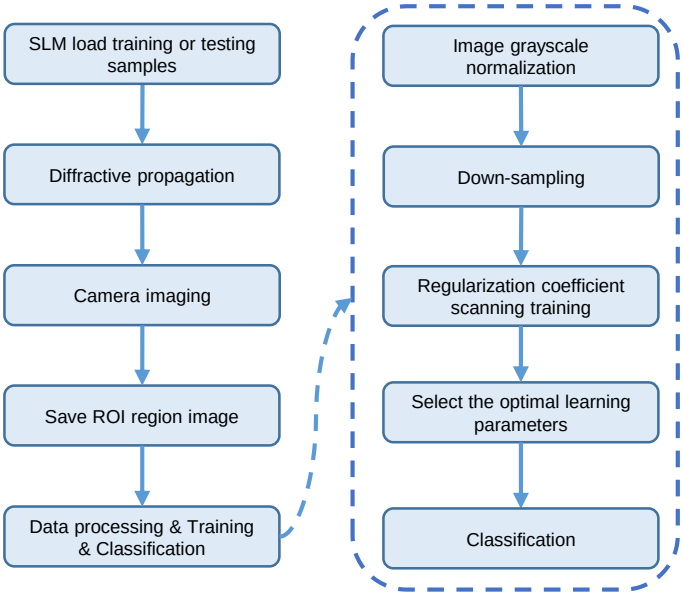

Fig. S5: The workflow of photonic neural network.

The phase-encoded image on SLM and the diffraction image collected from the ROI area (with an ROI window size of 400×400, an actual area of 4.512 mm×4.512 mm) on the camera sensing plane correspond to the input **x** and hidden neurons **h**, respectively, in the general ELM neural network. The diffraction images of all training samples after grayscale normalization and down sampling are used to construct the hidden layer matrix **H** for training. By scanning and selecting an appropriate regularization coefficient *C*, the network learns the parameters **β**. Finally, we use the trained network to classify all testing samples and perform accuracy statistics. It is important to emphasize that in the experiment, the input encoding of the CIFAR-10 color dataset and the collection of diffraction images are conducted separately for samples from each RGB channel. This process follows the experimental procedure of the grayscale image classification network, with the fusion of the three channels completed only after the summation operation.

Fig. S6 demonstrates the impact of different diffraction distances (9, 18, 27 cm) on test accuracy using the MNIST dataset with a 5×5 duplicate array and 10,000 hidden neurons. As the diffraction distance increases, the accuracy remains relatively stable. However, when the diffraction distance becomes too large, the high-frequency features of the image are gradually lost during propagation, leading to a decrease in accuracy. An optimized and suitable propagation distance of 18 cm has been identified for our experiments to enhance accuracy.

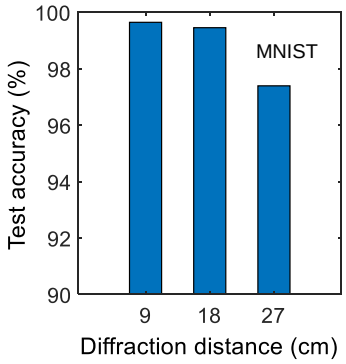

Fig. S6: Experimental impact of diffraction distance on test accuracy

Fig. S7(a)-(d) illustrate test accuracies of the MNIST, Fashion-MNIST, CIFAR-10 (RGB), and CIFAR-10 (grayscale) datasets under different duplicate array formats, all demonstrating the performance enhancement of multi-synaptic connections. The final network's trained weights matrix $\boldsymbol{\beta}$ are determined by the regularization coefficients C selected in Table.S3. Fig. S7(e)-(p) show the confusion matrices for the testing samples with 10,000 hidden neurons. Fig. S8 presents photonic network parameters, including input, hidden layer feature matrix, and trained weight matrix.

For more duplicate array formats, Fig. S9 presents the experimental performance on the CIFAR-10(grayscale) for array format of 1×1, 5×5, 9×9, 11×11, and 20×20. As the array format increases, the accuracy initially rises and then declines. The array formats of 1×1, 5×5, and 9×9 were identified as the optimized results in our experiments. The reason for the accuracy improvement with increasing array formats is discussed in Supplementary Note 4. However, as the array format grows further, the physical limitations of the selected ROI region appear to hinder the effective collection of information from the diffraction field. Therefore, larger array configurations are not further explored in this work. Nevertheless, the accuracy advantage of multi-synaptic connections over mono-synaptic connections remains evident.

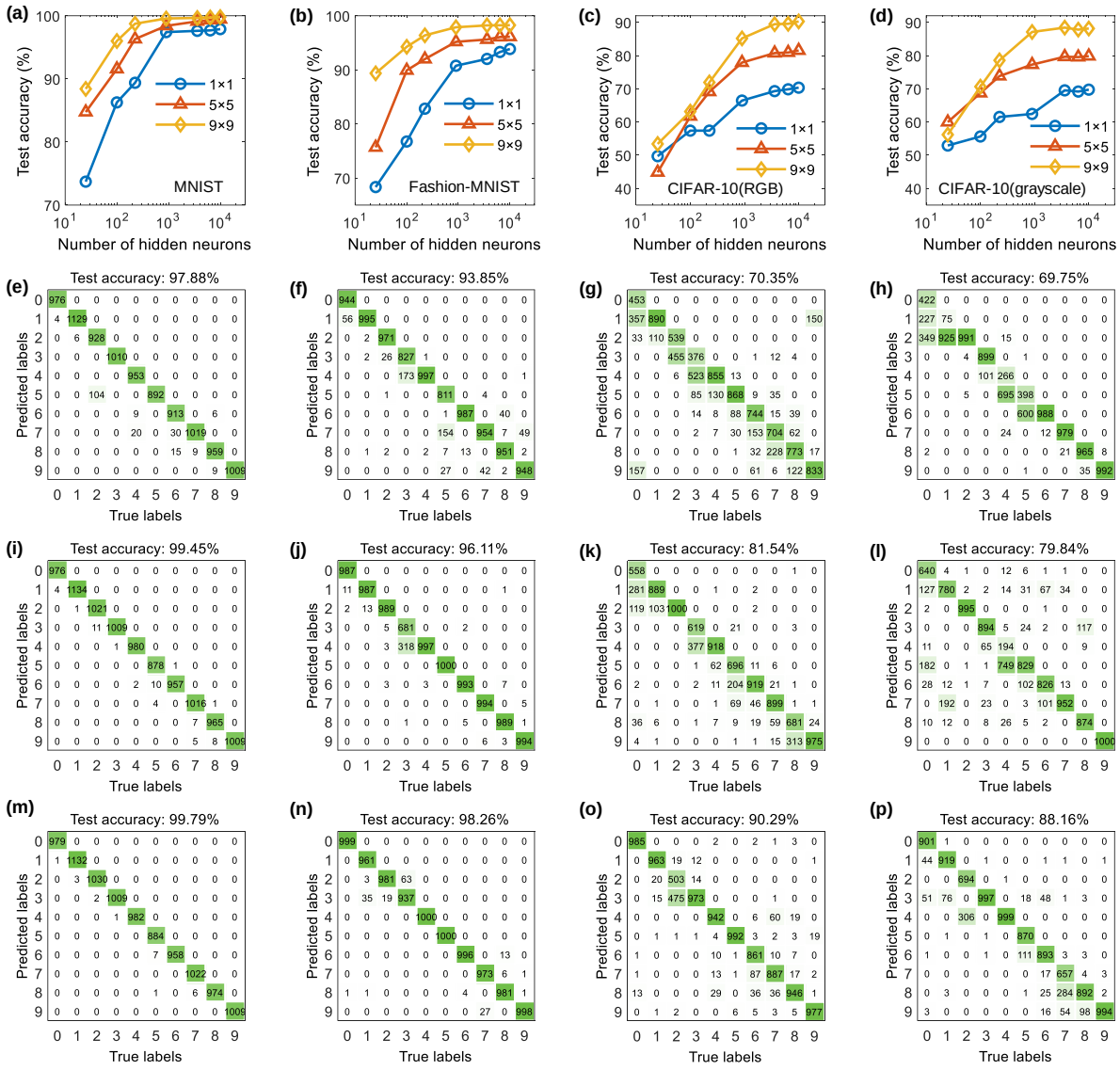

Fig. S7: Experimental results of image classification for photonic neural network. (a)-(d) Test accuracy under different numbers of hidden neurons and duplicate array formats for MNIST, Fashion-MNIST, CIFAR-10(RGB) and CIFAR-10(grayscale). Confusion matrices for the mentioned datasets in the 1×1 ((e)-(h)), 5×5 ((i)-(l)) and 9×9 ((m)-(p)) format with 10,000 hidden neurons.

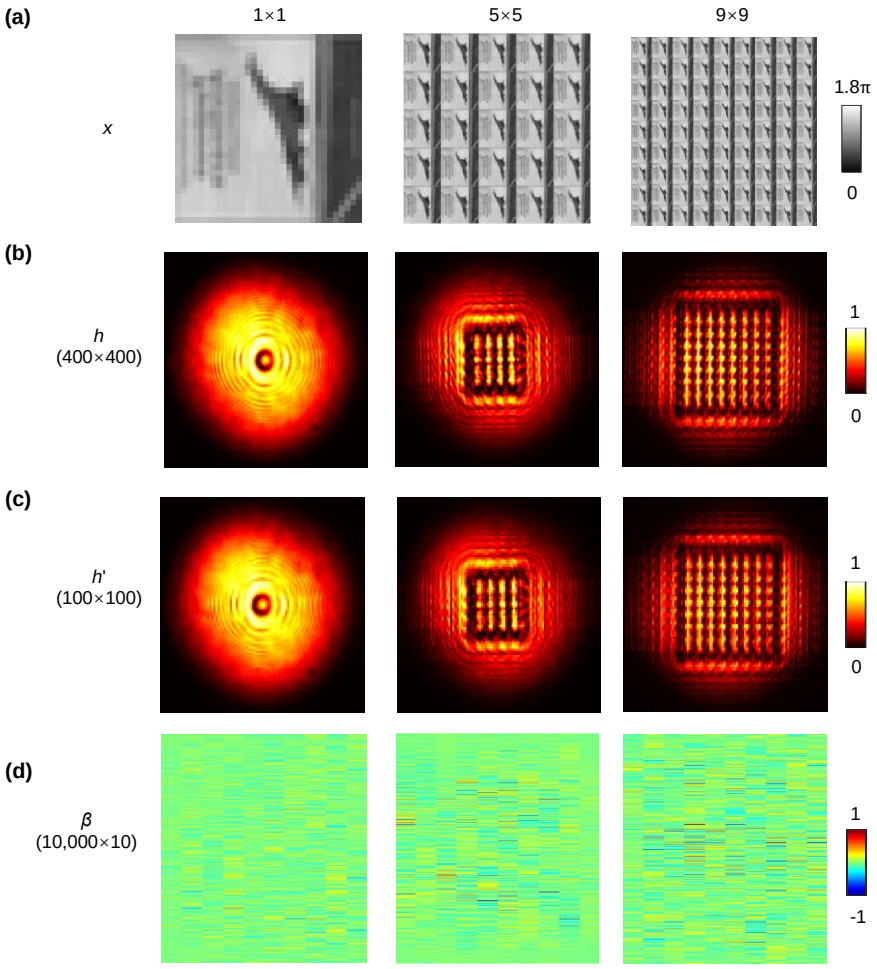

Fig. S8: Network parameters for photonic neural network. (a) Input image examples from CIFAR-10(grayscale) dataset in phase-encoding form. (b) Hidden layer feature matrix **h**. (c) Hidden layer feature matrix **h'** after down sampling. (d) Trained weight matrix **β** with 10,000 hidden neurons.

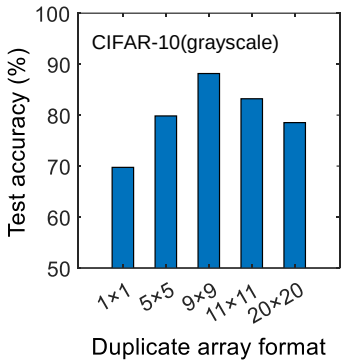

Fig. S9: Experimental impact of duplicate array format on test accuracy.

Supplementary Note 4: Noise Analysis for Multi-Synaptic Connections

The Fourier transform is typically an integral from negative infinity to positive infinity. To deal with real signals with finite lengths or sizes, the input Methods Eq. 4 is truncated by multiplication with a window function $h(u,v)$. In the spatial frequency domain, this operation corresponds to convolution with the window function's frequency spectral distribution $H(U,V)$:

$$A_z(U,V) = \left[A_0(U,V) * H(U,V)\right] S(U,V). \tag{2}$$

The difference between mono-synaptic and multi-synaptic connections lies in the width of the window function. In the spatial frequency domain, the spreading of $H(U,V)$ function causes crosstalk noise between frequency components, which is affected by the window width. To elucidate the mathematical reasons for the performance advantage of multi-synaptic connections, a quantitative analysis of crosstalk noise between adjacent frequencies is presented herein.

Consider two frequency components with a period of l and $l/2$ in an input signal $F_0(u)$ shown in Fig. S10(a), they have a frequency interval of $1/l$ in the frequency distribution of the signal $A_0(U)$ shown in Fig. S10(b). A window function that truncates the signal of an infinitely length is illustrated as a rectangular window $h(u) = \mathrm{rect}(u)$, as shown in Fig. S10(c), where w is the width of rectangular window. By Eq. 2, it causes a frequency component to spread the frequency domain in the form of $H(U) = w\mathrm{sinc}(wU)$ (Fig. S10(d)), leading to interference between frequencies, or so-called crosstalk noise, as shown in Fig. S10(e).

As shown in Figs. S10(d) and (e), the spreading of $H(1/l)$ is narrowed down with enlarging w, such that the crosstalk noise contribution is suppressed. To quantify the crosstalk noise, the maximum noise contribution of $H(1/l)$ that contributes noise to the frequency components beyond $2/l$ is calculated in logarithm[34], as shown in Fig. S10(f). Here, the signal length l is set to 32×8.3 μm (defined as one period width, using CIFAR-10 as an example), and the smallest window width to 32×8.3 μm. The window width is varied from 1 to 10 times the

period width, observing a decrease in crosstalk noise with increasing window. We suggest this reduction in crosstalk noise as a reason for improving accuracy by duplicate array formats of input for multi-synaptic connections in optical model-based and photonic neural networks. Fig. S10(f) shows a continued downward trend with increasing window width. However, due to other error sources (including mathematical approximations and digital noise), optical model-based networks cannot enjoy this trend as much as physical transformation based photonic networks.

For all duplicate array formats, we chose a duplicate interval of 0 as the optimized option. Here, we discuss the crosstalk noise for different duplicate intervals. As the duplicate interval increases, the rectangular function transitions from a wide rectangular window to many separated rectangular windows with a width of 1×1 duplicate array format (Fig. S11(a)). Taking a 5×5 duplicate array format as an example, it causes a frequency component to spread the frequency domain in the form of $H(U) = \sum_{n=-2}^{2} \mathrm{sinc}(wU)\mathrm{e}^{-j2\pi wn(w+d)}$, where d is duplicate interval. According to our definition of crosstalk noise (Fig. S11(b)-(d)), as the duplicate interval increases, the crosstalk noise also increases, which explains the observed decrease in test accuracy shown in the Fig. 2(c).

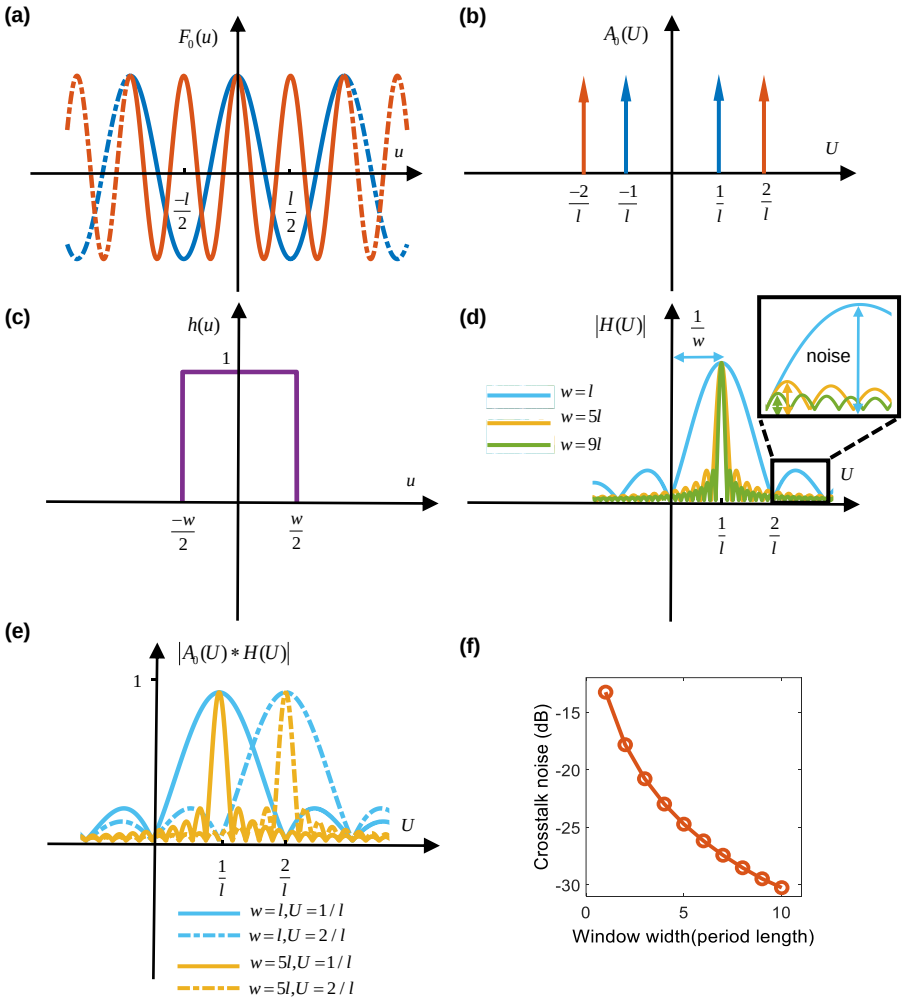

Fig. S10: Analysis of frequency crosstalk noise for different duplicate array formats. (a) The input signal with a period length of l and $l/2$. (b) The signal spatial frequency distribution. (c) Rectangular window truncation function. (d) Fourier transform function of rectangular function at $1/l$. The crosstalk noise of $1/l$ to $2/l$ is defined in the figure. (e) The frequency spectrum distribution of the truncation function at $1/l$ and $2/l$. (f) The crosstalk noise at $2/l$ due to the frequency spectrum diffusion at $1/l$ as the window width varies.

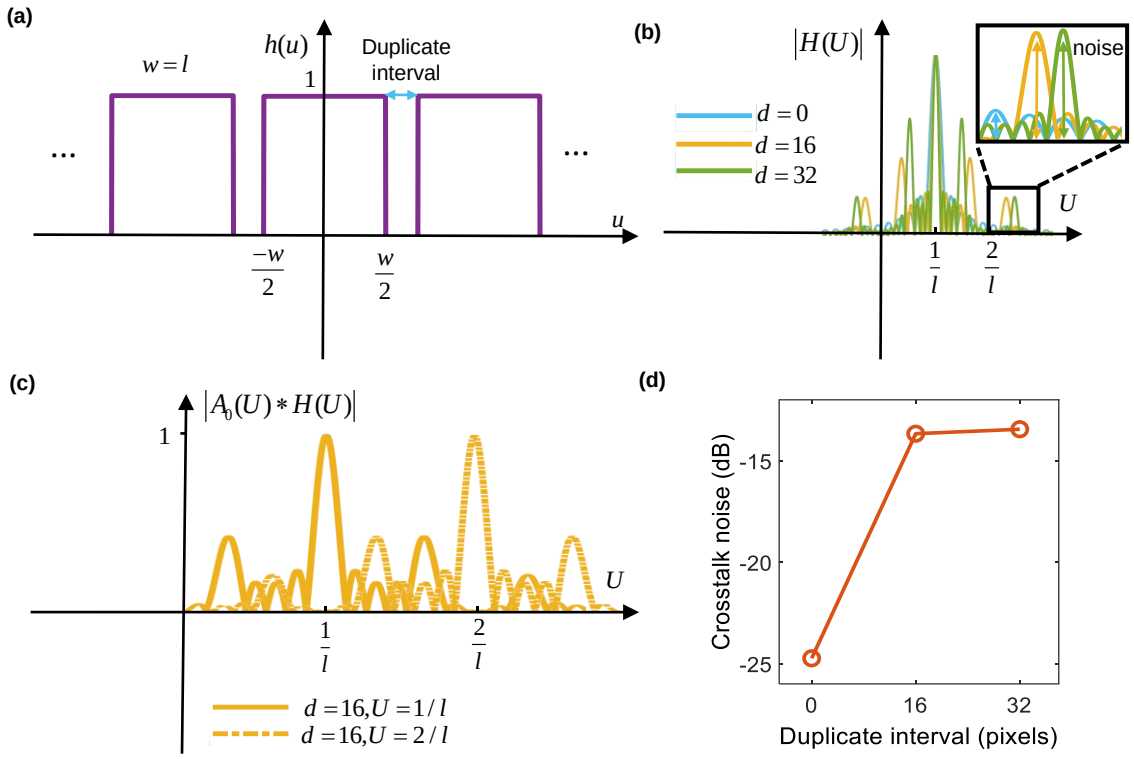

Fig. S11: Analysis of frequency crosstalk noise for different duplicate intervals. (a) Rectangular window truncation function. (b) Fourier transform function of rectangular function at 1/*l*. The crosstalk noise of 1/*l* to 2/*l* is defined in the figure. (c) The frequency spectrum distribution of the truncation function at 1/*l* and 2/*l*. (d) The crosstalk noise at 2/*l* due to the frequency spectrum diffusion at 1/*l* as the duplicate interval varies.

Supplementary Note 5: Classification Performance and Computing Efficiency Analysis

Table.S4 summarizes advanced works on digital and hardware networks, including electronic, hybrid and optical networks. The photonic multi-synapse neural network architecture competes with deep digital networks on the three datasets, outperforming most hardware networks. This demonstrates the superior performance of the photonic architecture, which can achieve high accuracy using only a single hidden layer.

The computing speed is assessed by dividing the number of operations by the computation time:

$$\xi = \frac{R_o + R_e}{t_o + t_e}, \tag{3}$$

where R_o and R_e represent the optical and electrical operation counts, respectively. $t_o \approx 3\,\mathrm{ms}$ is the optical projection time, primarily encompassing the camera's exposure time. $t_e \approx 6.2\,\mathrm{ms}$ is the runtime for data processing and classification on the digital end, measured using the same method described in Methods. For a CIFAR-10 image with 32×32 input neurons, the photonic multi-synapse neural network architecture forms a 9×9 duplicate array and projects it to 400×400 hidden neurons, the estimated number of optical projection operations is approximately 2.654×10^{10} OPs. This includes a nonlinear operation count of 400^2, which can be considered negligible. As an example, taking the down sampling to 900 hidden neurons, the electronic operations, which include down sampling and matrix multiplication, involve 1.7×10^5 OPs. The photonic multi-synapse neural network exhibits a computing speed of 2.89 TOPs/s.

To assess energy consumption, the energy expenditure directly used for computation, denoted as direct energy efficiency, is given as

$$\eta = \frac{R_o + R_e}{P_o t_o + R_e / \eta_e}, \tag{4}$$

which is calculated based on the optical and electronic energy consumption that is directly used for the computational operations in network architectures[35]. The experiment utilized a laser power of 0.86 μW, and the exposure time for a single image is 3 ms, indicating that optical projection operations from SLM plane to the camera sensing plane consumes 2.58 nJ of energy. There is potential for further optimization of the laser power or exposure time.

The electronic operations are processed with an Intel(R) Core(TM) i9-14900KF CPU @ 3.20 GHz, computing speed ξ_e= 1.23 TOPs/s, power consumption P_e= 125.0 W, and the calculate energy efficiency $\eta_e = \xi_e/P_e$= 9.83 GOPs/J. Therefore, this architecture has a direct energy efficiency of 1.53 POPs/J. Considering the optical projection only, the direct energy efficiency can reach 10287 POPs/J.

The computing speed of the cutting-edge GPU (NVIDIA H100 PCIe) is 51.2 TOPs/s with an energy efficiency of ~0.15 TOPs/J[67,68]. Our architecture consists solely of a single hidden layer network, with optical computations accounting for over 99% of the process. Expanding to deeper networks could increase the ratio of optical to electrical computations, thus further enhancing computational speed and reducing energy consumption.

For the optical projections, the MAC operations count in our architecture is 1.327×10^{10}. The energy consumption for one optical MAC operation is calculated to be 0.2 aJ. Compared to other works (Table.S5), this architecture demonstrates relatively low energy expenditure.

Furthermore, this architecture is capable of mapping high-resolution images and handling exceptionally large data dimensions. Additionally, on the detection plane, physical hidden neurons reduction can be achieved by using low-resolution pixel detectors, significantly alleviating the computational burden on electronic computers.

Except for the inference, our architecture has cost advantages over deep learning, especially for the training. Its training speed, in the order of seconds (Less than 1 second for $\boldsymbol{\beta}$ with 900 hidden neurons and one selected *C*, while approximately 40 seconds for 10,000 hidden

neurons), makes it suitable for tasks requiring frequent training, without the need for high-end training hardware.

Table.S1 Regularization coefficient *C* for RNT-based neural network

Datasets	Duplicate array format	Number of hidden neurons						
		25	**100**	**225**	**900**	**3,600**	**6,400**	**10,000**
MNIST	1×1	10^3	10^6	10^4	10^2	10^4	10^4	10^1
Fashion-MNIST	1×1	10^6	10^6	10^7	10^5	10^3	10^2	10^1
CIFAR-10(grayscale)	1×1	10^3	10^6	10^3	10^3	10^2	10^1	10^1
	5×5	10^3	10^5	10^3	10^4	10^2	10^1	10^1
	9×9	10^9	10^5	10^6	10^4	10^2	10^1	10^1

Table.S2 Regularization coefficient *C* for optical model-based neural network

Datasets	Duplicate array format	Number of hidden neurons						
		25	**100**	**225**	**900**	**3,600**	**6,400**	**10,000**
MNIST	1×1	10^8	10^6	10^7	10^9	10^5	10^5	10^4
	5×5	10^7	10^4	10^4	10^5	10^7	10^5	10^3
	9×9	10^6	10^5	10^4	10^6	10^6	10^2	10^5
Fashion-MNIST	1×1	10^7	10^8	10^7	10^{10}	10^8	10^7	10^6
	5×5	10^7	10^5	10^6	10^4	10^6	10^6	10^4
	9×9	10^4	10^8	10^7	10^6	10^5	10^7	10^4
CIFAR-10(grayscale)	1×1	10^5	10^4	10^4	10^2	10^2	10^1	10^2
	5×5	10^4	10^4	10^4	10^2	10^2	10^1	10^1
	9×9	10^2	10^3	10^3	10^2	10^1	10^1	10^1

Table.S3 Regularization coefficient *C* for photonic neural network

Datasets	Duplicate array format	Number of hidden neurons						
		25	**100**	**225**	**900**	**3,600**	**6,400**	**10,000**
MNIST	1×1	10^{2}	10^{1}	10^{0}	10^{1}	10^{0}	10^{-1}	10^{-1}
	5×5	10^{2}	10^{2}	10^{4}	10^{3}	10^{3}	10^{1}	10^{1}
	9×9	10^{3}	10^{3}	10^{4}	10^{3}	10^{2}	10^{2}	10^{2}
Fashion-MNIST	1×1	10^{3}	10^{2}	10^{1}	10^{0}	10^{-1}	10^{-1}	10^{-1}
	5×5	10^{5}	10^{5}	10^{3}	10^{3}	10^{1}	10^{1}	10^{4}
	9×9	10^{5}	10^{3}	10^{5}	10^{4}	10^{3}	10^{3}	10^{1}
CIFAR-10(RGB)	1×1	10^{4}	10^{2}	10^{1}	10^{1}	10^{0}	10^{0}	10^{0}
	5×5	10^{3}	10^{3}	10^{1}	10^{0}	10^{-1}	10^{-2}	10^{-2}
	9×9	10^{2}	10^{1}	10^{2}	10^{1}	10^{0}	10^{-1}	10^{-1}
CIFAR-10(grayscale)	1×1	10^{6}	10^{2}	10^{1}	10^{4}	10^{2}	10^{1}	10^{1}
	5×5	10^{4}	10^{2}	10^{1}	10^{-1}	10^{0}	10^{-2}	10^{-2}
	9×9	10^{4}	10^{3}	10^{2}	10^{2}	10^{1}	10^{0}	10^{0}

Table.S4 Test accuracy of different architectures on three benchmark datasets

Type	Architecture	MNIST(%)	Fashion-MNIST(%)	CIFAR-10(%)
Digital	LeNet[29]	99.40[36]	98.80[37]	66.43[38]
	AlexNet[39]	99.23[40]	88.82[41]	72.64[38]
	ELM[42]	97.39±0.1[42]	93.38[43]	44.09[44]
	ML-ELM[45]	99.03±0.04[10]	93.56[43]	/
	SVM[46]	94.00[47]	83.20[38]	37.13[38]
	ResNet[48]	99.51[38]	93.23[38]	95.55[49]
Electronic	PCM Memristor ResNet-32[50]	/	/	93.70
	NeuRRAM[51]	99.00	/	85.70
	MRAM-DIMA[52]	~98.00	/	~91.00
	STT-MRAM[53]	97.20	/	81.30
	Discrete Memristor NN[54]	97.16	/	/
	Perovskite Memristor NN[55]	/	90.10	/
	STELLAR[56]	92.30	/	/
	NWNs[57]	93.40	/	/
Hybrid	OE-ELM[24]	~98.00	/	/
	OE-CNN[9]	/	/	44.40
	OE-CNN[58]	98.00	/	54.00
	OE-PLNN[59]	96.36	/	/
	OE-ELM[23]	92.18	/	/
	OE-D^2NN[19]	89.10	81.70	/
	OE-D^2NN[60]	95.47	/	/
	OE-LOEN[61]	97.21	/	/

	ACCEL[20]	85.50	82.00	/
	VCSEL neuron[62]	96.10	/	/
	DPU[35]	96.60	84.60	
	DANTE ONN[63]	~96.00	/	51.00
Optical	D^2NN[4]	88.00	90.00	/
	D^2NN[18]	84.00	/	/
	ONN[64]	99.00	/	/
	PCM chip[15]	95.30	/	/
	PNN[22]	97.00	/	/
	D^2NN chip[65]	89.30	81.30	/
	Taichi[66]	/	/	93.7
	Photonic multi-synapse neural network (this work)	99.79	98.26	90.29

Table.S5 Energy efficiency of different architectures

Architecture	Energy efficiency(/MAC)
MZI mesh[10]	30fJ
DPU[35]	0.82fJ
Sub-λ Nanophotonics[69]	30.6aJ
WDM&PCM chip[15]	17fJ
Photonic multi-synapse neural network (optical projections)	0.2aJ