Science Bulletin 65 (2020) 1177–1183

Contents lists available at ScienceDirect

Science Bulletin

journal homepage: www.elsevier.com/locate/scib

Article

Inverse design of an integrated-nanophotonics optical neural network

Yurui Qu [a,b,1], Huanzheng Zhu [c,1], Yichen Shen [d], Jin Zhang [e], Chenning Tao [c], Pintu Ghosh [c], Min Qiu [a,b,*]

[a] Key Laboratory of 3D Micro/Nano Fabrication and Characterization of Zhejiang Province, School of Engineering, Westlake University, Hangzhou 310024, China
[b] Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310024, China
[c] State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China
[d] Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[e] Electrical and Computer Engineering, University of California, San Diego, CA 92093, USA

ARTICLE INFO

Article history:
Received 27 December 2019
Received in revised form 10 March 2020
Accepted 19 March 2020
Available online 1 April 2020

Keywords:
Optical neural networks
Deep learning
Inverse design
Integrated nanophotonics
Silicon photonics

ABSTRACT

Artificial neural networks have dramatically improved the performance of many machine-learning applications such as image recognition and natural language processing. However, the electronic hardware implementations of the above-mentioned tasks are facing performance ceiling because Moore's Law is slowing down. In this article, we propose an optical neural network architecture based on optical scattering units to implement deep learning tasks with fast speed, low power consumption and small footprint. The optical scattering units allow light to scatter back and forward within a small region and can be optimized through an inverse design method. The optical scattering units can implement high-precision stochastic matrix multiplication with mean squared error $< 10^{-4}$ and a mere $4 \times 4\ \mu m^2$ footprint. Furthermore, an optical neural network framework based on optical scattering units is constructed by introducing "Kernel Matrix", which can achieve 97.1% accuracy on the classic image classification dataset MNIST.

1. Introduction

The past few years have witnessed the great strides made in the field of deep learning and its applications in image classification [1], speech recognition [2] and decision-making [3]. Deep learning has also penetrated into a number of different areas of science such as drug design [4,5], genetics [6,7], materials science [8], high-energy physics [9] and photonic structure design [10–15]. Solving complex problems with high accuracy necessitates large and deep neural networks, which in turn requires huge computional resources and high-efficiency hardwares. A number of fast and energy-efficient electronic hardwares are developed such as graphical processing units (GPUs), field-programmable gate arrays (FPGAs) and application-specific integrated circuits (ASICs), including Google TPU. Besides, new computing architectures like neuromorphic computing [16–18] and quantum machine learning [19] with the potential for fast speed and low power consumption are demonstrated.

Optical neural networks (ONNs), in this regard, is a promising alternative method to electronic counterparts. ONNs present two distinctive advantages. (1) Artificial neural networks (ANNs) relies heavily on matrix multiplications which can be performed at speed of light and detected by at rates exceeding 50 GHz [20] in ONNs. (2) Once trained, ONN is passive and computation on the optical signals can be implemented with minimal power consumption [21]. ONNs based on diffractive optics [22] and free-space optics [23,24] have been reported but require 3D-printed diffractive elements or bulky optical components such as lenses, and thereby it makes scaling to large number of neurons challenging. Integrated photonic circuits [25] can provide a CMOS-compatible, scalable appoach to implement optical deep learning tasks, but the current footprint of on-chip Mach–Zehnder interferometers is larger than 100 μm and makes scaling to a large matrix multiplication (1000×1000) impossible. To date, the goal of an integrated ONN circuit that is scalable to a large matrix multiplication remains unrealized.

In this article, we introduce a new integrated ONN framework based on optical scattering unit (OSU) which is fast, energy-efficient and scalable. The OSU is an integrated-nanophotonics computing unit consisting of a multimode inteference (MMI) coupler with nanopatterned coupler region. Nanopatterning enables light scattering within a small coupler region and increases the degrees of freedom which can be optimized through an inverse design method. The OSU can be divided into coherent and incoherent architectures. Coherent OSU can, in principle, implement arbitary unitary matrix multiplications. Incoherent OSU, however,

* Corresponding author.
E-mail address: qiu_lab@westlake.edu.cn (M. Qiu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scib.2020.03.042
2095-9273/

can directly manipulate light intensity to achieve stochastic matrix multiplications. In general, operations on light intensity can be much easier because the phase coherence is not required. The studies of incoherent photonic architectures remains insufficient so far. We demonstrate that incoherent OSU can implement high-precision stochastic matrix function with less than 10^{-4} mean squared error and a mere 4×4 μm^2 footprint. By convolving input images with "Kernel Matrix", the ONN framework based on OSU can achieve 97.1% accuracy on the classic image classification dataset MNIST.

2. Framework and methods

2.1. ONNs framework

The schematic of ANN is presented in Fig. 1a. Each neuron is connected to the other neurons of the next layer (in the form of matrix multiplication) and then followed by a nonlinear activation function. The input data is fed into the ANN, and the output is calculated through the forward propagation. All parameters (weights) are optimized using back propagation [26] and embedded in the fabrication of the chip. The ONN framework is depicted in Fig. 1b. The input data is preprocessed to a high-dimensional vector which is encoded into optical signals. Forward propagation is implemented by multi-layer ONNs, each ONN consisting of a matrix multiplication unit and a nonlinear unit. The nonlinear unit is implemented using common optical nonlinearities, such as saturable absorption [27,28] or hybrid model [25] which have been demonstrated in previous photonic circuits. In principle, the ONN architectures can achieve arbitrary ANN functions in the optical domain. Next, we introduce a compact integrated ONN framework that can use both coherent light and incoherent light for computing.

2.2. Training methods of optical CNN models

The physical structures that can implement single matrix multiplication are optimized using the finite-difference frequency-domain (FDFD) simulation with adjoint method [29]. The optical CNN model is constructed and trained using the open-source machine learning framework TensorFlow. The optical CNN contains two sets of alternating 3×3 convolution layers and 2×2 max pooling layers followed by a fully connected layer. Max pooling can be regarded as a nonlinear function. We first randomly initialize the weights with a normal distribution and then add extra restrictions to the weights such as the special matrices implemented by OSUs. We use Adam optimizer with a mini-batch size of 400. Loss function is the cross entropy loss between the prediction and the label. Here, the model is chosen without regularization and still performs quite well.

3. Results

3.1. Coherent photonic architecture

To realize a compact and scalable optical computing unit on an integrated photonic platform, we propose an OSU consisting of a patterned single-layer MMI couple on a silicon-on-insulator (SOI) substrate, as shown in Fig. 2a. The patterned OSU allows light to scatter back and forward, and thereby it provides a vast number of degrees of freedom for optimization within a small region. Several methods such as steepest-gradient descent [30] or direct-binary search [31] have been applied to perform such optimization. Here, we use the adjoint method to explore the full design space of the coupler region of OSU, which allows us to implement matrix multiplications using only one MMI coupler. The adjoint method can compute the gradient of an objective function with respect to the refractive index of the design region using only two full-field simulations [29]. Here, we assume that the optical signals from different input ports have the same frequency and are coherent at the output ports. Considering the optical loss to be negligible, the OSU, in principle, can implement arbitrary unitary matrix multiplication (see the Supplementary materials). To realize an optical computing unit that can implement any real-valued matrix, we note that an arbitrary, real-valued matrix M can be decomposed as $M = U\Sigma V^*$ through singular value decomposition [32]. Σ is a diagonal matrix without negative real numbers which can be implemented using Mach Zehnder interferometers [25] or optical attenuators/ amplification materials such as semiconductors and dyes [33]. U, V^* are unitary matrices, which can be implemented using OSUs.

We choose rotation matrix as the optimization goal of the OSU. Rotation matrix is one special kind of unitary matrix and can perform a rotation of one vector in Euclidean space. Fig. 2b shows an example of the rotation matrix that can rotate the vector (6,5,4) to the vector (1,2,3) in three-dimensional space. The vectors (6,5,4) and (1,2,3) have been normalized to unit vectors based on light intensity. For optimizing OSU, 500 examples that satisfy the object rotation matrix requirement are chosen, each example consisting of a pair of input and output vectors. Among them, 400 examples are training data and 100 examples are test data. The input and output vectors are normalized first (the summation of light intensities equals one) and then encoded on the amplitude of the electric field. The prediction can be calculated based on the light intensities of output ports. The mean squared error (MSE) is used to evaluate the difference of light intensities between the exact output and the prediction as $\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(I_{i,\mathrm{prediction}} - I_{i,\mathrm{exact}})^2$, where N represents the number of ports. The OSU is first optimized using training data and then used to calculate the MSE using test data. The MSE decreases over training epochs, which demonstrates the OSU truly learns about the features of the object rotation

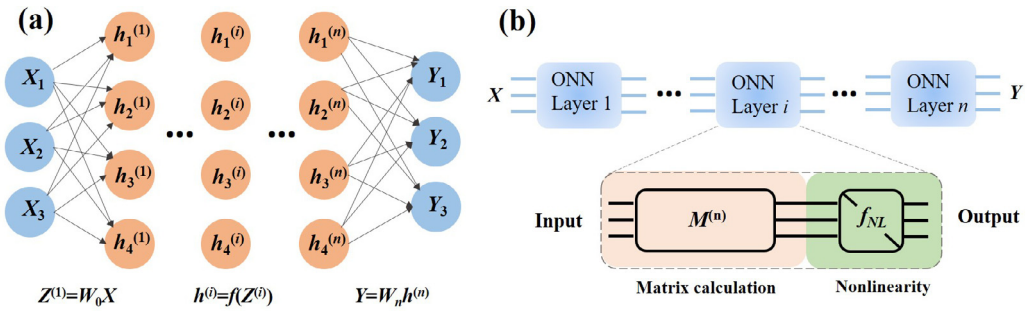


Fig. 1. (Color online) The schematic diagrams of general ANN and ONN architectures. (a) ANN consists of an input layer, several hidden layers and an output layer. (b) A matrix multiplication unit and a nonlinear unit that compose each layer of ONN.

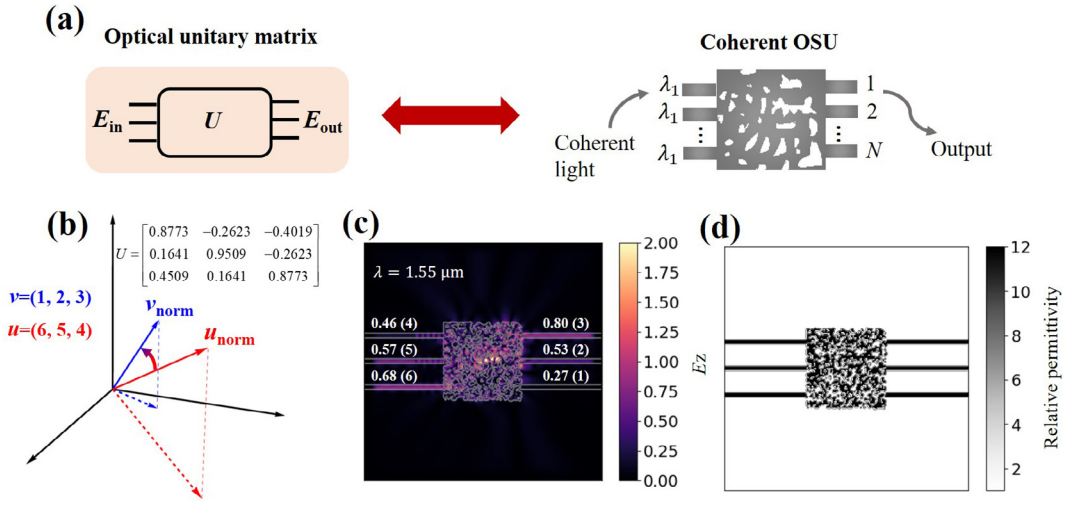


Fig. 2. (Color online) Optical unitary matrix and coherent OSU. (a) Optical unitary matrix multiplication can be implemented by coherent OSU with a patterned coupler region. (b) Rotation matrix, which is one special unitary matrix, can rotate the unit vector in three-dimensional space. (c) The electric field distribution of the optimized OSU that can rotate the vector $u = (6, 5, 4)$ to the vector $v = (1, 2, 3)$. Two vectors have been normalized to unit vectors based on light intensity. (d) The optimized refractive index distribution for the object rotation matrix.

matrix (Fig. S1 online). We plot the field distribution which shows that the OSU rotates the normalized vector (6,5,4) to (1,2,3). The result matches well with the object rotation matrix, as shown in Fig. 2c. The footprint of the optimized coupler region is 4×4 μm^2 which is discretized into 80×80 pixels. The final optimized refractive index distribution of the OSU is shown in Fig. 2d. The refractive index of each pixel is constrained to lie between 1 (air) and 3.47 (Si at 1.55 μm). The wavelength of light source is 1.55 μm.

3.2. Incoherent photonic architecture

The incoherent photonic architecture has an key advantage over its coherent counterpart because it does not require phase coherence. Practically, the systems of incoherent photonic architectures are much simpler to realize compared to the coherent photonic architectures. Incoherent photonic architecture based on OSUs is proposed by adding optical signals of various wavelengths to different ports (shown in Fig. 3a). Here, the patterned coupler region of the OSU can be regarded as a "beam splitter" that splits one beam of light to several output waveguides. At each output waveguide, the light intensities of different wavelengths can be summed directly due to their incoherent nature (see details in the Supplementary materials). The stochastic matrix (also called probability matrix, Markov matrix, etc.) can be used to describe the relation between input and output intensities in an incoherent OSU if the law of energy conservation is satisfied. The stochastic matrix P, which is represented by P_{ij} as the i^{th} row and j^{th} column entry, can be described as

$$\sum_{i=1}^{S} P_{ij} = 1. \tag{1}$$

Each column of the stochastic matrix P represents the light intensities of one specific wavelength in all output waveguides. Here, we assume that the input intensity of each wavelength equals to one, and therefore the summation of each column of matrix P has to be equal to one as well. More details about why this incoherent OSU can implement the stochastic matrix multiplication are described in the Supplementary materials.

We begin with a special 4×4 stochastic matrix as our optimization goal, as shown in Fig. 3b. The output intensity, which is calculated by multiplying the input matrix I_{in} with the stochastic matrix P, can be expressed as

$$I_{out} = PI_{in} = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.25 & 0.33 & 0 & 0 \\ 0.25 & 0.33 & 0.5 & 0 \\ 0.25 & 0.33 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = I_1 \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} + I_2 \begin{bmatrix} 0 \\ 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} + I_3 \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 0.5 \end{bmatrix} + I_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{2}$$

The first column of the matrix P requires that the light from input port 1 is splitted into four output waveguides with equal intensity. The second column of the matrix P requires that the light from input port 2 is split into output waveguides 2', 3' and 4' with equal intensity, and so on. Through the inverse design of the coupler region, the function of the OSU can gradually approach the objective stochastic matrix multiplication. The MSE descreases to 1×10^{-4} as the training epoch increases to 200, which demonstrates that the OSU can implement the object stochastic matrix with negligible errors (Fig. 3c). The optimized refractive index distribution of the OSU is shown in Fig. 3d. The footprint of the optimized coupler region is 4×4 μm^2 which is discretized into 80×80 pixels. The performances of different footprints of the optimized coupler regions are compared and shown in Fig. S2 (online). The field distributions at different wavelengths are plotted in Fig. 3e. Here, four wavelengths of optical signals are 1.55, 1.56, 1.57 and 1.58 μm. The electric field is simulated using the FDFD method. The electric field distributions and the objective stochastic matrix are well matched. The OSU is robust to the noises of the input signals (wavelength, phase and intensity) and exhibits good fabrication tolerance (Figs. S3 and S4 online).

3.3. Optical convolution and deep learning

Our ONN framework is constructed based on the aforementioned incoherent OSU and performs image classification task on the classic dataset MNIST [34]. First, we introduce optical convolutional neural network (CNN). Conventional CNN contains a number

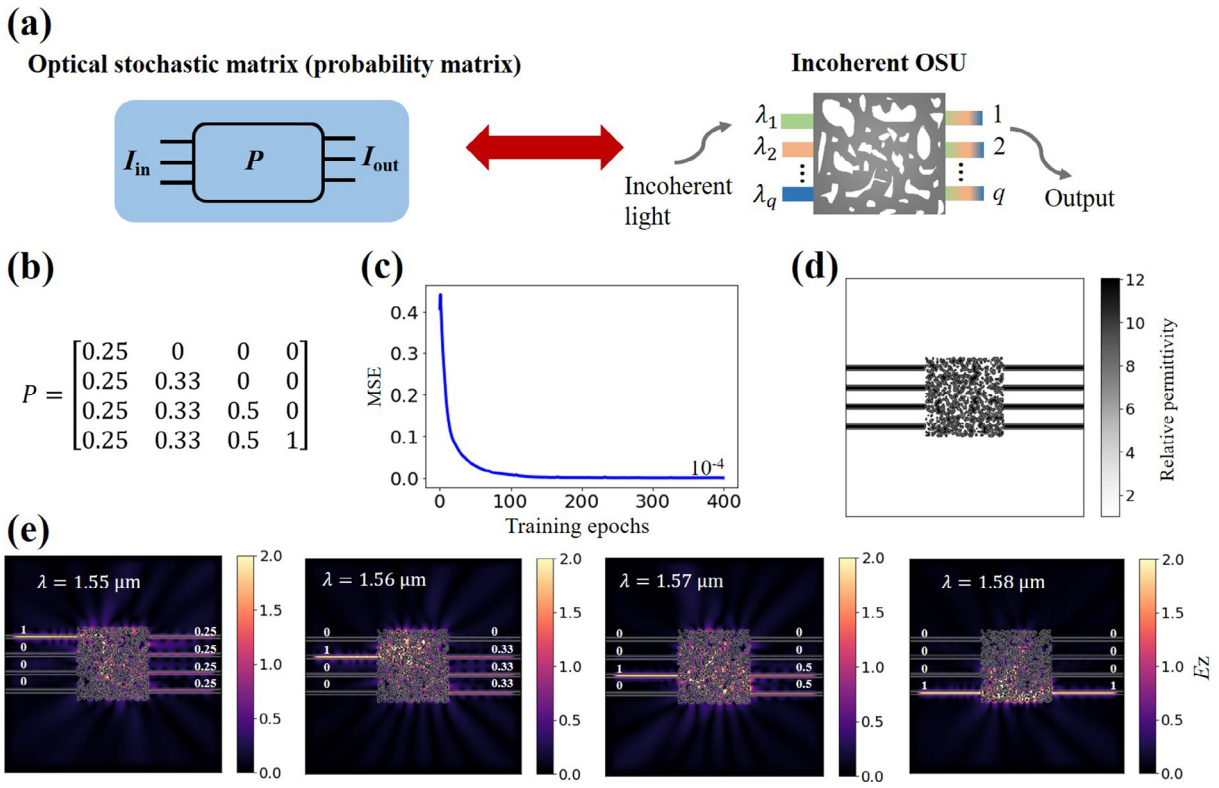


Fig. 3. (Color online) Incoherent OSU and optical stochastic matrix. (a) Incoherent OSU with input light of various wavelengths in different ports. Such OSU can implement optical stochastic matrix multiplication. (b) A special 4 × 4 stochastic matrix as the goal of optimization. (c) Mean square error (MSE) during training process. (d) The optimized refractive index distribution of the OSU. (e) The electric field distributions corresponding to four different wavelengths. Each subgraph represents one column of the objective stochastic matrix.

of convolutional layers, pooling layers and several fully connected layers. Among them, the convolutional layers are the most computation-intensive layers, especially for deep neural networks. For example, VGG19 has 16 convolutional layers and 3 fully connected layers [35], and ResNet can have 152 convolutional layers at most [36]. Here, we propose a CNN algorithm whose convolutional layers can be implemented in optical domain. The CNN algorithm contains two sets of alternating 3 × 3 convolution layers and 2 × 2 max pooling layers followed by a fully connected layer (Fig. 4a). Optical convolutional operations can be performed in photonic circuits using "Kernel Matrix" [37], which implements a transformation from the convolutional operation to the optical kernel matrix multiplication. In Fig. 4b, an image is segmented into a set of "patches" with the same dimension as the convolutional kernels. These patches are convolved with the set of kernels and the outputs are a set of matrices whose entries are kernel-patch dot products, which is the general convolutional operation. To implement convolutional operations in OSU, the image patches are vectorized and then sent into an ONN in sequence (Fig. 4c). The set of kernels are converted to "Kernel Matrix" by vectorizing and stacking each kernel. The one-dimensional image patches can thus be efficiently multiplied by the "Kernel Matrix", which is equivalent to convolutional operations.

The conventional CNN on a 64-bit computer and the ONN architecture are trained together using the same 55,000 handwritten digit images and tested using another 5000 images for comparison. The conventional CNN algorithm has the same architecture as that shown in Fig. 4a. We apply 3 × 3 kernels with 1 input channel and 4 output channels, as well as 3 × 3 kernels with 4 input channels and 8 output channels to the first and the second convolutional layer, respectively. For the ONN architecture, 3 × 3 kernels can be transformed into the "Kernel Matrix" with nine input and output channels, which can be implemented by a nine-port OSU (see the inset in Fig. 4c). The first and second convolutional layers can be implemented by one and three layers of 9 × 9 optical "Kernel Matrix", respectively. The max pooling operation can be regarded as a nonlinear function because it takes the max of selected region and creates a new output matrix which is not linearly related to the input matrix. In the simulation, the proposed ONN architecture can achieve the classification accuracy of 97.1%, which is comparable to that of a 64-bit computer (97.3%). Fig. 4d and e report the confusion matrices of the 64-bit computer and the ONN architecture for the same 5000 test images, respectively. The results clearly demonstrate that the ONN architecture can learn handwritten digit classifications as well as the conventional CNN algorithm on a 64-bit computer.

3.4. The footprints and performances of patterned and unpatterned MMI coupler architectures.

To demonstrate the reduced footprint do not sacrifice a lot in the performance of our ONN architecture, we compare the footprints along with the number of parameters and performances of patterned and unpatterned MMI architectures in Table 1. The patterned MMI coupler is the aforementioned OSU, while the unpatterned one is the common MMI coupler which is widely used as a splitter, combiner, switch and multiplexer. A general unpatterned $N \times N$ MMI coupler contains N incoming and N outcoming waveguides, as well as a multimode coupler in the middle. The field distribution of input waveguides can be reproduced at the output waveguides, which is called N self-images. These input and output waveguides have equal magnitudes but different phases. In general, the length of the coupler region has to be much larger than its width to satisfy the self-imaging condition.

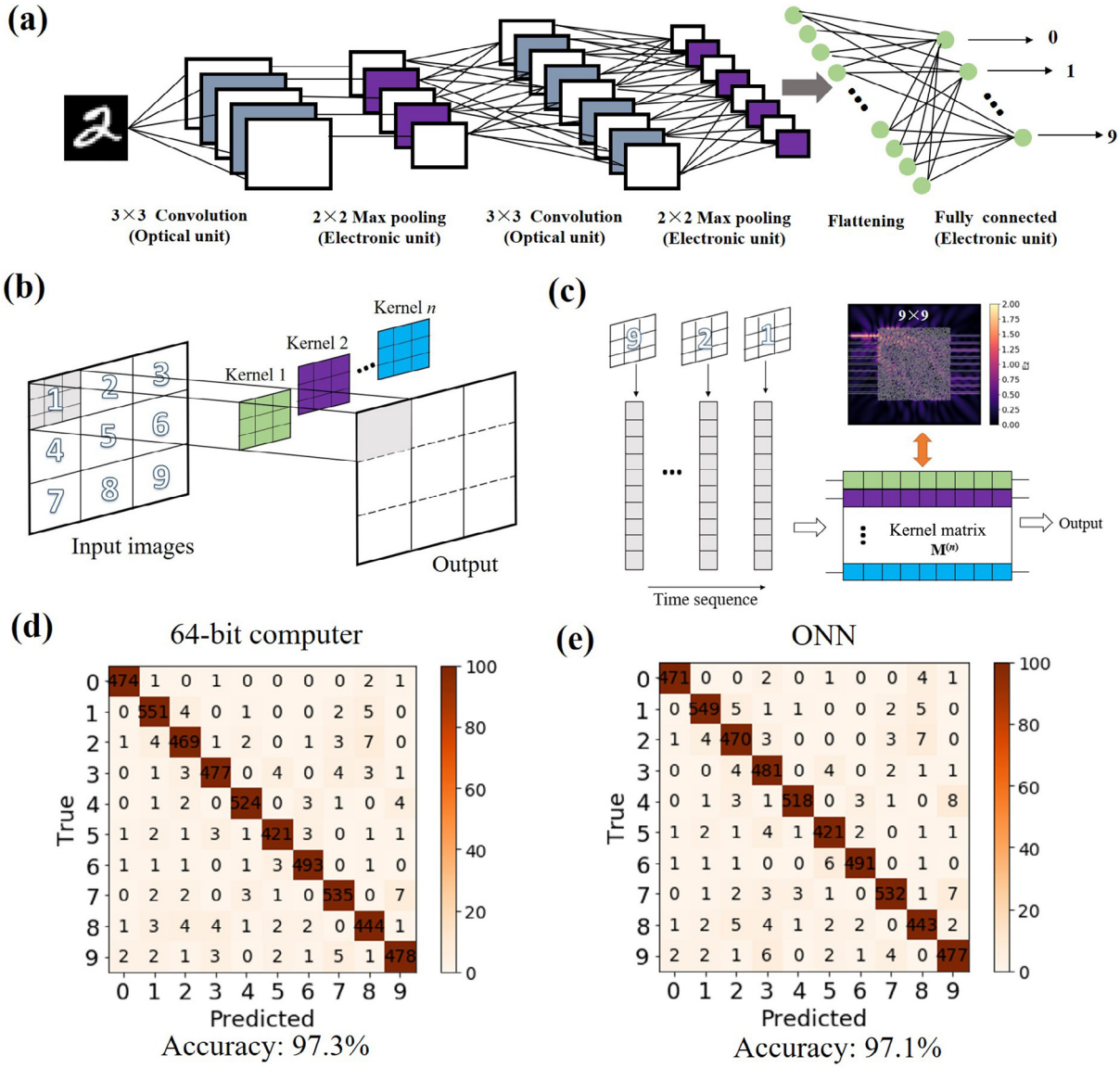


Fig. 4. (Color online) Optical convolutional neural network and MNIST classification confusion matrix. (a) Optical convolutional neural network consisting of two sets of convolutional layers and max pooling layers followed by a fully connected layer. (b) The convolutional operation that the images are convolved with a set of kernels. (c) The image patches and the set of kernels are vectorized. The one-dimensional image patches can be multiplied by the "Kernel Matrix". The inset is a 9×9 optical hardware implementation of the "Kernel Matrix". (d) Test accuracy and confusion matrix for a 64-bit computer. (e) Test accuracy and confusion matrix for the ONN architecture.

Table 1
The footprints and performances of unpatterned and patterned MMI architectures.

	The footprint of one layer	Total footprint	The number of parameters	Classification accuracy
$N \times N$ cascaded MIM couplers (unpatterned)	$W \times (4n_{ver}/N\lambda_0)W^2$	$W \times (4n_{ver}/\lambda_0)W^2$	N^2	98.6%
$N \times N$ OSU (patterned)	$W \times W$	$W \times W$	$(N-1) \times N$	97.1%

To provide enough trainable parameters in ONN, a number of MMI couplers ($m = N + 1$) are cascaded in sequence with N phase shifters inserted between two MMI couplers (shown in Fig. S5 online). The performance of such ONN architecture can be improved by increasing the number of MMI couplers and phase shifters, as it leads to more degrees of freedom and trainable parameters in the optical computing matrix. The whole transfer matrix M^n, which is implemented by combining alternating couplers and phase shifters, can be expressed as

$$M^n = (M_{coupler} \cdot M_{shift})^{m-1} \cdot M_{coupler}, \quad (3)$$

where m indicates the number of MMI couplers. $M_{coupler}$ is a fixed matrix whose entries represent the relative phases between the input and output waveguides. M_{shift}, however, is a diagonal matrix with trainable parameters. Each diagonal entry of M_{shift} represents the phase shift that is applied on the corresponding phase shifter. The expressions of $M_{coupler}$ and M_{shift} are described in the Supplementary materials.

We first compare the footprints of optical computing units based on unpatterned and patterned MMI couplers. The footprint of single-layer self-imaging MMI coupler is $W \times (4n_{ver}/N\lambda_0)W^2$, where n_{ver} is the slab effective index (vertical analysis) in the coupler region and

λ_0 is light wavelength in vacuum [38]. Here, we omit the sizes of waveguides and phase shifters. The total footprint of N-layer cascaded MMI couplers is $W \times (4n_{ver}/\lambda_0)W^2$, as shown in Table 1. For example, if the 4×4 optical computing unit contains four cascaded unpatterned MMI couplers and the gap between two waveguides is 1 μm, the total footprint is 4×156 μm^2. In contrast, the total footprint of the optical computing unit based on OSU is only 4×4 μm^2, which is 39 times smaller than unpatterned cascaded MMI architecture. The scattering loss of both OSU and conventional MMI is also compared (Figs. S6 and S7 online). $N \times N$ ONN architecture based on unpatterned cascaded MMI couplers has N^2 trainable parameters, while $N \times N$ ONN architecture based on incoherent OSU only has $(N-1) \times N$ trainable parameters because the summation of each column of stochastic matrix has to be one.

Next, we compare the image classification accuracies of ONN architectures based on unpatterned and patterned MMI couplers on the dataset MNIST. The ONN architecture based on unpatterned cascaded MMI couplers can be constructed using optical CNN algorithm shown in Fig. 4a. The optical computing unit consistes of five layers of MMI couplers ($m = 5$) with 9 phase shifters in each layer. The first and second convolutional layers consists of 3×3 kernels with 1 input channel and 4 output channels, as well as 3×3 kernels with 4 input channels and 8 output channels, respectively. The sets of kernels can be transformed to four individual 9×9 optical "Kernel Matrix". The classification accuracy of ONN architecture based on unpatterned MMI couplers is 98.6% (Fig. S5 online). Although the patterned MMI architecture reduces the footprint by a factor of tens, or even hundreds, it can still achieve classification accuracy of 97.1%.

4. Discussion and conclusions

In summary, we propose an integrated-nanophotonics ONN framework based on OSU with micrometer-level footprint for optically accelerated deep learning. OSU is single-layer MMI coupler whose coupler region is patterned through the inverse design method. The nanopatterning enables one to engineer the refractive index in space at a deep sub-wavelength scale to reduce the footprint of optical computing unit. We demonstrate that a 4×4 OSU with the footprint of 4×4 μm^2 can be optimized to implement stochastic matrix multiplication with less than 10^{-4} MSE. The incoherent OSU can be trained and optimized without the need for considering the phase coherence. By performing optical convolutional operations using "Kernel Matrix", the proposed ONN architecture can achieve an accuracy of 97.1% on the classic image classification dataset MNIST.

This work provides a systematic prototypical demonstration of a new ONN architecture based on OSU. Further consideration and adaption of such ONN architecture need to be taken into account. One natural step is towards realizing re-programmable ONN architecture by incorporating phase changing materials such as GST [39,40]. Moreover, the proposed ONN architectures still rely on computers at the training phase, which makes the training process inefficient. *In-situ* training on the integrated photonics platform needs to be further studied. Recently, a training shceme based on the adjoint variable method involving in situ intensity measurements has been proposed [41]. As a final remark, we envision that our results open a new direction in optical neural networks and this method can be easily applied to other neural network architectures to drastically decrease their footprints without compromising the efficiency or functionality.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2017YFA0205700) and the National Natural Science Foundation of China (61927820). Yurui Qu was supported by Zhejiang Lab's International Talent Fund for Young Professionals.

Author contributions

Yurui Qu initiated the original idea of optical scattering units; Yurui Qu, Huanzheng Zhu, Yichen Shen, Jin Zhang and Chenning Tao developed the theory and performed the computations; Yurui Qu wrote the manuscript with Huanzheng Zhu, Yichen Shen, Pintu Ghosh and Min Qiu; Min Qiu supervised the project and coordinated all the work; All the authors discussed the results and contributed to the final manuscript.

Appendix A. Supplementary materials

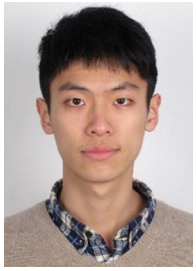
Supplementary materials to this article can be found online at https://doi.org/10.1016/j.scib.2020.03.042.

References

[1] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012:1097–105.
[2] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Process Mag 2012;29:82–97.
[3] Silver D, Huang A, Maddison CJ, et al. Mastering the game of go with deep neural networks and tree search. Nature 2016;529:484–9.
[4] Ma J, Sheridan RP, Liaw A, et al. Deep neural nets as a method for quantitative structure–activity relationships. J Chem Inf Model 2015;55:263–74.
[5] Gawehn E, Hiss JA, Schneider G. Deep learning in drug discovery. Mol Inf 2016;35:3–14.
[6] Leung MK, Xiong HY, Lee LJ, et al. Deep learning of the tissue-regulated splicing code. Bioinformatics 2014;30:i121–9.
[7] Xiong HY, Alipanahi B, Lee LJ, et al. The human splicing code reveals new insights into the genetic determinants of disease. Science 2015;347:1254806.
[8] Ramprasad R, Batra R, Pilania G, et al. Machine learning in materials informatics: recent applications and prospects. npj Comput Mater 2017;3:54.
[9] Baldi P, Sadowski P, Whiteson D. Searching for exotic particles in high-energy physics with deep learning. Nat Commun 2014;5:4308.
[10] Qu Y, Jing L, Shen Y, et al. Migrating knowledge between physical scenarios based on artificial neural networks. ACS Photon 2019;6:1168–74.
[11] Peurifoy J, Shen Y, Jing L, et al. Nanophotonic particle simulation and inverse design using artificial neural networks. Sci Adv 2018;4:eaar4206.
[12] Liu D, Tan Y, Khoram E, et al. Training deep neural networks for the inverse design of nanophotonic structures. ACS Photon 2018;5:1365–9.
[13] Malkiel I, Mrejen M, Nagler A, et al. Plasmonic nanostructure design and characterization via deep learning. Light Sci Appl 2018;7:1–8.
[14] Ma W, Cheng F, Liu Y. Deep-learning-enabled on-demand design of chiral metamaterials. ACS Nano 2018;12:6326–34.
[15] Liu Z, Zhu D, Rodrigues SP, et al. Generative model for the inverse design of metasurfaces. Nano Lett 2018;18:6570–6.
[16] Mead C. Neuromorphic electronic systems. Proc IEEE 1990;78:1629–36.
[17] Merolla PA, Arthur JV, Alvarez-Icaza R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. Science 2014;345:668–73.
[18] Wang Z, Joshi S, Savelév SE, et al. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. Nat Mater 2017;16:101–8.
[19] Biamonte J, Wittek P, Pancotti N, et al. Quantum machine learning. Nature 2017;549:195–202.
[20] Vivien L, Polzer A, Marris-Morini D, et al. Zero-bias 40Gbit/s germanium waveguide photodetector on silicon. Opt Express 2012;20:1096–101.
[21] Khoram E, Chen A, Liu D, et al. Nanophotonic media for artificial neural inference. Photon Res 2019;7:823–7.
[22] Lin X, Rivenson Y, Yardimci NT, et al. All-optical machine learning using diffractive deep neural networks. Science 2018;361:1004–8.
[23] Chang J, Sitzmann V, Dun X, et al. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. Sci Rep 2018;8:12324.
[24] Bueno J, Maktoobi S, Froehly L, et al. Reinforcement learning in a large-scale photonic recurrent neural network. Optica 2018;5:756–60.
[25] Shen Y, Harris NC, Skirlo S, et al. Deep learning with coherent nanophotonic circuits. Nat Photonics 2017;11:441.
[26] Rumelhart DE, Hinton GE, Williams RJ, et al. Learning representations by back-propagating errors. Nature 1986;323:533–6.

[27] Bao Q, Zhang H, Ni Z, et al. Monolayer graphene as a saturable absorber in a mode-locked laser. Nano Res 2011;4:297–307.
[28] Selden A. Pulse transmission through a saturable absorber. Br J Appl Phys 1967;18:743.
[29] Hughes TW, Minkov M, Williamson IA, et al. Adjoint method and inverse design for nonlinear nanophotonic devices. ACS Photon 2018;5:4781–7.
[30] Piggott AY, Lu J, Lagoudakis KG, et al. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. Nat Photonics 2015;9:374–7.
[31] Shen B, Wang P, Polson R, et al. An integrated-nanophotonics polarization beamsplitter with 2.4 × 2.4 μm^2 footprint. Nat Photonics 2015;9:378.
[32] Lawson CL, Hanson RJ. Solving least squares problems. Siam 1995. https://doi.org/10.1137/1.9781611971217.
[33] Connelly MJ. Semiconductor optical amplifiers. Springer Science & Business Media; 2007.
[34] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:2278–324.
[35] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
[36] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. p. 770–8.
[37] Bagherian H, Skirlo S, Shen Y, et al. On-chip optical convolutional neural networks. arXiv:1808.03303, 2018.
[38] Paiam MR, MacDonald RI. Design of phased-array wavelength division multiplexers using multimode interference couplers. Appl Opt 1997;36:5097–108.
[39] Ríos C, Stegmaier M, Hosseini P, et al. Integrated all-photonic non-volatile multi-level memory. Nat Photonics 2015;9:725.
[40] Qu Y, Li Q, Cai L, et al. Thermal camouflage based on the phase-changing material GST. Light Sci Appl 2018;7:26.
[41] Hughes TW, Minkov M, Shi Y, et al. Training of photonic neural networks through *in situ* backpropagation and gradient measurement. Optica 2018;5:864–71.

Yurui Qu received his B.Sc. degree in Optical Engineering from the Zhejiang University in 2014, and Ph.D. degree in Laboratory of Integrated Photonics and Nanophotonics from the same university in 2019. He also obtained honors degrees of Chu Kechen Honors College in Zhejiang University. His research interest focuses on nanophotonics, thermal emission and optical neural network.



Huanzheng Zhu received the B.Sc. degree from Zhejiang University in 2016. He also obtained honors degrees of Chu Kechen Honors College in Zhejiang University. He received the M.Sc. degree from Imperial College London, UK, in 2017. He is now a Ph.D. student in Laboratory of Integrated Photonics and Nanophotonics, Zhejiang University. His research interest focuses on thermal camouflage, high temperature thermal management and optical neural network.

Min Qiu received B.Sc. degree (1995) and Ph.D. degree (1999) in Physics, both from Zhejiang University, China, and Ph.D. degree (2001) in Electromagnetic Theory from Royal Institute of Technology (KTH), Sweden. He joined KTH as an assistant professor in 2001 and became Professor of Photonics in 2009. He moved to Zhejiang University, China, in 2011. Since 2018, he is the Chair Professor of Photonics and Vice President for Research, Westlake University, China. His research interests include nanofabrication technology, nanophotonics, and green photonics.