Astronomical Techniques and Instruments, Vol. 3, January 2026, 26–42

Review Open Access

Optical neural networks: principles, challenges, and future prospects in computing and astrophotonics

Xiaoqian Zhang[1], Huan Wang[1], Qichang An[2*], Hongchao Zhao[1*]

[1]*School of Advanced Manufacturing*, *Shenzhen Campus of Sun Yat-sen University*, *Shenzhen* 518107, *China*

[2]*Changchun Institute of Optics*, *Fine Mechanics and Physics*, *Chinese Academy of Sciences*, *Changchun* 130033, *China*

*Correspondences: anjj@mail.ustc.edu.cn; zhaohongch@mail.sysu.edu.cn

Received: March 18, 2025; Accepted: June 9, 2025; Published Online: June 10, 2025; https://doi.org/10.61977/ati2025031; https://cstr.cn/32083.14.ati2025031

Citation: Zhang, X. Q., Wang, H., An, Q. C., et al. 2026. Optical neural networks: principles, challenges, and future prospects in computing and astrophotonics. *Astronomical Techniques and Instruments*, **3**(1): 26−42. https://doi.org/10.61977/ati2025031.

Abstract: The rapid development of optical neural networks (ONNs) has led to the introduction of new research avenues for computing power enhancement. Because of the characteristics of optical signals, which include low power consumption, low latency, high parallelism, and large bandwidths, optical computing based on neural network architectures is showing promise for processing of spatial signals, temporal signals, and on-chip information. At present, there is a lack of a unified ONN computing architecture, and because of the limitations of the physical characteristics of these networks, different application scenarios have led to proposals of different requirements for the structural design, device selection, integration method, and signal processing method of the network. In this paper, we systematically elaborate on the practical value of ONNs, analyze their computational fundamentals in depth, discuss the challenges faced in computational and astrophotonics applications in detail, and simultaneously emphasize the important position and broad prospects of optical computing in the future information society.

Keywords: ONNs; Astronomy; Artificial intelligence

1. INTRODUCTION

Artificial intelligence (AI) has now pervaded most sectors of society and is widely used in fields including medical image analysis, the molecular and material sciences, and language identification. In particular, AI has revolutionized the astronomy field, enabling automated celestial object classification, exoplanet detection, gravitational wave signal processing, and large-scale sky surveys to be performed with unprecedented accuracy. AI-driven models have enhanced astronomical data processing significantly, thus aiding researchers in uncovering new cosmic phenomena and improving our understanding of the universe. The research into and application of artificial neural networks (ANNs) have attracted considerable attention because of their powerful data processing and pattern recognition capabilities. However, traditional electronic neural networks are hampered by certain inherent defects, including the susceptibility of their electrical signals to interference and processing speeds that are proportional to their energy loss, which have limited further development and applications of ANNs. The demand for training computing power has also grown with the boom in both big data and large models. However, there is a mismatch between the computing power provided by existing electronic chips and the computing power requirements of AI; current computing capabilities cannot meet the growing computational demands of modern technology. As the integrated circuit manufacturing process advances to technology nodes below 5 nm, problems including noise, crosstalk, energy consumption, and costs caused by the high circuit integration density are becoming increasingly serious. The traditional route to performance improvement by shrinking the transistor size is difficult to sustain because of factors such as the slowdown of Moore’s Law, chip heat dissipation requirements, and transmission bandwidths. Therefore, there is an urgent need for a new hardware architecture with high computing power and high energy efficiency to provide an alternative.

Against this backdrop, all ONNs have been gaining widespread attention because of their prominent advantages, comprising high bandwidths, high interconnectivity, parallel internal processing, low latency, and low energy consumption. The research and development of ONNs includes advances in both applied and theoretical fields, e.g., deep neural networks, ultrasensitive perception, and topological photonics, and it also helps to improve learning processes by several orders of magnitude. In astronomy, ONNs have the potential to process the massive datasets acquired from space telescopes and radio arrays, thus accelerating real-time image reconstruction, transient event detection, and gravitational wave data analysis. Optical computing offers an efficient way to handle the enormous computational requirements of both astrophysical simulations and multi-messenger astronomy. For example, researchers at Tsinghua University have capitalized on the symmetry of photon propagation to develop a fully forward-mode (FFM) learning method. Their approach enables direct design and adjustment of the optical parameters of physical optical systems, thereby allowing deep ONNs to be trained using millions of parameters.

Optical computing has been used previously in numerous applications. The Mach-Zehnder interferometer (MZI) is an important optical interferometer that is widely used in optical signal modulation and demodulation in optical communications. The MZI encodes and decodes optical signals by altering the optical path difference within the interferometer, thus improving the transmission efficiency and the capacity of optical communication systems. By combining singular value decomposition of the weight matrix with unit matrix decomposition, these MZI arrays can implement a weight matrix optically and perform specific optical calculations.

In astronomy, interferometric techniques are used widely in high-resolution imaging, e.g., in very-long-baseline interferometry, which synthesizes data acquired from multiple telescopes to achieve superior angular resolution for observation of distant galaxies and black hole shadows. The principles of optical interferometry have also played a crucial role in adaptive optics by correcting atmospheric distortions in real time to enhance ground-based telescope performance. The lens Fourier transform is used in radar imaging systems to convert acquired radar data into images. Time-domain signals can be converted into frequency-domain signals using the Fourier transform, and frequency-domain signals can be converted into images by applying the inverse Fourier transform. This method is widely used in synthetic-aperture radar (SAR) and inverse SAR to generate high-resolution radar images for use in target identification and terrain mapping. In terms of modulators, micro-ring resonators can modulate optical signals by varying the refractive index or optical path difference of the resonator to control the intensity or phase of the optical signals, thereby achieving both encoding and transmission of optical signals. Additionally, micro-ring resonators can be used to construct optical devices, e.g., filters and switches, to implement signal filtering, routing, and switching functions in optical communication networks. The rise in AI and gradual broadening of the possible application scenarios for optical computing have given new vitality and purpose to this field.

ONNs are computational models that use optical devices, including lasers, optical modulators, filters, and detectors, to simulate and implement inference functions for neural networks. The development of ONNs can be traced back to the 1960s with the birth of lasers, and the increasing depth of optical coherence research contributed to the rapid advancement in information optics. In 1982, Hopfield[1] introduced the concept of the Hopfield network architecture. In 1985, Farhat et al.[2] implemented the first ONN based on the Hopfield model. In 2017, Shen et al.[3] used silicon photonics technology to construct a fully optical convolutional architecture based on an MZI optical switch array. In 2018, Lin et al.[4] combined spatial light modulators with a three-dimensional (3D) printed dielectric material to construct a multilayer perceptron to enable implementation of a diffractive deep ONN[5]. With the continuous progress being made in AI, photonic computing, and astronomical instrumentation, ONNs promise to revolutionize the astrophysical data processing field. From deciphering cosmic microwave background data to enhancing exoplanet detection via transit photometry, the use of ONNs can pave the way for more efficient and accurate astronomical discoveries. Since the work of Lin et al.[4], the types of ONNs have grown in diversity and complexity, causing ONNs to enter a period of vigorous development.

2. ADVANCES IN OPTICAL NETWORKS FOR HIGH-PRECISION DETECTION

2.1. Conventional Memristor-Based ONNs

Memristor-based neural networks (MNNs) have been identified as a promising hardware solution for neural network implementation, primarily because of their low power consumption and accelerated matrix operations. However, during hardware deployment, the selection of both the memristor update thresholds and the in situ update scheme can affect the MNN's effectiveness significantly. To address these important issues, Shen et al.[6] introduced a novel scheme that combined dynamic threshold (DT) and gradient accumulation (GA) methods with the threshold properties. The authors simulated realistic memristor characteristics, including both pulse-to-pulse and device-to-device behavior, by introducing random noise into the Vteam memristor model. They proposed a DT scheme to improve in situ training accuracy by using the intrinsic properties of the memristors. The accumulation of gradients during backpropagation was used to regulate the memristor updates with greater precision and thus enhanced the accuracy of the in situ training. To address the high-resolution requirements for accurate weight updates and the high durability requirements for the frequent set/reset operations, a hybrid accuracy-based training approach was investigated, whereby the weight updates are accumulated in software and the memristor device is only updated when the accumulated update value exceeds the programmed granularity. This approach has been shown to relax the conductance update resolution and durability requirements significantly while achieving accuracy comparable to that of the software. Their experimental results demonstrated a substantial enhancement in test accuracy when using the dynamic threshold and gradient accumulation (DTGA) scheme on the Modified National Institute of Standards and Technology (MNIST) dataset (from 82.98% to 96.15%) and the Fashion-MNIST dataset (from 75.58% to 82.53%). Robustness analysis demonstrated that the DTGA scheme tolerated random noise factors of 0.03 and 0.02 for the MNIST and the Fashion-MNIST datasets, respectively, thus demonstrating the scheme's reliability under various conditions. Notably, when using the Fashion-MNIST dataset, the DTGA scheme improved overall performance by 7% with a corresponding 7% reduction in training time. This study therefore, offered a compelling solution for hardware implementation of various neuromorphic systems.

Owing to the presence of physiological noise in brain signals, previous studies have encountered difficulties in attempts to achieve high classification accuracy, and these studies have often neglected to evaluate the reliability of the computed classification accuracy. To address these classification problems, Bak et al.[7] proposed a memristor-based convolutional neural network (M-CNN) that was capable of classifying stress states. The weight updating process of this model involves a stochastic gradient descent with momentum, in which the normalized memristor conductance is used as a weight proxy. The conductance values are then adjusted to categorize the stress state. The classification accuracy between the control and the stress groups was calculated using the M-CNN, and the latest convolutional neural network (CNN) model, DenseNet, was used to simulate the accuracy under the same conditions. The results from this study indicate that the accuracy of the M-CNN (93.33%) is higher than that of the CNN (87.50%), and this outcome is supported by the precision, the recall, and the F-score calculated from the confusion matrix. A similar outcome was observed when M-DenseNet (92.38%) was compared with C-DenseNet (90.00%), with M-DenseNet demonstrating higher accuracy than C-DenseNet, though both networks exhibited lower accuracy than M-CNN. The study's findings contributed to the implementation and optimization of an M-CNN, which integrates hardware-based memristors with software-based CNNs with the aim of enhancing the classification accuracy of the stress states.

To date, hardware neural networks have mainly focused on shallow networks (with two to five layers). Hardware implementation of deep neural networks remains challenging because of their layer-by-layer structures, which can lead to long training times, signal interference, and low accuracy being caused by gradient explosion and vanishing behavior. To address these issues, Chen et al.[8] proposed a hybrid ultra-deep photoelectric neural network and an ultra-deep super-resolution reconstruction neural network that can achieve optical cross-layer transmission of electronic information of the photodetector by using negative ultraviolet photoconductive light-emitting memristors (N-LEMs) with intrinsic parallelism and software-hardware co-design. This approach uses the strong correlation between the optical and electrical outputs of the light-emitting memristors effectively. Furthermore, deep neural network training requires a long reset time, but the use of N-LEMs can reduce the reset time required for deep neural network training effectively. The ultra-deep neural network designed for multi-task recognition and the neural network intended for super-resolution image reconstruction both effectively avoided gradient vanishing (explosion) and increased their layer numbers to 54 and 134, respectively, while also demonstrating strong transfer learning capabilities and near software-level performance. The proposed N-LEM and the optical cross-layer transmission strategy filled the gap for efficient, accurate, and highly robust construction of deep neural networks with low power consumption, thus offering a new solution for high-precision multi-functional hardware neural networks and edge AI.

Transistors with photonic sensing capabilities can be used as machine vision system components, but they face challenges in encoding and processing optical data. The lack of materials, processing, and functional compatibility between memristors and phototransistors makes it difficult for the sensing, storage, and computation functions of the integrated arrays to work together. For example, the single operating mode of conventional memristors cannot support the dynamic encoding requirements of optical signals. Therefore, Dang et al.[9] proposed a multi-phototransistor and single memristor array based on niobium oxide memristors, which used a silicon-based compatibility process and a Mott oxide heterojunction-based memristor in combination with the blue-light enhancement/red-light suppression characteristics of the phototransistor to realize direct coupling of the optical signals with conductance modulation. It is difficult for existing hardware systems to support the dynamic demands of both bio-heuristic (pulse coding) and machine learning (analog coding) algorithms, and there is insufficient flexibility in the encoding of spatio-temporal information in optical signals. The re searchers designed the multi-mode memristor such that it could be switched among three operating modes: the linear resistive mode (analog signal processing), the threshold switching mode (pulsed signal generation), and the short duration memory mode (dynamic temporal encoding). These modes can be switched flexibly through peripheral circuitry programming, which is compatible with both machine-learning (analog) and bio-inspired (spiking) neural network architectures and supports dynamic adaptation of various algorithms. The array senses and processes optical images and synchronizes their spatio-temporal data in different coded formats. When the array is coupled to a classifier network using a transistor and a thyristor-based nonvolatile memory array, it can support various ONNs (including optical CNNs, recurrent neural networks, and spiking neural networks). Optical signals must undergo multiple photoelectric conversions during the sense-storage-accounting integration process, but the nonlinear responses of conventional photodetectors are insufficient to realize the activation functions of neural networks directly (e.g., ReLU or pulse triggering). Therefore, the researchers combined the output current of the photodetector with a threshold comparator to realize nonlinear activation in the electric domain and also introduced materials such as sulfur-based glass, which uses its light intensity-dependent nonlinear refractive index properties to realize the required activation function directly in the optical domain. The hardware system must be compatible with various neural network architectures (e.g., convolutional and impulse neural networks), but the existing algorithms are difficult to map directly to the hybrid optical-electrical-memristor system. This study optimized this mapping relationship between the algorithms and hardware by constructing a joint optical-electrical simulation model. In situ updating of the weights was realized through the nonvolatile storage property of the memristor to support dynamic loading of the different network structures.

During the development of memristor-based ONNs, their structures gradually evolved from early shallow networks that contained only inputs, outputs, and simple intermediate connection layers into multilayer structures. The structure of the multilayer perceptron neural network based on memristors is shown in Fig. 1. By drawing on traditional electronic deep-learning network architectures, layer-by-layer abstraction and extraction of complex features were implemented by adding hidden layers. In materials terms, most ONNs were initially based on traditional metal oxides, e.g., TiO_2, which had some limitations despite mature preparation processes. Later, new materials, including two-dimensional materials (graphene and molybdenum disulfide) and organic materials, began to emerge that improved the expressive power and learning depth of these networks while also reducing power consumption, improving integration, and enhancing their long-term stability. Together, these technological advances and material innovations have propelled memristor-based ONNs toward more complex and efficient applications. Memristor-based ONNs are very promising for high-precision detection, but challenges remain. For example, existing memristor materials and devices may suffer performance degradation during long-term operation or high-frequency use and are limited in the depth and scale of their construction.

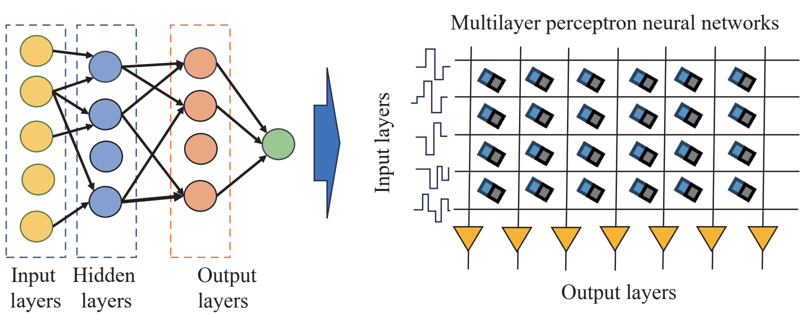

Fig. 1. Schematics of memristor-based multilayer perceptron neural networks[6].

2.2. On-Chip Integrated Photonic Neural Networks

2.2.1. *On-chip photonic neural network based on MZI*

Photonic computing has drawn considerable attention for its great potential to accelerate ANN tasks at much higher clock rates than digital electronic alternatives. In particular, reconfigurable photonic processors consisting of MZI grids are very promising for photonic matrix multipliers. Traditionally, three cascaded MZI grids (two generic $N \times N$ unitary MZI grids and one diagonal MZI grid) are required to represent the $N \times N$ weight matrix, which severely limits scalability. Therefore, Tian et al.[10] proposed a hybrid optoelectronic training framework that combined optical forward propagation with electrical backpropagation. The gradient information from the optical computation results was captured accurately by introducing a low-noise photoelectric conversion module with digitally-assisted calibration. As the number of components increased, the crosstalk and the signal loss between the optical paths increased significantly, leading to reduced computational accuracy. Additionally, the traditional gradient descent (GD) algorithm has high computational complexity when used in large-scale matrix operations, which makes it difficult to optimize the hardware parameters in real time. To this end, the research team used a stochastic parallel GD (SPGD) algorithm, which replaces parameter-by-parameter optimization of the traditional GD with parallel evaluation of randomly perturbed parameters, reducing the computation required significantly. By using the real part of a nongeneric $N \times N$ unitary MZI lattice to represent the real-valued matrices, a 4×4 photonic neural chip is realized that suffers from a lower loss of learning capacity when compared with conventional architectures using O (N^2) MZIs. In applications such as photonic neural networks, it reduces the number of MZIs required to the O ($N\log_2 N$) level while reducing the cost of the learning capacity loss. This architecture offers comprehensive advantages in terms of optical losses, chip size, power consumption, and coding errors.

In situ training of ANNs using photonic accelerators remains challenging. To this end, Tang et al.[11] proposed a silicon micro-ring resonator (MRR) optical crossbar array with a symmetric structure that realizes both forward and backward propagation seamlessly on the same optical chip. This array can potentially accelerate the inference and training phases of deep learning, effectively resolve the insertion loss imbalance problem of the previous structure, and eliminate the discrepancy between the forward and backward propagation directions. The structure uses an MRR and an MZI to implement matrix-vector multiplication, which solves the unbalanced insertion loss problem for optical paths in previous asymmetric structures by balancing the number of waveguide crossings in the forward and backward signal paths, causing all optical paths to have a uniform length and insertion loss. Although the thermo-optic phase shifter used in their study suffers from high power consumption (~19.3 mW/π) and has a potential for thermal crosstalk, the team anticipates that use of an ultra-low-power electro-optic phase shifter will reduce power consumption significantly and eliminate thermal crosstalk to enable a more compact, energy-efficient circuit. The team fabricated a 4 × 4 circuit on a silicon-on-insulator platform and used it in an iris classification inference task with a simple neural network, achieving classification accuracy of 93.3%; the neural network was then trained using simulated on-chip backpropagation, and its accuracy for the same inference task reached 91.1% after training. Additionally, the team simulated a CNN based on a 9 × 9 MRR cross-switch array for handwritten digit recognition, which achieved test accuracy of 93.4% after 10 training cycles on the MNIST dataset. This work contributed to the realization of compact, energy-efficient photonic accelerators for deep learning.

2.2.2. *On-chip diffractive neural networks (DNNs)*

The structure of the diffractive deep neural network (D^2NN) is shown in Fig. 2. Wang et al.[12] constructed an integrated photonic meta-system on a silicon photonic platform based on directional diffraction and dispersion engineering methods that were realized using sub-wavelength structures to resolve the problems of image classification, wavelength demultiplexing, and hyperspectral imaging at communication wavelengths. Three functional meta-systems were demonstrated: a 15-pixel spatial pattern classifier (with femtosecond input accuracy of nearly 90%), a multi-channel wavelength demultiplexer, and a hyperspectral image classifier. The system realized high-throughput vector matrix multiplication by integrating nearly 10^3 nanoscale phase shifters as weighting elements within an area of 0.135 mm^2 and using the diffraction and interference effects of light between multilayer hypersurfaces. In particular, the diffraction process not only realized propagation and coupling of the light field between the layers but also incorporated the manufacturing errors and any random phase fluctuations during the measurements (e.g., nanofabrication-induced phase shifts) into the training process, which allowed the pre-trained meta-system to handle the input uncertainty without post-tuning. Experiments showed that the classification accuracy of the 15-pixel spatial pattern classifier was nearly 90% under femtosecond-scale inputs, and the classification accuracy of the two-layer element system for “X”, “Y”, and “Z” letter images with a continuous wave input reached 92%. The researchers pre-trained the geometric parameters of the rectangular slots in the silicon layer using a deep diffraction neural network. With the support of the diffraction characteristics, the three-layer meta-system realized demultiplexing of signals at 1 490 nm, 1 530 nm, and 1 570 nm, and the optical field was converged accurately to the corresponding output waveguide after diffraction. The dual-wavelength pattern classification system used the wavelength selectivity of diffraction to achieve high spectral classification accuracy of 70% for the “X”, “Y”, and “Z” patterns at different wavelengths. The diffractive element system offers an alternative machine learning architecture for photonic integrated circuits with densely integrated phase shifters, spatial multiplexing throughput, and data processing capabilities.

Existing architectures can only handle data with regular structures and are unable to generalize graph-structured data beyond the Euclidean space. Yan et al.[13] proposed the diffractive graph neural network (DGNN), an all-optical graph representation learning architecture based on diffractive photonic computing units (DPUs) and on-chip optical devices, to address this shortcoming of the existing architectures. The DGNN is constructed using integrated DPUs to generate optical node features, where each DPU consists of successive diffractive layers implemented using metalines to convert node attributes into optical neural messages. In this architecture, strip optical waveguides are used to encode the input node attributes and then output the converted results, and the optical neural messages sent from node neighborhoods are aggregated using optical couplers. The DGNN architecture permits horizontal cascading of the DPUs, thus expanding the receptive field and enabling capture of the complex dependencies between the node neighborhoods during implementation of optical message-passing over graph-structured data. Furthermore, the multi-head strategy used in numerous contemporary deep-learning models (e.g., transformers, graph attention networks) has been adopted during vertical stacking of the DPUs with the objective of extracting the higher-dimensional optical node features and enhancing their learning capabilities. During fabrication, it is necessary to consider unavoidable manufacturing defects, e.g., misalignment between layers or other imperfections, which may affect the robustness of the free-space optical processors. The researchers improved the robustness of these free-space optical processors to unavoidable misalignments or other imperfections in the fabrication process significantly by introducing these variations as stochastic parameters during the training phase. The results obtained showed that this optical DGNN

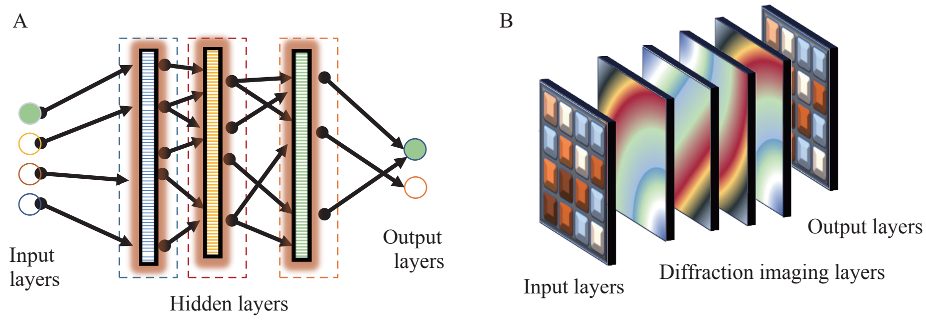

Fig. 2. The D²NN comprises multiple transmissive (or reflective) layers, where each point on a given layer acts as a neuron with a complex-valued transmission (or reflection) coefficient. These transmission/reflection coefficients for each layer can be trained using deep learning to perform functions between the input and output planes of the network. After the learning phase, the D²NN design is fixed, and when it has been fabricated or 3D-printed, it then performs the learned function at the speed of light. We trained and implemented different types of D²NNs experimentally. A comparison between (A) a CNN and (B) the D²NN is presented above.

achieved competitive or superior classification performances when compared with electronic graph neural networks in terms of synthetic graph models and three real-world graph benchmark datasets. Additionally, the optical DGNN supports graph-level classification, where additional DPUs are used to aggregate all the optical node features into a graph-level representation for classification. The results of a skeleton-based human action recognition test demonstrated the effectiveness of this architecture for use in graph-level classification tasks.

The DGNN architecture proposed by Yan et al.[13] only uses a linear model for optical information transfer and does not use any nonlinear activation functions that may enhance the model's learning capability. Yildirim et al.[14] proposed a programmable framework to generate optical nonlinearities using multiple scattering in a low-power optical device called nonlinear processing with linear optics only (nPOLO), which eliminates the need for electronic components to achieve higher-order nonlinearities. This framework uses the nonlinear relationship between the scattering potential and the scattered field to synthesize programmable linear and nonlinear transformations simultaneously at low optical powers to achieve all-optical processing in neural networks using low-power continuous-wave lasers and diffractive layers. Furthermore, the nPOLO framework enables simultaneous linear and nonlinear operations to be performed in the optical domain, thus unifying the multilayer light modulation with structural nonlinearities such that the collective effect of the data-modulated layers on the propagating light produces a higher-order nonlinear transformation of the data. These data are embedded repeatedly in the modulation layers and are combined with trainable parameters to create the desired linear and nonlinear relationship between the data and the output field. Their results showed that increasing the numbers of layers and data repetitions produced higher-order nonlinearities, e.g., polynomial orders, which included cross terms between the different input data elements. In experiments with different datasets and layer configurations, problems with the method's robustness to experimental imperfections and noise were observed. Therefore, the researchers modeled the experimental imperfections by introducing phase noise into the simulation; they found that when the two systems had equal degrees of freedom for the displayed pixels within the modulation layer, configurations with data repetition showed greater robustness to the experimental imperfections and simulated noise than those without such data repetition. By creating multiple paths leading from the input data to the output plane, the data repetition scheme produces higher polynomial orders and cross terms that couple with the different optical paths to reach the detector plane, which may make this data repetition scheme less noise-sensitive.

Liu et al.[15] proposed on-chip diffractive ONNs (DONNs) and used a structural reparameterization algorithm to design ultra-compact DONNs that addressed the problems of the increased power and resource requirements of electronic devices during complex tasks; they demonstrated 93.3% accuracy experimentally on the Iris plant dataset, in line with numerical predictions. They encountered challenges when modeling the on-chip DONNs directly, particularly in portraying the interactions between the silicon grooves and the light field accurately. To address this issue, the team used a deep complex neural network to model the complex interactions in each metal line, thus enabling accurate numerical characterization of the on-chip light propagation. This work increased the computational density by more than an order of magnitude, raising the possibility of mapping more neurons and connections onto optical devices.

On-chip ONNs are able to convert large numbers of parameters into optical form while performing passive computations, but they still face challenges in scalability and multitasking. Liu et al.[16] proposed a novel network architecture for multitasking using in-memory optical computation by fabricating two ultra-compact memory diffraction-based chips with an integration level of over 60000 parameters/mm^2 to increase the computational density and the energy efficiency of the on-chip neuromorphic hardware. This approach uses the transfer learning principle to embed most of the parameters into fixed optical components and some into tunable electronic components. In addition, the physical propagation process is modeled using a deep regression algorithm, and the compact ONN can handle a variety of tasks. In the classification task, the shared optical core+lightweight transfer network (SOC+LTN) architecture was used to process datasets including iris, penguin, and wheat seeds with an average accuracy of 95.9%, which is comparable to that of large-scale electronic neural networks and simultaneously reduces the amount of high-power numerical computation by 90%. In the regression task, the SOC was used as a multi-kernel convolutional unit by the structure reparameterization algorithm; different classical convolutional kernels were combined with LTN to realize the migration, and the CNN model was constructed during MNIST handwritten digit recognition with accuracy of 95.0%. In terms of model training, traditional light propagation modeling depends on approximate numerical methods that are inefficient and are limited by conditions such as layer spacing. The team therefore introduced a deep regression neural network (DRNN) to replace the traditional Rayleigh-Sommerfeld diffraction model; it mapped the complex light propagation process into differentiable neural network layers using the structural reparameterization algorithm, trained the DRNN using finite-difference time-domain (FDTD) simulation data to shorten the optimization time significantly while maintaining accuracy, and improved the chip integration density and computational energy efficiency. This work demonstrated the considerable potential of in-memory optical computing frameworks and next-generation AI platforms.

ONN functionality is limited by the scale of the on-chip integration and the requirement for a coherent light source. Cui et al.[17] proposed the spectral convolutional neural network (SCNN) with material meta-imaging functionality for this purpose. The optical convolutional layer is realized by integrating an ultra-large-scale, pixel-aligned spectral filter on a complementary metal-oxide-semiconductor (CMOS) image sensor. This layer enables realization of a highly parallel spectral vector inner product of the incident incoherent natural light (i.e., direct information carriers), thus enabling in-sensor optical analogue computation with extremely high energy efficiency. Integration of the large-scale spectral modulation structures (e.g., hypersurfaces, pigment layers) onto the surfaces of CMOS image sensors requires strict alignment of the modulation units with the sensor pixels, which is extremely demanding on the manufacturing process. Additionally, conventional ONNs rely on high-precision but high-cost hypersurface structures that are difficult to mass-produce. Therefore, the research team used two different spectral modulation schemes: a high-precision modulation based on optical hypersurfaces and a low-cost pigment-based mass production scheme. The former realized the potential for full light-field sensing by optimizing the hypersurface design, while the latter could be flowed on 12-in wafers using a proven semiconductor process, reducing the integration costs significantly and improving production feasibility. This represents the first integrated optical computing scheme using natural light. The study also used the same SCNN chip for a completely different real-world complex task and achieved accuracy of more than 96% in pathology diagnosis and almost 100% accuracy in facial anti-spoofing at video rates. These results demonstrate that in-sensor edge computing chips using natural light are feasible and scalable for application to a variety of portable terminals.

The DNN proposed by Fu et al.[18] was designed to operate in the near-infrared region and was nano-printed on a CMOS chip via vibrational mirror two-photon nanolithography using 10 nm axial nanosteps to achieve a neuron density of 10^8 neurons/mm^3. Their aim was to resolve challenges for conventional sensors and electronic neural networks with optical problems, including symmetric and asymmetric decryption performed by single-layer perceptrons and direct phase retrieval performed by multilayer DNNs. The compact form factors and the lithographic fabrication techniques of DNNs can be integrated directly into the optoelectronic sensors, thus allowing for co-integration of an optical diffraction layer with an additional electronic neural network layer or for use of the sensor's nonlinear response as a nonlinear activation function in such a way that deep neural networks are formed. This energy-efficient combination of machine learning and on-chip integration may provide a basis for new highly integrated systems based on deep neural networks, which can then perform the task of reasoning about information in the optical domain more quickly, more efficiently, and more robustly than conventional digital neural networks. As a result, this development may have a transformative effect on sensing, microscopy, high-precision laser nanolithography, quantitative phase imaging methods, and optical computing applications.

2.3. Free-Space DNNs

Free-space DNNs are built using bulk optical elements (e.g., 4*f* systems, spatial light modulators (SLMs), digital micromirror devices). A typical example is the optoelectronic hybrid CNN based on the 4*f* system, which realizes the convolutional kernel function by placing the phase mask in the Fourier plane, and although the convolutional kernel cannot be reconfigured, it has demonstrated low training costs and efficient prediction performance. In addition, the D^{2}NN based on free-space diffractive elements realizes all-optical learning through phase modulation of the multilayer diffractive surfaces, and the linear relationship between the thickness of the diffractive units and the phase difference ($h = \lambda\phi/(2\pi\Delta n)$) provides the theoretical basis for their hardware implementation[19].

When an object is located behind a random and unknown scattering medium (e.g., a random phase diffuser), its image information will be distorted severely, thus rendering traditional imaging and classification methods ineffective. To solve these classification problems, Bai et al.[20] proposed an all-optical processor that uses broadband illumination detected by a single pixel to classify unknown objects directly through an unknown random phase diffuser. To construct a DNN that can process information about objects behind an unknown random diffuser, the diffractive layer design and optimization are necessary to ensure that the network operates at the speed of light and achieves efficient object classification. Therefore, Bai et al.[20] used deep-learning techniques to optimize a set of transmissive diffractive layers that formed a physical network that was capable of all-optical mapping of the spatial information of an input object located behind a random diffuser into the output light power spectrum detected via a single pixel on the output plane of the DNN. During training, several randomly generated phase diffusers were used to aid the generalization capability of the DNN. After completion of this single-pass deep-learning-based training process, the resulting diffractive layers could be fabricated physically to form a single-pixel network. The accuracy of the framework was demonstrated numerically using broadband radiation to classify unknown handwritten digits through random new diffusers that were never used in the training phase, with a blind-test accuracy of 87.74 ± 1.12% being achieved. This single-pixel, all-optical object classification system using random diffusers is based on passive diffractive layers that can process broadband input light and operate in any part of the electromagnetic spectrum by simply scaling the diffraction features based on the relevant wavelength range.

Existing spectral reconstruction methods require either bulky equipment or complex electronic reconstruction algorithms, which have limited the performances and applications of these systems. Wang et al.[21] proposed a scalable all-optical opto-intelligence spectrometer (OIS) for high-precision spectral reconstruction of spatially coherent or incoherent multispectral input sources using DNNs to solve the image reconstruction problem. The OIS establishes a mapping relationship between the inputs and outputs and then transforms the spectral amplitude of the input source into the detected output plane intensity, where each detector corresponds to a specific spectral band. Additionally, Wang et al.[21] optimized the modulation layer phases using two types of mean-squared-error (MSE) loss functions, which enabled high-quality reconstruction of the input spectrum and ensured a reasonable high-contrast intensity distribution at the output plane. Random phases were introduced at the input plane and spatially incoherent inputs were simulated through a combination of multiple forward propagations and averaging at the output plane. Although the reconstruction quality decreased gradually with increasing aperture, the OIS could still reconstruct the required spectrum effectively. Therefore, these researchers designed a simple but effective electronic calibration module located at the back end of the optical reconstruction module to further improve the spectral reconstruction quality. After correction by the electrical calibration module, the reconstruction results matched the input results closely and surpassed those obtained via spatially coherent spectral reconstruction tasks. In the optical reconstruction module, the optical modulation layer phase had to be optimized for high-quality spectral reconstruction. The researchers corrected the optical reconstruction results via an electronic neural network to obtain new spectral intensity vectors. The optical modulation layer parameters remained fixed during weight optimization of the electronic neural network. The weight update process was guided by the loss function. Simulation experiments indicated that the OIS could implement high-precision spectral reconstruction under spatially coherent and incoherent light source conditions without a reliance on complex electronic algorithms, and integration with a simplified electrical calibration module improved its performance further. This framework considers spatial and partial temporal coherence carefully, thus solidifying its status as an important reference for the deployment of photonic neural networks in the spectral analysis of real-world objects. The all-optical DNN is complemented by a simple electronic module that makes the architecture exceptionally efficient, thus enabling fast inference and reducing the system complexity.

Yu et al.[22] proposed an ONN accelerator (the time-wavelength multiplexed photonic neural network accelerator, or TWM-PNNA) based on use of a time-wavelength multiplexing mechanism to combine optical computing with a distributed acoustic sensing (DAS) system to overcome the performance bottleneck caused by traditional electronic signal processing in real-time data processing. The system maintained more than 90% classification accuracy when the full-connectivity parameter retention was no less than 60%, thus laying a foundation for the construction of an all-optical, low-power, real-time monitoring system. In the case of unmodulated chirp, each bit sequence remains independent; however, in the case of interaction between modulated chirp and fiber dispersion, a pulse broadening phenomenon occurs that results in interference between adjacent bits, which then triggers inter-symbol interference within the channel, degrading the signal transmission quality and increasing the bit error rate (BER). When the photodetector superimposes power on signals at different wavelengths, the modulation chirp exacerbates the signal distortion, which then affects the computational accuracy of the convolution kernel in the TWM-PNNA system. In addition to pulse broadening, the modulation chirp causes spectral broadening of the signal, which increases the probability of spectral overlap occurring between neighboring channels in a wavelength division multiplexing system significantly. This spectral overlap increases the inter-channel crosstalk, which reduces the signal-to-noise ratio (SNR), and the combination of these effects affect the computational accuracy and the BER of the system. To reduce the chirp effect in the system, the researchers improved the TWM-PNNA system performance by 17.9% when compared with the original method for complex pattern recognition tasks by using a two-arm modulation method at a wavelength spacing of 0.4 nm. The parameter redundancy and the computational burden of the fully connected layer were other challenges faced by the team. The fully connected layer contained a large number of weighting parameters and direct implementation of these parameters would increase both hardware complexity and computational cost. The team optimized the model by applying a post-training pruning-aware tuning (PAT) technique that uses a mask matrix to eliminate any weakly correlated neuron connections. The experimental results showed that when the retention rate for the all-connection parameters was no less than 60%, the classification accuracy could be maintained at more than 90%, which offers a basis for model lightweighting. In addition, the team proposed use of modulators to realize optical full-connectivity computation; this approach breaks through the limitations of the fixed weights and numbers of ports in the full-connectivity operation of traditional wavelength selective switches by dynamically adjusting the attenuation coefficients of the modulators for each channel to load the weights and thus it improves the flexibility and scalability of the system. This study realized automatic optimization of ONN design using pruning and chirp modulation, which improved the accuracy and stability of optical computing, laid a foundation for the deployment and optimization of ONNs in practical applications, provided an innovative computational framework for DAS systems, and thus paved the way for all-optical convergence of DAS systems with computing systems.

Hua et al.[23] presented a large-scale integrated photonic accelerator composed of more than 16000 photonic components. The accelerator was designed to implement standard linear matrix-add-column (MAC) functionality using an innovative 2.5D hybrid advanced packaging methodology to design the logic, storage, and control functions that support the photonic matrix MAC operation in a co-integrated electronic chip. The aim was to achieve seamless integration of the electronic and photonic chips on a commercial scale to address the technical challenges faced by integrated photonics systems during the scaling-up process, e.g., ensuring scaled integration of the device clusters, consistent performance improvements, establishing a standard design and verification process for complex circuits, and providing large-scale system packaging. The system achieved high-speed computation at up to 1 GHz with a low latency of 3 ns per cycle, an average 7.61-bit computational accuracy, throughput of 8.19 TOPS, and energy efficiency of 2.38 TOPS/W (including the lasers), and addressed a combinatorial optimization problem, e.g., the Ising model, with higher throughput than the commercial NVIDIA A10 chip. This system achieved a two-orders-of-magnitude improvement in both latency and computation time when compared with commercial NVIDIA A10 GPUs. Responding to the challenge of developing computational models and algorithms to fit their new hardware, the team designed a heuristic recursive algorithm based on MAC operation of the photonic loop, which introduced laser noise, analog drive noise, and digital control noise into the circuit. They then tuned the SNR actively to achieve a controllable bit-flip such that the algorithm converged to the Ising model's ground state distribution through iterative matrix MAC calculations, while simultaneously taking advantage of the optical range length limitations in the photonic circuit. The optical range length limiting delay property in photonic circuits enables nanosecond-scale low-delay computation, which significantly accelerates the solution process of the combinatorial optimization problem.

In multiple scattering scenarios, the wavefront structure of the light passing through a complex medium is destroyed, producing a highly randomized scattering pattern in the output signal, and the original structure can then no longer be traced from the output image. Most conventional methods to cope with these problems are based on the optical memory effect. However, there is a natural limitation on the applicability of the optical memory effect: it only holds when the scatterer is sufficiently thin and the medium structure changes little, but when the system thickness increases or the number of layers of the medium increases, the angular correlation decreases dramatically and is ultimately lost; this leads to a complete failure of the traditional image reconstruction strategy, which relies on angular retention. To break through this physical limitation, Zhang et al.[24] constructed a three-layer ONN that operated in the all-optical domain with a field-of-view (FOV) that was enlarged by a factor of 271. They realized image reconstruction in strongly scattering media without an optical memory effect. The first optical convolution module in the network receives the input random scattering map and then maps it into multiple feature sub-maps by encoding the phase structure, realizing both decoupling of the primary information and channel separation; the second layer then extracts the higher-order spatial modes by using the composite phase kernel to enhance texture and structure perception within the image; finally, the third layer is the fully connected output layer, which is responsible for fusion of the features of the various channels to enable output of the reconstruction results for the target image obtained by interferometric imaging. On the basis of the Fourier optics theory, the device uses a fully optical structure to accomplish multilayer convolution, feature extraction, and image reconstruction, thus eliminating the reliance on angular correlation and reference wavefronts experienced in traditional imaging systems. The device can be reconfigured dynamically for ultra-fast multi-task image reconstruction with computational power of 15.7 million operations per second (POPS). In the multilayer strong scattering case, the angular range of the optical memory effect decreases as the number of scattering layers and the layer spacing increase, which makes it difficult for conventional methods to extract features from the images effectively. The researchers thus used the vortex phase function in the convolutional kernel to extract the edge information that was embedded in the scattering pattern. Experiments showed that the convolutional ONN with the vortex kernel has higher Pearson correlation coefficient values than that without a vortex kernel, and that it can effectively extract the object edges in any direction, which improves its feature extraction ability in strong scattering scenarios. Additionally, misalignment between the layers of the convolutional ONN can cause its performance to degrade significantly. Therefore, the team used a symmetric Gaussian phase distribution to calibrate the system's misalignment error to mitigate the effects of this problem on the model performance. This research not only realized the paradigm leap from memory-dependent to autonomous perception in theory, but also demonstrated the practical feasibility of using ONNs as information processing systems in engineering implementations. The architecture of the network is detached completely from the electronic computing platform, which fundamentally breaks through the speed and power consumption bottlenecks of existing image reconstruction systems. Furthermore, the researchers noted that their network can be integrated on the nanometer scale via advanced processing technologies such as 3D laser direct writing to construct a chip-on-chip photonic neural network device, thus laying the processing foundation for development of a new type of computing system that replaces computation with light in the future.

Optical single instruction, multiple data (SIMD) logic operations have been implemented using waveguide-based photonic circuits. However, the scalability of the implementation method is limited because of the difficulty involved in achieving accurate, large-scale fabrication. Because of the inherent parallelism of free-space propagation, the use of DNNs offers the potential to solve this computational problem. However, existing approaches to DNN-based logic operations typically involve performing only one or a few logical operations on a small number of bits. Furthermore, these methods still require computational encoding and decoding of the inputs and outputs, which represent a significant challenge for end-to-end optical computing. Mashiko et al.[25] proposed an optical computing architecture called diffraction casting (DC) to perform flexible and scalable parallel logic operations. By incorporating the shadow casting (SC) scheme and a DNN, DC can perform 16 switchable optical SIMD logic operations on 256 bits, thus resolving the computational problem for multi-data logic operations. Unlike the SC scheme, which is based on geometrical optics, DC is based on wave optics. The optical cascade for DC comprises reconfigurable illumination, use of diffractive optical elements (DOEs), and input layers. The illumination pattern and the DOEs are designed to perform 16 logic operations on any binary input image pair, and the output intensities of the optical cascade are binarized to generate the final result. This method only requires the illumination pattern to be changed to alter the logic operations. DC eliminates the need for computational encoding and decoding of the inputs and outputs, which forms an inherent bottleneck in SC schemes and in the DNN-based logic operations described previously. Mashiko et al.[25] enabled end-to-end all-optical computing by introducing a buffer around the input region that eliminated the need to encode and decode the input and output, respectively. One problem with the use of DC in practical applications is its low energy efficiency. This issue can be resolved by using illumination with phase modulation and optimizing the physical conditions (including the layer interval, the buffer, and the image sensor). Additionally, alignment errors may occur in physical system implementations. The researchers proposed several methods to compensate for these alignment errors, including introducing alignment errors or model errors during computational training to enhance system robustness, use of a closed-loop process to feed alignment errors or model mismatches back to the controllable optics, and use of integrated chip fabrication techniques to reduce the alignment errors and model mismatches significantly. In previous studies, the optical diffraction processes for each spatial region of the logic operations and the bits had to be designed separately. This method, however, allows for interference occurring among the spatial regions of the bits and the logic operations, and thus allows these regions to be densely positioned based on the illumination pattern and the end-to-end design of the DOEs. This advantage provides high scalability and integration capabilities for all-optical operations, thus eliminating the need for computational encoding and decoding.

Most existing photonic computing accelerators are focused on discriminative neural networks. Large-scale generative photonic computing machines have remained largely unexplored, partly because of poor data accessibility, accuracy, and hardware feasibility. Zhan et al.[26] used random light scattering in disordered media as a native noise source and then applied large-scale diffractive optical computing techniques to generate images from that noise, thereby addressing the information processing and computation problems by pursuing the spatial parallelism of the light alone to realize hardware consistency. This large-scale spatial speckle pattern was combined with DONNs to provide hardware-platform consistency for the photonic generative neural network (all spatial pattern modulation), and the excellent statistical properties of the scattered signal ensure the high quality of the noisy signal. However, efficient mapping from random speckle patterns to target images requires correlation of the two using precise optical encoding strategies. The researchers therefore proposed two encoding strategies: random position encoding and physics-aware position encoding. The latter approach improved the quality and consistency of the generated image significantly by reducing the target image dimensions using a variational autoencoder (VAE) and then correlating the image with the illumination position of the scattering medium. Additionally, the optoelectronic generator component used two ONN architectures:

cascaded and parallel. The cascaded network implemented deep learning via layer-by-layer feature extraction, and the parallel network performed feature aggregation using independent feature extraction modules to adapt to the requirements of the different datasets and generation tasks. In the experiments, the scanning range of the noise input was constrained by the system aperture. To enhance the sample capacity, the researchers proposed replacement of the phase grating used for scanning with specially designed grayscale patterns. Unlike the previous binary scanning methods, this grayscale wavefront shaping strategy provided additional channels for noise sampling. Compared with the recent generative network models deployed on various items of optical computing hardware, this scheme improved the image generation capability by an order of magnitude, from generating a single class to generating 10 classes, and thus further expanded the application prospects of optical computing in the generative modeling field. This work has therefore paved the way for more sophisticated applications, including real-world data augmentation and multimodal generation.

Among the various ONNs studied in the optical neuromorphic engineering field, spiking neural networks (SNNs) have been remarkably successful in simulating the computational principles of the human brain. The event-based spiking properties of optical SNNs offer capabilities that are difficult to achieve or are unmatched by other neural networks in terms of their low-power operation, speed, temporal processing, analogue computation, and hardware efficiency. Ahmadi et al.[27] proposed the pioneering free-space optical deep spiking convolutional neural network (OSCNN), which uses free-space optics to improve energy efficiency and processing speeds while also maintaining high pattern detection accuracy. Specifically, the model applies a Gabor filter at its initial layer for efficient feature extraction and uses optical components such as intensity-delay conversion devices and synchronizers, which are designed using off-the-shelf optics. The OSCNN consumes only 1.6 W and has a processing time of 2.44 ms, which is considerably higher than that of conventional electronic CNNs on GPUs (which typically consume 150–300 W and have processing times of 1–5 ms), and is competitive with other free-space ONNs. To ensure nanoscale component alignment accuracy, an advanced micropositioning system and an active feedback control mechanism were proposed. To improve the signal integrity, high-quality optical components, error correction algorithms, adaptive optics, and anti-noise coding schemes were used. The integration of the optical and electronic components was optimized by designing high-speed photoconverters, custom integrated circuits, and advanced packaging techniques. In addition, high-efficiency, compact semiconductor laser diodes were used and new cooling strategies were developed to minimize power consumption and reduce the device footprint.

The dependence of conventional DONNs on fully coherent light sources limits their application to real and complex optical environments. Although the implementation of DONNs under partially coherent light by considering DONNs with coherent and incoherent light is possible, this approach ignores the coherence structures. To address this problem, Dong et al.[28] proposed the partially coherent DONN (PC-DONN). This approach introduces the spatial coherence of the light as an auxiliary dimension (i.e., the parameter related to the spatial coherence length) and develops a training algorithm that is applicable to a light source of any spatial coherence, thereby expanding the algorithm's practical application to low-coherence or even incoherent environments and addressing the computational problem that traditional DONNs only operate under fully coherent light. The team constructed a unified coherence representation framework based on the Gaussian-Schell model, which uses the coherence length l of the light field to characterize the spatial extent of the phase correlations in the light field, thus enabling a precise description of a wide range of light source conditions from fully coherent to fully incoherent. They demonstrated the effectiveness of their proposed PC-DONN via recognition of handwritten digits in the MNIST database through both simulations and experiments. A PC-DONN trained with l = 0.2 mm maintained accuracy above 82% when the light coherence length was reduced to l = 0.05 mm and reached 90% with further optimization. Conversely, the accuracy of conventional coherent DONNs decreased significantly from 91% to 26% under these conditions, thus fully demonstrating the robustness and effectiveness of the proposed method under partially coherent light conditions.

Table 1 summarizes the core hardware parameters of the optical networks related to high-precision detection, including physical hardware type, detector selection, operating wavelength band, and bandwidth. As can be seen from the comparison, the hardware configurations of different systems are highly matched to their application scenarios: for example, the all-optical processor designed for scattering medium imaging adopts a wide-bandwidth design to accommodate broadband illumination, while the on-chip integrated DGNN selects a narrow-bandwidth communication wavelength to adapt to the photonic chip architecture. These parameter differences provide important references for hardware selection and scenario adaptation in optical neural networks.

3. APPLICATION PROSPECTS

Optical nonlinear effects allow optical signals to be encoded and decoded using specific physical rules; this property is not limited to performing complex computational tasks and opening multiple novel application scenarios in sensing technologies based on physical laws. When simulating complex physical scenarios while following the laws of physics, optical coding techniques may be required to work in concert with simple electrical networks to convert, amplify, or transmit signals to ensure sta-

Table 1. Hardware, detector, and band parameters of the optical computing system

System	Physical hardware	Detector	Band	Bandwidth
Ultra-deep neural network[8]	N-LEM arrays (new device with dual optoelectronics outputs)	Light-emitting diode (LED)	Ultraviolet range	–
All-optical processor[20]	3D-printed transmissive diffractive layer	Single-pixel spectral detector	0.6 –1.2 mm	60 mm
Diffractive graph neural network (DGNN)[13]	On-chip optocouplers	Optoelectronic detector	1.554 μm	Narrow bandwidth
nPOLO framework[14]	Liquid crystal SLM, mirrors, continuous wave lasers	Camera (CAM)	Low-power continuous-wave lasers (visible or near-infrared band)	–
OIS[21]	3D-printed diffractive layer	Camera (CAM) Single-pixel mixer/AMC	Visible band Terahertz band	–

ble operation and precise control of the entire system.

In an innovative scheme[29] to implement a nonlinear ONN using linear optical paths, a joint team from Yale University and the École Normale Supérieure Paris used a spherical chamber of the size of a golf ball, which was partially lined internally with a reconfigurable array of tiny mirrors, with the remainder lined with a diffuse reflectance coating. When they launched a laser beam into a hole in the chamber, the light bounced around the inside before emerging from the other hole as a speckle pattern. The researchers realized that the chamber could act as a "reservoir computer"–a special type of neural network. On average, the light bounces off the surface of the chamber thousands of times before exiting to produce a speckle pattern that contains rich information about the correlation between the pixels in the input data, which serves as the basis by which the neural network achieves intelligence. Rather than random scattering, the École Polytechnique Fédérale de Lausanne (EPFL) team crafted a zigzag optical path that allowed the beam to be scattered over an SLM with four different copies of the input. The pixels in each copy could be tuned using trainable parameters, thus allowing the system to be trained like a traditional neural network. Although the light was reflected only four times, this was sufficient to produce the effective nonlinear effects required to implement a simple image classification task successfully. Furthermore, the multiple encoding design makes the system highly noise-resistant. A team at the Max Planck Institute for the Science of Light designed a network structure based on bidirectional scattering. Unlike traditional neural networks, which have a unidirectional information flow, the light waves in their design can propagate through the network in both directions. The system naturally generates the desired nonlinear dependences by encoding the data and parameters in neurons. Preliminary simulations indicate that this optical implementation may reduce the training process, which normally takes hours, to the millisecond level, thus indicating great potential for practical applications. All three teams relied on changing how the data were encoded, rather than on nonlinear optical materials, to achieve nonlinear functionality. The essence lies in not encoding the data in the light field, but instead encoding it in some parts of the system that interact with the light. In this way, nonlinear functions can be computed easily, and neural network implementation becomes simple. Because these optical computations rely on changing the way the data are encoded to achieve nonlinear functionality, they can provide new approaches for novel, efficient physical neural network computing architectures in the future.

Notably, traditional reflective imaging techniques and imaging methods that rely on the laws of physics, e.g., diffraction and interference, can undergo innovation and optimization through a unifying optical computing framework. Optical computing uses the parallel processing power and high speed of light signals to integrate these diverse imaging mechanisms into an efficient and flexible information processing system. This means that both the simple reconstruction of object contours and the complex visualization of physical processes can be realized through well-designed optical systems without a reliance on traditional imaging equipment or algorithms. Within this framework, imaging and computation processes are no longer clearly demarcated but are deeply integrated, i.e., computation and detection in one. This approach greatly improves information processing efficiency and accuracy, and opens new avenues to explore complex physical phenomena, develop new sensing technologies, and promote the advancement of optical computing technology.

In the wavefront detection field, the current development focus for adaptive optics technology is mainly on measurements contaminated by a single isoplanatic aberration within a relatively small FOV and does not address the problem of anisoplanatism effectively. Moreover, extension of the measurement FOV beyond this scale while also maintaining high accuracy remains challenging, and the complex optical configurations used in these technologies often produce bulky and costly system designs. Digital adaptive optics estimates and corrects for spatially nonuniform aberrations via a meta-imaging sensor. However, the iterative reconstruction process involved incurs significant computational costs, which hinders its practical application in real-time observations of dynamic turbulence. Guo et al.[30] proposed a light-field-based plug-and-play wide-field wavefront sensor (WWS), which can observe atmospheric turbulence directly at a frequency of

30 Hz over a wide FOV spanning 1 100". Their experimental measurements were consistent with the von Karman turbulence model, and the root-mean-square error of the predicted wavefront at the central wavelength of 525 nm was reduced significantly by 195 nm for the WWS when compared with a conventional wavefront sensor. Optical coherence tomography (OCT), which is based on the scattering principle, has a wide range of potential wavefront and wireless sensing applications. Combining optical networks can help to address problems in OCT-based wavefront sensing while also achieving high-resolution determination and sensing. Applying deep-learning techniques to wavefront sensing can improve the sensitivity and dynamic range significantly. For example, optical differentiation wavefront sensors use deep-learning-driven optical differentiation techniques to realize higher spatial resolution and scalable dynamic ranges by increasing the filter size. In terms of robustness improvement, combining a nonlinear amplitude filter with a CNN-based reconstruction method produced improvements in the sensitivity, dynamic range, and robustness of wavefront sensing. The ability of the CNN architecture to accurately predict hybrid wavefronts that combine Zernike aberrations and random-pattern phase profiles highlights the versatility of zonal-reconstruction-based CNNs for complex wavefront shape handling. OCT relies on light backscattered from different sample regions to generate 3D images and uses different positioning techniques to obtain axial and transverse information. By optimizing the optical network structure and improving the light source coherence, imaging speeds can be improved by reducing the optical signal attenuation during transmission, and imaging depths can be increased by adjusting the light source parameters or using multi-mode fibers. The convergence and innovation of these technologies will drive further OCT development in areas that include biomedical imaging and industrial inspection.

In terms of target recognition and classification, Xu et al.[31] proposed a hardware-aware training and pruning method that enhances the robustness of photonic neural networks (PNNs) by introducing a hardware-algorithm co-design approach that focuses on encouraging the weights to move to the variance-insensitive region and training the neural network parameters toward the noise-robust and energy-efficient region. Furthermore, the variance-insensitive region coincides with the region that requires minimal tuning power for weight assignment and thus reduces power consumption. The experimental classification accuracy of a two-layer CNN based on this method reached 95.0% for the MNIST dataset classification, which is comparable to the theoretical value.

Optical computing is inherently characterized by high speeds and low power consumption, which give optical networks significant advantages when handling large-scale data and high-performance computational tasks. Because of the characteristics of optical computing, including its high bandwidth, low latency, high energy efficiency, and parallelism, it has broad application prospects in fields including wireless signal processing, high-speed data centers, and edge computing. The ability of optical network computing to process data at the speed of light increases processing speeds and reduces energy consumption, thus making it uniquely advantageous for modern computing and communications.

4. CONCLUSIONS AND OUTLOOK

The rapid development of ONNs has opened a new research avenue for computing power enhancement. Because optical signal characteristics include low power consumption, low latency, high parallelism, and large bandwidths, optical computing based on neural network architectures is promising for processing of spatial signals, temporal signals, and on-chip information. This review has addressed the practicality of ONNs, introduced the fundamental principles of ONN-based computing, described the challenges involved in different application scenarios, and highlighted the important role of optical computing in the future information society.

The development of optical computing in terms of the four dimensions of scale, node, depth, and power consumption is highly consistent with Moore's Law. With its expanding computational scale, continuously improving node integration, increasing algorithmic depth and architectural hierarchy, and gradually decreasing power consumption, optical computing exhibits a pattern similar to the increasing numbers of transistors, improving performance, and decreasing energy consumption described by Moore's Law, thus opening a new path to high-efficiency computing. However, the current methods for training these models efficiently are limited by computer simulation techniques on digital computers. Xue et al.[32] developed an in situ machine learning method called FFM learning in free-space and integrated optical systems that implements a computationally-intensive training process on a physical system by mapping the optical system onto a parameterized in situ neural network and then enabling self-learning oriented toward application goals. Most machine learning operations are performed in parallel and efficiently in the field, thus alleviating the limitations of numerical modeling. Traditional GD algorithms rely on backpropagation mechanisms, but it is difficult to implement backward light propagation directly in optical systems because of the unidirectional nature of light propagation and the difficulty of physical calibration. This problem is particularly prominent during experiments, where the complexity of the light propagation paths and their susceptibility to environmental disturbances lead to deviations between gradient calculations from offline modeling and the response of the actual physical system. By using the spatial symmetry and Lorentz reciprocity of photon propagation, Xue's team physically circumvented the need for backpropagation by designing the physical system symmetrically, and calculating the gradient based on the data and the output light field measured during error prediction. They then updated the parameters efficiently using a GD algorithm to ensure that the gradient calculations for the error propagation do not need to be performed by reversing the optical path, but rather by using the symmetrical forward propagation path. FFM learning shows that training of the deepest ONN with millions of parameters can achieve accuracy comparable to that of the ideal model. The system supports all-optical focusing through scattering media with resolution ranging up to the diffraction limit; it also enables parallel imaging of out-of-sight hidden objects at frame rates in excess of kHz and all-optical processing at light intensities as low as the subphoton/pixel (5.40×10^{18}-energy-efficiency-per-watt-per-second) level at room temperature.

Emerging physical platforms offer increased speed, energy efficiency, and scalability through the use of new technologies intended to overcome the traditional Moore's Law barriers for more powerful and efficient smart applications. Sedov et al.[33] proposed a novel neuromorphic network architecture based on exciton-polariton condensed lattices that were intricately connected to each other and pumped using nonresonant optical energy. Their network uses a binary framework in which each neuron performs binary operations enabled by the spatial coherence of pairwise coupled condensates. This coherence, which is generated by the propagation of polarized bullets, ensures efficient network-wide communication. The binary neuron switching mechanism driven by nonlinear repulsion of the exciton component of the polarizers offers both computational efficiency and scalability advantages over sequentially weighted neural networks. When compared with sequential or pulse-coded binary systems, the network realizes parallel processing and improves computational speeds. In image recognition applications, the predicted classification accuracy of 97.5% is impressive. In speech recognition, the system achieves a classification accuracy of ~68% on a subset of 10 classes to exceed the performance of the traditional benchmark hidden Markov model mixed with a Gaussian model, thus showing its potential to outperform existing polar neuromorphic systems.

Attainment of optical computing with higher computing power awaits further investigation. The upsurge in the number of parameters used in neural network models poses a significant challenge for traditional tensor computing hardware, whereas optical chip performance is limited by the device size and the photonic integration scale. To address these challenges, Meng et al.[34] proposed an ultra-high computing density optical tensor processing unit. The unit uses a thermally tuned-free MRR as its core unit and uses wavelength division multiplexing to perform real-value weight adjustments through wavelength tuning and modulator bias-point changes. The device achieved ultra-high computational density of 34.04 TOPS/mm^2 and offers a new technological path toward optical computing with high computing power. This scheme extends the data input dimension effectively by introducing the optical/microwave multidomain multiplexing technique, which enables optical tensor operations, improves information access flexibility, and reduces the dependence on high-speed analog-to-digital and digital-to-analog converters.

When exploring applications for complex ONNs, consideration of their applicability and necessity is essential. Not all simple problems need to be processed using complex ONNs, which may lead to the phenomenon of overfitting, where the model fits the training data too perfectly to be generalized to new data. Therefore, our approach to specific problems should be selected rationally on the basis of the scale and complexity of the problem. For problems that require high flexibility and scalability, an optical network architecture with online connectivity can be used. This architecture can better adapt to variations in the problem while maintaining system efficiency by adjusting the connection weights dynamically. When addressing problems that involve spatial information or require high-precision light-field modulation, a spatial MZI may be a more appropriate choice. An integrated MZI network can effectively meet the requirements of application scenarios that do not require high computational accuracy but do require high system stability. For complex problems that need to use the free-space propagation properties of light, free-space scattering networks may provide an effective solution. These networks enable efficient optical signal processing through the design of specific scattering structures. Specifically, analyzing the problems initially in detail helps to improve system performance and efficiency while avoiding resource wastage and overfitting problems caused by overcomplexity.

One significant challenge in the future of computing is in-memory computing, i.e., the use of photonic memory to enable almost instantaneous operations and responses. However, this field is currently constrained by several factors, including slow switching speeds and limited programmability. A team of researchers from the University of Cagliari in Italy, the University of California at Santa Barbara, the University of Pittsburgh, and the Institute of Science Tokyo has developed a new magneto-optical memory by using cerium-substituted yttrium-iron garnet miniature magnets as data storage units and by controlling the optical signal paths through the materials. This device can achieve switching speeds that are up to 100 times faster than the current state-of-the-art photonic integration technology while consuming only approximately 1/10th of the energy of its predecessor. Moreover, it can be reprogrammed multiple times to adapt to the requirements of different tasks. Conversely, although existing high-end optical storage devices can typically withstand only up to 1 000 write operations, this magneto-optical memory can support more than 2.3 billion rewrites, indicating that it may have a near-infinite service life. The magneto-optical material also allows use of an external magnetic field to modulate how the light travels, which led the team members to program the miniature magnets using an electric current to store information. These magnets, in turn, determine how the light travels inside the material, enabling implementation of complex operations, including matrix-vector multiplication, an important component of all neural network architectures.

Solving time-evolving partial differential equations (PDEs) using on-chip photonics remains a great challenge because of the difficulties of large coefficient matrix photonic computation, high accuracy requirements, and error accumulation. Yuan et al.[35] overcame these challenges by realizing a microcomb-driven photonic chip and then introducing time-division multiplexing and matrix partition techniques into PDE photonic solving; these techniques can solve PDEs with a large coefficient matrix on a photonic chip of limited size. In densely integrated photonic chips, thermal crosstalk may cause microloop resonance wavelength drift and arithmetic errors. The researchers reduced thermal coupling through thermal isolation design and strict temperature control mitigation by using silicon-based waveguide structures prepared via deep-ultraviolet lithography, while the temperature fluctuations are controlled to within ±0.1°C via an active temperature control system in the experimental environment. In high-temperature sensitivity experiments, the solution errors for Poisson's and Laplace's equations were reduced to approximately 4% by the measures above. To support the solution of multiple PDEs (e.g., the heat conduction equation, Burgers' equation) on the same chip, a flexible wavelength allocation and operational unit multiplexing scheme must be designed. The researchers realized dynamic adaptation of different equation types via reconfigurable microcomb channel spacing techniques (e.g., by selecting free spectral range spacing to enhance the signal-to-noise ratio of the carrier) and modular silicon photonic engine design. By adjusting the pump laser detuning parameters, the microcomb channel spacing can be configured flexibly to optimize the efficiency for solving specific equations. Time-evolving PDEs, including the heat equation with the first-order time derivative, the wave equation with the second-order time derivative, and the nonlinear Burgers' equation, are solved with accuracy ranging up to 97%. Furthermore, parallel solving of the Poisson's equation and Laplace's equation is demonstrated experimentally on a single chip, with accuracies of 95.9% and 95.8%, respectively. This study demonstrated the potential of photonic computing for high-performance numerical problem solving and lays a foundation for the construction of high-precision, ultra-high-speed next-generation computing platforms that will promote the development of on-chip photonic computing.

Integrated optical computing systems are gradually dominating the development trend. Integrated computing systems based on silicon photonics platforms offer many advantages, including CMOS process compatibility and a high degree of integration. Therefore, they now represent a mainstream solution and are becoming increasingly diversified. For example, MZI arrays can be integrated into silicon photonics platforms to implement matrix operations, optical waveguide delay lines can be used to implement reservoir operations, and sub-wavelength metamaterials can form optical operators. In the future, the heterogeneous integration process will need to be upgraded to improve system integration further for device miniaturization. Existing approaches use ANNs or SLMs to reconstruct images that were scrambled after propagation through fibers. Despite these advances, achieving direct optical image transportation through multi-mode fibers (MMFs) using integrated optical elements with micron-scale footprints remains challenging. Yu et al.[36] developed a miniature DNNs that was integrated on the distal face of an MMF for direct transmission of all-optical images over optical fibers. The DNNs has a footprint of 150 μm by 150 μm and were fabricated on the facet of a 0.35-m-long MMF using 3D two-photon nanolithography. The system achieved a minimum image reconstruction feature size of approximately 4.90 μm over a FOV of 65 μm by 65 μm when imaging handwritten digits. Additionally, exploration of optical computing using all-optical architectures should also be performed in parallel, and will mainly include exploring nonlinear optical device solutions to implement all-optical neurons, examining optical interconnection technologies to realize flexible data transmission, and investigating optoelectronic memristor devices to achieve low-power memory and real-time information processing. The goal of optical computing technologies, whether performed using all-optical or intelligent optoelectronic hybrid architectures, is to solve real-world tasks. Therefore, reconfigurable optical devices must be explored to solve multiple problems flexibly. In the future, the tasks that are achievable by optical computing will not be limited to simple reasoning tasks. These tasks will extend to finance, energy, astrophotonics, in-vehicle computing, medical mobile Internet, and other fields with high demands for computing power and a broad market. This will produce a positive feedback loop for research, further promoting the development of optical computing and ushering in a new era of AI.

In astronomy, optical computing presents revolutionary opportunities for use in large-scale sky surveys, exoplanet detection, and gravitational wave signal processing. The ability of ONNs to process massive datasets in parallel with minimal energy consumption is particularly advantageous for the analysis of astronomical images and time-series data. Cui et al.[37] trained a CNN model using all-sky camera data from the Xuelong (China's third generation polar icebreaker and research vessel) to classify all-sky space-based cloud images automatically based on image quality, application context, and other factors; they aimed to realize an automatic classification algorithm with high robustness and adaptability that will provide important aid in astronomical site selection and to produce a highly maneuverable automatic classification method for all-sky camera data. ONNs can enhance real-time data analysis in radio interferometry significantly, improving both the resolution and accuracy of astronomical observations.

Additionally, the ability of optical computing to perform high-speed spectral analysis will enable the detection of faint exoplanetary signals by distinguishing rapidly between noise and astrophysical sources. Because next-generation telescopes will produce increasingly large datasets, the integration of optical computing into astronomical research could lead to groundbreaking discoveries and a deeper understanding of the universe.

Optical networks play important roles in the astronomy field, promoting progress in astronomical research because of their efficient data transmission and remote observation assistance. Wang et al.[38] analyzed and discussed typical applications of machine learning, ANNs, CNNs, and generative adversarial networks in candidate identification. Additionally, they proposed that AI technology offers significant advantages in pulsar candidate identification when compared with traditional manual screening; this technology reduces the misjudgment rate under noise interference significantly by performing multi-scale feature extraction (e.g., wavelet transforms in the time-frequency domain) with an integrated learning strategy, which results in an F1-score improvement by >30%, and completes initial screening of the candidate within milliseconds, thus reducing the manual analysis workload and the time cost. Simultaneously, these techniques can also deal with complex nonlinear relationships and discover patterns and features that are difficult to detect using traditional methods, thus improving the identification accuracy. In addition, the application of AI technology brings new perspectives and tools for pulsar survey data processing, using the t-distributed stochastic neighbor embedding (t-SNE) algorithm to visualize identification results, demonstrating the distributions of pulsar and non-pulsar samples in multidimensional space, and providing researchers with a more intuitive in-depth data analysis, thus promoting the development of pulsar research and the entire astronomy field.

Optical computing is expected to play an essential role in astronomical data processing. The vast amounts of data generated by modern telescopes, e.g., the Square Kilometre Array (SKA) and the Vera C. Rubin Observatory, require highly efficient computing solutions. The advantages of optical computing, which include parallel processing, high bandwidths, and energy efficiency, make it an ideal candidate technology for handling large-scale astronomical datasets. Future research will explore how photonic computing architectures can accelerate image reconstruction, signal processing, and real-time data analysis for astrophysical discoveries, thus potentially revolutionizing our understanding of the universe.

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2021YFC2202100) and the National Natural Science Foundation of China (12473090).

AI DISCLOSURE STATEMENT

The artificial intelligence tool used is DeepSeek. AI is used by uploading references and sending commands. After AI output, the original literature is read thoroughly and its validity is judged manually. The summary of the challenges encountered and their solutions in the research process of the references shown in Section 2 of the paper was influenced by the use of artificial intelligence tools and verified for accuracy by our team. This is the place where DeepSeek is used in this paper. The authors carefully reviewed, edited, and revised the DeepSeek-generated texts to their own preferences, assuming ultimate responsibility for the content of the publication.

AUTHOR CONTRIBUTIONS

Qichang An conceived the ideas, designed and implemented the study. Huan Wang collected data and performed statistical analyses. Xiaoqian Zhang wrote the paper. Hongchao Zhao revised the paper. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

[1] Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, **79**(8): 2554−2558.

[2] Farhat, N. H., Psaltis, D., Prata, A., et al. 1985. Optical implementation of the Hopfield model. *Applied Optics*, **24**(10): 1469−1475.

[3] Shen, Y. C., Harris, N. C., Skirlo, S., et al. 2017. Deep learning with coherent nanophotonic circuits. *Nature Photonics*, **11**: 441−446.

[4] Lin, X., Rivenson, Y., Yardimci, N., et al. 2018. All-optical machine learning using diffractive deep neural networks. *Science*, **361**(6406): 1004−1008.

[5] Zhang, N., Huang Z. Q., Zhang, Z., et al. 2024. Progress in optical computing and optoelectronic intelligent computing. *Chinese Journal of Lasers*, **51**(18): 1800001. (in Chinese)

[6] Shen, S. Y., Guo, M. J., Wang, L. D., et al. 2025. DTGA: an in-situ training scheme for memristor neural networks with high performance. *Applied Intelligence*, **55**: 167.

[7] Bak, S., Park, J., Lee, J., et al. 2024. Memristor-based CNNs for detecting stress using brain imaging signals. *IEEE Transactions on Emerging Topics in Computational Intelligence*, **8**(1): 140−149.

[8] Chen, Z. J., Lin, Z. Y., Yang, J., et al. 2024. Cross-layer transmission realized by light-emitting memristor for constructing ultra-deep neural network with transfer learning ability. *Nature Communications*, **15**: 1930.

[9] Dang, B. J., Zhang, T., Wu, X. L., et al. 2024. Reconfigurable in-sensor processing based on a multi-phototransistor–one-memristor array. *Nature Electronics*, **7**:

991−1003.
[10] Tian, Y., Zhao, Y., Liu, S. P., et al. 2021. Scalable and compact photonic neural chip with low learning-capability-loss. *Nanophotonics*, **11**(2): 329−344.
[11] Tang, R., Ohno, S., Tanizawa, K., et al. 2024. Symmetric silicon microring resonator optical crossbar array for accelerated inference and training in deep learning. *Photonics Research*, **12**(8): 1681−1688.
[12] Wang, Z., Chang, L., Wang, F. F., et al. 2022. Integrated photonic metasystem for image classifications at telecommunication wavelength. *Nature Communications*, **13**: 2131.
[13] Yan, T., Yang, R., Zheng, Z. Y., et al. 2022. All-optical graph representation learning using integrated diffractive photonic computing units. *Science Advances*, **8**: eabn7630.
[14] Yildirim, M., Dinc, N. U., Oguz, I. et al. 2024. Nonlinear processing with linear optics. *Nature Photonics*, **18**: 1076−1082.
[15] Liu, W. C., Huang, Y. Y., Sun, R., et al. 2024. Diffraction-based on-chip optical neural network with high computational density. In Proceedings of JSAP-Optica Joint Symposia 2024.
[16] Liu, W. C., Huang, Y. Y., Sun, R., et al. 2025. Ultra-compact multi-task processor based on in-memory optical computing. *Light: Science & Applications*, **14**: 134.
[17] Cui, K. Y., Rao, S. J., Xu, S., et al. 2025. Spectral convolutional neural network chip for in-sensor edge computing of incoherent natural light. *Nature Communications*, **16**: 81.
[18] Fu, T. Z., Huang, Y. Y., Sun, R., et al. 2023. Integrated diffractive optical neural network with space-time interleaving. *Chinese Optics Letters*, **21**(9): 091301.
[19] Fu, T. Z., Zhang, J. F., Sun, R., et al. 2024. Optical neural networks: progress and challenges. *Light: Science & Applications*, **13**(1): 263.
[20] Bai, B. J., Li, Y. H., Luo, Y., et al. 2023. All-optical image classification through unknown random diffusers using a single-pixel diffractive network. *Light: Science & Applications*, **12**: 69.
[21] Wang, Z., Chen, H., Li, J. N., et al. 2024. Opto-intelligence spectrometer using diffractive neural networks. *Nanophotonics*, **13**(20): 3883−3893.
[22] Yu, F. H., Di, K. J., Chen, W. J., et al. 2025. Time-wavelength multiplexed photonic neural network accelerator for distributed acoustic sensing systems. *Advanced Photonics*, **7**(2): 026008.
[23] Hua, S. Y., Divita, E., Yu, S. S., et al. 2025. An integrated large-scale photonic accelerator with ultralow latency. *Nature*, **640**: 361−367.
[24] Zhang, Y. C., Zhang Q. M., Yu, H. Y., et al. 2024. Memory-less scattering imaging with ultrafast convolutional optical neural networks. *Science Advances*, **10**(24): eadn2205.
[25] Mashiko, R., Naruse, M., Horisaki, R. 2024. Diffraction casting. *Advanced Photonics*, **6**(5): 056005.
[26] Zhan, Z. Y., Wang, H., Liu, Q., et al. 2024. Photonic diffractive generators through sampling noises from scattering media. *Nature Communications*, **15**(1): 10643.
[27] Ahmadi, R., Ahmadnejad, A., Koohi, S. 2024. Free-space optical spiking neural network. *PLOS One*, **19**(12): e0313547.
[28] Dong, B. W., Brückerhoff-Plückelmann, F., Meyer, L. et al. 2024. Partial coherence enhances parallelized photonic computing. *Nature*, **632**: 55−62.
[29] Miller, J. L. 2024. Nonlinear optical computing doesn't need nonlinear optics. *Physics Today*, **77**(10): 12−14.
[30] Guo, Y. D., Hao, Y. H., Wan, S., et al. 2024. Direct observation of atmospheric turbulence with a video-rate wide-field wavefront sensor. *Nature Photonics*, **18**: 935−943.
[31] Xu, T. J., Zhang, W. P., Zhang, J. W., et al. 2024. Control-free and efficient integrated photonic neural networks via hardware-aware training and pruning. *arXiv:2401.08180* .
[32] Xue, Z. W., Zhou, T. K., Xu, Z. H., et al. 2024. Fully forward mode training for optical neural networks. *Nature*, **632**: 280−286.
[33] Sedov, E., Kavokin, A. 2025. Polariton lattices as binarized neuromorphic networks. *Light, Science & Applications*, **14**: 52.
[34] Meng, X. Y., Shi, N. N., Zhang, G. J., et al. 2025. High-integrated photonic tensor core utilizing high-dimensional lightwave and microwave multidomain multiplexing. *Light: Science & Applications*, **14**: 27.
[35] Yuan, H. Y., Du, Z. C., Qi, H. X., et al. 2025. Microcomb-driven photonic chip for solving partial differential equations. *Advanced Photonics*, **7**(1): 016007.
[36] Yu, H. Y., Huang, Z. H., Lamon, S., et al. 2025. All-optical image transportation through a multimode fibre using a miniaturized diffractive neural network on the distal face. *Nature Photonics*, **19**: 486−493.
[37] Cui, S., Xu, Y. F., Su, L. Y., et al. 2019. Classification of all-sky camera data based on convolutional neural network. *Astronomical Research and Technology*, **16**(2): 225−235. (in Chinese)
[38] Wang, W. Q., Wang, J., Ye, X. C., et al. 2025. Application of AI technology in pulsar candidate identification. *Astronomical Techniques and Instruments*, **2**(1): 27−43.