PHOTONICS Research

In situ optical backpropagation training of diffractive optical neural networks

Tiankuang Zhou,[1,2,3,†] Lu Fang,[2,3,†] Tao Yan,[1,2] Jiamin Wu,[1,2] Yipeng Li,[1,2] Jingtao Fan,[1,2] Huaqiang Wu,[4,5] Xing Lin,[1,2,4,7] and Qionghai Dai[1,2,6,8]

[1]Department of Automation, Tsinghua University, Beijing 100084, China
[2]Institute for Brain and Cognitive Science, Tsinghua University, Beijing 100084, China
[3]Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China
[4]Beijing Innovation Center for Future Chip, Tsinghua University, Beijing 100084, China
[5]Institute of Microelectronics, Tsinghua University, Beijing 100084, China
[6]Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China
[7]e-mail: lin-x@tsinghua.edu.cn
[8]e-mail: qhdai@tsinghua.edu.cn

Received 4 February 2020; revised 27 March 2020; accepted 27 March 2020; posted 30 March 2020 (Doc. ID 389553); published 28 May 2020

Training an artificial neural network with backpropagation algorithms to perform advanced machine learning tasks requires an extensive computational process. This paper proposes to implement the backpropagation algorithm optically for *in situ* training of both linear and nonlinear diffractive optical neural networks, which enables the acceleration of training speed and improvement in energy efficiency on core computing modules. We demonstrate that the gradient of a loss function with respect to the weights of diffractive layers can be accurately calculated by measuring the forward and backward propagated optical fields based on light reciprocity and phase conjunction principles. The diffractive modulation weights are updated by programming a high-speed spatial light modulator to minimize the error between prediction and target output and perform inference tasks at the speed of light. We numerically validate the effectiveness of our approach on simulated networks for various applications. The proposed *in situ* optical learning architecture achieves accuracy comparable to *in silico* training with an electronic computer on the tasks of object classification and matrix-vector multiplication, which further allows the diffractive optical neural network to adapt to system imperfections. Also, the self-adaptive property of our approach facilitates the novel application of the network for all-optical imaging through scattering media. The proposed approach paves the way for robust implementation of large-scale diffractive neural networks to perform distinctive tasks all-optically.

https://doi.org/10.1364/PRJ.389553

1. INTRODUCTION

Artificial neural networks (ANNs) have achieved significant success in performing various machine learning tasks [1], diverse from computer science applications (e.g., image classification [2], speech recognition [3], game playing [4]) to scientific research (e.g., medical diagnostics [5], intelligent imaging [6], behavioral neuroscience [7]). The explosive growth of machine learning is due primarily to the recent advancements in neural network architectures and hardware computing platforms, which enable us to train larger-scale and more complicated models [8,9]. A significant amount of effort has been spent on constructing different application-specific ANN architectures with semiconductor electronics [10,11], the performance of which is inherently limited by the fundamental tradeoff between energy efficiency and computing power in electronic computing [12]. As the scale of an electronic transistor approaches its physical limit, it is necessary to investigate and develop the next-generation computing modality during the post-Moore's law era [13,14]. Using photons instead of electrons as the information carrier to perform optical computing has potential properties to provide high energy efficiency, low crosstalk, light-speed processing, and massive parallelism. It has the potential to overcome problems inherent in electronics and is considered to be the disruptive technology for modern computing [15,16].

Recent works on the optical neural network (ONN) have made substantial progress in performing large-scale complex computing and high optical integrability by using state-of-the-art intelligent design approaches and fabrication techniques [17–19]. Various ONN architectures have been proposed, including the optical interference neural network [20,21], diffractive optical neural network [22,23], photonic reservoir

Corrected: 22 June 2020

computing [24,25], photonic spiking neural network [26–30], optical recurrent neural network [31,32], etc. Among them, constructing diffractive networks with diverse diffractive optical elements provides an extremely high degree of freedom to train the model and facilitates important applications in a wide range of fields, such as object classification [22,33–38], segmentation [23], pulse engineering [39], and depth sensing [40]. It has been demonstrated that the all-optical machine learning framework using diffractive ONN [22,23], i.e., diffractive deep neural networks (D^2NNs), can successfully classify the Modified National Institute of Standards and Technology (MNIST) handwritten digits dataset [41] with classification accuracy quite approaching electronic computing. These diffractive ONN models are physically fabricated with 3D printing or lithography for different inference tasks, where the network parameters are fixed once the network is created. The approach proposed in this paper adopts the cascading of spatial light modulators (SLMs) as the diffractive modulation layers, which can be programmed to train different network models for different tasks.

Proper training of the ANN with algorithms, such as error backpropagation [41], is the most critical aspect of making a reliable model and guarantees accurate network inference. Current ONN architectures are typically trained *in silico* on an electronic computer to obtain its designs for physical implementation. By modeling the light–matter interaction along with computer-aided intelligent design, the network parameters are learned, and the structure is determined to be deployed on photonic devices. However, due to the high computational complexity of the network training, such *in silico* training approaches fail to exploit the speed, efficiency, and massive parallel advantage of optical computing, which results in long training time and limited scalability. For example, it takes approximately 8 h to train a five-layer diffractive ONN configured with 0.2 million neurons as a digit classifier running on a high-end modern desktop computer [22]. Furthermore, different error sources in practical implementation will deviate the *in silico* trained model and degenerate inference accuracy. *In situ* training, in contrast, can overcome these limitations by physically implementing the training process directly inside the optical system. Recent works have demonstrated the success of *in situ* backpropagation for training the optical interference neural network [42] and physical recurrent neural network [43,44]. Nevertheless, these approaches either require strict lossless assumptions for calculating the time-reversed adjoint field or work only for a real-valued network by modeling the amplitude of the field, which cannot be applied to the diffractive ONN due to the complex-valued inherency and the presenting of diffractive loss. Another line of work based on the volumetric hologram [45,46] requires an undesirable light beam in the hologram recording and size-1 training batch, which dramatically restricts the network scalability and computational complexity. In this work, we propose an approach for *in situ* training of the large-scale diffractive ONN for complex inference tasks that can overcome the lossless assumption by modeling and measuring the forward and backward propagations of the diffractive optical field for its gradient calculation.

The proposed optical error backpropagation for *in situ* training of the diffractive ONN is based on light reciprocity and phase conjunction principles, which allow the optical backpropagation of the network residual errors by backward propagating the error optical field. We demonstrate that the gradient of the network at individual diffractive layers can be successively calculated highly parallel to measurements of the forward and backward propagated optical fields. We design a reprogrammable system with off-the-shelf photonic equipment by simulation for implementing the proposed *in situ* optical training, where phase-shifting digital holography is used for optical field measurement, and the error optical field is generated from a complex field generation module. Different from *in silico* training, by programming the multilayer SLMs for iteratively updating the network diffractive modulation coefficients during training, the proposed optical learning architecture can adapt to system imperfections, accelerate the training speed, and improve the training energy efficiency on core computing modules. Also, diffractive ONNs implemented with multilayer SLMs can be easily reconfigured to perform different inference tasks at the speed of light. The numerical simulations on the proposed reconfigurable diffractive ONN system demonstrate the high accuracy of our *in situ* optical training method for different applications, including light-speed object classification, optical matrix-vector multiplier, and all-optical imaging through scattering media.

2. OPTICAL ERROR BACKPROPAGATION

The diffractive ONN framework proposed in Ref. [22] comprises the cascading of multiple diffractive modulation layers, as shown in Fig. 1(a), where an artificial neuron on each layer modulates the amplitude and phase of its input optical field and generates a secondary wave through optical diffraction for connecting to other neurons of the following layers. The modulation coefficients of neurons are iteratively updated during the training that tunes the network towards a specific task. Despite that the network configurations in Ref. [22] for proof-of-concept experiments adopt only the linear diffractive optical neuron for processing a complex optical field, the detector on the output plane of the network measures its intensity (square of the amplitude) distribution, which performs the activation function of the diffractive computing result. Also, the optical nonlinearity can be incorporated to achieve the activation function for neurons at individual layers [23] that can accomplish more complicated inference tasks. Instead of training in an electronic computer and fabricating with 3D printing, we propose to implement the phase-only diffractive modulation layers with the phase SLM, e.g., liquid crystal on silicon (LCOS), which can be programmed to update network weights and enables *in situ* training of both linear [22] and nonlinear [23] diffractive ONNs.

In this section, we derive the optical error backpropagation model for linear diffractive ONN, and its extension to the nonlinear diffractive ONN can be found in Section 6 of Appendix A. To implement the error backpropagation optically, we build the *in situ* training optimization model and demonstrate that the gradient of the loss function of diffractive ONNs can be calculated by measuring the forward propagated input optical field and backward propagated error optical field. The forward propagation model of the diffractive ONN in Fig. 1(a) can be established based on the Rayleigh–Sommerfeld

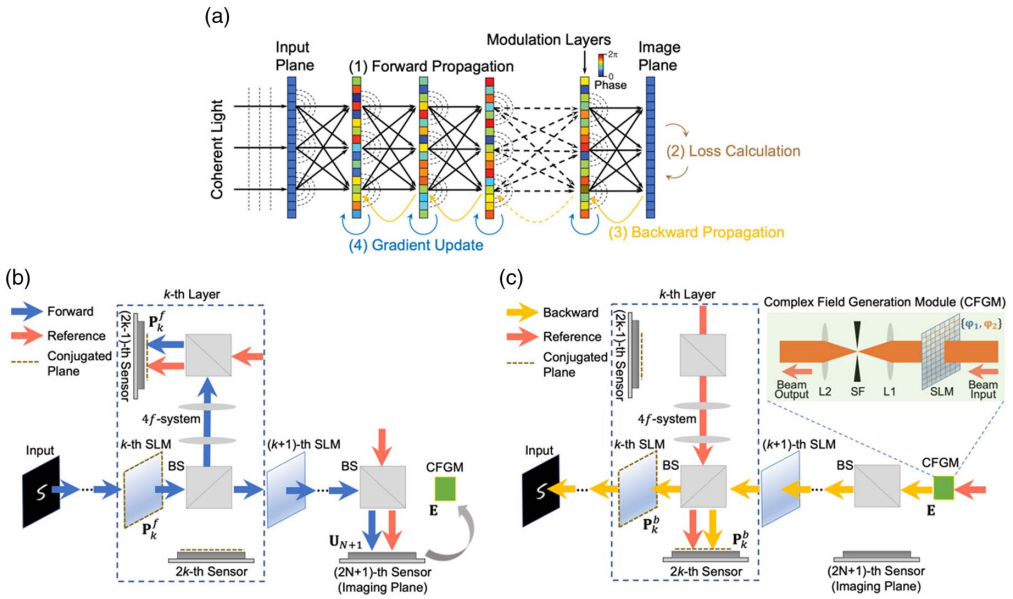

Fig. 1. Optical training of diffractive ONN. (a) The diffractive ONN architecture is physically implemented by cascading spatial light modulators (SLMs), which can be programmed for tuning diffractive coefficients of the network towards a specific task. The programmable capability makes it possible for *in situ* optical training of diffractive ONNs with error backpropagation algorithms. Each iteration of the training for updating the phase modulation coefficients of diffractive layers includes four steps: forward propagation, error calculation, backward propagation, and gradient update. (b) The forward propagated optical field is modulated by the phase coefficients of multilayer SLMs and measured by the image sensors with phase-shifted reference beams at the output image plane as well as at the individual layers. The image sensor is set to be conjugated to the diffractive layer relayed by a 1:1 beam splitter (BS) and a $4f$ system. (c) The backward propagated optical field is formed by propagating the error optical field from the output image plane back to the input plane with the modulation of multilayer SLMs. The error optical field is generated from the complex field generation module (CFGM) by calculating the residual errors between the network output optical field and the ground truth label. With the measured forward and backward propagated optical fields, the gradients of the diffractive layers are calculated, and the modulation coefficients of SLMs are successively updated from the last to first layer.

diffraction principle [22], where the complex transform of the optical field between successive diffractive modulation layers can be formulated as

$$\mathbf{U}_k = \mathbf{M}_k \mathbf{W}_k \mathbf{U}_{k-1}, \tag{1}$$

where $\mathbf{U}_k$ represents the vectorized output optical field of a k-th layer of the network; $\mathbf{W}_k$ is the diffractive weight matrix with the forward light propagation from the $(k-1)$-th to the k-th layer; and $\mathbf{M}_k = \mathrm{diag}(e^{j\boldsymbol{\phi}_k})$ represents the diagonalization of a vectorized diffractive modulation at the k-th layer with a phase coefficient of $\boldsymbol{\phi}_k$, with j being the imaginary unit, i.e., $j^2 = -1$. Under coherent illumination, the optical field of an object at the input layer $\mathbf{U}_0$ propagates through multiple diffractive optical layers and generates the optical field of the output layer of the network $\mathbf{U}_{N+1}$ at the imaging plane:

$$\mathbf{U}_{N+1} = \mathbf{W}_{N+1}\left(\prod_{k=N}^{1} \mathbf{M}_k \mathbf{W}_k\right)\mathbf{U}_0, \tag{2}$$

where we assume that the diffractive ONN is composed of N layers excluding the input and output layers. The detector at the imaging plane measures the intensity distribution of the resulting optical field and obtains the network inference result:

$$\mathbf{O} = |\mathbf{U}_{N+1}|^2 = \left|\mathbf{W}_{N+1}\left(\prod_{k=N}^{1} \mathbf{M}_k \mathbf{W}_k\right)\mathbf{U}_0\right|^2. \tag{3}$$

The *in situ* optical training of the network involves successively updating phase coefficients of diffractive layers with the calculated gradient to minimize a loss function. Let $L(\mathbf{O}, \mathbf{T})$ represent the loss function of diffractive ONN that measures the differentials between network outputs $\mathbf{O}$ and ground truth labels $\mathbf{T}$. The gradient of the defined loss function with respect to the phase modulation coefficient of a k-th layer $\boldsymbol{\phi}_k$ can be derived as

$$\begin{aligned}\frac{\partial L}{\partial \boldsymbol{\phi}_k} &= \frac{\partial L}{\partial \mathbf{U}_{N+1}}\frac{\partial \mathbf{U}_{N+1}}{\partial \boldsymbol{\phi}_k} + \frac{\partial L}{\partial \mathbf{U}_{N+1}^*}\frac{\partial \mathbf{U}_{N+1}^*}{\partial \boldsymbol{\phi}_k} \\ &= 2\mathrm{Re}\left\{\left(\frac{\partial L}{\partial \mathbf{O}} \odot \mathbf{U}_{N+1}^*\right)^T \frac{\partial \mathbf{U}_{N+1}}{\partial \boldsymbol{\phi}_k}\right\},\end{aligned} \tag{4}$$

where $\odot$ represents element-wise multiplication, and $*$ means conjugation of the optical field. The gradient of $\mathbf{U}_{N+1}$ with respect to the $\boldsymbol{\phi}_k$ can be derived as

$$\begin{aligned}\partial \mathbf{U}_{N+1}/\partial \boldsymbol{\phi}_k &= j\mathbf{W}_{N+1}\left(\prod_{i=N}^{k+1} \mathbf{M}_i \mathbf{W}_i\right)\mathrm{diag}\left(\left(\prod_{i=k}^{1} \mathbf{M}_i \mathbf{W}_i\right)\mathbf{U}_0\right) \\ &= \left(j \cdot \mathrm{diag}\left(\left(\prod_{i=k}^{1} \mathbf{M}_i \mathbf{W}_i\right)\mathbf{U}_0\right)^T \right. \\ &\quad \left. \times \left(\left(\prod_{i=k+1}^{N} \mathbf{W}_i^T \mathbf{M}_i\right)\mathbf{W}_{N+1}^T\right)\right)^T,\end{aligned} \tag{5}$$

Corrected: 22 June 2020

where diag represents the diagonalization operation. Substituting Eq. (5) into Eq. (4) and letting $\mathbf{E} = \frac{\partial L}{\partial \mathbf{O}} \odot \mathbf{U}_{N+1}^{*}$ denote the error optical field, we have

$$\begin{aligned}\frac{\partial \mathrm{L}}{\partial \boldsymbol{\phi}_k} &= 2\mathrm{Re}\{\mathbf{E}^T \partial \mathbf{U}_{N+1}/\partial \boldsymbol{\phi}_k\} \\ &= 2\mathrm{Re}\left\{j\left(\left(\prod_{i=k}^{1}\mathbf{M}_i\mathbf{W}_i\right)\mathbf{U}_0\right)\right. \\ &\quad \left.\odot\left(\left(\prod_{i=k+1}^{N}\mathbf{W}_i^T\mathbf{M}_i\right)\mathbf{W}_{N+1}^T\mathbf{E}\right)\right\}^T \\ &= 2\mathrm{Re}\{(j\mathbf{P}_k^f \odot \mathbf{P}_k^b)\}^T, \end{aligned} \tag{6}$$

where

$$\begin{cases}\mathbf{P}_k^f = \left(\prod_{i=k}^{1}\mathbf{M}_i\mathbf{W}_i\right)\mathbf{U}_0, \\ \mathbf{P}_k^b = \left(\prod_{i=k+1}^{N}\mathbf{W}_i^T\mathbf{M}_i\right)\mathbf{W}_{N+1}^T E. \end{cases} \tag{7}$$

The underlying physical meaning of $\mathbf{P}_k^f$ is the output optical field of a k-th layer by propagating the optical field of an object from the input layer to layer k. According to the reciprocity principle of light propagation [44], $\mathbf{W}_k^T$ represents the diffractive weight matrix between the $(k-1)$-th layer and the k-th layer with backward light propagation in the opposite direction of $\mathbf{W}_k$. Therefore, $\mathbf{P}_k^b$ is the optical field corresponding to the backward propagation of the error optical field $\mathbf{E}$ from the output plane of the network to the k-th layer. The error optical field $\mathbf{E}$ comprises the multiplication of two terms, i.e., $\partial L/\partial \mathbf{O}$ and $\mathbf{U}_{N+1}^{*}$. $\partial L/\partial \mathbf{O}$ represents the gradient of the loss function L with respect to output intensity $\mathbf{O}$, where we use the mean squared error as the loss function, i.e., $L(\mathbf{O},\mathbf{T}) = \|\mathbf{O}-\mathbf{T}\|_2^2$, where $\|\cdot\|_2$ is the L^2 norm, with which $\partial L/\partial \mathbf{O} = 2(\mathbf{O}-\mathbf{T})$. $\mathbf{U}_{N+1}^{*}$ represents the conjugation of the network output optical field that ensures the success of backpropagating the residual error optically according to the phase conjunction technique [42]. By measuring the forward and backward propagated optical fields $\mathbf{P}_k^f$ and $\mathbf{P}_k^b$ at the k-th layer, the gradient of the loss function with respect to the phase coefficient of the k-th layer can be efficiently calculated based on Eq. (7).

As shown in Fig. 1(a), the proposed *in situ* optical training procedure for measuring the gradient of a diffractive layer in the network can be summarized as follows: (1) measuring the forward optical field at the network output layer $\mathbf{U}_{N+1}$ and the k-th layer $\mathbf{P}_k^f$ by forward propagating the input optical field through diffractive modulation layers to the output image plane; (2) calculating the error optical field $\mathbf{E}$ with the network output optical field and the ground truth label; (3) measuring the backward optical field at the k-th layer $\mathbf{P}_k^b$ by backward propagating the error optical field $\mathbf{E}$ from the output imaging plane through diffractive modulation layers to the input object plane; (4) calculating the gradient of the current layer with the measured forward and backward optical fields. During each iteration of the training, given an input–output example, the network phase modulation coefficients are updated with the calculated gradient by the chain rule that backward adjusts one layer at a time from the last layer to the first layer. The pseudo-code of the proposed optical error backpropagation algorithm can be found in Section 1 of Appendix A. In the next section, we demonstrate the system design to implement the abovementioned *in situ* optical training procedure for diffractive ONN with off-the-shelf photonic equipment.

3. EXPERIMENTAL SYSTEM DESIGN AND CONFIGURATION

We propose an optical system design composed of off-the-shelf photonic equipment to implement the optical error backpropagation for *in situ* training of diffractive ONN, as shown in Figs. 1(b) and 1(c). The proposed system configuration adopts multilayer programmable SLMs as the diffractive modulator for high-speed updating of network coefficients during the training towards performing distinct inference tasks. To measure the forward and backward propagated optical field at the individual modulation layer, we adopt a pair of image sensors and conjugate their sensor planes to the SLM modulation plane in the forward and backward optical paths, respectively. Conjugation of the planes is achieved by using a beam splitter (BS) and a $4f$ system. Different transmission rates of BS can be applied by multiplying constant coefficients during the gradient update. We use the 1:1 BS in this implementation for the sake of simplicity. Since the image sensor can measure only the intensity distribution of an optical field, we adopt the phase-shifting digital holography approach [47] to measure the interference of the forward and backward optical fields with the phase-shifted reference beam to obtain the network optical field effectively. The output optical field of the network is simultaneously measured for calculating the error optical field given the inputs and ground truth labels, which is used to propagate the residual errors backward and minimize a loss function by iteratively updating the phase modulation coefficients of SLMs. We show that the error optical field can be generated by an interleaving of two complementary phase modulation patterns with low-pass filtering, which can be physically implemented with an SLM and a $4f$ system. In the following, we detail the measurement and generation of the complex optical field for the network gradient calculation.

A. Measuring the Network Optical Field

We adopted four-step phase-shifting digital holography [47] to measure the optical field at individual layers. Assume $\mathbf{P}_k = \mathbf{A}_k e^{j\boldsymbol{\theta}_k}, k = 1,\ldots,N$ represents the forward propagated optical field $\mathbf{P}_k^f$ or the backward propagated optical field $\mathbf{P}_k^b$ at the k-th layer, where $\mathbf{A}_k$ refers to its amplitude, and $\boldsymbol{\theta}_k$ refers to its phase. The network optical field was interfered with a four-step phase-shifted reference beam $U_R = A_R e^{j\theta_R}$, i.e., the phase value of the wavefront $\theta_R = 0, \pi/2, \pi, 3\pi/2$, where A_R is the amplitude with a constant value. The corresponding intensity distributions of the interference results $\mathbf{I}_0, \mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3 = |\mathbf{U}_R + \mathbf{P}_k|^2$ were sequentially measured with an image sensor, with which the amplitude and phase of the optical field at the k-th layer can be accurately calculated as

$$\begin{cases}\boldsymbol{\theta}_k = \arctan((\mathbf{I}_3 - \mathbf{I}_1)/(\mathbf{I}_0 - \mathbf{I}_2)), \\ \mathbf{A}_k = \alpha((\mathbf{I}_1 - \mathbf{I}_0)/2(\sin\boldsymbol{\theta}_k - \cos\boldsymbol{\theta}_k)), \end{cases} \tag{8}$$

where $\alpha = 1/A_R$ is a constant value.

B. Generating the Error Optical Field

The error optical field was generated with a complex field generation module (CFGM) that acts as the source of a backward propagated optical field. In this paper, we implemented it with a phase-only field generator SLM and a low-pass filtering $4f$ system [48], as shown in Fig. 1(c). The field generator SLM was set to be conjugated to the imaging plane in order to backpropagate the residual error between network output fields and ground truth labels for updating the network weights. To generate the complex optical field with a phase-only SLM, the error optical field, i.e., $\mathbf{E} = \mathbf{A}_\xi e^{j\boldsymbol{\theta}_\xi}$, was decomposed into two optical fields with a constant amplitude, i.e., $\mathbf{E} = \beta(e^{j\boldsymbol{\varphi}_1} + e^{j\boldsymbol{\varphi}_2})$, so that two phase patterns $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ can be interleaved multiplexed on the SLM. Let $\mathbf{M}_1$ and $\mathbf{M}_2$ represent a pair of a complementary pixel-wise binary checkerboard pattern, i.e., $\mathbf{M}_1 + \mathbf{M}_2 = 1$, and then the multiplexed phase pattern for the SLM can be formulated as $\boldsymbol{\varphi} = \mathbf{M}_1(\boldsymbol{\varphi}_1)\uparrow_2 + \mathbf{M}_2(\boldsymbol{\varphi}_2)\uparrow_2$, where $\uparrow_2$ represents the 2× nearest-neighbor spatial upsampling of two phase patterns. Since the zero order of the spectrum of the optical field generated by the multiplexed pattern is the desired error optical field, a spatial low-pass filter was placed at the Fourier plane of a $4f$ system to filter out the undesired high-order diffraction spectra. We further performed the nearest-neighbor upsampling (2× in this paper) on the multiplexed pattern to achieve a broader spectrum separation and reduce the spectral aliasing in the generated error optical field.

With the proposed system design to measure the forward propagated input optical field and backward propagated error optical field for *in situ* training of the network, the gradient at each diffractive layer is successively calculated, and the phase coefficients of SLMs are iteratively updated by adding the phase increment $\Delta\boldsymbol{\phi}_k = -\eta(\partial L/\partial\boldsymbol{\phi}_k)$, where η is a constant network parameter determining the learning rate of the training. Since there are constant scale factors during the measurement of the network optical field and the generation of the error optical field, i.e., α and β, respectively, we tune the value of η during the training so that the phase coefficients are updated at an appropriate step size. Different from the brute force *in situ* training approach [20] that computes the gradient of ONNs by sequentially perturbing the coefficient of individual neurons, the optical error backpropagation approach proposed in this paper allows for tuning the network coefficients in parallel. This enables us to effectively train the large-scale network coefficients and significantly enhances the scalability of the diffractive ONN. Furthermore, since our framework directly measures the optical field of a network at individual layers, it avoids performing the interference between the forward and backward optical fields and eliminates the assumption of losslessness used in Ref. [42]. This is important for the optical training of diffractive ONNs because of the inherent diffraction loss on the network periphery caused by freespace light propagation.

4. NUMERICAL SIMULATIONS AND APPLICATIONS

In this section, we numerically validate the effectiveness of the proposed optical error backpropagation and demonstrate the success of *in situ* optical training of simulated diffractive ONNs for different applications, including light-speed object classification, optical matrix-vector multiplication, and all-optical imaging through scattering media.

A. Light-Speed Object Classification

Object classification is a critical task in computer vision and is also one of the most successful applications of ANNs. The conventional object classification paradigm typically requires to capture and store large-scale scene information as an image by using an optoelectronic sensor and compute with artificial intelligence algorithms in an electronic computer. Such a storage and computing separation paradigm places significant limitation on the processing speed. Our all-optical machine learning framework based on diffractive ONNs performs the light-speed computing directly on the object optical wavefront so that the detectors need to measure only the classification result, e.g., 10 measurements for 10 classes on the MNIST dataset, as shown in Fig. 2(a). This dramatically reduces the number of measurements and enhances the classification response speed. The proposed *in situ* optical training in this paper allows for the robust implementation of diffractive ONNs and enables the reconfigurable capability by using programmable diffractive layers.

The effectiveness of the proposed approach was validated by comparing the performance between *in situ* optical training and *in silico* electronic training [22] of a 10-layer diffractive ONN (see Section 2 and Fig. 5 of Appendix A for classification performance w.r.t. the layer number) for classifying the MNIST dataset under the same network settings. The object can be encoded into different properties of the optical field, e.g., amplitude, phase, and wavelength. In this work, the diffractive ONNs were set to work under a coherent illumination at the SLM working wavelength of 698 nm (e.g., with CNI MRL-FN-698 laser source), and the input objects were encoded by using the amplitude of the optical field. Considering the practical implementation of the *in situ* optical training system, the pixel pitch of the SLM was set to 8 μm, e.g., with the Holoeye PLUTO-2 phase-only SLM. Since the CFGM module requires 4× nearest-neighbor spatial upsampling to reduce the spectrum aliasing, the neuron size was set to 32 μm by binning 4×4 SLM pixels. For the classification diffractive ONN, the network was configured with 10 diffractive modulation layers by packing 150×150 neurons on each layer, covering an area of 4.8 mm × 4.8 mm per layer on the SLM. The successive layer distance was optimized with grid search within the range of 0–50 cm and set to 20 cm. The MNIST handwritten digit $(0, 1, \ldots, 9)$ dataset has 55,000 training images, 5000 validation images, and 10,000 testing images with each image size of 28×28 pixels. The image size was upsampled four times with boundary padding to match the network size. To satisfy the boundary condition of freespace propagation numerically implemented with the angular spectrum method [22], the network periphery was further padded with a length of 0.8 mm. The classification criterion is to find the detector with the maximum optical signal among 10 detector regions (corresponding to 10 digits), where each detector width is set to 1.6 mm. This was also used as a loss function during network training. We used the stochastic gradient descent algorithm (Adam optimizer) [22] for the *in silico* electronic training;

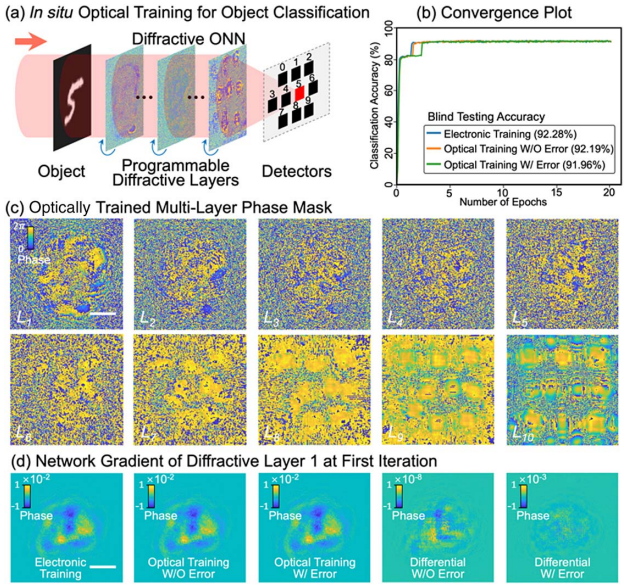

Fig. 2. *In situ* optical training of the diffractive ONN for object classification on the MNIST dataset. (a) By *in situ* dynamically adjusting the network coefficients with programmable diffractive layers, the diffractive ONN is optically trained with the MNIST dataset to perform object classification of the handwritten digits. (b) The numerical simulations on 10-layer diffractive ONN show the blind testing classification accuracy of 92.19% and 91.96% for the proposed *in situ* optical training approach without and with the CFGM error, respectively, which achieves a performance comparable to the electronic training approach (classification accuracy of 92.28%). (c) After the optical training (with CFGM error), phase modulation patterns on 10 different diffractive layers ($L_1, L_2, \ldots, L_{10}$) are shown, which are fixed during the inference for performing the classification at the speed of light. (d) The visualization of the network gradient reveals that the proposed optical error backpropagation accurately obtains the network gradient with accuracy comparable to the electronic training by calculating the differential between the electronic and optical gradients of the diffractive layer one at first iteration. Scale bar: 1 mm.

the *in situ* optical training was numerically implemented with the proposed optical error backpropagation algorithm. Both *in silico* electronic training and *in situ* optical training were implemented on the platform of Python version 3.6.7 with TensorFlow framework version 1.12.0 (Google Inc.) using a desktop computer (Nvidia TITAN XP Graphical Processing Unit, GPU, and Intel Xeon Gold 6126 CPU at 2.60 GHz with 64 cores, 128 GB of RAM, running with a Microsoft Windows 10 operating system).

The convergence plots of *in silico* electronic and *in situ* optical training of 10-layer diffractive ONN by using the MNIST training dataset and evaluating on the validation dataset are shown in Fig. 2(b). The optical training was assessed with and without incorporating the errors from the CFGM, which was caused by the spectrum aliasing in generating the error optical field. With the learning rate setting of 0.01 and batch size setting of 10 for both electronic and optical training, the electronic training converged after 15 epochs of iteration, and the optical training converged after 16 epochs and 24 epochs without and with the CFGM errors, respectively. For the *in situ* optical training, the training instances in each batch are sequentially fed into the system during implementation. Also, the phase value of each neuron was wrapped to the range between 0 and 2π in order to meet the modulation range of the SLM. The phase modulation layers of optical training with CFGM errors were converged to the patterns, as shown in Fig. 2(c), and the trained model was blind tested with the MNIST testing dataset. The numerical simulation results show that the proposed *in situ* optical training method achieves comparable performance with respect to the electronic training, i.e., blind testing accuracy of 92.19% and 91.96% for the optical training without and with the CFGM errors, respectively, compared with the blind testing accuracy of 92.28% for electronic training. To further demonstrate the accuracy of the network gradient during the *in situ* optical training, we compared and calculated the differential between the electronic and optical gradient of diffractive layer one at first iteration, as shown in Fig. 2(d). The differentials without and with CFGM error are six and approximately two orders of magnitude lower than the calculated gradient, respectively, which indicates that the proposed optical error propagation can accurately calculate the network gradient under the lossy condition. Despite that the CFGM error incorporates the stochastic noise to the gradient update that slows down the convergence speed and slightly decreases the classification accuracy, the proposed *in situ* optical training approach can adapt to its error as well as other system imperfections and successfully performs the object classification on the MNIST dataset.

B. Optical Matrix-Vector Multiplication

Matrix-vector multiplication is one of the fundamental operations in artificial neural networks, which is the most time- and energy-consuming component implemented with electronic computing platforms due to the use of a limited clock rate and large numbers of data movement. The intrinsic parallelism of optical computing allows large-scale matrix multiplication to be implemented at the speed of light with high energy efficiency without the use of the system clock or data movement. Previous works [20,49] on optical matrix-vector multiplication have limited degrees of freedom for constructing the matrix operator and required solving an optimization problem in electronic computers to derive the design before deploying with photonic equipment. Our *in situ* optical training approach eliminates the requirement for electronic optimization and has a much higher degree of freedom to achieve the desired matrix operator, which not only improves the optimization efficiency but also enhances the scalability of the operation.

The computational architecture of *in situ* optical training of diffractive ONN for matrix-vector multiplication is shown in Fig. 3(a). The elements of the input vector $\mathbf{X} \in \mathbf{R}^{M_1}$ are encoded into different regions at the input plane, the values of which are represented by the amplitude of input optical fields. Correspondingly, the elements of the output vector $\mathbf{Y} \in \mathbf{R}^{M_2}$ are encoded into different regions at the output plane of the network and are measured as the detector intensity. Given an arbitrary matrix operator $\mathbf{H} \in \mathbf{R}^{M_2 \times M_1}$, we demonstrate that the reconfigurable diffractive ONN can be optically trained as the desired matrix-vector multiplier, i.e., $\mathbf{Y} = \mathbf{HX}$, with high accuracy. For demonstration, we used the proposed optical error propagation framework to train a four-layer

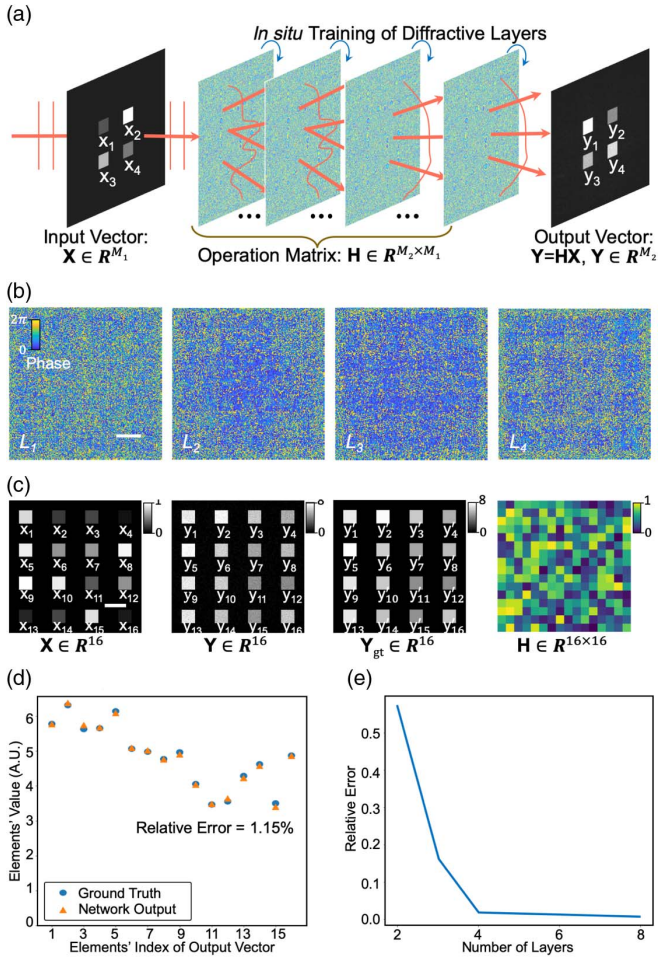

Fig. 3. *In situ* optical training of the diffractive ONN as an optical matrix-vector multiplier. (a) By encoding the input and output vectors to the input and output planes of the network, respectively, the diffractive ONN can be optically trained as a matrix-vector multiplier to perform an arbitrary matrix operation. (b) A four-layer diffractive ONN is trained as a 16×16 matrix operator [shown in the last column of (c)], the phase modulation patterns (L_1, L_2, L_3, L_4) of which are shown and can be reconfigured to achieve different matrices by programming the SLM modulations. (c) With an exemplar input vector on the input plane of the trained network (first column), the network outputs the matrix-vector multiplication result (second column), which achieves comparable results with respect to the ground truth (third column). (d) The relative error between the network output vector and ground truth vector is 1.15%, showing the high accuracy of our optical matrix-vector architecture. (e) By increasing the number of modulation layers, the relative error is decreased, and matrix multiplier accuracy can be further improved. Scale bar: 1 mm.

diffractive ONN to perform a 16×16 matrix operation, i.e., $M_1 = M_2 = 16$; the element values of a target matrix are shown in the last column of Fig. 3(c). For the *in situ* optical training of the matrix-vector multiplier network, we changed the neuron number on each layer to 200×200 (covering an area of 6.4 mm × 6.4 mm) and the detector width to 0.64 mm, while keeping other network settings the same as that in light-speed object classification. We used the target matrix operator $\mathbf{H}$ to generate input–output vector pairs as the training, validation, and testing datasets. The trained model performance with respect to the size of the training dataset can be found in Fig. 6 of Appendix A. To guarantee the model accuracy and minimize the computational resource of data generation, the training dataset was generated with 500 input–output vector pairs. Increasing the training set size can further improve model accuracy. Both validation and testing datasets were generated with 1000 input–output vector pairs for sufficiently evaluating the generalization of the network. The input vectors of the dataset were randomly sampled with a uniform distribution between zero and one. The optical training process converged after 30,000 iterations (60 epochs) under CFGM errors, and the trained diffractive modulation patterns are shown in Fig. 3(b). Figure 3(c) shows that the trained matrix-vector multiplier diffractive ONN successfully generated the output vector (third column) with accuracy comparable to the ground truth vector (second column) by taking the exemplar vector from the testing dataset as an input (first column). Different element values of the output vector were obtained by averaging the intensity of different detector regions [Fig. 3(d)]. The relative error, calculated as $|\mathbf{Y} - \mathbf{Y}_{gt}|_2 / |\mathbf{Y}_{gt}|_2$,

Corrected: 22 June 2020

was used to quantitatively evaluate the accuracy of the output vector $\mathbf{Y}$ with respect to the ground truth vector $\mathbf{Y}_{gt}$, which was found to be 1.15% on the exemplar vector and 1.37% over the testing vectors. We further demonstrate in Fig. 3(e) that increasing the number of diffractive modulation layers, e.g., from two layers to four layers, allows training the network with a higher degree of freedom, which significantly reduces the relative error of the output vector and improves the accuracy of the *in situ* optically trained diffractive ONN for performing the optical matrix-vector multiplication.

C. All-Optical Imaging Through Scattering Media

Imaging through scattering media has been one of the difficult challenges with essential applications in many fields [50–52]. Previous approaches typically performed object reconstruction in an electronic computer with the captured speckle intensity measurements, which use only limited input information due to missing the optical phase and hinder instantaneous observation of dynamic objects behind the scattering media due to limited electronic processing speed. In this work, we applied the proposed architecture for all-optical imaging through scattering media so that the detector can directly measure the de-scattered results. The *in situ* optical training of diffractive ONN provides an extremely high degree of freedom to control the distorted wavefront and reconstruct the object optical field with high scalability. Since the architecture performs the optical computing directly on the distorted optical field, the input of the diffractive network contains both amplitude and phase information, which facilitates high-quality reconstruction. Also, *in situ* training with optical error propagation characterizes the property of the scattering media and allows the network to be adapted to the medium perturbation with high efficiency.

The numerical simulation results of using diffractive ONN for imaging through translucent scattering media are demonstrated in Fig. 4. The scattering media were emulated with random phase patterns with a uniform distribution sampling of the phase value ranging from 0 to 2π, which strongly distorted the object wavefront and generated speckle patterns on the detector under conventional imaging [Fig. 4(a), top]. To demonstrate the application of our architecture for all-optical imaging through scattering media that allows to provide instantaneous reconstruction of objects [Fig. 4(a), bottom], we *in situ* optically trained the diffractive ONN with errors of CFGM by using the neuron number of 200 × 200 on each layer and keeping other network settings unchanged. We trained and tested the network with both MNIST and Fashion-MNIST datasets under a fixed distance of 90 cm between the object and scattering media. The MNIST dataset was trained by using a two-layer diffractive ONN (converged after five epochs of iteration), the performance of which calculated as the peak signal-to-noise ratio (PSNR) on the testing dataset with respect to the layer distance is shown in Fig. 4(b). Increasing the layer distance

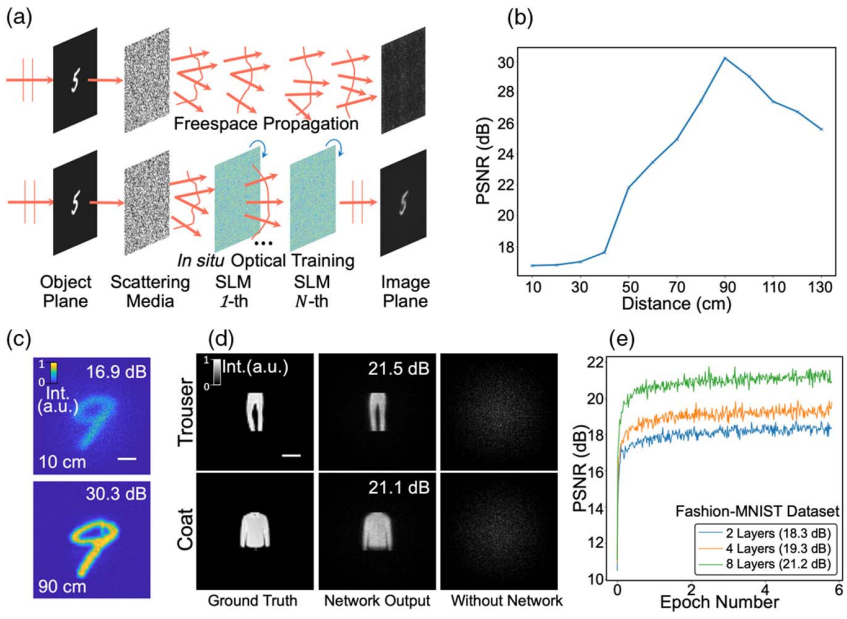

Fig. 4. Instantaneous imaging through scattering media with *in situ* optical training of the diffractive ONN. (a) The wavefront of the object is distorted by the scattering media and generates the speckle pattern on the detector under freespace propagation (top row). The diffractive ONN is *in situ* optically trained to take the distorted optical field as an input and perform the instantaneous de-scattering for object reconstruction (bottom row). (b) The MNIST dataset is used to train a two-layer diffractive ONN. The performance of the trained model is evaluated by calculating the peak signal-to-noise ratio (PSNR) of the de-scattering results on the testing dataset, which increases with the reasonably increasing layer distance. (c) The network de-scattering result on the handwritten digit "9" from the MNIST testing dataset shows PNSRs of 16.9 dB and 30.3 dB at layer distances of 10 cm and 90 cm, respectively. (d) An eight-layer diffractive ONN trained with the Fashion-MNIST dataset successfully reconstructs the objects of "Trouser" and "Coat" (images of the testing dataset) from their distorted optical wavefront. (e) Convergence plots of the two-, four-, and eight-layer diffractive ONN trained with the Fashion-MNIST dataset, which achieves PSNRs of 18.3 dB, 19.3 dB, and 21.2 dB on the testing dataset, respectively. Scale bar: 1 mm.

Corrected: 22 June 2020

increases the number of neuron connections between the successive diffractive layers, which leads to the improvement of the network in performing the task of imaging through scattering media. However, the network performance starts to decrease after the layer distance of 90 cm, since significantly increasing the layer distance decreases the effective diffractive modulation resolution at the same time. The network de-scattering results of the digit "9" of the testing dataset at distances of 10 cm and 90 cm are shown in Fig. 4(c), the PSNRs of which are 16.9 dB and 30.3 dB, respectively. For the Fashion-MNIST dataset, we evaluated the network performance by using the layer number of 2, 4, and 8 with the layer distance of 90 cm. The convergence plots of the network trained with three different layer numbers are shown in Fig. 4(e), where the PSNRs of the trained model on the testing dataset are 18.3 dB, 19.3 dB, and 21.2 dB, respectively. The de-scattering results of the trained eight-layer diffractive ONN on the "Trouser" and "Coat" images of the testing dataset are shown in the middle column of Fig. 4(d), the PSNRs of which are 21.4 dB and 21.1 dB, respectively. As a comparison, the freespace propagation of the strongly distorted optical wavefront of two objects without using the diffractive ONN generated speckle patterns on the detector [third column of Fig. 4(d)]. The results demonstrate the effectiveness of our approach for the *in situ* reconstruction of the object from its distorted wavefront and achieving all-optical imaging through scattering media.

5. DISCUSSION

A. Optical Training Speed and Energy Efficiency

The optical error backpropagation architecture proposed in this paper allows us to accelerate the training speed and improve the energy efficiency on core computing modules compared with electronic training. Without considering the power consumption of peripheral drivers, the theoretical calculation of training speed and energy efficiency for the proposed optical training of diffractive ONN can be found in Section 4 of Appendix A, where the optical training time per iteration and its energy efficiency are formulated in Eqs. (A1) and (A2), respectively. We evaluated the computational performance of the proposed optical training with the network settings of three different applications detailed in Section 4. As increasing the framerate of the adopted sensors and SLMs will reduce the optical training time of each iteration [Eq. (A1)], we consider the state-of-the-art framerate of an off-the-shelf sCMOS sensor (e.g., Andor Zyla 5.5), which provides a 100 fps (frames per second) frame rate at a spatial resolution of 5.5 megapixels, and off-the-shelf phase-only SLM (e.g., Holoeye PLUTO-2), which provides a 60 fps frame rate at a spatial resolution of 2.1 megapixels. Under such settings, according to Eq. (A1), the theoretical optical training time of each iteration is limited by the forward and backward measurement of the optical field, i.e., a total of eight frames of measurements at each sensor for the gradient update. Thus, the optical training time of each iteration under the batch size of one is $t = 0.08$ s, which is independent of network scale and allows training a large-scale diffractive ONN. With network configurations and training platforms detailed in Section 4.A, the *in silico* electronic training time of each iteration with a batch size of 10 for the classification network is 1.32 s, which is 1.65 times slower than the optical training (feeding in training instances of each batch sequentially).

To calculate the optical energy efficiency, as the power range of CNI MRL-FN-698 laser source at a working wavelength of 698 nm is 0–200 mW, we set the power of the laser source to 1 mW, i.e., $P = 1$ mW. It offers 0.12 μW light power even after 13 diffractive layers considering a 50% transmission rate of BS, which can still provide hundreds of photons per pixel per microsecond for the Andor Zyla 5.5 sCMOS sensor and can achieve sufficient SNR measurements for *in situ* computing. According to the formulation in Eq. (A2), the optical energy efficiencies of the proposed *in situ* optical training of diffractive ONN architecture under the light source power of 1 mW for the applications in classification, matrix-vector multiplication, and imaging through scattering media were calculated to be 7.86×10^{11} MAC/(s · W), 5.85×10^{11} MAC/(s · W), and 1.17×10^{12} MAC/(s · W), respectively, as shown in Table 1 of Appendix A. The energy efficiency of the core computing module is 2–3 orders of magnitude higher compared with the current GPU and CPU, which are on the order of 10^9 MAC/(s · W) [14].

B. System Calibration Under Misalignment Error

The architecture of *in silico* electronic training of diffractive ONN is confronted with the great challenge of physical implementation of the trained model, since different error sources in practice will deteriorate the model. For example, with an increasing layer number, the alignment complexity of diffractive layers will be significantly increased, which restricts the network scalability. To address this issue, we propose the *in situ* optical training architecture for physically implementing the optical error backpropagation directly inside the optical system, which enables the network to adapt to system imperfections and avoids the alignment between successive layers. Nevertheless, at each layer, the gradient calculation in optical error backpropagation requires measurements of the forward and backward propagated optical fields; the misalignment between the forward and backward measurements will lead to errors in the calculated gradient and deteriorate the training model. For example, the numerical evaluation demonstrates that the misalignment of 8 μm on the measurements at each layer decreases the classification accuracy of *in situ* optical training from 91.96% to 89.45% with CFGM error. Different from the *in silico* electronic training, the alignment needs to be performed only within the layer, and the alignment complexity is independent of the network layer number.

Furthermore, such misalignment can be calibrated out by optically calculating the gradient of each layer to estimate the amount of misalignment, as demonstrated in Section 5 of Appendix A (Fig. 7). We adopted the symmetrical Gaussian phase profile as the calibration pattern of the SLM and calculated the gradient by using the uniform input pattern as well as the uniform ground truth measurement. If the measurements are aligned, then the calculated gradient is a symmetrical pattern under the symmetrical Gaussian phase modulation. The misalignment in x and y will correspondingly

lead to the asymmetry of the gradient patterns in x and y, the amount of which determines the amount of misalignment error. By aligning the forward and backward measurement systems to minimize the asymmetry of the gradient pattern at each layer, the *in situ* optical training system can be calibrated.

6. CONCLUSION

In conclusion, we have demonstrated that the diffractive ONN can be *in situ* trained at high speed and with high energy efficiency with the proposed optical error backpropagation architecture. Our approach can adapt to system imperfectness and achieve highly accurate gradient calculation, which offers the prospect of reconfigurable and robust implementation of large-scale diffractive ONN. The numerical evaluations by using the simulated experimental system, configured with multilayer programmable SLMs, for three different applications, including light-speed object classification, optical matrix-vector multiplication, and all-optical imaging through scattering media, demonstrate the effectiveness of the proposed approach. The architecture can be easily extended to nonlinear diffractive ONNs by measuring the optical field at nonlinear layers and calculating additional nonlinear gradients (details in Section 6 of Appendix A). By incorporating additional optical elements, e.g., a microlens array, the proposed approach can potentially be extended to implement optical convolutional neural networks, and batch normalization or dropout may also be incorporated by multiplying the factors to or turn off the SLM coefficients.

Limitations of the proposed *in situ* optical training system include the sequential read-in mode and the relatively high cost of the existing SLM. These could be alleviated with integrated photonics: with the emergence of programmable on-chip optoelectronic devices, e.g., tunable metasurface SLMs [53], the proposed architecture could potentially be implemented at the chip scale to achieve the in-memory optical computing machine learning platform with high-density integration and be more cost effective. Due to the ubiquitous use of analog devices and the imperative trending of *in situ* learning architecture in modern neuromorphic computing [54], we believe the proposed optical error backpropagation approach for *in situ* training of ONNs provides essential support in neuromorphic photonics for building next-generation high-performance large-scale brain-inspired photonic computers.

APPENDIX A

1. Pseudo Code of Optical Error Backpropagation

Objective Function: $L(\mathbf{O}, \mathbf{T})$

Initialization: Number of layers N, batch size B, learning rate η, SLM phase coefficient $\boldsymbol{\phi}_k$ at k-th layer

WHILE NOT CONVERGED

1. Getting a batch size of input–output samples from the training dataset
2. Gradient calculation with an input–output example
   FOR i in range (1, B) DO
   1) Generating the ground truth target $\mathbf{T}^i$
   2) Forward propagation
      Propagate $\mathbf{U}_0^i$ from input plane
      For k in range (1, N) DO
         Record $\mathbf{P}_k^f$ at $2k$-th camera
      END FOR
      Record $\mathbf{U}_{N+1}^i$ at the output layer
   3) Error optical field calculation
      $$\mathbf{E}^i = \left(\frac{\partial L}{\partial \mathbf{O}}\right)_{\mathbf{O}=\mathbf{U}_{N+1}^i} \odot \mathbf{U}_{N+1}^{i*}$$
   4) Error optical field generation
      Generate $\mathbf{E}^i$ with CFGM
   5) Backward propagation
      Propagate $\mathbf{E}^i$ from the output plane
      For k in range (1, N) DO
         Record $\mathbf{P}_k^b$ at $(2k+1)$-th camera
      END FOR
   6) Gradient calculation
      For k in range (1, N) DO
      $$\frac{\partial L^i}{\partial \boldsymbol{\phi}_k} = 2\mathrm{Re}\{(j\mathbf{P}_k^f \odot \mathbf{P}_k^b)\}^T$$
      END FOR
   END FOR
3. Gradient averaging
   $$\frac{\partial \mathbf{L}}{\partial \boldsymbol{\phi}_k} = \frac{1}{B}\sum_i \frac{\partial \mathbf{L}^i}{\partial \boldsymbol{\phi}_k}$$
4. SLM phase coefficient update
   For k range (1, N) DO
   $$\boldsymbol{\phi}_k = \boldsymbol{\phi}_k - \eta \frac{\partial \mathbf{L}}{\partial \boldsymbol{\phi}_k}$$
   END FOR

END WHILE

2. Classification Performance w.r.t. Number of Layers

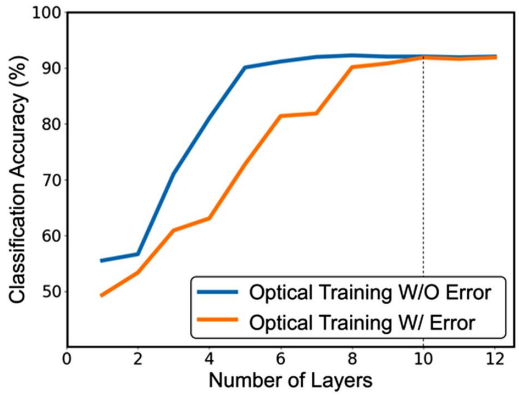

Fig. 5. Performance of the *in situ* optically trained MNIST classifier with respect to the number of diffractive layers. The classification accuracy increases with the increase in number of layers. For demonstration and comparison, the layer number of the classification network is set to 10, as shown in Fig. 2 of the main text.

3. Training Set of Optical Matrix-Vector Multiplier

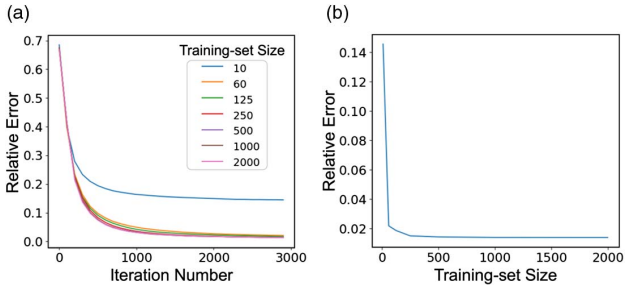

Fig. 6. Performance of the trained optical matrix-vector multiplier with respect to the size of the training set. The training, testing, and validation datasets are generated in an electronic computer by using the target matrix operator as shown in the last column of Fig. 3(c) of the main text, which is used for *in situ* optical training of the diffractive optical neural network (ONN) to perform the optical matrix-vector multiplier. The dataset's input vectors are randomly sampled with a uniform distribution between zero and one. With the network settings detailed in Section 4.C of the main text, the convergence plots of the training with different training set sizes are shown in (a), where the relative errors are evaluated over the validation dataset. The performance of the optically trained diffractive ONN with respect to the training set size evaluated on the testing dataset is shown in (b). Although increasing the size of the training set reduces the relative error and improves network performance, it requires more computational resources in an electronic computer. The numerical experimental results show the comparable model accuracy and convergence speed when the training set size is larger than 500, which is therefore adopted for this application. To sufficiently evaluate the generalization of the network, both the validation and testing datasets are set to have a size of 1000.

4. Optical Training Speed and Energy Efficiency

Processing speed and energy efficiency are considered as two major advantages of optical computing. In the following, without considering the power consumption of peripheral drivers, we analyze the theoretical training speed and energy efficiency of the proposed optical error backpropagation architecture for *in situ* optical training of the diffractive ONN.

During the derivation of optical error backpropagation in Section 2 of the main text, we have shown that the complex transform of the optical field between the successive diffractive ONN layers includes freespace propagation and modulation of a diffractive layer. To make a fair comparison with electronic computing, the freespace propagation was numerically calculated with the angular spectrum method [22], where the output optical field after freespace propagation was calculated by Fourier transforming of the input optical field, multiplying with the propagation kernel, and conducting the inverse Fourier transform on the multiplication result. For a diffractive layer with neuron numbers of $M \times M$, the numbers of complex-valued multiply-accumulate (MAC) operations of both the Fourier transform and inverse Fourier transform operators are $2M^2\log_2 M$, and the multiplication of the propagation kernel requires M^2 complex-valued MAC operations. Also, the modulation of the optical field with a diffractive layer requires additional M^2 complex-valued MAC operations. Therefore, the total numbers of complex-valued MAC operations at each diffractive ONN layer are $4M^2\log_2 M + 2M^2 = 2M^2(2\log_2 M + 1)$.

For a network with N diffractive layers, the network gradient at each iteration was calculated by measuring the forward and backward propagated optical fields at individual layers. We adopted the four-step phase-shifting digital holography (detailed in Section 3 of the main text) to measure the optical field, which performs four times of optical interference and corresponds to $4M^2$ complex-valued MAC operations at each layer. By summing the operations of the forward and backward optical field propagations as well as the measurements at all diffractive layers, the total numbers of computational operations of gradient calculation at each iteration are $R = 4NM^2(2\log_2 M + 3)$ complex-valued MAC, which corresponds to $R = 16NM^2(2\log_2 M + 3)$ real-valued MAC. Since each iteration of the optical training requires both forward and backward optical field measurements, i.e., a total of eight frames of measurements at each sensor for the gradient update, the theoretical optical training time of each iteration can be formulated as

$$t = \max\{8/F_s, 1/F_m\}, \tag{A1}$$

where F_s refers to the framerate of the sensor, and F_m refers to the framerate of the SLM determining the speed of gradient updating. Let P denote the power of the light source for the system; then the energy efficiency for the proposed optical error backpropagation architecture can be formulated as

Table 1. Computational Performance of the Proposed Optical Training Architecture[a]

In situ Optical Training Applications	MNIST Classification	Matrix-Vector Multiplication	De-scattering (Fashion-MNIST)
Performance	Accuracy: 91.86%	Relative error: 1.13%	PSNR: ~22.00 dB
Number of layers (N)	10	4	8
Neurons per layer ($M \times M$)	150 × 150	200 × 200	200 × 200
Total parameters	225,000	160,000	320,000
Training time per iteration (s)	**0.08**	**0.08**	**0.08**
Energy efficiency [MAC/(s·W)]	$\mathbf{7.86 \times 10^{11}}$	$\mathbf{5.85 \times 10^{11}}$	$\mathbf{1.17 \times 10^{12}}$

[a]The optical training time per iteration under the batch size of one is 0.08 s, which is independent of the network scale. The energy efficiencies were calculated on three different diffractive ONNs designed for three different applications, including the MNIST dataset classification, matrix-vector multiplication, and imaging through scattering media. Without considering the energy consumption of peripheral drivers, the calculation results show that the energy efficiency of optical training on the core computing modules achieves 2–3 orders of magnitude higher than the current GPU and CPU that are on the order of 10^9 MAC/(s·W) [14].

$$E = R/Pt = 16NM^2(2\log_2 M + 3)/Pt. \quad \textbf{(A2)}$$

5. System Calibration Under Misalignment Error

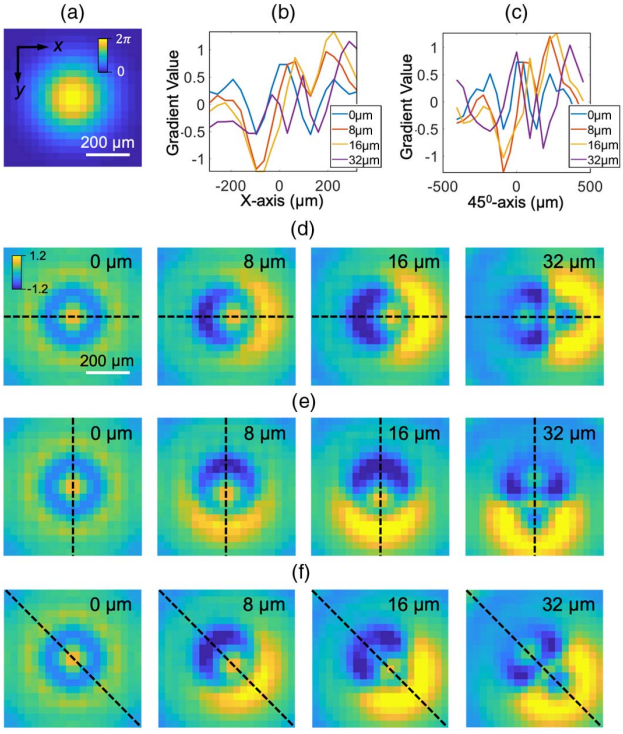

Fig. 7. Gradient calculation for system calibration under misalignment error. The proposed *in situ* optical training avoids the accumulation of misalignment error from layer to layer, and the alignment complexity is independent of the network layer number. The misalignment of our *in situ* optical training is evaluated by including different amounts of misalignment between the measurements of the forward and backward optical fields at each layer. To calibrate the system at each layer, the symmetrical Gaussian phase profile (a) is used as the calibration pattern on the spatial light modulator. The calibration process is to optically calculate the gradient of the diffractive layer given the uniform input pattern as well as the uniform ground truth measurement for determining the amount of misalignment. Due to the use of symmetrical Gaussian phase modulation, the calculated gradient should also be symmetrical if there is no misalignment error, as shown in the first columns of (d)–(f). The misalignment on the x axis and y axis, e.g., 8 μm, 16 μm, and 32 μm, will cause the corresponding asymmetry of the gradient patterns on the x axis and y axis, as shown in the second, third, and fourth columns of (d)–f), respectively, with the cross-section profiles shown in (b) and (c), where the amount of asymmetry can be used for estimating the amount of misalignment error. The *in situ* optical training system can be calibrated by minimizing the asymmetry of the gradient pattern at each layer. Scale bar: 200 μm.

6. Optical Training of Nonlinear Diffractive ONN

We extend the proposed architecture for *in situ* optical training of the nonlinear diffractive ONN by incorporating the optical nonlinearity layers in between the diffractive layers considering the physical implementation, as shown in Fig. 8(a). Let $\boldsymbol{f}(\cdot)$ represent the element-wise nonlinear operator performing the activation function for each layer; then Eq. (3) of the main text can be generalized as

$$\begin{aligned}\mathbf{O} &= |\mathbf{U}_{N+1}|^2 \\ &= \left|\mathbf{W}_{N+1}\left(\prod_{k=N}^{1}\boldsymbol{f}_k\mathbf{W}_k\mathbf{M}_k\mathbf{W}_{k-\frac{1}{2}}\right)\mathbf{U}_0\right|^2, \end{aligned} \quad \textbf{(A3)}$$

where $\mathbf{W}_{k-\frac{1}{2}}$ represents the diffractive propagation between the $(k-1)$-th nonlinear layer and k-th diffractive layer, and $\mathbf{W}_k$ represents the diffractive propagation between the k-th diffractive layer and the k-th nonlinear layer.

With nonlinear layers in the diffractive propagation model, the gradient calculation in Eq. (5) of the main text can be formulated as

$$\begin{aligned}&\partial\mathbf{U}_{N+1}/\partial\boldsymbol{\phi}_k \\ &= \begin{pmatrix} j\mathrm{diag}\left(\mathbf{M}_k\mathbf{W}_{k-\frac{1}{2}}\left(\prod_{i=k-1}^{1}\boldsymbol{f}_i\mathbf{W}_i\mathbf{M}_i\mathbf{W}_{i-\frac{1}{2}}\right)\mathbf{U}_0\right)^T \\ \left(\mathbf{W}_k^T\mathrm{diag}(\mathbf{g}_k)\left(\prod_{i=k+1}^{N}\mathbf{W}_{i-\frac{1}{2}}^T\mathbf{M}_i\mathbf{W}_i^T\mathrm{diag}(\boldsymbol{g}_i)\right)\mathbf{W}_{N+1}^T\right)\end{pmatrix}^T, \end{aligned} \quad \textbf{(A4)}$$

where $\boldsymbol{g}_k = \boldsymbol{f}_k'(\mathbf{W}_k\mathbf{M}_k\mathbf{W}_{k-\frac{1}{2}}(\prod_{i=k-1}^{1}\boldsymbol{f}_i\mathbf{W}_i\mathbf{M}_i\mathbf{W}_{i-\frac{1}{2}})\mathbf{U}_0)$ is the gradient of the nonlinear function in forward propagation. Finally, the nonlinearity will result in the modification of the updating rule of $\mathbf{P}_k^f$ and $\mathbf{P}_k^b$ in Eq. (7):

$$\begin{cases}\mathbf{P}_k^f = \mathbf{M}_k\mathbf{W}_{k-\frac{1}{2}}\left(\prod_{i=k-1}^{1}\boldsymbol{f}_i\mathbf{W}_i\mathbf{M}_i\mathbf{W}_{i-\frac{1}{2}}\right)\mathbf{U}_0, \\ \mathbf{P}_k^b = \mathbf{W}_k^T\mathrm{diag}(\boldsymbol{g}_k)\left(\prod_{i=k+1}^{N}\mathbf{W}_{i-\frac{1}{2}}^T\mathbf{M}_i\mathbf{W}_i^T\mathrm{diag}(\boldsymbol{g}_i)\right)\mathbf{W}_{N+1}^T E.\end{cases} \quad \textbf{(A5)}$$

Figure 8 illustrates the optical forward and backward propagations of diffractive ONN with nonlinear layers. For ease of formulation, we denote the optical field of the nonlinear layer in the forward and backward propagations as $\mathbf{P}_k^{f\text{-}NL}$ and $\mathbf{P}_k^{b\text{-}NL}$, which can be formulated as

$$\begin{cases}\mathbf{P}_k^{f\text{-}NL} = \mathbf{W}_k\mathbf{M}_k\mathbf{W}_{k-\frac{1}{2}}\left(\prod_{i=k-1}^{1}\boldsymbol{f}_i\mathbf{W}_i\mathbf{M}_i\mathbf{W}_{i-\frac{1}{2}}\right)\mathbf{U}_0, \\ \mathbf{P}_k^{b\text{-}NL} = \left(\prod_{i=k+1}^{N}\mathbf{W}_{i-\frac{1}{2}}^T\mathbf{M}_i\mathbf{W}_i^T\mathrm{diag}(\boldsymbol{g}_i)\right)\mathbf{W}_{N+1}^T E.\end{cases} \quad \textbf{(A6)}$$

The differences between the nonlinear and linear *in situ* optical training are as follows: (1) in forward propagation, the optical field of the nonlinear layer $\mathbf{P}_k^{f\text{-}NL}$ should be additionally measured; (2) the backpropagation of each layer is divided into two steps, i.e., measuring the optical field $\mathbf{P}_k^{b\text{-}NL}$ first and then using $\mathbf{P}_k^{f\text{-}NL}$ and $\mathbf{P}_k^{b\text{-}NL}$ to generate the modulation field $\boldsymbol{g}_k \odot \mathbf{P}_k^{b\text{-}NL}$ for performing the second step of backpropagation.

To demonstrate the optical training of diffractive ONN with nonlinearity, ferroelectric thin films [23] are adopted for the diffractive ONN, which are placed between the successive diffractive layers. The nonlinear activation function can be formulated as

$$\boldsymbol{f}(\mathbf{E}_{\mathrm{in}}) = \mathbf{E}_{\mathrm{in}}e^{j\pi\frac{\langle|\mathbf{E}_{\mathrm{in}}|^2\rangle}{1+\langle|\mathbf{E}_{\mathrm{in}}|^2\rangle}}. \quad \textbf{(A7)}$$

Corrected: 22 June 2020

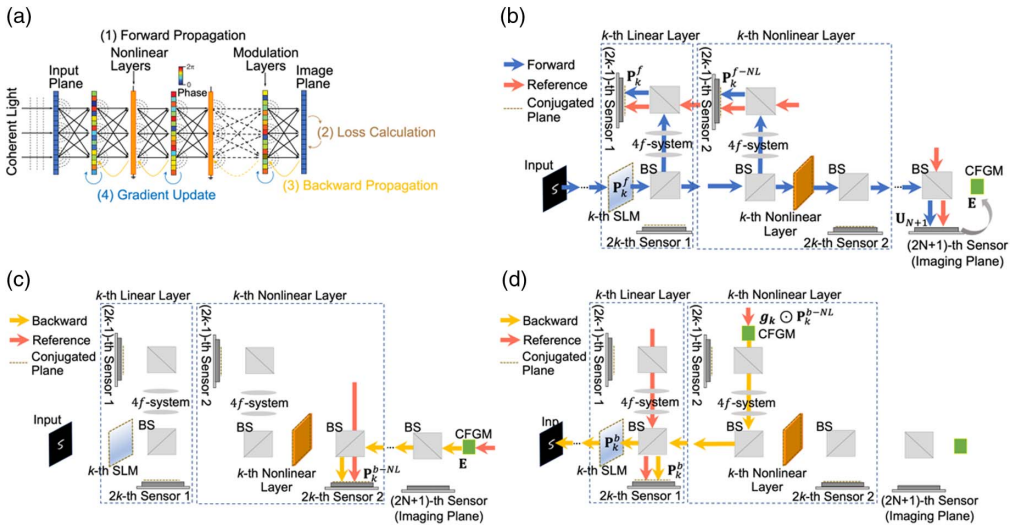

Fig. 8. *In situ* optical training of nonlinear diffractive ONN. (a) The optical nonlinearity layer is incorporated into the proposed architecture by using the ferroelectric thin film [23] to perform the activation function for individual layers. (b) To calculate the optical gradient for the nonlinear diffractive ONN, the optical fields are measured for both diffractive and nonlinear layers during forward propagation. (c), (d) Backward propagation is divided into two steps, i.e., backward propagating the error optical field and modulation field separately.

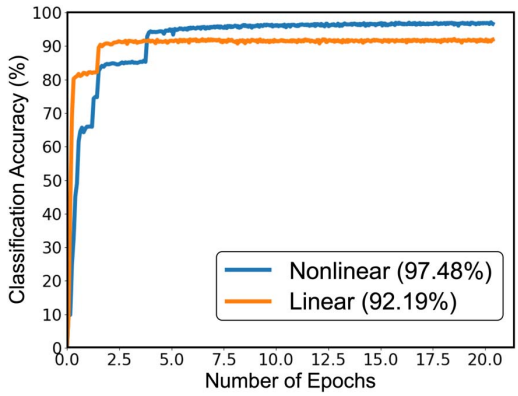

Fig. 9. Convergence plot of the nonlinear diffractive ONN for object classification on the MNIST dataset in comparison with the linear diffractive ONN in Section 4.A of the main text.

The *in situ* training result of nonlinear diffractive ONN for classifying the MNIST dataset with the same network setting as the corresponding linear diffractive ONN is shown in Fig. 9. After the incorporation of nonlinearity, the classification accuracy of the MNIST dataset increases from 92.19% to 97.48%.

Funding. Beijing Municipal Science and Technology Commission (No. Z181100003118014); National Natural Science Foundation of China (No. 61722209); Tsinghua University Initiative Scientific Research Program.

Disclosures. The authors declare no conflicts of interest.

†These authors contributed equally to this paper.

REFERENCES

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
2. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105.
3. G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition," IEEE Signal Process. Mag. **29**, 82–97 (2012).
4. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, and M. Lanctot, "Mastering the game of go with deep neural networks and tree search," Nature **529**, 484–489 (2016).
5. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," Nat. Med. **25**, 24–29 (2019).
6. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**, 921–943 (2019).
7. A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, *DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning* (Nature, 2018).
8. C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," arXiv:1705.09792 (2017).
9. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.
10. P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, and Y. Nakamura, "A million spiking-neuron integrated circuit with a scalable communication network and interface," Science **345**, 668–673 (2014).
11. J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, and W. He, "Towards artificial general intelligence with hybrid Tianjic chip architecture," Nature **572**, 106–111 (2019).
12. B. Marr, B. Degnan, P. Hasler, and D. Anderson, "Scaling energy per operation via an asynchronous pipeline," IEEE Trans. Very Large Scale Integr. Syst. **21**, 147–151 (2012).
13. J. M. Shainline, S. M. Buckley, R. P. Mirin, and S. W. Nam, "Superconducting optoelectronic circuits for neuromorphic computing," Phys. Rev. Appl. **7**, 034013 (2017).

14. P. R. Prucnal and B. J. Shastri, *Neuromorphic Photonics* (CRC Press, 2017).
15. D. Woods and T. J. Naughton, "Optical computing: photonic neural networks," Nat. Phys. **8**, 257–259 (2012).
16. D. R. Solli and B. Jalali, "Analog optical computing," Nat. Photonics **9**, 704–706 (2015).
17. Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, "Artificial neural networks enabled by nanophotonics," Light Sci. Appl. **8**, 1 (2019).
18. X. Luo, "Engineering optics 2.0: a revolution in optical materials, devices, and systems," ACS Photon. **5**, 4724–4738 (2018).
19. P. Minzioni, C. Lacava, T. Tanabe, J. Dong, X. Hu, G. Csaba, W. Porod, G. Singh, A. E. Willner, and A. Almaiman, "Roadmap on all-optical processing," J. Opt. **21**, 063001 (2019).
20. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, and D. Englund, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**, 441–446 (2017).
21. T. W. Hughes, R. J. England, and S. Fan, "Reconfigurable photonic circuit for controlled power delivery to laser-driven accelerators on a chip," Phys. Rev. Appl. **11**, 064014 (2019).
22. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004–1008 (2018).
23. T. Yan, J. Wu, T. Zhou, H. Xie, F. Xu, J. Fan, L. Fang, X. Lin, and Q. Dai, "Fourier-space diffractive deep neural network," Phys. Rev. Lett. **123**, 023901 (2019).
24. G. Van der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," Nanophotonics **6**, 561–576 (2017).
25. L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, "High-speed photonic reservoir computing using a time-delay-based architecture: million words per second classification," Phys. Rev. X **7**, 011015 (2017).
26. J. Feldmann, N. Youngblood, C. Wright, H. Bhaskaran, and W. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," Nature **569**, 208–214 (2019).
27. R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," Phys. Rev. X **9**, 021032 (2019).
28. I. Chakraborty, G. Saha, and K. Roy, "Photonic in-memory computing primitive for spiking neural networks using phase-change materials," Phys. Rev. Appl. **11**, 014063 (2019).
29. T. Deng, J. Robertson, Z.-M. Wu, G.-Q. Xia, X.-D. Lin, X. Tang, Z.-J. Wang, and A. Huartado, "Stable propagation of inhibited spiking dynamics in vertical-cavity surface-emitting lasers for neuromorphic photonic networks," IEEE Access **6**, 67951–67958 (2018).
30. J. Robertson, T. Deng, J. Javaloyes, and A. Hurtado, "Controlled inhibition of spiking dynamics in VCSELs for neuromorphic photonics: theory and experiments," Opt. Lett. **42**, 1560–1563 (2017).
31. T. W. Hughes, I. A. Williamson, M. Minkov, and S. Fan, "Wave physics as an analog recurrent neural network," arXiv:1904.12831 (2019).
32. J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large-scale photonic recurrent neural network," Optica **5**, 756–760 (2018).
33. E. Khoram, A. Chen, D. Liu, L. Ying, Q. Wang, M. Yuan, and Z. Yu, "Nanophotonic media for artificial neural inference," Photon. Res. **7**, 823–827 (2019).
34. A. S. Backer, "Computational inverse design for cascaded systems of metasurface optics," arXiv:1906.10753 (2019).
35. S. Maktoobi, L. Froehly, L. Andreoli, X. Porte, M. Jacquot, L. Larger, and D. Brunner, "Diffractive coupling for photonic networks: how big can we go?" IEEE J. Sel. Top. Quantum Electron. **26**, 7600108 (2019).
36. Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, "All optical neural network with nonlinear activation functions," Optica **6**, 1132–1137 (2019).
37. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," Sci. Rep. **8**, 12324 (2018).
38. H. Chen, S. Jayasuriya, J. Yang, J. Stephen, S. Sivaramakrishnan, A. Veeraraghavan, and A. Molnar, "ASP vision: optically computing the first layer of convolutional neural networks using angle sensitive pixels," in *IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 903–912.
39. Y. Luo, D. Mengu, N. T. Yardimci, Y. Rivenson, M. Veli, M. Jarrahi, and A. Ozcan, "Design of task-specific optical systems using broadband diffractive neural networks," arXiv:1909.06553 (2019).
40. J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3D object detection," arXiv:1904.08601 (2019).
41. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE **86**, 2278–2324 (1998).
42. T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of photonic neural networks through *in situ* backpropagation and gradient measurement," Optica **5**, 864–871 (2018).
43. M. Hermans, J. Dambre, and P. Bienstman, "Optoelectronic systems trained with backpropagation through time," IEEE Trans. Neural Netw. Learn. Syst. **26**, 1545–1550 (2014).
44. M. Hermans, M. Burm, T. Van Vaerenbergh, J. Dambre, and P. Bienstman, "Trainable hardware for dynamical computing using error backpropagation through physical media," Nat. Commun. **6**, 6729 (2015).
45. K. Wagner and P. Demetri, "Multilayer optical learning networks," Appl. Opt. **26**, 5061–5076 (1987).
46. D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals," Appl. Opt. **27**, 1752–1759 (1988).
47. I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," Opt. Lett. **22**, 1268–1270 (1997).
48. O. Mendoza-Yero, G. Mínguez-Vega, and J. Lancis, "Encoding complex fields by using a phase-only optical element," Opt. Lett. **39**, 1740–1743 (2014).
49. M. W. Matthès, P. del Hougne, J. de Rosny, G. Lerosey, and S. M. Popoff, "Optical complex media as universal reconfigurable linear operators," Optica **6**, 465–472 (2019).
50. A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," Nat. Photonics **6**, 283–292 (2012).
51. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**, 1181–1190 (2018).
52. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," Optica **5**, 1–9 (2018).
53. G. K. Shirmanesh, R. Sokhoyan, P. C. Wu, and H. A. Atwater, "Electro-optically tunable universal metasurfaces," arXiv:1910.02069 (2019).
54. Z. Wang, C. Li, P. Lin, M. Rao, Y. Nie, W. Song, Q. Qiu, Y. Li, P. Yan, and J. P. Strachan, "*In situ* training of feed-forward and recurrent convolutional memristor networks," Nat. Mach. Intell. **1**, 434–442 (2019).

Corrected: 22 June 2020