Liu et al. PhotoniX (2021) 2:5
https://doi.org/10.1186/s43074-021-00026-0

PhotoniX

REVIEW Open Access

Research progress in optical neural networks: theory, applications and developments

Jia Liu[1], Qiuhao Wu[1], Xiubao Sui[1*], Qian Chen[1], Guohua Gu[1], Liping Wang[1] and Shengcai Li[2]

* Correspondence: sxbhandsome@163.com
[1]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
Full list of author information is available at the end of the article

Abstract

With the advent of the era of big data, artificial intelligence has attracted continuous attention from all walks of life, and has been widely used in medical image analysis, molecular and material science, language recognition and other fields. As the basis of artificial intelligence, the research results of neural network are remarkable. However, due to the inherent defect that electrical signal is easily interfered and the processing speed is proportional to the energy loss, researchers have turned their attention to light, trying to build neural networks in the field of optics, making full use of the parallel processing ability of light to solve the problems of electronic neural networks. After continuous research and development, optical neural network has become the forefront of the world. Here, we mainly introduce the development of this field, summarize and compare some classical researches and algorithm theories, and look forward to the future of optical neural network.

Keywords: Optical neural network, Deep learning, Optical linear operation, Optical nonlinearity, Training method

Introduction

As one of most active fields in computer science, artificial intelligence is focus on simulating structure of the nervous system through constructing artificial neural network (ANN) which establish connections between neurons in various layers of the neural network and make it good generalization ability and robustness. Since the 1980s, the research work of ANN has made great progress. Also, it has successfully solved many practical problems that are difficult to be solved by modern computers in the fields of pattern recognition, intelligent robot, automatic control, prediction and estimation, biomedicine, economy, etc., with good intelligence characteristics.

At present, electronic computing is still the most important computing power support for the implementation of artificial intelligence algorithms, especially deep ANN model. Although the specific hardware architectures are different, in a word, they all adopt the von Neumann type computing principle to complete the computing task with complex logic circuits and processor chips [1]. The original neural network architecture used CPU for computing, but it could not meet the requirements of a large

number of floating-point operations in deep network, especially training phase. Moreover, the parallel computing efficiency was too low, and it was quickly replaced by GPU with strong parallel computing capability. It can be said that GPU promoted the development of deep learning.

However, the demand for computational power in deep learning is endless. Limiting by the interference of electrical signals, energy consumption and physical limits [2, 3], although electronic components base on silicon can still support it now, the traditional deep learning has quietly appeared a bottleneck. The academia and industrial circles attempt to seek alternative methods to solve electronic defects that can take precautions on computing power. As the speed of light as high as 300,000 km per second, which is 300 times faster than that of electron, the information carrying ability and variety which is 2*10^4 times more than that of electric channels, as well as high parallelism and strong anti-interference [4, 5], it has great advantages in information transmission and optical computing. Replacing electricity with light has become a potential and promising work mode, which is the trend of the times.

Therefore, people try to build neural networks by optical way to achieve deep learning architecture. Optical neural network (ONN) emerges as the times require. It has the characteristics of high bandwidth, high interconnection and internal parallel processing, which can accelerate the partial operation of software and electronic hardware, even up to the "light speed", is a promising method to replace artificial neural network. In the photonic neural network, matrix multiplication can be performed at the speed of light, which can effectively solve the dense matrix multiplication in the artificial neural network, so as to reduce the consumption of energy and time. Moreover, the nonlinearity in ANN can also be realized by nonlinear optical elements. Once the training of the optical neural network is completed, the entire structure can perform the optical signal calculation at the speed of light without additional energy input. In 1978, Goodman of Stanford University first proposed the theoretical model of optical vector-matrix multiplier [6], which became an important step in optical calculation [7, 8], and promoted the development of optical matrix multiplier (OMM) [9, 10] and photonic neural network.

In this paper, we are going to discuss the hot topic in the field of deep learning——optical deep learning, that is to build neural network by optical method instead of traditional artificial neural network and train it. It has a large number of linear layers and is connected with each other. The specific structure of the paper is as follows: in the first chapter, it briefly introduces how the artificial neural network developed into optical neural network. ANN is mainly composed of two core components——linear part and nonlinear activation, and then is trained to adjust and optimize the weights of each connection, make the network converge in the end. Therefore, the second and third chapters start from the two core operations respectively, describe in detail how the researchers realize the linear operation and nonlinear activation function in the optical way after introducing the basic principles, so as to successfully build the optical neural network. The fourth chapter, according to different training methods, elaborates the particular training process of optical neural network, and carries out experiments and results comparison for some typical applications. Finally, in the fifth chapter, we analyze and discuss the optical neural network, describe the possible future research

direction and development of ONN; and a brief and to the point summary is given in the sixth chapter.

The optical realization of linear operation

In the introduce we mentioned ONN is the optical implementation of both of linear and nonlinear operations of ANN. According to the structure of ANN [11] and the working principle of neurons [12]——linear operation $z_i = b_i + \sum_j W_{ij}x_j$ and nonlinear activation $a_i = \phi(z_i)$ in Fig. 1, it can be seen that the neural network requires a lot of linear multiplication and summation operations. The most direct embodiment of such a multiplication and summation operation in the algorithm is to give two groups of data and carry out multiplication and addition operations in the "for" loop. If we think about this problem simply, we will find that many iterations are needed to complete this operation, which will waste a lot of computing resources. Thus, people begin to seek a faster method——vectorization method, which can make it into the multiplication of two matrices namely the input matrix and the weight matrix.

We know that it's easy to achieve the computation between two matrices by using an electronic computer, but it's also difficult to realize when the matrix dimension is very large. For example, to realize the multiplication of two matrices with a size of n*n, n^3 multiplication and n^3 addition operations need to be performed, which is $2n^3$ operations in total. If n is very large, assuming it is 1024, it requires to take 214,7483,648 calculations that is a huge number, up to millions of times. It can be seen that using computer to achieve multiplication operations is very time-consuming. However, if the high speed, high parallelism and anti-interference of light are used to achieve this operation by optical means, it is likely to require only a few or even only one operation. In the training of neural network, the data we need to process and analyze is extremely large. At this time, the characteristics of optics are extremely important, which can bring great convenience for calculation. The appearance of optical matrix multiplier lays the foundation of optical calculation, and provides a development path for the optics of neural network.

Next, we will briefly introduce optical matrix multiplier, which is the basic optical realization of linear multiplication and summation operation, namely matrix

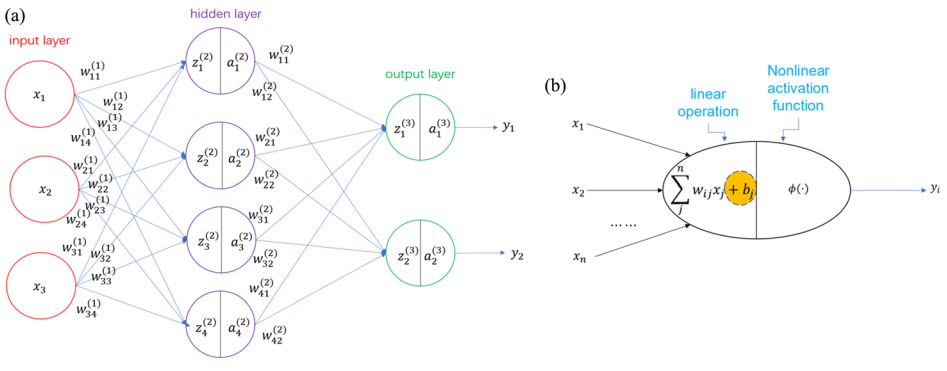

Fig. 1 **a** The structure of ANN [11]. **b** The principle of neurons [12]

multiplication, and then explain how to realize linear operation in optical neural network from the different principles of implementing the multiplication operation.

Optical matrix multiplier

Matrix multiplication is a very important operation in matrix operation, and its calculation process is complicated. Simply put, the multiplication between two matrices is that the corresponding elements of row i of the first matrix and column j of the second matrix are multiplied and added one by one, and then get result matrix element c_{ij}, which is also called inner product operation. The multiplication result matrix can be obtained by traversing the rows or columns of the two matrices once. If $A = (a_{ij})_{m \times s}$ and $B = (b_{ij})_{s \times n}$, the matrix multiplication operation is defined as follows:

$$A \times B = C, where\ C = \left(c_{ij}\right)_{m \times n}, c_{ij} = \sum_{k=1}^{s} a_{ik} b_{kj}, (i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$$

In fact, multiplication is a process of number accumulation for many times. Correspondingly, matrix multiplication is the sum of several different numbers after accumulation for many times. In the electronic computer, the accumulator as the core arithmetical unit, can be used to achieve matrix multiplication operation. Similarly, such an optical multiplier can be designed as the core of the photonic computing system, which has two-dimensional parallelism. Optical multiplication is the process in which optical information is loaded and converted, and the optical multiplier is responsible for realizing this process. The principle of the optical multiplier is simply described in Fig. 2(a).

If the function f in the graph is replaced by a matrix, the graph can be simply represented as multiplication between matrices. And matrices can be considered as a combination of vectors, so we can start from multiplying vectors by matrices to multiplying matrices by matrices. The vector-matrix multiplication system model was first proposed by Goodman [6]. After continuous research and improvement by scholars, the final structure of vector-matrix multiplier is shown in Fig. 2(b).

Let's take an m*n matrix A multiplied by an n-dimensional vector B to get an m-dimensional vector C. Firstly, vector B is realized by using linear array light source, and the light intensity of n light sources of linear array is corresponding to the input vector B. Then, the light beam emitted by the linear array source passes through the collimation lens L1 to form a parallel light and irradiates on the cylinder lens CL1. Due to its fan-out effect in the horizontal direction, B is duplicated by CL1 in the vertical direction to form a light band. After that, the beam reaches SLM, which is controlled by the computer to load the matrix A, and the two are multiplied. Then, the beam passes through the collimating lens CL2. Due to its fan-in effect in the vertical direction, the light of all pixels in the i-th row of SLM will be concentrated on the i-th detector of CCD. It can be found that in vector-matrix multiplication, the optical system will first copy and paste the vector and expand it into a matrix, and then multiply it with another matrix. From another point of view, this is a special kind of matrix-matrix multiplication. In 1993, an optical 4f system was proposed to realize the multiplication between matrices, as shown in Fig. 2(c), which mainly uses the Fourier transform of lens and the convolution principle [13].

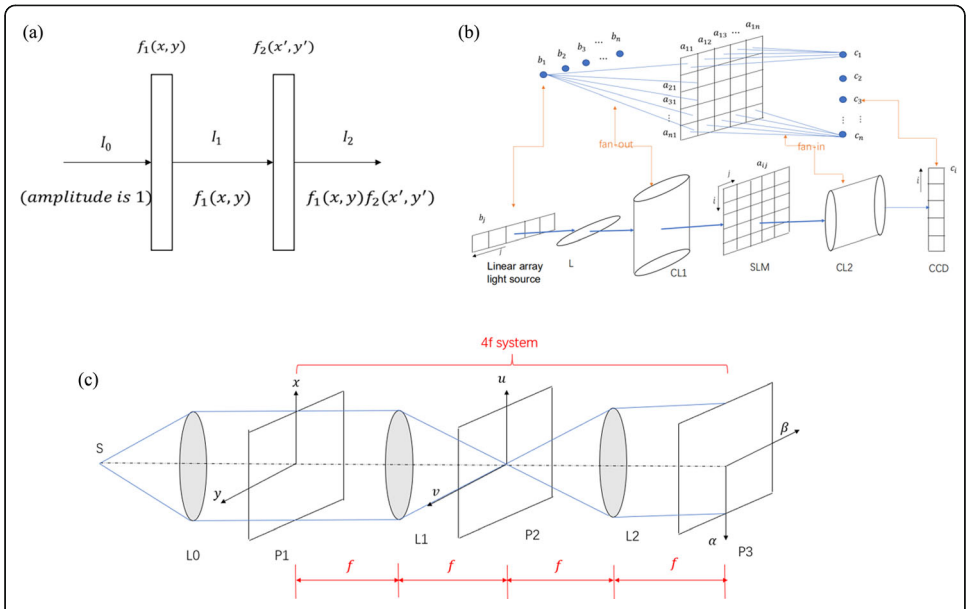

Fig. 2 Optical Matrix Multiplier. **a** The optical multiplier. **b** The structure of Vector-Matrix multiplication system [6]. **c** Matrix-matrix multiplication realized by 4f system [13]

The optical matrix multiplier fully embodies the parallel computing power of light, and the optical linear operation completed by OMM is essentially to realize the modulation of information-carrying light by means of certain approaches and some properties of light, such as diffraction, interference and so on.

Diffraction of light to realize linear operation

Light travels along a straight line in the air. When encountering an obstacle or a small hole, light will deviate from the straight-line propagation path, resulting in the phenomenon of uneven distribution of light intensity, which is called diffraction. After the discovery of diffraction in 1665, it attracted the attention of many scholars who invested a great deal of efforts in this field, and formed a complete system theory after long-term development.

In 1678, The Dutch physicist Huygens proposed that every point on the wave surface could be regarded as the wave source of the emitted secondary wave, emitting spherical secondary wave respectively. At a certain time in the future, the envelopment surface of these secondary waves would be the new wave surface at that time, which is the Huygens principle [14]. Although Huygens principle well explains refraction and reflection and birefringence of light, it does not involve the analysis of light wave intensity and wavelength, and cannot well explain diffraction phenomenon. After the appearance of the Young's Double-Slit Interference experiment in 1810 [15], Fresnel supplemented Huygens principle with the help of wavelet coherent superposition in 1815, and developed qualitative Huygens principle into semi-quantitative principle with mathematical proof, which is called Huygens-Fresnel principle [16], expressed as $\tilde{E}(P) = \frac{A}{i\lambda}\iint_{\Sigma} \frac{\exp(ikR)}{R} K(\theta) d\sigma$. However, this principle is only a semi-quantitative principle, and there is no specific function representation for the tilt factor, and the meaning of

proportionality coefficient is not clear, so it has limitations. Therefore, Kirchhoff and Sommerfeld derived the diffraction formula according to the general wave theory, and gave the specific form of the tilt factor and proportional coefficient. Kirchhoff used Green Theorem [17] to solve the Helmholtz equation [18], obtained the complex amplitude of monochromatic light in free space, and finally concluded the Kirchhoff integral theorem [19], specifically expressed the basic concepts of Huygens-Fresnel principle. The Kirchhoff's diffraction formula is as follows: $U(P_0) = \frac{A}{j\lambda}\iint_{\Sigma}\frac{e^{jk(r+l)}}{rl}\frac{cos<\vec{n},\vec{r}> - cos<\vec{n},\vec{l}>}{2}ds$. Although Kirchhoff's diffraction formula gives a good practical effect, the boundary conditions of Kirchhoff hypothesis violate the potential field theorem [20]. Therefore, Sommerfeld adopted another Green's formula to overcome the problem that Kirchhoff boundary condition assumption violate the potential theory theorem, making it self-consistent in theory. Its specific form is: $\begin{cases} U_{\mathrm{I}}(P_1) = \frac{A}{j\lambda}\iint_{\Sigma}\frac{e^{jk(r+l)}}{rl}\,cos<\vec{n},\vec{r}>ds \\ U_{\mathrm{II}}(P_1) = -\frac{A}{j\lambda}\iint_{\Sigma}\frac{e^{jk(r+l)}}{rl}\,cos<\vec{n},\vec{l}>ds \end{cases}$, this is the Rayleigh-Sommerfeld equation [21].

The above equations are all based on Fresnel diffraction. In addition, Fraunhofer diffraction was also discovered [22], which is a special case of Fresnel diffraction and belongs to far-field diffraction. Because the Fraunhofer diffraction field is easy to calculate theoretically, has great application value and it is not difficult to realize experimentally, people pay more attention to it. In particular, the rise of Fourier optics in modern transform optics endows classical Fraunhofer diffraction with new modern optical significance. With the rise of optical Fourier transform, the transformation from space domain to frequency domain is realized. Light can represent more contents, and the distribution of light in Fresnel diffraction is also analyzed in more detail. Kirchhoff and Rayleigh-Sommerfeld diffraction both discuss the propagation of light in the spatial domain, and the propagation of light in the frequency domain is summarized as angular spectrum theory [23].

Diffraction is a very extensive optical phenomenon, which contains a lot of content. The theories related to diffraction can be collectively called diffraction theory. But because the light is electromagnetic wave, the diffraction problem cannot be separated from the classical electromagnetic field theory based on Maxwell's equations, and electromagnetic field is also a vector field, so the strict diffraction theory should be the vector diffraction theory. When the light vector is only one component, or does not involve the diffraction light propagation, polarization state and the case that aperture wavelength is much larger than light wavelengths, the light can be regarded as a scalar, accordingly it is the scalar diffraction theory.

The implementation based on Rayleigh-Sommerfeld equation

Any obstacle can cause light to diffract, but only when the size of the obstacle or hole is smaller than or similar to the wavelength of light, obvious diffraction phenomenon can be observed. Diffraction produces numerous wavelets at the small aperture. These

wavelets superimpose each other when they reach the viewing screen. The degree of mutual weakening becomes lighter or heavier regularly during the overlapping, thus forming the light and dark streaks. In fact, diffraction is the coherent superposition of infinite continuous wavelet, which is mathematically represented as an integral problem. Therefore, the optical diffraction phenomenon can be used to design the linear operation of the optical neural network and realize the linear multiplication and summation operation in the neural network.

According to Rayleigh-Sommerfeld equation of diffraction theory, we can regard each neuron of a given diffraction layer as a secondary source of wave consisting of the optical model: $w_i^l(x,y,z) = \frac{z-z_i}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right)$, which is also the basic of many diffraction network architectures. In these networks, transmittance was taken as a learnable parameter W, and then training and learning were carried out to complete the task of identification and classification. Under normal circumstances, when the network using the diffraction principle modulates light waves and conducts diffraction light analysis, there will be such a premise that the vibration direction of the light vector in the whole light wave field doesn't change, or only one component of the light vector is considered, so vector diffraction is generally simplified to scalar diffraction for using.

In June 2018, Lin Xing, a researcher from the University of California, Los Angeles (UCLA), and other researchers, innovatively proposed an all-optical diffraction deep learning framework based on light diffraction, which they called the diffraction deep neural network (D^2NN) [24]. D^2NN is composed of multi-layer diffraction surfaces to form the physical layer. By cooperating with these diffraction surfaces, the linear operation function of neural network can be performed in the form of light. The principle and structure of the whole network are shown in Fig. 3(a), which is composed of input layer, several diffraction layers and output layer. In the input layer, the information is

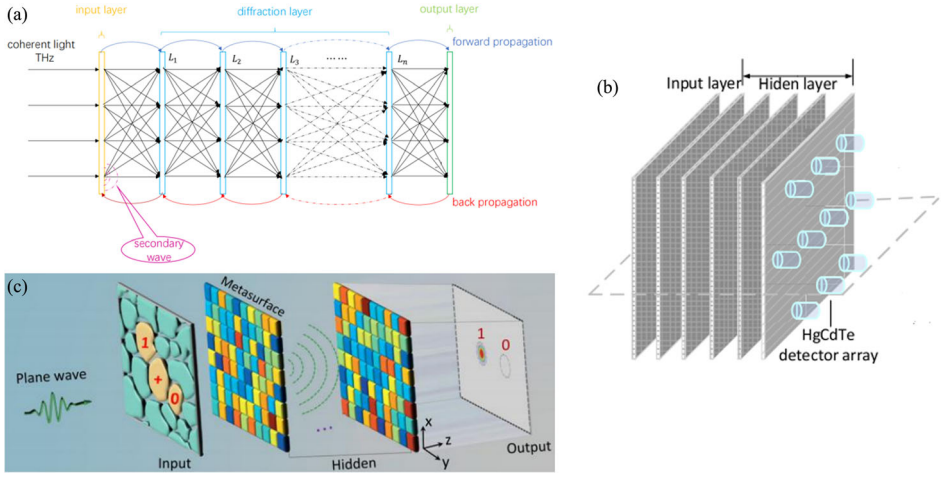

Fig. 3 Optical neural networks using light diffraction to realize linear operation. **a** Schematic diagram of deep diffraction neural network D^2NN [24]. **b** Diffraction grating network system [25]. **c** Metasurface implements optical logical operations [26]

encoded into the amplitude channel or phase channel of the input surface by irradiating with coherent light. There are several holes on the input surface, and diffraction of the beam occurs on the input surface, which results in coherent superposition of wavelet, changes the amplitude and phase of input wave, and completes the coding process. Through optical diffraction, light goes from the input layer through the diffraction layers to the output layer, achieving layer-by-layer connections. Similar to the input layer, the diffraction layer has certain parameters. Light of terahertz frequency can transmit through the diffraction layer. After being modulated by parameters through the diffraction layer, the coherent superposition of wavelet is carried out, so as to realize the modulation of light wave and complete the process of forward propagation, that is, the optical linear calculation process of neural network. In the output layer there will be a photoelectric detection array to detect the output light intensity. In 2019, this research group also proposed a broadband diffraction neural network based on the same architecture [27], which makes the model's demand for light sources no longer limited to monochromatic coherent light sources, and can process information modulated by time-incoherent light sources, expanding the application range of ONN realized by this architecture.

The architecture of 3D printed deep diffractive optical neural network achieves high-speed and low-power calculation, which is unique and innovative, but it still has some big problems. The first is the diffraction layer. Although the manufacturing cost of the diffraction layer is relatively low and the accuracy rate can reach 91.75%, it is difficult to achieve miniaturization of devices, process complex data and image analysis. Moreover, all parameters cannot be reprogrammed after 3D printing. The second problem is the light source, the THz light source used in this study. Such a system is expensive and bulky. The third is the surrounding experimental environment. In this study, an optical platform is required to carry out the network architecture. Due to the existence of optical diffraction, the requirements on the surrounding environment, such as vibration and optical environment, will be quite severe. Ozcan said, although the research uses light at terahertz frequencies, it is also possible to make light at visible, near-infrared or other frequencies in the future, and such networks could also be made by photolithography or other techniques. Therefore, inspired by the diffraction deep neural network, more and more scholars have begun to devote themselves to the study of variants based on D^2NN.

In December 2019, a team from Tianjin University developed a matrix grating to replace the 3D-printed diffraction layer [25], and used a carbon dioxide laser tube to emit 10.6um infrared light for detection by the HgCdTe detector array, as shown in Fig. 3(b). Similarly, the superposition of light waves is realized through the diffraction of each slit of the grating and the interference between slits, thereby achieving the optical linear operation. It is worth mentioning that infrared light source is used in the network which has the following advantages: Firstly, it can reduce the cost of the whole network architecture; Secondly, the size of a single neuron can be reduced to 5 μm, and the characteristic size is reduced by 80 times compared with the previous network. In this way, the matrix grating of 1 mm*1 mm can contain 200*200 neurons, and the distance between layers can also be shortened. The miniaturized matrix grating will be very helpful to integrate into the silicon photonic platform and acquire more extensive application.

In 2020, a team from Zhejiang University proposed to realize optical logic operation using metasurfaces based on diffraction neural network [26]. The optical logic computation is equivalent to the classification task, optical logic unit is designed based on the diffraction neural network, and finally realize the logic operation. The feasibility and completeness of this method is proved theoretically. Figure 3(c) shows the layout of the diffraction neural network based on the optical logical operation. Each region of the input layer is assigned a specific logical operator or an input logical state, which has two different states for light transmittance. In other words, the input layer only needs to set the transmission state of each region, then the input plane wave can be spatially coded for specific optical logic operation. The hidden layer is composed of the metasurfaces. According to Huygens-Fresnel diffraction principle, taking AND, OR, and NOT-logical units as examples, the hyperparameters and weight coefficients of the diffraction neural network are obtained through learning and training. Then, according to these parameters, an efficient medium metasurface is used to construct the phase mask. As a hidden layer, it is designed to decode the encoded input light and generate the output light logic state. Two regions are set in the output layer and light passing through the hidden layer is directionally scattered by the metasurface to one of the two designated regions in the output layer. Compared with 3D-layer diffraction system and matrix-grating network system, this method does not require complex optical control system, and only need simple plane wave as input. By selectively activating sub-region of input layer, different logical calculation functions can be realized.

To sum up, D^2NN based on the Rayleigh-Sommerfeld equation, is able to perform various complex functions that traditional computer neural networks can achieve at a speed close to the speed of light and without energy consumption. It opens up new opportunities for using passive components based on artificial intelligence to quickly analyze data, images and object classification, so as to realize all-optical image analysis, feature detection and object classification. For example, a driverless car using this technology can immediately respond to a stop sign. As soon as it receives light from the sign diffraction, D^2NN can read the sign information; the technology can also be used to categorize a large number of targets, such as looking for indications of disease in millions of cell samples. In addition, new camera designs and optical components using D^2NN to perform tasks can be implemented, passively used in medical technology, robotics, security, and any application that requires image and video data. For example, all-optical diffraction neural network can be used to construct holograms that can realize "THz" imaging at a very low cost through 3D printing [28], reconstructing high-quality images at a high speed.

The implementation based on the Fourier transform

The Fourier transform of light is also a member of the great family of diffraction, which grows out of Fraunhofer diffraction and plays an extremely important role in modern optics due to some of its special properties, such as convolution theorem. Based on Fourier optics, the Fourier lens of optical element can realize the Fourier transform and complete the conversion of time-space domain and frequency domain. According to the convolution theorem [29], the convolution of two two-dimensional continuous functions in the space domain can be obtained by the inverse transformation of the

product of their corresponding two Fourier transforms. On the contrary, convolution in the frequency domain can be obtained by Fourier transform of product in the space domain. Hence, the multiplication operation can be done by convolution in the frequency domain, and then by the inverse Fourier transform.

$$f(x,y)*h(x,y) <=> F(u,v)H(u,v) \tag{1}$$

$$f(x,y)h(x,y) <=> 1/2\pi[F(u,v)*H(u,v)] \tag{2}$$

Not only that, one of the simplest and most basic functions of a lens is to converge light beam, which can be similar to a summation operation to a certain extent. Therefore, we can use the function of Fourier transform and light wave convergence and superposition of lens wave to realize the function of linear multiplication and summation function of optical neural network.

In 1989, Tai Wei Lu proposed a two-dimensional programmable optical neural network [30], which is based on the interconnection structure of lens array. Linear summation is realized by using lens array, and has good parallel computing and programming abilities. But because of the influence of imaging aberrations and light detection, the number of neurons is severely limited. In 1997, Yang's team used a coaxial lens array to build an optical neural network with 32*32 neurons [31], which significantly reduced aberration and improved light efficiency. In 1993, Yasunori Kuratomi proposed an optical neural network with vector feature extraction [32]. The network structure consists of four layers, namely the input layer, the two hidden layers and the output layer. In the input layer, a flat plate is used to convert letters to binary grid pattern. In the hidden layer 1, a 2 * 2 lens array is applied to realize four feature extraction layers, used to extract feature line segments, and focus on feature-extracting optical neuron device (FEOND) as the hidden layer 2 to extract feature vector. Ultimately, the FEOND output is obtained from readout beam by crossed polarizers, which is detected by CCD for recognition tasks. Neurons in the output layer are fully connected with the neurons in the hidden layer 2.

In the early stage, the lens network reflected its focusing and gathering function, so as to achieve linear summation operation. With the gradual maturity of Fourier theory, the convolution theorem began to be discovered and used.

In August 2018, Julie Chang et al. from Stanford University proposed an optoelectronic hybrid neural network based on diffractive optical elements [33]. A layer of optical convolution operation is added to the network before the electronic calculation, including a "4f system" composed of two convex lenses with both focal length of f which realizes two cascaded Fourier transforms, as shown in Fig. 4(a). Due to optical convolution, the computation of the whole network is greatly reduced.

In September 2019, researchers from Hong Kong University of science and technology, demonstrated an all-optical neural network (AONN) [34], with tunable linear operation and the optical nonlinear activation function. Figure 4(b) shows the experimental implementation schematic diagram of an optical neuron, the linear operation of which is programmable implemented by the spatial light modulator and the Fourier lens. During the linear operation, the laser spot is used to represent the vector, and the laser beam is divided into different directions by using SLM. The incident light power in different regions of SLM represents different input layer nodes. By

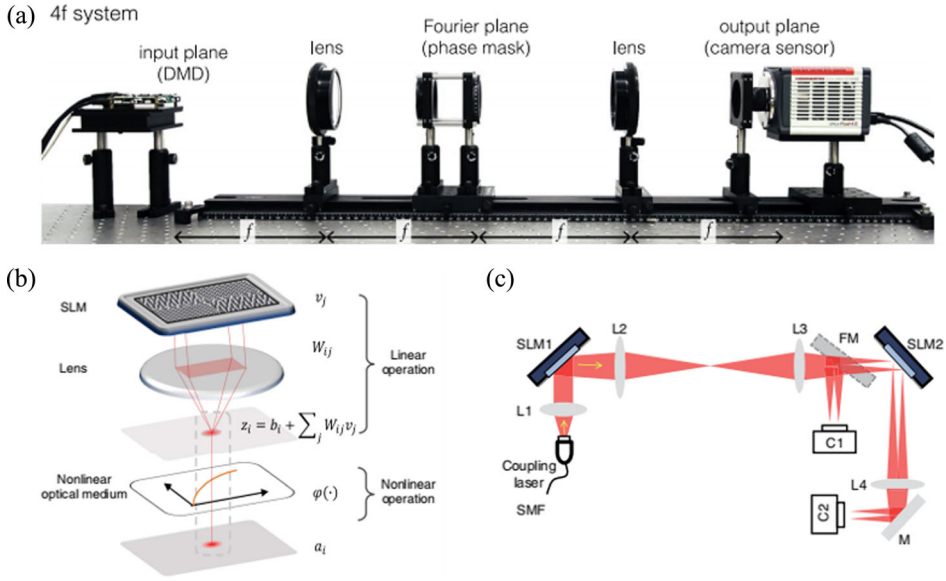

Fig. 4 Optical neural networks with linear operation realized by Fourier transform of light [33]. **a** Optical convolution operation is realized by 4f system. **b** Experimental realization of optical neurons in AONN [34]. **c** The linear operating system of AONN [34]

superposing multiple phase gratings, the incident light will illuminate different directions and have certain weights. Then, the Fourier transform of the lens is used to superimpose all diffraction beams in the same direction onto the points on its front focal plane, so as to realize the linear summation function. The specific linear operating system is shown in Fig. 4(c).

Interference of light to realize linear operation

When multiple beams with the same frequency, same vibration direction and fixed phase difference are superimposed in a certain space, there will be the phenomenon that distribution of light intensity is different from the sum of the original intensity of multiple beams, which is called interference [35]. Interference and diffraction are essentially same, both are superpositions of waves, and the spatial distribution of light and dark is not uniform. However, there are differences between them in terms of forming conditions, distribution rules and mathematical treatment methods. Diffraction is the superposition of numerous small element amplitudes, which is calculated by integration. While interference is a superposition of a finite number of beams, calculated by summation. It can be said that diffraction is a complex interference, and in fact interference and diffraction often go hand in hand. Both interference and diffraction can achieve linear summation.

Shen. Y et al. proposed a new photonic chip system for a new all-optical neural network, as shown in Fig. 5 [36]. The calculation method of the beam in the photonic chip is similar to the basic principle of interference, and the linear operation is realized by a cascaded array with 56 programmable Mach-Zehnder interferometers. The network consists of a cascade of multiple OIUs and ONUs. In OIU, the principle of matrix multiplication is singular value decomposition (SVD). As we all know, any real matrix

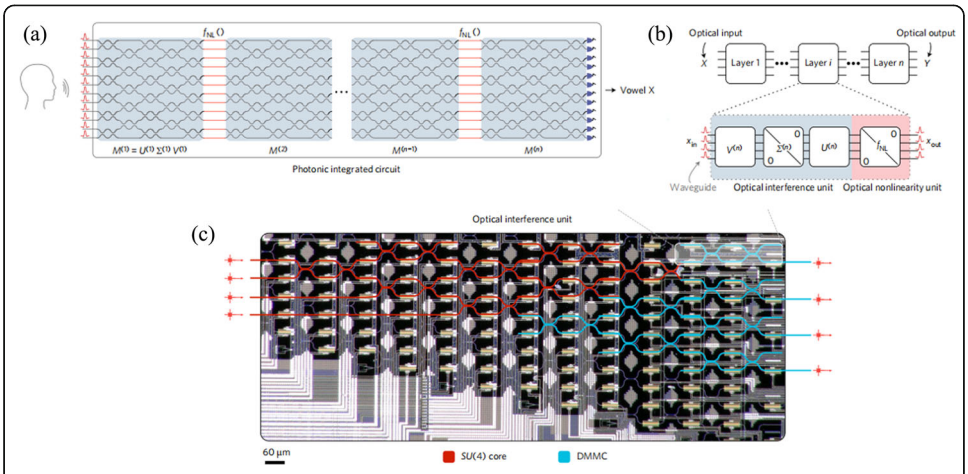

Fig. 5 Using the principle of interference to realize the linear operation in the neural network [36]. **a** Coherent nanophotonic circuit. **b** Each layer of the neural network is composed of optical interference unit OIU and optical nonlinear unit ONU. **c** The internal structure of OIU unit

M can be decomposed through SVD into M = U Σ V †. U, V † can be achieved by optical beam splitter and phase shifter, Σ can be realized by optical attenuator. By tuning the phase shifter integrated in MZIs, you can perform any size of operation on the input. This new method uses multiple beams to propagate and produce interference pattern with using interaction of wave, thereby conveying the desired operation results. In principle, the optical chip with this architecture can run on traditional artificial intelligence algorithm, which is much faster than the traditional electronic chip, with less than one thousandth of the energy.

Scattering of light to realize linear operation

When light meets obstacles or holes, diffraction will occur; When multiple beams of light meet, interference will occur; If light is incident on an opaque surface or random medium, it will be reflected from all aspects by tiny particles, which is known as scattering and was first discovered by scientists in the early 1960s. The so-called scattering [37] is the phenomenon that the spatial distribution, polarization state or frequency of light intensity is changed by the action of molecules or atoms in the propagation medium. The scattering medium is the propagation medium that causes the scattering phenomenon.

In 1990, The random scattering medium has been proved theoretically that it can be used as a thin lens to image the target [38]. In 2007, I. M. Vellekoop et al. from Twente University in the Netherlands verified the I. Freund's point of view experimentally, used feedback control technology to control the spatial light modulator (SLM) to modulate wavefront phase of the incident light in the scattering medium, making wavefront phase distortion compensation caused by optical scattering. As a result, originally chaotic scattering light is focused to the specified location. The technology is the wavefront modulation focusing technology [39]. In 2010, I. M. Vellekoop combined wavefront modulation focusing technology with the optical memory effect of random scattering medium [40], and successfully observed the fluorescent structure located behind the

random scattering medium through scanning imaging [41]. Based on the technology, E. G. Putten et al. achieved the super diffraction limit scanning microscopic imaging of gold nanoparticles by using "'random scattering lens" made of gallium phosphide (GaP) [42], with a resolution of 97 nm. It's the first time to realize the super diffraction limit imaging based on random scattering medium, which has opened a new page in the field of far-field super diffraction limit imaging, and set off a research upsurge of random scattering imaging technology in the world. Next, optical coherence tomography technology [43], speckle correlation imaging technology [44], optical phase conjugation technology [45] and other technologies emerged successively, providing more choices for observing targets through random scattering media such as biological tissues.

With the continuous development of deep learning and scattering imaging, as well as the strong learning ability of deep learning, researchers attempt to combine them to try to make new breakthroughs and development. For instance, the realization of scattering medium target recognition based on the direct machine learning of speckle intensity image [46]. In the experiment, the camera captures the speckle intensity image of the amplitude or phase object on the spatial light modulator, and classifies the acquired face and non-face speckle intensity images by using support vector machine.

In the D^2NN-type network mentioned above, the diffraction modulation layer is similar to the scattering medium. After the light wave passes through, the optical parameters such as the spatial distribution, polarization state will change, and finally the specular pattern with fine-sized particles will be obtained. Moreover, according to the results of computer training simulation, each diffraction pattern is very similar to speckle pattern. Therefore, such a network modulation layer can be analogous to the scattering medium. In fact, we still according to parameters we have learned, design 3D printing layers gratings, etc. Each parameter or pixel on it can be regarded as the neuron in the network, but the number of neurons is limited. The scattering medium is different, its internal disorder dielectric particle assembly can provide thousands of optical computing neurons, even with larger scattering loss.

The deep ONN constructed by active tumor cells in 2018 well embodies this design [47]. It uses a living three-dimensional tumor brain model to demonstrate the morphological dynamics of tumor detected by a trained random neural network through image transmission. Tumor brain cells act as scattering mediators and play a role of hidden layers, and the number of waveguide hybrid nodes, namely neurons, is tens of thousands. In this three-dimensional tumor brain model, each cell is a scattering center with a complex transfer function. By training SLM weights of the input layer, a design using scattering media to construct ONN is presented. The specific structure of ONN is shown in Fig. 6(a). The input layer is realized by a spatial light modulator after iterative training, the middle layer is a three-dimensional spherical layer, and the output is composed of CCD, which detect the intensity distribution.

In fact, whether the diffractive modulation layer of the diffraction network or the scattering ONN of tumor cells, their networks are layered. In previous studies, researchers have found that neural networks need an appropriate number of layers to complete specific tasks, so as to achieve low loss, high accuracy and good performance. If the number of network layers is too few, its training inference ability cannot reach the desired results; If the number of layers is too much, problems of gradient decline and overfitting are likely to occur, resulting in poor results and extremely long training

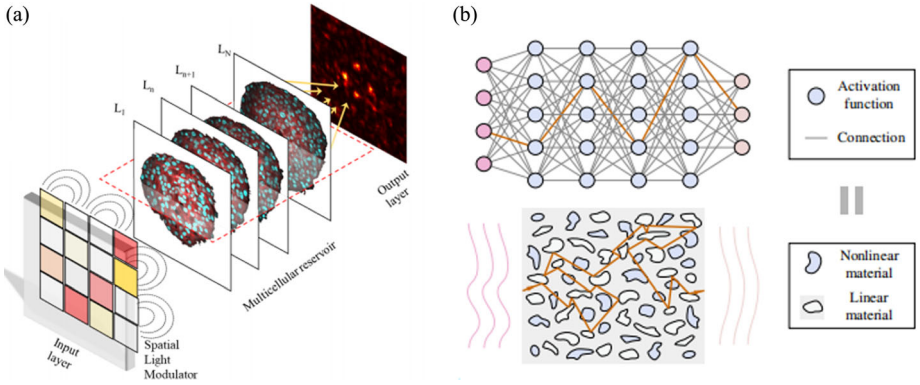

Fig. 6 Optical neural networks using scattering to realize linear operation. **a** Deep ONN constructed by active tumor cells [47]. **b** Nanophotonic neural medium NNM [48]

time. Of course, in the photonic neural network, since the task is completed at the speed of light, we hope that under the premise of ensuring the experimental effect, the number of layers is as much as possible, so that we can train a better network and get more accurate results. Thus, there can be an assumption that exists an optical neural network with an infinite number of layers.

In August 2019, Erfan Khoram et al. designed a new type nano-medium, called nano-photonic neural medium NNM [48], which is composed of matrix material silicon dioxide and a large number of dopants. The dopants may be either pores or materials with different refractive index from the matrix material. A large number of dopants can strongly scatter the incident light in both positive and negative directions. The position and shape of dopants are equivalent to the weight parameters in the traditional neural network. Scattering makes the incident light mix in space, and the incident light contains the information of the input image, which is similar to the linear matrix multiplication in the traditional neural network. NNM is shown in Fig. 6(b), linear materials complete linear matrix multiplication and nonlinear materials complete activation function. This nanostructure, which allows light energy to be redistributed in different directions in space, can be used for computation between neurons and has a stronger expression capability than layered optical networks. In fact, the layered network is a subset of NNM because the medium can be molded into connected waveguides just like a layered network. The light only needs to pass through the scattering medium, which may surpass the previous hierarchical feedforward network and become a very deep neural network, and there is no problem of gradient decrease in the deep neural network. In addition, NNM does not need to follow any specific geometry, so it can be easily integrated into existing visual or communication devices, and will have a wider range of applications.

In 2020, Y. Qu et al. from Oregon State University, inspired by the NNM, proposed an integrated ONN framework based on optical scattering elements based on optical scattering unit [49], taking the network structure of coherent nanophotonic circuits as a prototype, which integrated optical interference unit and optical nonlinear unit. The core structure of the optical network framework is an integrated nano-photonic computing unit——Optical Scattering Unit OSU, which be comprised of a multi-mode

interference (MMI) coupler with a nanometer-pattern coupler region to implement matrix multiplication. It has the same function as the matrix multiplication unit OIU. OSU can be designed as coherent architecture like OIU to realize arbitrary unitary matrix multiplication. Similarly, a more advantageous noncoherent architecture can be designed which directly manipulates the light intensity to achieve random matrix multiplication. In addition, the researchers also realized the optical convolution operation of CNN based on noncoherent OSU. The core of the realization of convolution operation in OSU is to use "kernel matrix" to execute in the photonic circuit, so as to realize the conversion from convolution operation to optical kernel matrix multiplication. The image is divided into blocks and vectorized. By vectorizing and stacking each kernel, the kernel set is converted into a "kernel matrix", so that the one-dimensional image blocks can be effectively multiplied by the "kernel matrix", which is equivalent to the convolution operation. Since nano-imaging makes light scatter within a small region of the coupler and increases the degree of freedom, it can be optimized by an inverse design approach.

Wavelength division multiplexing (WDM) to realize linear operation

Using the principle of diffraction to achieve the optical linear operation, the optical signals propagate in the air. Specific transmission medium, such as scattering medium, can also be used for signal transmission, corresponding to the content in Section 2.2.4; or optical fiber, with wide transmission bandwidth, low transmission loss, strong anti-interference, light weight and low cost, has obvious advantages in light transmission. In optical fiber transmission, at present, WDM is often relied on [50]. WDM can effectively improve the transmission capacity and realize the separation and composition of light. Therefore, optical fiber can be used for the calculation of huge data.

In 2012, Y. Paquot et al. [51] successfully constructed an optoelectronic hybrid serial recurrent neural network based on optical fiber system, whose structure is shown in Fig. 7(a). Signals are injected from an arbitrary waveform generator (AWG) and modulated on light by amplifiers and modulators. The reservoir layer in the middle consists of variable optical attenuators, delay lines, feedback photodiodes, mixers, amplifiers, and a Mach-Zehnder modulator. The photodiode converts the outputs of the system into electrical signals and reads them out. By training and controlling the output weights, the system can realize the recognition of square wave and sine wave. This network can achieve the equalization of communication channel and is the scene expansion of photonic neural network in the field of communication. In the same year, F. Duport et al. also used optical fiber system to construct an all-optical circulating neural network [52], adopting optical fiber delay switching of single nonlinear node for offline training, and its structure was shown in Fig. 7(b). In addition to directly using delay lines to obtain the delay function, devices such as micro-ring array and multimode interference separator array can also realize the delay [55]. At the same time, multistage or more complex time division multiplexing are adopted, which greatly improve the information processing speed and gain better information processing results [55]. In order to explore the multiplexing ability of light, it is verified in [56] that two optical

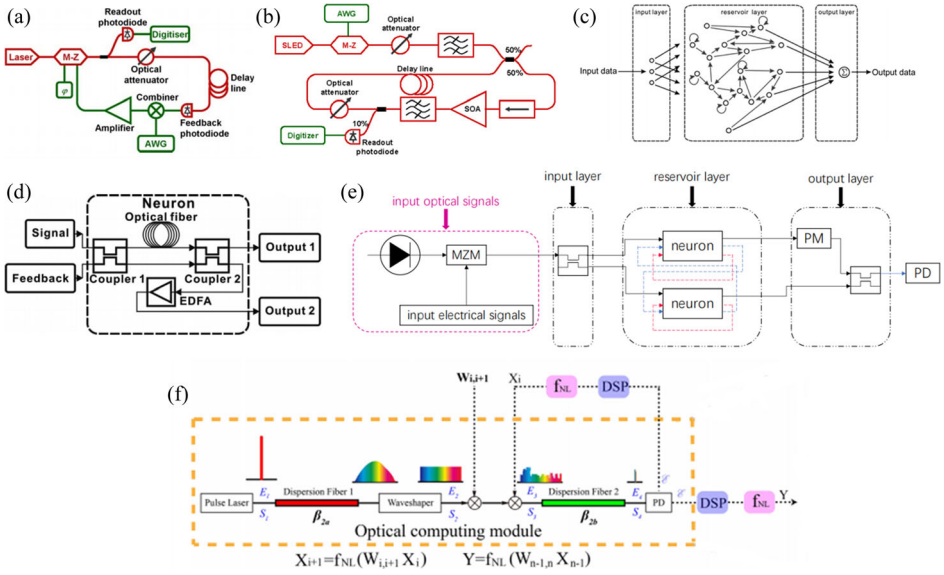

Fig. 7 The neural networks with reservoir computing function realized by optical fiber. **a** Optoelectronic reservoir computing [51]. **b** All-optical reservoir computing [52]. **c**, **d**, **e** respectively describes the principle of reservoir computing, and the structures of the neuron and neural network constructed based on the optical fiber communication system [53]. **f** A serial electro-optic neural network based on time-domain stretching [54]

modes can be used to carry out two independent information processing tasks simultaneously in the same reservoir pool.

Based on the above researches, T. Cheng et al. proposed an optical neural network system for optical reservoir computing based on optical fiber communication system in October 2019 [53]. The RC optical system is composed of input layer, reservoir layer and output layer. The schematic diagram is shown in Fig. 7(c). The input weight matrix of the input layer is W^{in}, which is implemented by the directional coupler. The input layer is used to scale the size of the input data to the size of the reservoir layer corresponding to the matrix W. The reservoir layer is composed of multiple neurons, and its function is similar to the hidden layer in the neural network. Each neuron is composed of two directional couplers, optical fiber and EDFA. The structure diagram is shown in Fig. 7(d). It has two outputs, one of which serves as the output of optical neurons, and the other can be connected back to the feedback of the same optical neurons or to other optical neurons to achieve signal reproduction and interconnection between optical neurons. The specific structure is shown in Fig. 7(e). The output layer has an optocoupler consisting of a Mach-Zehnder phase modulator and a directional coupler to implement the readout matrix W^{out}, which converts the results of the reservoir layer into the output of the RC system. The directional coupler realizes the weight setting among neurons, and the fiber establishes the connection among neurons, realizing the linear operation among neurons together. However, due to the existence of optical fiber, directional coupler and EDFA, such a fiber network is limited in dimension and scale. In order to expand the dimension of the photonic neural network, time can be exchanged for space and scale of the neural network can be expanded in the case that the computing speed is not reduced too much. A serial electro-optical neural network (TS-NN) based on time-domain stretching is proposed by Chen Hongwei's research

group from Tsinghua University [54]. The system structure is shown in Fig. 7(f). This system is a loop system, and n-1 loop times can realize a n-layer network. In each cycle, two operations are involved——linear computation (matrix multiplication) and nonlinear transformation. Linear operation mainly adopts the time stretching method, which make the ultrashort period pulse broadened and flattened by means of dispersive fiber and wavelength converter. Then the weight matrix is used to modulate the processed pulse, and the output of the last time is used as the input of this time to modulate the modulated pulse again, so as to realize the optical multiplication of input vector and weight matrix. Finally, DSP uses signal processing algorithm to process the results. This method realizes the photoelectric hybrid fully connected neural network through the parallel-to-serial scheme, which can realize the large-scale neural network. Although it is not an optical neural network, such an idea of expanding the network scale by exchanging time for space is worthy of our reference.

Whether as optical wave transmission or communication, optical fiber is a very potential development direction in the future. At present, optical fiber has a very mature performance in WDM technology, broadband amplifier technology such as erbium-doped fiber amplifier EDFA, dispersion compensation technology, soliton WDM transmission technology and so on. In the aspect of optical network, the traditional optical networks have realized all-optical among nodes, but electrical devices are still used at network nodes, which limits its development. All-optical network that replaces electrical nodes with optical nodes will be the important development direction of optical fiber in the future. The combination of optical fiber and optical network in 5G era to realize a real all-optical network is an engineering technology that can be further studied.

In the neural network, another typical application of WDM technology is the all-optical spiking neural network in 2019. The working mechanism of neurons used in this network is similar to synapses mechanism of human brain neurons, it can simulate the spike discharge and naturally reflect the actual situation of biological neurons, known as spiking or pulse neuron. The proposal of spiking neurons began in 1997, when W. Maass first proposed spiking neural network [57], which used impulse function to simulate signal as the way of information transmission between neurons. Neuromorphic silicon photonics was put forward by Alexander N. Tait from Princeton University in December 2017, which is the world's first photonic neural network [58], as shown in Fig. 8(a). Each node in the network works under a specific wavelength of light, the light from each node will be detected and summed by total power before it is sent to the laser, then the laser output will be fed back to the node to create a feedback loop with nonlinear characteristics. Such a photonic network can be used to solve differential equations and is demonstrated in which nodes is similar to the trigger mechanism of human brain neurons, called pulse or spiking neurons.

In fact, the neural network mentioned above abstracts the input of the network into matrix or vector, and the neuron mainly performs matrix multiplication operation. Whereas biological neurons process information in the form of impulses, so these networks only retain the structure of neural networks, greatly simplifying the neuron model, which is better described as "units" rather than neurons. In contrast, pulse/spiking neurons are closer to the biological model of human brain neurons, which exist in two states——activated and inactive. They are activated only when their membrane

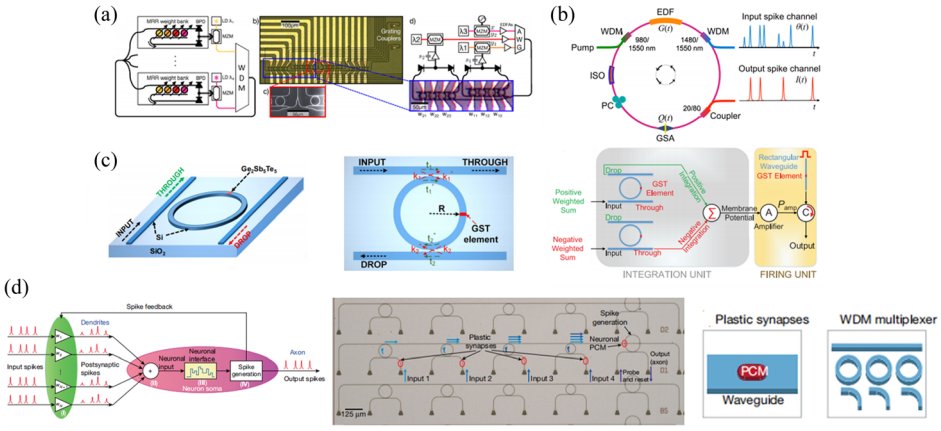

Fig. 8 Optical spiking neural networks realized by WDM principle. **a** The structure of the first photonic neural network [58]. **b** Graphene spiking neuron [59]. **c** Bipolar integration-firing neuron with GST [60]. **d** The all-optical spiking neurosynaptic network structure formed based on the principle of WDM realized by PCM and MRR array [61]

potential reaches a threshold, thus they are not activated in every iteration propagation, a bit like dropout regularization in artificial neural networks. When a neuron is activated, it produces a signal and transmits it to other neurons, raising or lowering membrane potential of its cascaded neurons. In a pulse/spiking neural network, the current activation level of the neuron is usually modeled as some kind of differential equation, and it will rise and continue for a period of time after the arrival of the stimulus pulse and then gradually decline. Spiking neural network enhances the ability to process spatial-temporal data: on the one hand, the neurons in such neural network are only connected with nearby neurons to process input blocks respectively, thereby enhancing the processing capacity of spatial information; on the other hand, because the training depends on the time interval information of pulses, the information lost in the binary encoding can be retrieved in the pulse time information, thus enhancing the processing capacity of the time information. It turns out that pulse/spiking neuron is more powerful computing unit than traditional artificial neuron, which is a major development trend in the future. However, owing to the difficulties in the training method and hardware implementation of pulsed neural network, it has not been widely used yet, and most of the researches about pulsed neural network is still focused on the theoretical research and the verification of simple structure. But more and more researchers are now devoting themselves to training algorithms and hardware (optical) implementations of pulsed neural networks.

In 2016, Prucnal's research team in the Princeton University proposed a spike processing system based on the activated graphene fiber laser, in which the activated graphene fiber laser plays the role of spiking neuron [59], as a basic component of spike information processing. In 2018, a neural mimicry photonic integrated circuit based on distributed feedback (DFB) laser structure was proposed [62]. The laser has two photodetectors, which can generate both inhibitory and excitatory stimuli at the same time. The system is compatible with the broadband-and-weight (B&W) protocol [63]. In the same year, superconducting photoelectric spike ring neurons were designed, known as Loop Neurons [64]. These neurons are composed of single-photon detectors, Josephson junction and light-emitting diodes. Josephson junction detects event integration and

converts it into supercurrent, and finally store it in the superconducting circuit. Also in 2018, a new spiking neuron equipment——a bipolar integration-firing neuron [60] was introduced, including integral unit that consists of two double bus ring resonator with embedded phase change materials (PCM) (i.e., $Ge_2Sb_2Te_5$ (GST)), which controls the propagation in the loop, sums the output of the resonator, and is used to stimulate the ignition unit that is composed of a photon amplifier, a circulator and a rectangular waveguide with a GST component on the top. This spiking neuron can be connected with photon synapse to form an all-photon spiking neural network. We show graphene spiking neurons and bipolar integration-firing neurons in Fig. 8(b) and (c), respectively.

Based on this, Feldmann J et al. mentioned an all-optical spiking synaptic realization using PCM in May 2019 [61]. The structure of this photonic neural network is shown in Fig. 8(d). It's a fully connected network. When inputting the pulse, the PCM unit on the waveguide is used for weighting in each neuron, and the MRR array is used as WDM for summation; The spiking mechanism is implemented by the PCM on the ring resonator. The PCM crystal is a special unit in which has two states, crystalline and amorphous, have different effects on the input pulse. Because of this, PCM can modulate the pulse and realize the weighting operation. For amorphous PCM cells, the synaptic waveguide is highly transmissive and can achieve strong connections between neurons. In the crystalline state, most of the light transmitted to the PCM is absorbed, leading to weak connections among neurons. After the pulses are weighted by PCM, they are integrated and sent together by WDM to a ring resonator integrated with PCM cells to realize the summation. In this way, the linear operation in ONN is achieved.

The optical realization of nonlinear activation function

In the previous chapter, we discussed the optical realization of linear operations in neural networks, but it is not enough to have linearity in neural networks. It also requires the processing of nonlinear activation functions, similar to the function of synapses in the brain nervous system. Nonlinear function can accelerate the convergence speed of the network and improve the recognition accuracy, which is an indispensable part of the neural network. Without it, no matter how much the network layer is, it can be attributed to a huge linear operation, however, most problems are nonlinear. The introduction of the activation function provides the nonlinear factors for neurons, making the neural network approximate any nonlinear function, so that neural network can be applied to many nonlinear models.

In the electronic neural network, we can use the existing nonlinear activation function, or define a function to carry out the nonlinear operation. However, in the photonic neural network, this becomes a bottleneck for its development. The reason is that nonlinear optical components need to match the high-power laser, which is more difficult to realize nonlinear functions than electronic devices, and the nonlinear functions realized by them have many non-ideal characteristics. In 1967, Seldon et al. proposed a saturated absorber model or an electronic module [65] to realize nonlinear operations in the photonic neural network, but this method is difficult to accurately control and requires the conversion of optical signals into electrical signals through photodiodes, thus reducing the computing speed. At present, there are two ways to realize the nonlinear operation in the photonic neural network: one is to use the electronic or

photoelectric methods, and the other is to use the nonlinear effects of some special materials. In the following chapter, we will first describe nonlinear optical effects in detail, and then introduce different activation implementations and corresponding optical neural networks according to different effects.

Nonlinear optical effect

Nonlinear optical effect is the effect caused by the nonlinear polarization of the medium under the action of strong light, which originates from the nonlinear polarization of molecules and materials, and is manifested as the nonlinear relationship between the effect of light on the medium and the response of the medium [66]. Under the action of incident light field, the motion state and charge distribution of the atoms, molecules or ions that make up the medium must change in a certain form to form an electric dipole, generate an electric dipole moment and then radiate a new light wave. In this process, the electric polarization intensity vector P of the medium is an important physical quantity. P has a nonlinear relationship with incident light vector E:

$$P = \varepsilon_0 \chi^{(1)} E + \varepsilon_0 \chi^{(2)} E^2 + \varepsilon_0 \chi^{(3)} E^3 + \cdots$$

where $\chi^{(1)}$ 、 $\chi^{(2)}$ 、 $\chi^{(3)}$ respectively referred to the first order (linear), second order and third order (nonlinear) polarizability of the medium. The studies showed that $\chi^{(1)}$ 、 $\chi^{(2)}$ and $\chi^{(3)}$ were reduced in turn.

In the case of ordinary incident light, the second- or higher- order electric polarization intensity can be ignored, the medium only shows linear optical properties, and its electric polarization intensity P has a simple linear relationship with the incident light field intensity E. while is incident with a strong monochromatic laser, the order of magnitude of the light field intensity E can be compared with or close to the average electric field intensity $|E_0|$ within the atom. The contribution of the second-order or third-order electric polarization intensity cannot be ignored; the electric polarization intensity P and the incident light field intensity E show a power series relationship, and the nonlinear optical effect occurs at this time.

There are many kinds of nonlinear optical effects, which can be divided into second-order, third-order and higher-order nonlinear optical effects according to the relationship between electric polarization intensity and electric field. Of course, we generally only study second-order and third-order nonlinear optical effects. According to the interaction mode between laser and medium, that is, whether there is energy exchange between them, it can be divided into active nonlinear optical effect and passive nonlinear optical effect; according to the changed parameters, it can also be divided into optical frequency conversion effect, optical nonlinear absorption, optical Kerr effect and self-focusing, optical bistability effect, optical phase conjugation effect, stimulated scattering effect, etc.

Implementation of nonlinear activation in photonic neural network

In the current studies on photonic neural networks, we find that optical nonlinear activation does not exist in some networks or is simulated electronically. For example, in diffraction network D^2NN, there is no activation function. In the serial photonic neural network based on time-domain stretched, its nonlinear transformation is realized by nonlinear functions such as non-negative s-type function simulated by the electronic

devices in the system. There are also some networks taking advantage of nonlinear effects for nonlinear activation designs. At present, saturation absorption, optical bistability and Kerr effect have been considered as potential activation functions in ONNs.

Nonlinear optical absorption

Optical absorption means that when a photon enters a medium, atoms and molecules absorb the energy of the photon and happen the energy level transition [67]. In this process, if the photon energy is strong enough, the absorption coefficient of the medium will change with the light intensity. The change can be linear or nonlinear, namely linear and nonlinear optical absorption, respectively. The two main optical mechanisms of nonlinear optical absorption are saturation absorption, anti-saturation absorption and two-photon absorption.

When the laser is incident into the medium, the absorption coefficient of the medium decreases with the increase of the light intensity in the medium. When the intensity of the input light wave exceeds the threshold value, the absorption property of the medium begins to become saturated. This nonlinear optical behavior is called saturable absorption. Saturation absorption is caused by the transition of the particles constituting the medium from the ground state level to the first excited state level. In the case of saturated absorption, the relationship between the absorption coefficient of the medium and the light intensity I in the medium can be expressed as: $\alpha(I) = \frac{\alpha_0}{1+I/I_c}$, as shown in Fig. 9(a). Correspondingly, the relationship curve between transmittance and light intensity is opposite, similar to the curve of sigmoid function, as shown in Fig. 9(b).

On the contrary, anti-saturated absorption is the effect of increasing the absorption coefficient with the increase of light intensity. Its absorption characteristic curve is somewhat similar to the sigmoid function curve, and it is not commonly used in nonlinear activation. Two-photon absorption, as the name suggests, is that an atom in a medium absorbs two photons at the same time, then goes from the ground state to the excited state. When two light beams with frequencies ω_1 and ω_2 pass through a nonlinear medium, if value of $\omega_1 + \omega_2$ is close to a certain transition frequency in the medium, the two beams will attenuate at the same time. Two-photon absorption is a third-order nonlinear optical effect.

Optical nonlinear absorption can be realized by both optoelectronic devices and optical methods. Saturation of optoelectronic devices such as optical attenuation amplifiers, erbium-doped fiber amplifiers, and semiconductor optical amplifiers, etc. can be used as nonlinear activation. In the all-optical reservoir computing implementation based on semiconductor optical amplifier array on the chip proposed by F. Duport et al., the saturation gain effect of semiconductor optical amplifier is utilized to realize the network nonlinear function. In the reservoir computing based on the optical fiber communication system, each neuron is composed of two directional couplers, optical fiber and erbium-doped fiber amplifier, while the erbium-doped fiber amplifier realizes the nonlinear function and each neuron has such nonlinear activation function. In terms of optics, saturable absorbers, such as optical dyes, graphene, C60, etc., can be used to play the role of nonlinear activation. In the 2014 optical reservoir computing [69], graphene saturable absorbers or two-photon absorption [70] were used as optical

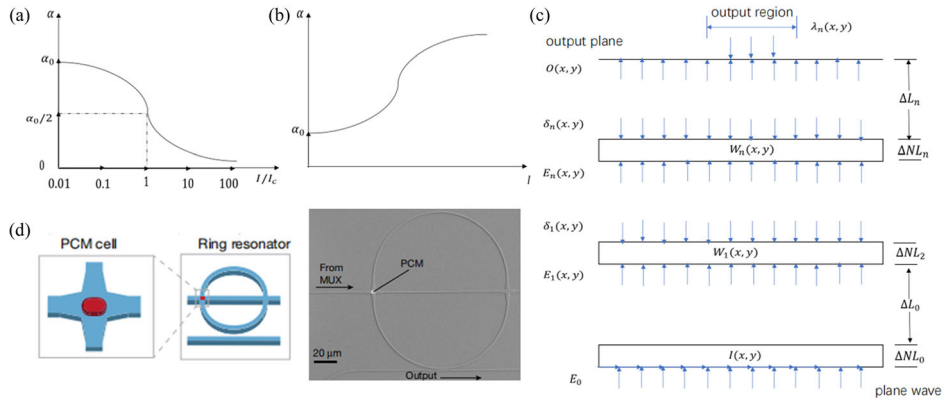

Fig. 9 **a**, **b** Absorption coefficient and transmittance curves of saturation absorption. **c** The structure of Kerr-type network [68]. **d** Phase change material PCM realizes nonlinear activation [61]

nonlinear units. In 2016, the Prucnal's research group in the Princeton University proposed a spike processing system based on an activated graphene fiber laser, which also uses graphene as a saturated absorber to perform nonlinear activation functions. In the coherent nanophotonic circuit in 2017, its nonlinear unit ONU is realized by a saturation absorber that can be integrated into the nanophotonic circuit, such as fuel molecules, semiconductors, graphite saturated absorbers and saturation amplifiers. For the incident light I_{in}, emergent light I_{out} is given by nonlinear equations: $I_{out} = f(I_{in})$, using the model of saturated absorber is $\sigma\tau_s I_0 = \frac{1}{2}\frac{\ln(T_m/T_0)}{1-T_m}$. Once I_0 is given, $T_m(I_0)$ can be solved by this formula, and the emission intensity can be obtained by $I_{out} = I_0\ T_m(I_0)$. In 2020 Nanophotonic media network system, the nonlinear activation function is also achieved by making dopants composed of dye semiconductors or graphene saturable absorbers. These dopants can perform distributed nonlinear activation, which mainly reflects the ReLU function, allowing signals with an intensity higher than the set threshold to pass and obstructing signals below the threshold.

Optical Bistability

When the light beam passes through the optical system, there is a nonlinear relationship between the incident light intensity and the transmitted light intensity, thereby achieving the optical switch. For instance, optical restriction, optical bistability, various interference switches and so on. In electronics, bistability is a unit circuit that has two different resistance values for the same input electrical signal. In photonics, bistable state is an optical element, which has two transmittances with different levels for the same incident light intensity, which is called optical bistable state. It is of great significance for understanding the storage, operation and logical processing of optical information.

In a nonlinear optical system, when the input light intensity is small, the output light intensity of the system is also small. When the input light intensity increases to a certain critical light intensity value, the output light intensity of the system will jump to a certain high light intensity state, as if a switch is turned on. After that, if the input light intensity is further reduced, the system will no longer return to the low light intensity state at the original critical value, but there will be another critical value at the lower

light intensity, making the system jump from high state to the low state. In this process, the "hysteresis" phenomenon appears in the input-output transfer relationship in the optical system, similar to the hysteresis loop in electromagnetism.

Optical bistable equipment may be used in high-speed optical communication, optical image processing, optical storage, optical limiter and optical logic elements. In particular, optical bistable devices made of semiconductor materials, with the characteristics of small size, low power and short switching time (10–12 s) and so on, are likely to become the logic components of optical computers in the future. Optical bistability has become a very active research field because of its great potential application value.

In reservoir computing, in addition to the saturated absorption being used to design as the activation function, we can also combine the bistability [71] with ring resonator according to the characteristics of bistability, to realize nonlinear activation structure of neural network [72]. This point is reflected in [73]. In coherent nanophotonic networks, the nonlinear activation function of ONU element can be realized by bistable nonlinear effect in addition to saturation absorption.

Optical Kerr effect

Kerr effect [74], is the third-order nonlinear effect. Under the action of electric field, the refractive index $n_{//}$ and $n_{\perp}$ of polarized light waves along parallel and perpendicular to the electric field direction change differently in the medium, and the difference Δn between them is proportional to quadratic power of the electric field, resulting in induced birefringence. Generally, the applied electric field is a direct current or low frequency alternating electric field. If the light/optical frequency electric field replaces the applied electric field, the same phenomenon will occur when the light is strong enough. At this time, Δn is proportional to the intensity of laser beam acting in the medium, where Δn is a nonlinear phase shift, which is called optical Kerr effect. If the parameter to be optimized is the phase, the optical Kerr effect can be used to realize nonlinear activation. Aiming at the nonlinear of Kerr medium, S.R. Skinner proposed an innovative structure of all-optical neural network using Kerr-type nonlinear optical materials in 1994 [68]. Figure 9(c) depicts the all-optical feedforward artificial neural network structure with Kerr media, which uses thin material layers separated by free space to realize weighted connection and nonlinear neuron processing, that is, the network consists of thin layer of nonlinear medium and thick layer of linear medium, namely free space. Linear layer in which light propagates to realize weight connection; nonlinear optical layer, used as a weight layer except the first one is the input layer, and performs nonlinear processing. Hence, there are two formulas as follows: 1. $E_{i+1}(\beta) = \frac{jC_i}{\pi}\int_{\Omega_i} F_i(\alpha)\, e^{-jC_i(\beta-\alpha)^2} d\alpha, where = C_i = \frac{k_0}{2\Delta L_i}$, which describes transmission of light from the coordinate $\alpha = (x, y)$ at the beginning of the i-th layer to the coordinate $\beta = (x^*, y^*)$ before the nonlinear layer of (i + 1)-th layer; 2. $F_i(\alpha) = E_i(\alpha) e^{-jk_0\Delta N L_i n_2(|\Gamma_i(\alpha)|^2+|E_i(\alpha)|^2)}, where\ \Gamma_0(\alpha) = I(\alpha), \Gamma_{i>0}(\alpha) = W_i(\alpha)$, describing the effect of the nonlinear layer, $E_i(\alpha)$ is the light entering the i-th nonlinear layer at the coordinate of $\alpha = (x, y)$.

Such a hierarchical network can not only process forward computation signal, but also realize the error backward propagation. This nonlinear method has advantages over other optical implementations because of the fast response speed of the Kerr-type

nonlinearity in the material, and the network can be proved to be simple. Other optical networks usually require separate and specific optical hardware for weighted connections and neuron processing.

Taking into account of the fast response of Kerr nonlinear materials and the third-order nonlinear optical effect of two-photon absorption, the researchers combined the Kerr effect with two-photon absorption to establish a nonlinear mechanism and used it in conjunction with the InGaAsP ring resonator to realize all-optical reservoir computing [75].

In the example of quantum optical neural network (QONN) architecture proposed by G. R. Steinbrecher in 2019 [76], the input is the single-photon Fock state, the unit nonlinearity is assigned to the Kerr-type interaction, and the quadratic phase is applied to the number of photons. The readout is provided by the photon number resolution detector, which measures the number of photons in each output mode. The single-mode Kerr interaction achieves photon coherence nonlinearity.

Other nonlinear activations

In 2018, R. Amin and J. George et al. pointed out that electro-optic absorption modulation could realize nonlinear modulation of light waves [77–79], discussed the method of mapping nonlinear activation function to transfer function of electro-optic modulator, and also pointed out that different functional activation functions could be implemented by making use of different electro-optic materials. For example, the ReLU function can be realized by utilizing the inverted filling light absorption mechanism of quantum dots (QD) [80].

In the same year, M. Iscuglio et al. designed a kind of optical nonlinear, which depends on the reversible transparent sensitivity caused by Fano resonance in the plasma oscillator subsystem and the nonlinear response of Buckyball (C60) membrane——anti-saturation absorption [81], realizing the fast and effective all-optical nonlinear, improving the throughput of the neural network, and reducing the delay and power consumption.

In May 2019, in the paper called “an all-optical spiking neurosynaptic network” on the “Nature”, it is mentioned that the construction of the network with the help of phase change material nonlinear PCM many times [61]. PCM combines with MRR achieve weight modulation, and integrates with ring resonator to realize peak function as nonlinear activation function. If the power of input pulse summation exceeds a certain threshold, the state of the PCM will change, producing peak/impulse. Otherwise, the probe pulse resonances with ring resonator, which is similar to the nonlinear response represented by ReLU function.

In addition, the two-layer AONN [34], designed by the Hong Kong team, proposed a special nonlinear activation function based on electromagnetic induced transparency (EIT)——a photo-induced quantum interference effect between atomic transitions, in laser-cooled atoms with electromagnetic induced transparency. The EIT nonlinear optical activation function is implemented by laser-cooled ^{85}Rb atoms in the dark-line two-dimensional magneto-optical trap (MOT), as shown in Fig. 10(a). The atomic energy level is shown in Fig. 10(b). Atoms are prepared in the ground state $|1\rangle$. The output beam after linear operation——the circular polarized coupled beam ω_c resonates

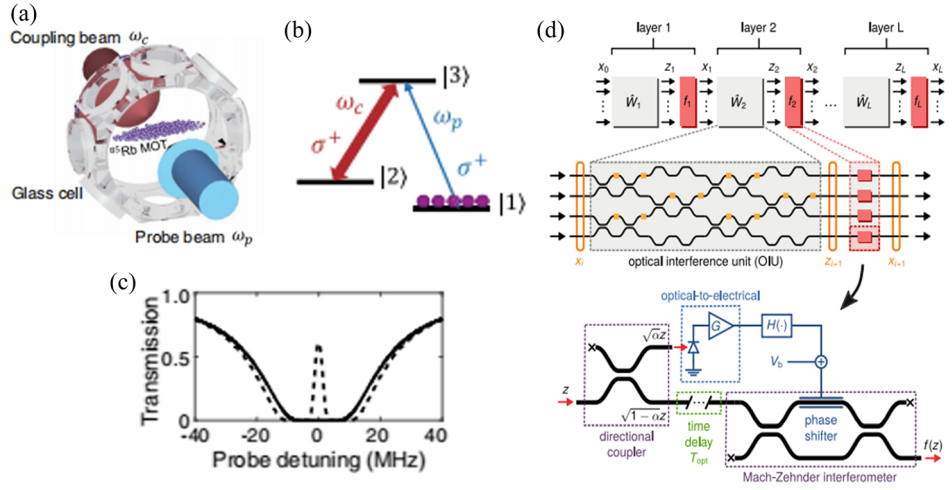

Fig. 10 **a**, **b**, **c** represents the preparation, energy level diagram, and transmission diagram of laser-cooled ^{85}Rb atom, respectively [34, 82]. **d** An electro-optical nonlinear activation function structure can be used to realize the optical nonlinear element ONU [83]

with $|2>$ $|3>$, incident to the electron cloud transversely, and back-propagation probe beam ω_p resonates with $|1>$ $|3>$. In the absence of a coupled beam, the atomic medium is opaque to the resonant detection beam, which is absorbed to the maximum extent by the atom as shown in the implementation of transmission spectrum in Fig. 10(c). In contrary, in the presence of a coupled beam, quantum interference between the transition paths leads to the EIT [82] spectral window, as shown in the dashed curve in the figure. The resonance peak transmission and bandwidth are controlled by coupling laser intensity, and the output of the resonance probe laser beam can be expressed as: $I_{p,out} = I_{p,in} e^{-OD\frac{4\gamma_{12}\gamma_{13}}{\Omega_c^2+4\gamma_{12}\gamma_{13}}} = \phi(\Omega_c^2)$. It can be seen that the probe output beam is a nonlinear realization of the coupled nonlinear input beam.

In January 2020, a nonlinear activation function structure of the optical neural network was proposed [83], which achieves the optical-to-optical nonlinearity by converting a small part of the optical input power into voltage. As the original optical signal passes through the interferometer, the remainder of it is modulated by the phase and amplitude of this voltage. For the input signal with an amplitude of z, the resulting nonlinear optical activation function $f(z)$ is the response of the interferometer under modulation and the result of elements in the electrical signal path. The schematic structural diagram is shown in Fig. 10(d). In addition, he demonstrated another implementation of the activation function, which could include a nonlinear MZI in which an arm has a material with a Kerr nonlinear optical response. Two different kinds of implementation methods were also conducted experimental demonstration analysis and comparison, highlighting that the lower activation threshold can be achieved by the electro-optic activation structures.

Training, experimental demonstration and analysis

For a neural network, training is a crucial and indispensable step, which affects the performance of network. The process of training is to calculate the target loss function by gap between the network output and the actual output and make it

minimize to optimize the network parameters and achieve the effect of network convergence. When the network performs the prediction task finally, the desired results can be achieved. In electronic neural networks, training is divided into two categories: supervised learning and unsupervised learning. In the process of optimizing parameters, we can make use of back propagation, adopt gradient descent, Adam, momentum and other methods to minimize the cost function of the network. Since training involves gradient calculation and even more complex calculation, how to train the network is a difficult and important step in optical neural network. At present, training in most of ONNs is implemented through software to obtain weight parameters, so as to complete inference in ONN architecture. However, the training in the electrical field has shortcomings of pertinence and dependence. We can customize the training methods according to the optical architecture of ONN, making full use of the photonic technology, although it will be more complicated. Next, training methods in ONNs will be introduced through different training or gradient calculation methods.

Backpropagation algorithm

Backpropagation algorithm is a learning algorithm suitable for multi-layer neural networks, based on gradient descent method. The main idea is: after ANN has completed the forward propagation process, the error between the estimated value and the actual value of the network is calculated, and the error is back propagated from the output layer to the hidden layer until it is propagated to the input layer; In the process of back propagation, the values of various parameters are adjusted according to the errors, and the above process is iterated continuously until the network converges. In the training of D^2NN type network, the function of back propagation algorithm is greatly exerted.

In the all-optical machine learning D^2NN structure [24], the output layer will have a photoelectric detector array to detect the output light intensity, which makes difference with the target light intensity. The loss function is defined by mean square error, with the help of back propagation algorithm and stochastic gradient descent to update the amplitude or phase of the entire network. This is the training process of this network, which is completed on the electronic computer. And then, the trained parameters of each layer are modeled and 3D printed out. Finally, the light source, the fabricated 3D diffraction modulation layer and the detector array are used to construct the photonic neural network for reasoning and prediction. In order to test the inference ability and performance of the network, the author carried out experiments on the MNIST dataset and the Fashion-MNIST dataset, and reached a high accuracy on the network structure which had designed the 5-layer D^2NN and increased the number of diffraction layers on this basis. The specific experimental results are shown in Table 1. Later, the researchers of the research group analyzed in detail the architecture of the diffraction neural network and different parameter designs, and used five phase-only diffraction modulation layers for handwriting number recognition and fashion product recognition, achieving 97.18% and 89.13% recognition accuracy respectively. In addition, the influence of learning loss

Table 1 Experimental Results for D^2NN

	five layers	seven layers	ten layers
MNIST	91.75%	93.39%	–
Fashion-MNIST	phase-only 81.13% complex-valued 86.63%	–	86.60%

function on the performance of optical neural network and the mitigation of gradient disappearance in error back propagation are also analyzed [84].

Replacing 3D layers with diffraction gratings also realized the network training by the above method [25]. Finally, optimized parameters were used for grating design, and the corresponding diffraction grating was etched through semiconductor processing technology. The phase values of neurons have the following relationship with ladder thickness of etched diffraction grating, that is, the height of Ge: $\Delta z = \frac{\lambda\phi}{2\pi\Delta n} = 0.5618\phi$. In order to train and test the D^2NN classifier, MNIST dataset was used for experiments, with obtaining higher recognition accuracy. Table 2 shows the specific experimental results.

In the neural network system which implements logical operation, metamaterial is used as the diffractive modulation layer [26]. Each layer of metasurfaces is composed of the array of scatterers, the size of which can control the amplitude and phase of the scattered light. The above network architecture is still used for training in the same way, and then the parameters trained are converted to the size of the scatterer, to modulate the amplitude or phase of transmitted light after each layer. At First, three basic logical operations, NOT, OR, and AND are experimentally demonstrated, and the accuracy can reach 100%. Then, a three-layer phase-only diffraction neural network is used to realize all seven optical logic gates in an optical system. By calculating the intensity distribution of two specified areas, the accuracy is still satisfactory. In addition, the team proposed a possible scheme for cascading optical logic gates and pointed out that expect for multilayer metasurfaces, there could be other platforms to promote optical logic gates, such as metamaterials/nanophotonics. As shown in Fig. 11(a).

From the above description, it can be concluded that for deep diffraction network D^2NN, computer learning and training of network hyperparameters, no matter using 3D printing diffraction layer, matrix grating or metasurface, are identical in physical essence; and the network parameter training is all completed by computer, using the same set of complete architecture. Besides, once the task you want to accomplish

Table 2 Experimental Results of Diffraction Grating Network System

	MNIST data set, recognition of number 7	
Number of neurons (10 mm,5 layers)	100*100	85.75%
	200*200	88.2%
Layers of network (200*200 neurons, 10 mm)	2	74.1%
	3	79.1%
	5	88.2%
Distance between layers (200*200 neurons, 5 layers)	5 mm	74.1%
	10 mm	88.2%
	20 mm	79.1%

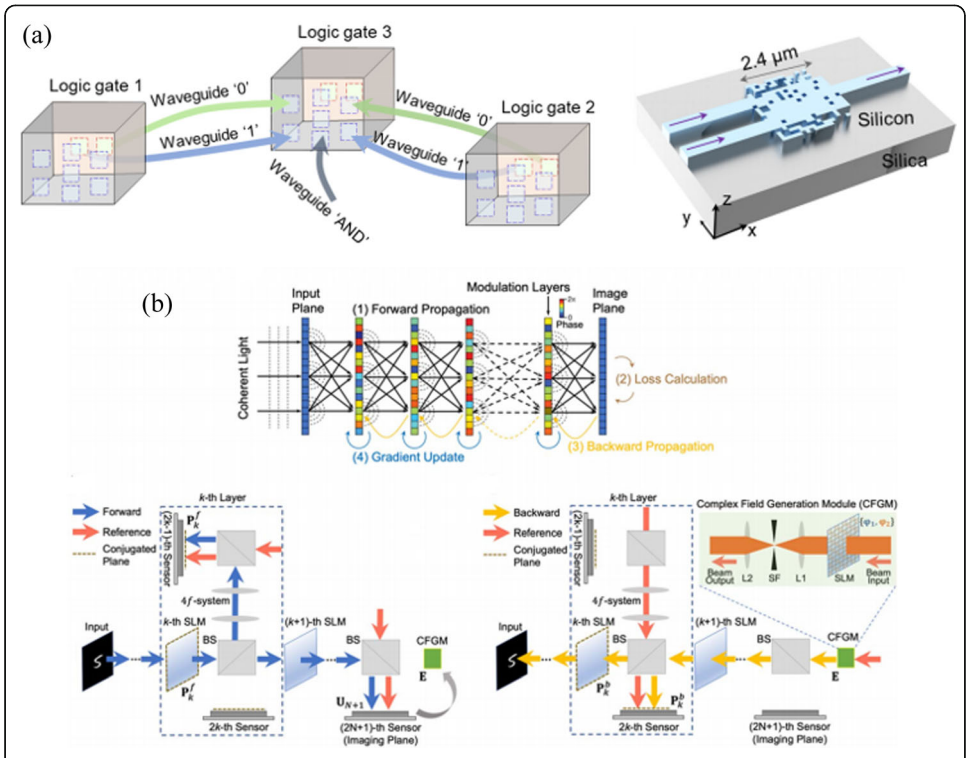

Fig. 11 a Possible solutions for optical logic gate cascading and other possible implementation platforms [26]. **b** Describes a SLM cascade neural network, and uses 4f system and SLM to realize optical field measurement, thereby carrying out optical training of diffraction ONN [85]

changes, the network needs to be trained again. According to the computer configuration parameters given in the paper, it will spend a long time training the network and remake the diffraction layers. These operations will consume a lot of time and resources.

In June 2020, Dai Qionghai team proposed a SLM cascaded neural network, which uses 4f system and SLM to achieve optical field measurement, and utilizes the error measurement module to realize network training [85], as shown in Fig. 11(b). The cascaded SLM is used as the hidden layer. Based on the principle of light reciprocity and phase combination, the gradient of the loss function relative to the weight of the diffraction layer is accurately calculated by measuring the forward and backward propagating light fields. The high-speed spatial light modulator is then programmed to update the diffraction modulation weight to minimize the error between the prediction and the target output, and perform inference tasks at the speed of light. This study not only realizes the SLM diffraction modulation layer, but also has one of the biggest features. In the realization of the back propagation algorithm, that is, it uses optical methods to carry out the back propagation algorithm to train linear and nonlinear diffractive optical neural networks in situ, thereby speeding up the training speed and improving the energy efficiency of the core computing module. Therefore, it not only realizes the optics of the network structure, but also realizes the optics of the training process and real-time programming.

In addition to the diffraction network, in the training of coherent nanophotonic circuit, the author also used the traditional back propagation algorithm and stochastic

gradient descent method to update the parameters, and constructed a two-layer fully connected neural network for speech recognition experiment. The recognition accuracy is only 76.7%, while equivalent electronic neural network can achieve accuracy of 91.7%. There is still much room for improvement in this method.

Forward propagation on Chip

Despite the backpropagation algorithm is widely used, and it is currently the most commonly used and most effective algorithm for training artificial neural networks. However, for some ANNs, when the number of effective parameters greatly exceeds the number of different parameters, especially RNNs and CNNs, the use of backpropagation for training is notoriously inefficient. To be exact, due to the recurrent nature of RNN, ANN becomes an extremely deep neural network, which the depth of the network is equal to sequence length, hence the problem of gradient disappearance is more common and especially serious. Meanwhile, in the CNN, the parameter sharing method of extracting features by using the same weight parameters repeatedly in different parts of the image runs through the whole network.

In addition to using back propagation for training in the coherent nanophotonic circuit, the research team has also proposed a way to directly obtain the gradient of each different parameter by only using forward propagation and finite difference methods on ONN [36]. The method of obtaining the gradient can be specifically described as the following process: first of all, calculating two forward propagation steps $J(W_{ij})$ and $J(W_{ij} + \delta_{ij})$ in a constant time, then calculating $\Delta W_{ij} = (J(W_{ij} + \delta_{ij}) - J(W_{ij}))/\delta_{ij}$, that the gradient of different weighting parameters ΔW_{ij} can be acquired only by forward propagation.

Of course, this kind of on-chip forward propagation method is essentially a simple finite difference method. Although it is simple in form and convenient in use, it requires to carry out a forward propagation for each independent parameter, including two calculations of loss function and one calculation of division. When there are many parameters, the efficiency is very low.

In-situ Back propagation and Adjoint method

As an all-optical neural network, the coherent nanophotonic circuit mentioned above, whose linear operation and nonlinear activation can be effectively completed by optical path, has good forward propagation speed and power efficiency, and has a good development prospect. Its training can either use the traditional back propagation algorithm or only using the forward propagation to directly train the neural network on the photonic chip, so as to realize the programmable optical neural network. However, Mach-Zehnder interferometer, directional coupler and phase modulator occupy a large space, so it is difficult to construct the optical network with more than 1000 neurons. In addition, due to the precision encoding phase, phase shift between thermal crosstalk and optical detection noise of MZI and other factors, the identification accuracy cannot reach the expected effect, and the accuracy is far lower than the equivalent electronic neural network. Therefore, this kind of inefficient training method cannot be applied to the neural network based on integrated photonic platform, and it is difficult to achieve the goal of large-scale, fast, programmable and high-precision photonic neural network.

In 2018, Tyler W. Hughes et al., from Stanford University, proposed a method of training neural network efficiently and locally to obtain parameters of optical path in backward propagation through the method of adjoint variables, which is similar to the means of calculating gradient in the common neural network [86]. Moreover, these gradients can be obtained by measuring the strength of the device.

As shown in Fig. 12(a), the transmission matrix W between the input and output of each layer is determined by the dielectric constant ε_l of the phase shifter of that layer. Using the mean square error (MSE) as the loss function L of the system, we first calculate derivative of the dielectric constant ε_l of the last layer corresponding to the loss function, and then compute recursively the gradient of each layer by the chain rule. The next is to calculate the gradient by electromagnetic adjoint variable method. The derivative of dielectric constant ε_l corresponding to the loss function of can be expressed as another form including the original quantity oj and the adjoint quantity aj: $\frac{dL}{d\varepsilon_l} = k_0^2 R\{\sum_{r\in r_\phi} e_{aj}(r)e_{oj}(r)\}$.

The last term in the intensity pattern due to the interference of e_{og} and e_{aj} is the amount needed to calculate the gradient: $I = |e_{og}|^2 + |e_{aj}|^2 + 2R\{e_{og}e_{aj}\}$, thus as long as e_{aj} in OIU can be generated, the measurement of gradient can be achieved simply by measuring light intensity. Figure 12(b) shows the experimental method for measuring the gradient. First, in step (1), input the original field X_{l-1} in the forward direction and record the intensity at each phase shifter, i.e. $|e_{og}|^2$. Then, input the difference between the actual output and the ideal output in step (1) in the reverse direction and record the intensity of each phase shifter, namely $|e_{aj}^*|^2$. In step (2), reverse output is a time-reverse adjoint field, which can be calculated by $X_{TR}^* = \hat{W}_l^T \delta_l$. As shown in step (3), when inputting the original field and the time-

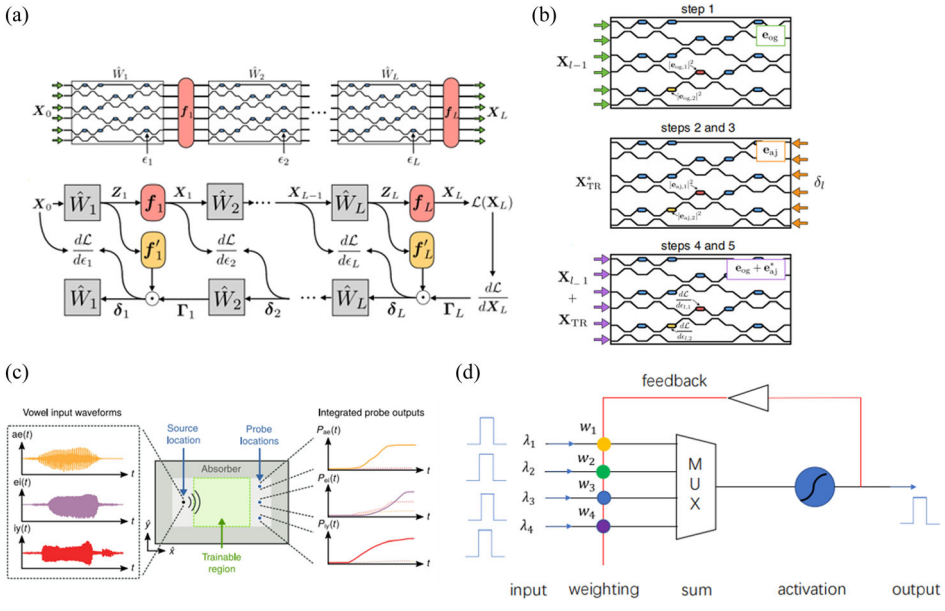

Fig. 12 a, **b** A kind of efficient and locally raining method for neural network [86]. **c** Wave physical simulation of recursive network layout [87]. **d** Neuron pattern with unsupervised learning [61]

inverse field at the same time, interference occurs, and record the intensity of each phase shifter, subtract the constant intensity term in steps (1) and (2) and multiply with them to obtain the gradient.

Furthermore, in situ backpropagation algorithm and adjoint method for gradient measurement are also used in the training of NNM [48], but its training parameters are different. The training of nanophotonic neural medium NNM is controlled by nonlinear Maxwell's equations, that is, an input image is used as the light source to solve the iterative process of nonlinear Maxwell's equations. Before the training, the electric field is randomly initialized to E_0, thus the dielectric constant could be calculated. The new electric field E_1 could be obtained by solving the equations with FDFD, then use E_1 to update ε and iterate continuously until the field convergences. Next, solve the gradient of the loss function to the dielectric constant until the structure of NNM is updated, and the training of one picture is also finished. In the end, repeat the above process with different pictures. At the beginning of training, dopants are randomly distributed; with the advancement of the training process, dopants begin to move, merge, and finally converge together. The dividing line is gradually generated during the training process. The training process seems to be training and updating dielectric constant ε, but it is actually changing the distribution of dopants in it, in other words, changing the material density of the entire material.

However, Maxwell's equations describe not only light waves, but also all types of waves belonging to electromagnetic waves. The discovery and development of electromagnetic wave are inseparable from the research of mechanical waves such as water waves and sound waves. Mechanical waves and electromagnetic waves have different generation mechanisms and their own characteristics. However, they are all waves, and there are many common rules, for instance, all of them can produce reflection, refraction, interference, diffraction and other phenomena; wave speed, wavelength and frequency have the same relationship; vibration law and energy distribution are similar to electromagnetic wave. Therefore, the training of light waves in NNM can be extended to the training of other waves with similar characteristics, such as sound wave, so as to realize the deep learning tasks in other fields.

In December 2019, Tyler W. Hughes et al. conducted an analysis and research on the neural network constructed by wave physical simulation [87]. First, they proved that the dynamics of the wave equation is conceptually equivalent to the dynamics of RNN, and then designed an inhomogeneous medium to demonstrate how to train the dynamics of the wave equation through the construction of the nonhomogeneous material distribution to classify vowels. The specific system layout is shown in Fig. 12(c). For demonstration, a binary system consisting of two materials is realized. As with NNM, the initial distribution of wave velocity is composed of uniform material area with velocity between the two materials that make up the system. When the system is trained, the wave equation model is used to carry out back propagation, and the gradient of the cross-entropy loss function of the measured output with respect to material density of each pixel in the trainable area is calculated. This method is mathematically equivalent to the adjoint method. Then, we used the Adam optimization algorithm to update the material density with this gradient information, and repeated the process until the final structure converged. The experimental results show that the structure can be used to

identify the vowel indeed. The average accuracy of system on the training data set is 92.6 ± 1.1%, and the average accuracy on the test data set is 86.3 ± 4.3%. The prediction performance of the system for acoustic emission vowels is almost perfect, the system can distinguish between iy vowels and ei vowels, but accuracy is poorer, especially in samples not shown in the test dataset.

Although the neural network system combining scattering and deep learning can perform classification and recognition tasks, there are some difficulties in material production. It is not easy to obtain active tumor slices, and the size of nanophotonic medium is also on the order of micron millimeter after calculation, and its fabrication is not simple. However, the materials of scattering media can be diversified and easy to obtain. For linear materials, in the optical platform, linear dopants such as pores can be used; in an acoustic environment, the distribution of materials may include air with a sound velocity of 331 m/s and porous silicone rubber with a sound velocity of 150 m/s [88]. For nonlinear materials, in the optical platform, utilizing the Kerr nonlinearity is the most direct method to realize the nonlinear wave velocity. Silicon (Si) and chalcogenide glass (such as As_2S_3) are two kinds of widely used nonlinear optical materials on the integrated platform, and chalcogenide glass has one of the highest damage thresholds [89]. Another commonly used optical nonlinearity is saturation absorption, which consists of the intensity-dependent absorption/damping, mathematically defined as: $(u) = \frac{b_0}{1+(\frac{u}{u_{th}})^2}$. One possible way to achieve this effect is to place graphene or other absorbent 2D materials on the linear optical circuit etched on a medium such as silicon. Acoustically, many fluids, especially those containing bubbles such as carbonated water, exhibit strong nonlinear responses. Not only light waves and sound waves, but also waves similar to Maxwell's equation can be used to construct network system by making use of inhomogeneous media for training and learning.

Training spiking neural network (SNN) with STDP mechanism

In all-optical spiking neural network, not only supervised learning of simple image recognition is carried out, but also unsupervised learning is demonstrated [61]. In the supervised learning experiment, the synaptic weight of the network is trained based on the computer, adopting the back propagation algorithm. Here, a set of training data consisting of input mode pair and expected output pair is given. According to the deviation between expected output and actual output, the synapse weight in the network is adjusted for optimization until the deviation is optimal and the network converges. In the unsupervised learning, the network can automatically update its weight through a feedback loop and adapt to specific patterns in this way, without the need for external computer control. The specific unsupervised neuron pattern is shown in Fig. 12(d). Unsupervised learning uses spiking timing dependent plasticity (STDP) criteria to update the weight, that is, the change in the weights of two synapses is related to the time difference between the pre-synaptic and post-synaptic neuron pulses [90]. If an input signal arrives just before the output peak, then the input signal is likely to have reached the trigger threshold and the corresponding weight will be increased. If the input pulse arrives after the output pulse, the weight of the synapse is reduced. The increase or

decrease of the weight is a function of the time difference between the input peaks and output peaks. Eq. (3) and (4) show the weight update of two mainstream unsupervised STDP learning algorithms:

$$\text{Two}-\text{phase STDP}: \Delta w_i = \begin{cases} A_+ \exp\left(\frac{\Delta t_i}{\tau_+}\right), \Delta t_i < 0 \\ A_- \exp\left(\frac{-\Delta t_i}{\tau_-}\right), \Delta t_i \geq 0 \end{cases} \tag{3}$$

$$\text{There}-\text{phase STDP}: \Delta w_i = A_+ \exp\left(\frac{-(\Delta t_i - 15)^2}{200}\right) - A_- \exp\left(\frac{-(\Delta t_i - 15)^2}{200}\right) \tag{4}$$

In 2020, Shuiying Xiang proposed the hardware architecture of a multi-layer photonic spiking neural network [91], which uses a vertical cavity surface emitting laser embedded with a saturated absorber as the pulse neuron, with two polarization modes. If the polarization pattern between the two is the same, it is considered as an excitatory synapse. If it is orthogonal polarization, it is considered as an inhibitory synapse. In addition, applying the photonic STDP criterion, a supervised learning algorithm based on Tempotron rule and photonic STDP rule is designed, which is suitable for multi-layer photonic spiking neural network. The neuromorphic neural network can solve the classical XOR problem, and consider the influence of physical parameters of photonic neurons on training convergence. Furthermore, the multi-layer photonic SNN is further extended to realize other logical tasks.

Pseudo-inverse matrix method for reservoir computing

Reservoir computing based on WDM technology is a special neural network system, in which training parameters and training methods are also different from other neural networks. RC system has a fixed reservoir, the so-called hidden layer, and its input matrix and internal connection matrix of the reservoir are given randomly and fixed. Thus, RC only needs to train the output matrix between the reservoir layer and the output layer. The existing training methods include pseudo-inverse matrix method, ridge regression, and least square method, the most commonly used training method is pseudo inverse matrix method. We take the RC system based on optical fiber communication as an example to illustrate the application of pseudo-inverse matrix method in the training of reservoir computing [53].

RC system is a recursive system subject to the time-limited internal state $x(n)$. Its neurons can be described as the function of input current and previous calculation results, expressed in the following way:

$$\tilde{x}(n) = f\left(W^{in}[1; u(n)], Wx(n-1)\right)$$

$$x(n) = (1-\alpha)x(n-1) + \alpha\tilde{x}(n)$$

The output of the network can be expressed as: $y(n) = W^{out}[1; u(n); x(n)]$. By collecting training data $[1; u(n); x(n)]$ and training target signal, the readout matrix waveform can be obtained by using the pseudo-inverse matrix method. When estimating the difference between the theoretical output and the system output, an indicator such as the normalized root mean square error (NRMSE) is used. The specific form is as follows:

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(Y' - Y\right)^2}{N}}}{\left(\max\left(Y'\right) - \min\left(Y'\right)\right)}$$

In order to evaluate the performance of the optical RC system, the simulation experiment was carried out to identify the input signal waveform using the commercial software of the optical fiber communication system, and the experiments of without interconnection and with interconnection were compared. The pseudo-inverse matrix method used in the training process is replaced by the optical delay tuning of the phase modulator (PM) between the optical neuron and the output layer directional coupler, and the output data is optimized by obtaining the lowest NRMSE. The final results show that ONN can provide better performance in recognition of input signal waveforms, and reservoir computation in the neural network with randomly connected optical neurons can provide better performance. In addition, the performance of RC system when performing linear and nonlinear operation in EDFA is also studied. Finally, it is proved that optical neurons should be activated by nonlinear activation function in the RC system, so as to obtain the ability of signal recognition.

Discussion and outlook

ONN is a promising alternative to ENN with two obvious advantages. Firstly, the matrix multiplication which ANN relies on can be performed at the speed of light in ONN and detected at a rate of over 50 GHz [92]; Secondly, after training, ONN is passive, and the calculation of optical signal can be realized with the minimum power consumption [48]. A large number of different types of ONNs have been reported at present, including ONNs based on diffraction optics and free space optics, integrated photonic circuits based on interference optics and synaptic mechanisms of spiking discharge, and even neural networks that utilize the principle of wavelength division multiplexing for reservoir computing. Here, we will summarize and discuss the network technologies involved in this paper, and point out the current challenges and the possible development directions in the future for the implementation of optical neural network.

In the study of diffraction neural network, it makes good use of the phenomenon of light diffraction, and realizes the full connection of neurons among all layers, so that the learning ability of the model will be better. But their research lacks an important part, which is nonlinear activation, and the researchers also suggest that their process does not involve nonlinear activation functions. In the future, we can try to implement such optical diffraction neural networks and add nonlinear work, such as using nonlinear optical media such as photorefractive crystals and magneto-optical traps, or using existing nonlinear activation functions that have been studied, to compensate for its absence through experiments. Beyond that, this kind of network and ONNs based on Fourier transform belong to free-space connection networks. Due to some heavy optical elements such as diffraction element and lens, it is challenging to scale and scale a large number of neurons. Scattering-based neural network is a type of neural network that is worth studying. Due to the disorder of scattering, light may be scattered from all directions. Therefore, the light passing through the scattering medium is equivalent to many computations, which is likely to surpass the previous hierarchical feedforward

network. And because of the particularity of nano scattering medium, we can realize many different real-time trainings. Optical neural network with chips as the mainstream, such as coherent nanophotonic circuit and spiking network, can offer a CMOS-compatible, scalable approach to achieve optical deep learning tasks, have huge advantages in device miniaturization and expanding the network size, and they work under light, with the strong computing power and minimal resource consumption. However, the cost of chip-type network is extremely expensive and the technical requirements are strict, requiring a lot of manpower material resources to support. So even though it has excellent development prospects, there are still technical challenges to be overcome.

Nonlinear operation is the root of the strong expression ability of ANN, it enables the neural network to learn complex mapping between the input and output, speeds up the convergence of the network and improves recognition accuracy. It's an indispensable component of the neural network. Graphene, PCM, EIT and other excellent nonlinear activations have emerged nowadays, however, there are huge challenges to implement the nonlinear function in the optical domain, Firstly, the optical nonlinearity is relatively weak, and its generation generally requires very high optical power, greatly increasing the energy consumption. At the same time, due to the high optical power, other optical devices in the system will be damaged. Secondly, the optical nonlinearity needs to be balanced with the working bandwidth, and the information processing capability of ONN will be limited. In addition, many elements in an optical circuit maintain a consistent resonant response between each other, requiring additional control circuits to calibrate each element [93]. Thirdly, in the architecture of photonic artificial intelligence chip, the flexibility of nonlinear activation function is high, and it is difficult to control the optical nonlinear effect. In the manufacturing process of nonlinear devices, the response tends to be stable, which cannot meet the need of flexibility. Meanwhile, there are many problems in the integration of nonlinear optical units on chip in terms of process compatibility and device consistency. In summary, how to realize the optical nonlinear activation function with low power consumption, high speed, easy realization and rich expression forms is a technical problem urgently to be solved by the technicians in this field.

Apart from seeking a breakthrough in the aspects of linear operation and nonlinear activation, we can also put our energy into the neural networks training. At present, many networks complete the training process on the computer, and then complete the identification and classification tasks in the optical system, such a method is inevitably too targeted. Therefore, it is extremely important to find a training method that can train in optical mode and realize real-time training. The coherent nanophotonic circuit realizes a forward-propagation programmable training and also proposes an efficient local training method. The training of nano scattering medium is a good example, too. It can change the dielectric constant of the material by controlling the electric field, thereby controlling the distribution of internal dopants and finally achieve stability. Furthermore, depending on the task or goal, the training can be repeated many times.

Of course, the successful implementation of ONN cannot be separated from the combination of technologies in other fields. For instance, the fabrication of 3D diffraction layer in the D^2NN uses 3D printing technology and Poisson surface reconstruction technology. Grating diffraction layer uses semiconductor processing technology for

etching, as well as silicon photonic integration technology is needed for coherent nanophotonic circuit, spiking network and nanometer neural medium carried on chip. There is even cooperation with the fields of metamaterials, scattering imaging, etc. Not only that, in order to realize nonlinear activation, it is also necessary to have knowledge reserves related to chemistry and materials.

Conclusions

In this paper, we introduce and analyze the advanced field of deep learning——optical neural network in detail. Firstly, we introduce how to realize the linear connection and optical nonlinear activation in ONN, and then describe how to train ONN in term of different training or gradient calculation methods. At last, we also conduct and discuss the optical neural network techniques, and point out the current challenges and future developments. For some typical applications, simple data analysis and comparison are also carried out. As an interdisciplinary product of photonic technology and artificial intelligence technology, photonic neural network can combine the advantages of photonic technology and artificial intelligence to build a high-speed, low-power, large-bandwidth network structure, breaking through the bottleneck of traditional electronic neural network. However, the photonic neural network still needs to overcome problems such as real-time training, implementation of nonlinear activation function, scale and application expansion, etc. It is believed that in the near future, the photonic neural network can better play the advantages brought by the combination of optoelectronic technology and artificial intelligence technology, so as to better build a green intelligent world.

Abbreviations

ANN: Artificial Neural Network; AWG: Arbitrary Waveform Generator; C60: Buckyball; CCD: Charge Coupled Device; D^2NN: Diffraction Deep Neural Network; DFB: Distributed Feedback; DSP: Digital Signal Processor; EDFA: Erbium Doped Optical Fiber Amplifier; EIT: Electromagnetic Induced Transparency; FDFD: Finite-Difference Frequency-Domain; FEOND: Feature-Extracting Optical Neuron Device; GS: Gale-Shapley algorithm; GST: Ge2Sb2Te5; MOT: Magneto-Optical Trap; MRR: Microring Resonator; MSE: Mean Square Error; MZI: Mach-Zehnder Interferometer; NNM: Nanophotonic Neural Medium; NRMSE: Normalized Root Mean Square Error; OMM: Optical Matrix Multiplier; ONN: Optical Neural Network; PCM: Phase Change Material; PM: Phase Modulator; QONN: Quantum Optical Neural Network; QD: Quantum Dots; RNN: Recurrent Neural Network; SLM: Spatial Light Modulator; SNN: Spiking neural network; STDP: Spike Timing Dependent Plasticity; SVD: Singular Value Decomposition; TS-NN: Time-Stretch Electro-Optical Neural Network; WDM: Wavelength Division Multiplexing

Acknowledgements

Not applicable.

Authors' contributions

Jia Liu finished the manuscript and prepared the figures, tables and references, and was a major contributor in writing the manuscript. Qiuhao Wu gave guidance and participated in the revision of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 11773018 and Grant 61727802, in part by the Key Research and Development programs in Jiangsu China under Grant BE2018126, in part by the Fundamental Research Funds for the Central Universities under Grant 30919011401 and Grant 30920010001, and in part by the Leading Technology of Jiangsu Basic Research Plan under Grant BK20192003.

Availability of data and materials

Data sharing is not applicable to this article as no new datasets were created in this review.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details
[1]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China.
[2]Institute of Armored Forces, Army Research Institute, Beijing, China.

Received: 23 December 2020 Accepted: 9 March 2021
Published online: 19 April 2021

References

1. Hines ML, Carnevale NT. The neuron simulation environment. Neural Comput. 1997;9(6):1179–209. https://doi.org/10.1162/neco.1997.9.6.1179.
2. Schwabe RJ, Zelinger S, Key TS, Phipps KO. Electronic lighting interference. IEEE Ind Appl Mag. 1998;4:46–8.
3. Markram H, Muller E, Ramaswamy S. Reconstruction and simulation of neocortical microcircuitry. Cell. 2015;163(2):456–92. https://doi.org/10.1016/j.cell.2015.09.029.
4. Tsai F-CF, O'Brien CJ, Petrović NS, Rakić AD. Analysis of optical channel cross talk for free-space optical interconnects in the presence of higher-order transverse modes. Appl Optics. 2005;44(30):6380–7. https://doi.org/10.1364/AO.44.006380.
5. Hu W, Li X, Yang J, Kong D. Crosstalk analysis of aligned and misaligned free-space optical interconnect systems. J Opt Soc Am A. 2010;27(2):200–5. https://doi.org/10.1364/JOSAA.27.000200.
6. Goodman JW, Dias AR, Woody LM. Fully parallel, high-speed incoherent optical method for performing discrete fourier transforms. Opt Lett. 1978;2(1):1–3. https://doi.org/10.1364/OL.2.000001.
7. Hu X, Wang A, Zeng M, Long Y, Zhu L, Fu L, et al. Graphene-assisted multiple-input high-base optical computing. Sci Rep. 2016;6:32911.
8. Caulfield HJ, Dolev S. Why future supercomputing requires optics. Nat Photon. 2010;4(5):261–3. https://doi.org/10.1038/nphoton.2010.94.
9. Mosca EP, Griffin RD, Pursel FP, Lee JN. Acoustooptical matrix-vector product processor: implementationissues. Appl Optics. 1989;28(18):3843–51. https://doi.org/10.1364/AO.28.003843.
10. Sun C-C, Chang M-W, Hsu KY. Matrix-matrix multiplication by using anisotropic self-diffraction in batio3. Appl Optics. 1994;33:4501X507.
11. Nasr MB, Chtourou M. A hybrid training algorithm for feedforward neural networks. Neural Process Lett. 2006;24(2):107–17. https://doi.org/10.1007/s11063-006-9013-x.
12. de Lima TF, Shastri BJ, Tait AN, Nahmias MA, Prucna PR. Progress in neuromorphic photonics. Nanophotonics. 2017;6(3):577–99. https://doi.org/10.1515/nanoph-2016-0139.
13. Chen Y. 4f-type optical system for matrix multiplication. Optim Eng. 1993;32.
14. PIAGGIO HTH. The mathematical theory of huygens' principle. Nature. 1940;145(3675):531–2. https://doi.org/10.1038/145531a0.
15. Young T. The Bakerian lecture. Experiments and calculations relative to physical optics. Abstr Pap Print Philos Transactions Royal Soc Lond. 1832;1:131–2.
16. Mandel L, Wolf E. Some properties of coherent light*. J Opt Soc Am. 1961;51(8):815–9. https://doi.org/10.1364/JOSA.51.000815.
17. Porter MB. Concerning green's theorem and the cauchy-riemann differential equations. Ann Math Sec Ser. 1905;7(1):1–2. https://doi.org/10.2307/1967189.
18. AL-Jawary MA, Wrobel LC. Numerical solution of the two-dimensional helmholtz equation with variable coefficients by the radial integration boundary integral and integro-differential equation methods. Int J Comput Math. 2012;89:1463–87.
19. Umul YZ. Young-kirchhoff-rubinowicz theory of diffraction in the light of sommerfeld's solution. J Opt Soc Am A. 2008;25(11):2734–42. https://doi.org/10.1364/JOSAA.25.002734.
20. Sommerfeld A. Optics. Lectures on theoretical physics, vol. iv. Am J Physiol. 1955;23(7):477–8. https://doi.org/10.1119/1.1934064.
21. Goodman J. Introduction to Fourier optics: 2rd Edition, Roberts and Company Publishers, Englewood; 1995. p. 35.
22. Karczewski B. Fraunhofer diffraction of an electromagnetic wave. J Opt Soc Am. 1961;51(10):1055–7. https://doi.org/10.1364/JOSA.51.001055.
23. Wang X, Xu Q, Liu E. Angular spectrum theory to calculate coupling efficiency in rectangular waveguide resonators. Opt Laser Technol. 2000;32(3):177–81. https://doi.org/10.1016/S0030-3992(00)00037-2.
24. Lin X, Rivenson Y, Yardimci NT, Veil M, Luo Y, Jarrahi M, et al. All-optical machine learning using diffractive deep neural networks. Science. 2018;361(6406):1004–8. https://doi.org/10.1126/science.aat8084.
25. Lu L, Zhu L, Zhang Q, Zhu B, Yao Q, Yu M, et al. Miniaturized diffraction grating design and processing for deep neural network. IEEE Photon Technol Lett. 2019;31(24):1952–5. https://doi.org/10.1109/LPT.2019.2948626.
26. Qian C, Lin X, Xu J, Sun Y, Li E, Zhang B, et al. Performing optical logic operations by a diffractive neural network. Light Sci Appl. 2020;9(1):59. https://doi.org/10.1038/s41377-020-0303-2.
27. Luo Y, Mengu D, Yardimci NT, Rivenson Y, Veli M, Jarrahi M, et al. Design of task-specific optical systems using broadband diffractive neural networks. Light Sci Appl. 2019;8(1):112. https://doi.org/10.1038/s41377-019-0223-1.
28. Liao D, Chan KF, Chan CH, Zhang Q, Wang H. All-optical diffractive neural networked terahertz hologram. Opt Lett. 2020;45(10):2906–9. https://doi.org/10.1364/OL.394046.
29. Blackwell CA, Simpson RS. The convolution theorem in modern analysis. IEEE Transact Educ. 1966;9(1):29–32. https://doi.org/10.1109/TE.1966.4321930.
30. Lu T, Wu S, Xu X, Yu FTS. Two-dimensional programmable optical neural network. Appl Optics. 1989;28(22):4908–13. https://doi.org/10.1364/AO.28.004908.
31. Gao S, Yang J, Feng Z, Zhang Y. Implementation of a large-scale optical neural network by use of a coaxial lenslet array for interconnection. Appl Optics. 1997;36(20):4779–83. https://doi.org/10.1364/AO.36.004779.
32. Kuratomi Y, Takimoto A, Akiyama K, Ogawa H. Optical neural network using vector-feature extraction. Appl Optics. 1993;32(29):5750–8. https://doi.org/10.1364/AO.32.005750.
33. Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. Sci Rep. 2018;8:12324.

34. Zuo Y, Li B, Zhao Y, Jiang Y, Chen Y-C, Chen P, et al. All-optical neural network with nonlinear activation functions. Optica. 2019;6(9):1132–7. https://doi.org/10.1364/OPTICA.6.001132.
35. Breit G. The interference of light and the quantum theory. Proc Natl Acad Sci. 1923;9(7):238–43. https://doi.org/10.1073/pnas.9.7.238.
36. Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, et al. Deep learning with coherent nanophotonic circuits. Nat Photon. 2017;11:44H46.
37. Elson JM, Rahn JP, Bennett JM. Light scattering from multilayer optics: comparison of theory and experiment. Appl Optics. 1980;19(5):669–79. https://doi.org/10.1364/AO.19.000669.
38. Rochon P, Bissonnette D. Lensless imaging due to back-scattering. Nature. 1990;348(6303):708–10. https://doi.org/10.1038/348708a0.
39. Vellekoop IM, Mosk AP. Focusing coherent light through opaque strongly scattering media. Opt Lett. 2007;32(16):2309–11. https://doi.org/10.1364/OL.32.002309.
40. Katz O, Small E, Silberberg Y. Looking around corners and through thin turbid layers in real time with scattered incoherent light. Nat Photon. 2012;6(8):549–53. https://doi.org/10.1038/nphoton.2012.150.
41. Vellekoop IM, Lagendijk A, Mosk AP. Exploiting disorder for perfect focusing. Nat Photon. 2010;4(5):320–2. https://doi.org/10.1038/nphoton.2010.3.
42. Bertolotti J, van Putten EG, Akbulut D, Vos WL, Lagendjk A, Mosk AP. Scattering optics resolve nanostructure. In: Proc. SPIE 8102, Nanoengineering: fabrication, properties, optics, and devices VIII; 2011. p. 810206.
43. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, et al. Optical coherence tomography. Science. 1991;254(5035):1178–81. https://doi.org/10.1126/science.1957169.
44. Katz O, Heidmann P, Fink M, Gigan S. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. Nat Photon. 2014;8(10):784–90. https://doi.org/10.1038/nphoton.2014.189.
45. Yaqoob Z, Psaltis D, Feld MS, Yang C. Optical phase conjugation for turbidity suppression in biological samples. Nat Photon. 2008;2(2):110–5. https://doi.org/10.1038/nphoton.2007.297.
46. Ando T, Horisaki R, Tanida J. Speckle-learning-based object recognition through scattering media. Opt Express. 2015;23(26):33902–10. https://doi.org/10.1364/OE.23.033902.
47. Pierangeli D, Marcucci G, Moriconi C, Perini G, Spirito MD, Papi EAM. Deep optical neural network by living tumour brain cells. Physis. 2018.
48. Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, et al. Nanophotonic media for artificial neural inference. Photon Res. 2019;7(8):823–7. https://doi.org/10.1364/PRJ.7.000823.
49. Qu Y, Zhu HZ, Shen YC, Zhang J, Tao CN, Ghosh P, et al. Inverse design of an integrated-nanophotonics optical neural network. Sci Bull. 2020;65(14):1177–83. https://doi.org/10.1016/j.scib.2020.03.042.
50. Koester CJ. Wavelength multiplexing in fiber optics. J Opt Soc Am. 1968;58(1):63–70. https://doi.org/10.1364/JOSA.58.000063.
51. Paquot Y, Duport F, Smerieri A, Dambre J, Schrauwen B, Haelterman M, et al. Optoelectronic reservoir computing. Sci Rep. 2012;2:287.
52. Duport F, Schneider B, Smerieri A, Haelterman M, Massar S. All-optical reservoir computing. Opt Express. 2012;20(20):22783–95. https://doi.org/10.1364/OE.20.022783.
53. Cheng T-Y, Chou D-Y, Liu C-C, Chang Y-J, Chen C-C. Optical neural networks based on optical fiber-communication. Neurocomputing. 2019;364:239–44. https://doi.org/10.1016/j.neucom.2019.07.051.
54. Zang Y, Chen M, Yang S, Chen H. Electro-optical neural networks based on time-stretch method. IEEE J Sel Top Quantum Electron. 2020;26(1):1–10. https://doi.org/10.1109/JSTQE.2019.2957446.
55. Zhang H, Feng X, Li B, Wang Y, Cui K, Liu F, et al. Integrated photonic reservoir computing based on hierarchical time-multiplexing structure. Opt Express. 2014;22(25):31356–70. https://doi.org/10.1364/OE.22.031356.
56. Nguimdo RM, Verschaffelt G, Danckaert J, der Sande GV. Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback. IEEE Transact Neur Netw Learn Syst. 2015;26(12):3301–7. https://doi.org/10.1109/TNNLS.2015.2404346.
57. Maass W. Networks of spiking neurons: the third generation of neural network models. Neural Netw. 1997;10(9):1659–71. https://doi.org/10.1016/S0893-6080(97)00011-7.
58. Tait AN, de Lima TF, Zhou E, Wu AX, Nahmias MA, Shastri BJ, et al. Neuromorphic photonic networks using silicon photonic weight banks. Sci Rep. 2017;7:7430.
59. Shastri BJ, Nahmias MA, Tait AN, Rodriguez AW, Wu B, Prucnal PR. Spike processing with a graphene excitable laser. Sci Rep. 2016;6:19126.
60. Chakraborty I, Saha G, Sengupta A, Roy K. Toward fast neural computing using all-photonic phase change spiking neurons. Sci Rep. 2018;8:12980.
61. Feldmann J, Youngblood N, Wright CD, Bhaskaran H, Pernice WHP. All-optical spiking neurosynaptic networks with self-learning capabilities. Nature. 2019;569(7755):208–14. https://doi.org/10.1038/s41586-019-1157-8.
62. Nahmias MA, Peng H, de Lima TF, Huang C, Tait AN, Shastri BJ, Prucnal PR. A TeraMAC neuromorphic photonic processor. In: 2018 IEEE photonics Conf. (IPC); 2018. p. 1–2.
63. Tait AN, Nahmias MA, Shastri BJ, Prucnal PR. Broadcast and weight: an integrated network for scalable photonic spike processing. J Light Technol. 2014;32(21):4029–41. https://doi.org/10.1109/JLT.2014.2345652.
64. Shainline JM, Buckley SM, McCaughan AN, Chiles J, Jafari-Salim A, Mirin RP, et al. Circuit designs for superconducting optoelectronic loop neurons. J Appl Phys. 2018;124(15):152130. https://doi.org/10.1063/1.5038031.
65. Selden AC. Pulse transmission through a saturable absorber. Br J Appl Phys. 1967;18(6):743–8. https://doi.org/10.1088/0508-3443/18/6/306.
66. Braunstein R. Nonlinear optical effects. Phys Rev. 1962;125(2):475–7. https://doi.org/10.1103/PhysRev.125.475.
67. Cotton A. Recherches Sur l'absorption et la dispersion de la lumiere par les milieux doues du pouvoir rotatoire. J Phys Theor Appl. 1896;5(1):237–44. https://doi.org/10.1051/jphystap:018960050023700.
68. Skinner SR, Steck JE, Behrman EC. Optical neural network using Kerr-type nonlinear materials. In: Proceedings of the fourth international conference on microelectronics for neural networks and fuzzy systems: IEEE; 1994. p. 12–5.

69. Dejonckheere A, Duport F, Smerieri A, Fang L, Oudar J-L, Haelterman M, et al. All-optical reservoir computer based on saturation of absorption. Opt Express. 2014;22(9):10868–81. https://doi.org/10.1364/OE.22.010868.
70. Cheng Z, Tsang HK, Wan X, Xu K, Xu J. In-plane optical absorption and free carrier absorption in graphene-on-silicon waveguides. IEEE J Sel Top Quant Electron. 2013;20:43–8.
71. Soljacic M, Ibanescu M, Johnson SG, Fink Y, Joannopoulos JD. Optimal bistable switching in nonlinear photonic crystals. Phys Rev E. 2002;66(5):055601. https://doi.org/10.1103/PhysRevE.66.055601.
72. Coarer FD, Sciamanna M, Katumba A, Freiberger M, Dambre J, Bienstman P, et al. All-optical reservoir computing on a photonic chip using silicon-based ring resonators. IEEE J Sel Top Quant Electron. 2018;24(6):1–8. https://doi.org/10.1109/JSTQE.2018.2836985.
73. Serber R. The theory of depolarization, optical anisotropy, and the Kerr effect. Phys Rev. 1933;43(12):1003–10. https://doi.org/10.1103/PhysRev.43.1003.
74. Weinberger P. John Kerr and his effects found in 1877 and 1878. Philos Mag Lett. 2008;88(12):897–907. https://doi.org/10.1080/09500830802526604.
75. Mesaritakis C, Kapsalis A, Syvridis D. All-optical reservoir computing system based on ingaasp ring resonators for high-speed identification and optical routing in optical networks. Quant Sens Nanophoton Devices XII. 2015;9370:608–14.
76. Steinbrecher GR, Olson JP, Englund D, Carolan J. Quantum optical neural networks. NPJ Quant Inf. 2019;5:60.
77. Amin R, George J, Khurgin J, El-Ghazawi T, Prucnal PR, Sorger VJ. Attojoule modulators for photonic neuromorphic computing. In: Conference on lasers and electro-optics: Optical Society of America; 2018. p. ATh1Q.4.
78. Amin R, Khan S, Lee CJ, Dalir H, Sorger VJ. 110 attojoule-per-bit efficient graphene-based plasmon modulator on silicon. In: Conference on lasers and electro-optics: Optical Society of America; 2018. p. SM1I.5.
79. George JK, Mehrabian A, Amin R, Meng J, de Lima TF, Tait AN, et al. Neuromorphic photonics with electro-absorption modulators. Opt Express. 2019;27(4):5181–91. https://doi.org/10.1364/OE.27.005181.
80. George J, Amin R, Mehrabian A, Khurgin J, El-Ghazawi T, Prucnal PR, Sorger VJ. Electrooptic nonlinear activation functions for vector matrix multiplications in optical neural networks. In: Advanced photonics 2018 (BGPP, IPR, NP, NOMA, sensors, networks, SPPCom, SOF): Optical Society of America; 2018. p. SpW4G.3.
81. Miscuglio M, Mehrabian A, Hu Z, Azzam SI, George J, Kildishev AV, et al. All-optical nonlinear activation function for photonic neural networks. Opt Mater Express. 2018;8:3851–63.
82. Fleischhauer M, Imamoglu A, Marangos JP. Electromagnetically induced transparency: optics in coherent media. Rev Mod Phys. 2005;77(2):633–73. https://doi.org/10.1103/RevModPhys.77.633.
83. Williamson IAD, Hughes TW, Minkov M, Bartlett B, Pai S, Fan S. Reprogrammable electro-optic nonlinear activation functions for optical neural networks. IEEE J Sel Top Quantum Electron. 2020;26(1):1–12. https://doi.org/10.1109/JSTQE.2019.2930455.
84. Mengu D, Luo Y, Rivenson Y, Ozcan A. Analysis of diffractive optical neural networks and their integration with electronic neural networks. IEEE J Sel Top Quantum Electron. 2020;26(1):1–14. https://doi.org/10.1109/JSTQE.2019.2921376.
85. Zhou T, Fang L, Yan T, Wu J, Li Y, Fan J, et al. In situ optical backpropagation training of diffractive optical neural networks. Photon Res. 2020;8(6):940–53. https://doi.org/10.1364/PRJ.389553.
86. Hughes TW, Minkov M, Shi Y, Fan S. Training of photonic neural networks through in situ backpropagation and gradient measurement. Optica. 2018;5(7):864–71. https://doi.org/10.1364/OPTICA.5.000864.
87. Hughes TW, Williamson IAD, Minkov M, Fan S. Wave physics as an analog recurrent neural network. Sci Adv. 2019;5(12):eaay6946.
88. Ba A, Kovalenko A, Aristegui C, Mondain-Monval O, Brunet T. Soft porous silicone rubbers with ultra-low sound speeds in acoustic metamaterials. Sci Rep. 2017;7:40106.
89. Qiu J, Si J, Hirao K. Photoinduced stable second-harmonic generation in chalcogenide glasses. Opt Lett. 2001;26(12):914–6. https://doi.org/10.1364/OL.26.000914.
90. Karmarkar UR, Najarian MT, Buonomano DV. Mechanisms and significance of spike-timing dependent plasticity. Biol Cybern. 2002;87(5-6):373–82. https://doi.org/10.1007/s00422-002-0351-0.
91. Xiang S, Ren Z, Zhang Y, Song Z, Guo X, Han G, et al. Training a multi-layer photonic spiking neural network with modified supervised learning algorithm based on photonic STDP. IEEE J Sel Top Quantum Electron. 2020;27:1–9.
92. Vivien L, Polzer A, Marris-Morini D, Osmond J, Hartmann JM, Crozat P, et al. Zero-bias 40Gbit/s germanium waveguide photodetector on silicon. Opt Express. 2012;20(2):1096–101. https://doi.org/10.1364/OE.20.001096.
93. Radulaski M, Bose R, Tran T, Van Vaerenbergh T, Kielpinski D, Beausoleil RG. Thermally tunable hybrid photonic architecture for nonlinear optical circuits. ACS Photon. 2018;5(11):4323–9. https://doi.org/10.1021/acsphotonics.8b00376.

Publisher's Note