

Review

Optical Convolutional Neural Networks: Methodology and Advances (Invited)

Xiangyan Meng^{1,2,3}, Nuannuan Shi^{1,2,3,*}, Guangyi Li^{1,2,3}, Wei Li^{1,2,3}, Ninghua Zhu^{1,2,3} and Ming Li^{1,2,3,*}

- ¹ State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China
- ² Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100190, China
- ³ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
- * Correspondence: nnshi@semi.ac.cn (N.S.); ml@semi.ac.cn (M.L.)

Abstract: As a leading branch of deep learning, the convolutional neural network (CNN) is inspired by the natural visual perceptron mechanism of living things, showing great application in image recognition, language processing, and other fields. Photonics technology provides a new route for intelligent signal processing with the dramatic potential of its ultralarge bandwidth and ultralow power consumption, which automatically completes the computing process after the signal propagates through the processor with an analog computing architecture. In this paper, we focus on the key enabling technology of optical CNN, including reviewing the recent advances in the research hotspots, overviewing the current challenges and limitations that need to be further overcome, and discussing its potential application.

Keywords: convolutional neural networks; optical computing; photonics signal processing



Citation: Meng, X.; Shi, N.; Li, G.; Li, W.; Zhu, N.; Li, M. Optical Convolutional Neural Networks: Methodology and Advances (Invited). *Appl. Sci.* **2023**, *13*, 7523. <https://doi.org/10.3390/app13137523>

Academic Editor: Habib Hamam

Received: 31 May 2023

Revised: 15 June 2023

Accepted: 19 June 2023

Published: 26 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Convolutional neural networks (CNNs), as an important category of deep neural networks, are inspired by the natural visual perceptron mechanism of living things [1]. Since the first modern sense framework of CNNs, known as LeNet-1 [2], emerged in 1989, numerous representative CNN frameworks have been developed, including LeNet-5 (1998) [3], AlexNet (2012) [4], ZFNet (2014) [5], VGGNet (2015) [6], GoogLeNet (2015) [7], and ResNet (2016) [8]. Meanwhile, abundant progress has been made to deepen CNNs' complexity and reduce the number of parameters [9–11]. Owing to the continuous optimization of network frames, CNNs have been widely used in image recognition [2–9,11–13], speech recognition [14–16], gaming [17,18], medicine [19,20], autonomous driving [21,22], and other fields.

The explosive increase in Internet data year by year has called for more intelligent and effective data processing [23]. As is well-known, there is a positive correlation between the accuracy of a CNN and the number of parameters [24]. Therefore, it has more stringent requirements on the computing hardware due to the demands of massive data processing and high-precision processing. For electrical hardware processors, performance improvements have followed Moore's Law over the past few decades [25,26]. As the chip manufacturing process has gradually approached its physical limitations in recent years, the growth rate of single-chip computing power has gradually slowed [27,28], and semiconductor technology has entered the post-Moore era. Additionally, using the Von Neumann computing paradigm in the traditional computing hardware, such as CPU, GPU, FPGA, ASIC, etc., it is an indisputable fact that the discrete architecture of processor and memory makes it inevitable to trade-off between bandwidth and power consumption [29–32]. Hence, it is an

obvious sharp conflict with the ever-increasing demand for high-performance processing and the slowing growth of computing power [28].

Optical devices, as an alternative, have been regarded as a competitive candidate in the “more than Moore” era [33] with the superiority of ultralarge bandwidth and ultralow power consumption. Compared with electrical vector-matrix multiplication (VMM), it is able to achieve better performance using optical devices, with the computing speed increasing by three orders of magnitude and the power consumption decreasing by two orders of magnitude [34]. In recent years, optical computing solutions, because of their intrinsic high computing speed [35], high computational density [36,37], and low power consumption [38,39], have been massively demonstrated by means of both discrete systems and integrated chips. Meanwhile, numerous review works have put more emphasis on optical matrix multiplication [40,41], special-technology-based optical computing [42–49], and optical neural networks [50–56]. CNNs, as one of the main branches, require more than 80% of full calculations to execute the convolution operation [57]. Accelerating the convolution process in the optical domain provides a subversive way to improve the computing speed and decrease the power consumption, which has not been systematically reviewed. Here, we focus on the key enabling technology of optical CNNs, review the recent advances in optical CNNs, discuss the realization mechanism of optical convolution, and make a prospect and an approach to further development toward the next generation of artificial neural networks.

2. Development of Optical Convolution Neural Network

Generally, a CNN is composed of convolutional layers, pooling layers, fully connected layers, and nonlinear activations. With the convolutional layer, convolution operations are conducted to extract the features of input images. The typical CNN architecture of LeNet-5 is shown in Figure 1a. The convolutional layer with multiple kernels concurrently performs the convolutional operations to extract various feature images. The pooling layer following the convolutional layer is used to subsample and compress features, reduce the amount of calculation, and alleviate overfitting. Immediately afterwards, the fully connected layers are able to realize the full connection of parameters and generate the final classification results. Additionally, various nonlinear activation functions are employed following the convolutional layers and the fully connected layers, aiming to lead the nonlinear properties to the network.

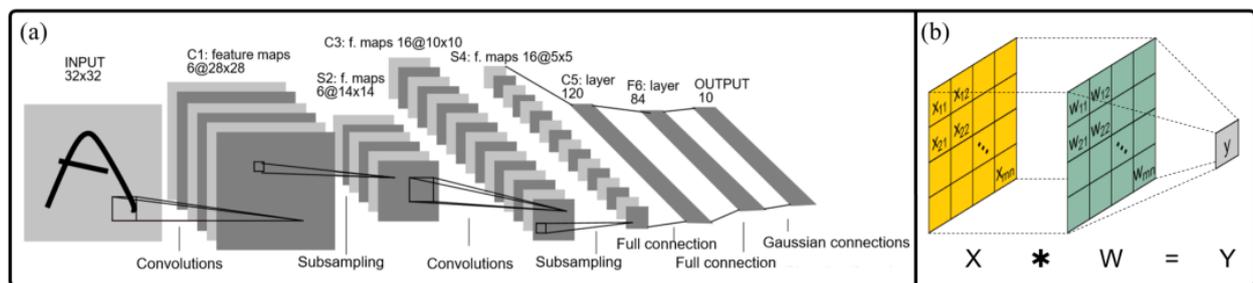


Figure 1. Basics of convolutional neural networks. (a) Scheme of typical LeNet-5 [3]. (b) Principle of one convolution operation. (a) Reprinted with permission from Ref. [3]. 1998, IEEE.

Figure 1b shows the principle of one convolution operation with an $m \times n$ kernel. One convolution operation can be divided into two processes: (1) multiplication between elements in the kernel matrix and data matrix and (2) addition of all multiplication results, which can be expressed as follows:

$$y = \sum_{i=1}^m \sum_{j=1}^n w_{ij} x_{ij}, \quad (1)$$

where w_{ij} is the element of the convolution kernel, x_{ij} is the element of the input data, m is the row count, and n is the column count in the convolutional kernel. From Equation (1), one convolution operation is converted into vector-vector multiplication (VVM), and parallel convolution of multiple kernels is represented as a VMM [58]. At the same time, the optical matrix operation has been widely investigated [59], which makes it convenient to accelerate the convolution operation process with optical methods.

The optical architecture enabling the convolution operation has been blooming. For most of the reported optical CNNs, the convolution operations of CNNs are accelerated in the optical domain, and the rest remain in the electrical domain, which absorbs both the respective advantages for the ultra-bandwidth and low loss of light and the high precision recognition of electricity. Optical CNNs based on the implementation principle are generally divided into four categories: diffraction-based optical CNNs, interference-based optical CNNs, wavelength division multiplexing-based (WDM-based) optical CNNs, and tunable optical attenuation-based optical CNNs.

2.1. Optical CNN Based on Optical Diffraction

Optical computing with free-space optical diffraction has been demonstrated over decades [60,61]. The basic scheme is the Fourier transform property of the lens [62]. When coherent light passes through the lens, a two-dimensional (2D) Fourier transform is performed without power consumption, and the focal plane behind the lens (which is also called the Fourier plane) presents the result of the Fourier transform. Since the convolution operation in the spatial domain is able to transform into a multiplication operation in the Fourier domain following the convolution theorem [63], the convolution operation is implemented with a “4F system” by adding a specially designed mask to the Fourier plane of the first lens.

As shown in Figure 2, the noted “4F system” is composed of two lenses placed coaxially with a distance of $2F$, and the distance between the input plane and output plane is $4F$. The Fourier transform of the input image $g(x, y)$ is accomplished in the focal plane of the first lens

$$G(u, v) = \mathcal{F}\{g(x, y)\}, \quad (2)$$

where \mathcal{F} represents the Fourier transform and u, v are the coordinates in the frequency domain. In the Fourier plane, the specially designed mask $M(u, v)$ introduces additional amplitude and phase modulation,

$$G'(u, v) = G(u, v)M(u, v). \quad (3)$$

Then, the inverse Fourier transform is realized with the second lens, which can be expressed as

$$\begin{aligned} g_o(x', y') &= \mathcal{F}^{-1}\{G'(u, v)\} \\ &= \mathcal{F}^{-1}\{G(u, v)M(u, v)\} \\ &= g(x, y) * m(x, y), \end{aligned} \quad (4)$$

where \mathcal{F}^{-1} is the inversive Fourier transform, $m(x, y)$ is the inverse Fourier transform of $M(u, v)$, and $*$ represents the convolution operation. From Equation (4), the convolution operation of the input image is simultaneously achieved when passing through the “4F system”, and then the feature maps are recorded by the camera sensor and converted into digital electrical signals to conduct nonlinear activation, pooling, and final recognition in the fully connected layer.

The CNN on optical diffraction was proposed and experimentally demonstrated in 2018 [64], and the schematic of the hybrid optoelectronic CNN as well as the experimental setup is shown in Figure 3a. The optical convolution operation was accomplished using the diffractive optical element (DOE) as the phase mask placed in the Fourier plane of the “4F system”. The optical feature image was converted into a digital electrical signal with a camera sensor, and the processes in terms of nonlinearity activation and full connection were still addressed in the electrical domain. The parallel convolution of multiple kernels was

realized via the optimization design of the phase mask. The experimental accuracy reached 44% on the grayscale CIFAR-10 dataset, which improved by approximately 50% relative to the single digital fully connected layer (accuracy of 30%). Apart from the diffractive optical element, the optical metasurface as the phase mask offered another approach using multiple physical layers or wavelength-sensitive material to conduct multichannel free-space convolution [65–68]. An amplitude-only weighting optical CNN in the Fourier plane was proposed in 2020 using a digital micromirror device (DMD) as a reconfigurable amplitude modulation device [69]. A typical schematic representation of a “ $4F$ system” on digital micromirror devices is shown in Figure 3b. Benefitting from the large-scale and high-speed properties of digital micromirror devices, classification tasks on two megapixel matrices at 10 kHz rates were performed. The classification accuracies were up to 98% on the MNIST database and 54% on the CIFAR-10 database. After that, multiple kernel parallel convolution operations based on the optical diffraction orders (Figure 3c) were realized by the same research group [70]. In 2023, using the two-dimensional Dammann grating as the key element for generating multiple displaced images, massive parallelism convolution acceleration was experimentally demonstrated with a computing accuracy of 8 bit [71]. A spatial light modulator (SLM) was used for data input, and the distance between the data input plane and Dammann grating corresponded to the kernel stride. The MNIST database was classified with an accuracy of 97.3%. Figure 3d shows the lensless architecture of the optical CNN [72]. In this scheme, the amplitude mask was placed close to the sensor. Different pixels of the input image were transmitted through different portions of the amplitude mask, underwent different attenuations, and finally focused on the camera sensor. The image sensor performed a weighted sum of the beams, where the weights were determined by the amplitude mask. The accuracy was up to 97.21% on the MNIST database.

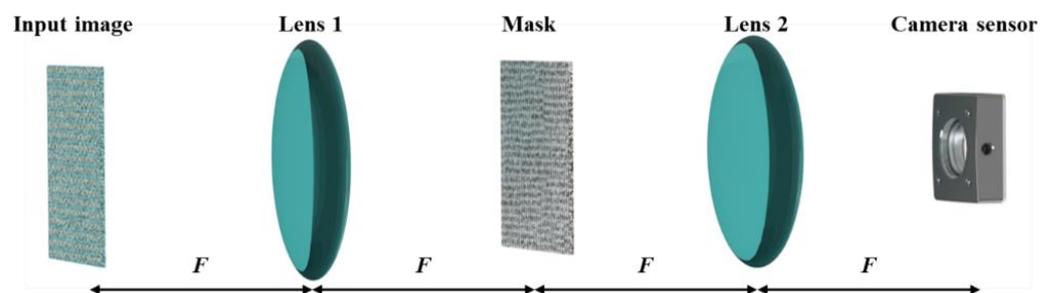


Figure 2. The typical structure of a “ $4F$ system” to realize an optical CNN.

The on-chip Fourier transform scheme was also investigated to realize the optical CNN [73–76]. The Fourier transform property of integrated star couplers was simulated to realize the optical CNN in 2020 [74]. In 2022, the optical integrated diffractive neural network (IDNN) chip in Figure 3e was fabricated on the silicon-on-insulator (SOI) platform [73]. Two on-chip cascaded diffractive cells were used to conduct the operation of the Fourier transform and inverse Fourier transform, where 10 Mach–Zehnder interferometers (MZIs) were used to load the input data and 10 MZIs were used for kernel adjustment. A convolution operation was performed on the proposed integrated diffractive neural network chip, and the recognition of 1D sequences and 2D images was carried out experimentally. In addition to optical CNN, direct classifications of Iris flower, MNIST dataset, and Fashion-MNIST dataset were conducted using the integrated diffractive neural network chip with experimental accuracies of 98.0%, 91.4%, and 80.4%, respectively. In addition to the above realization, other interesting schemes have also been reported, such as in situ training [77], position robustness [78], incoherent light-based [79], and so on [80,81], which contributed to the further development of optical CNNs.

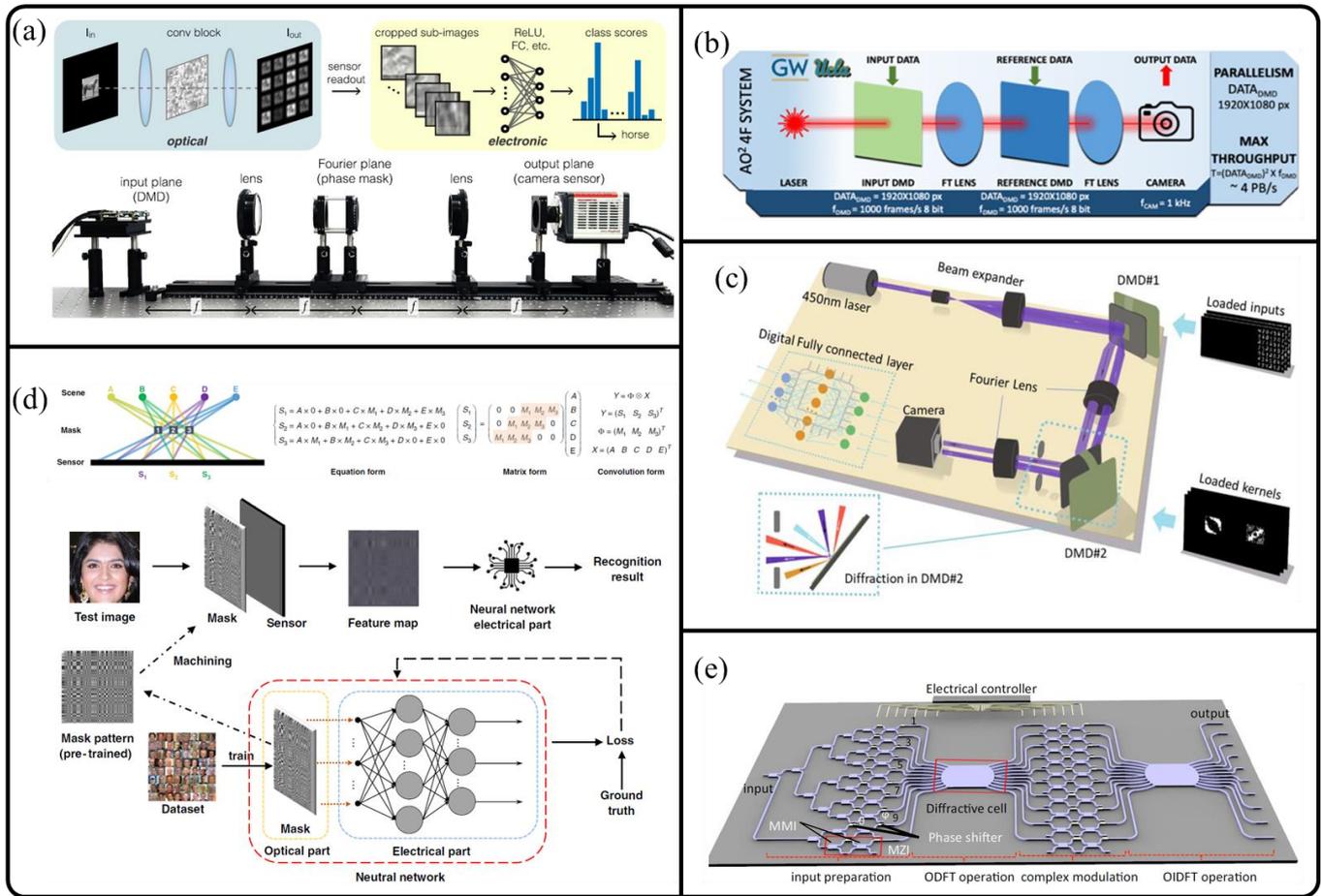


Figure 3. The realization of an optical CNN based on optical diffraction. (a) Phase-only modulation implementation of optical CNN using a diffractive optical element (DOE) [64]. (b) Amplitude-only modulation implementation of an optical CNN using a digital micromirror device (DMD) [69]. (c) Multichannel parallelized diffraction optical CNN [70]. (d) The lensless architecture of optical CNN [72]. (e) Integrated diffraction-based optical CNN structure [73]. (a–e) Reprinted with permission from Refs. [64,69,70,72,73] under a CC BY 4.0 and CC BY-NC-ND 4.0 license.

2.2. Optical CNN Based on Optical Interference

Benefitting from the development of photonic integration technology, matrix multiplication has been extensively focused on light interference [82–84]. Since the convolution operation can be mathematically converted to matrix multiplication, it has become a very popular scheme to realize optical CNNs by exploiting the interference properties of light [85–91]. As the basic unit to build a coherent unitary matrix for discrete convolution operation, the structure diagram of the MZI is shown in Figure 4a, where the splitting ratio of two optical couplers is fixed at 50:50, the internal phase shifter provides an additional phase θ ($0 \leq \theta \leq \pi$) to control the splitting ratio of the MZI, and the external phase shifter adds a relative phase φ ($0 \leq \varphi \leq 2\pi$) to two output arms. Ignoring the optical propagation loss, the transfer matrix $T(\theta, \varphi)$ for a single MZI unit is written as

$$\begin{aligned}
 T(\theta, \varphi) &= P_{\varphi} C P_{\theta} C \\
 &= \frac{1}{2} \begin{bmatrix} e^{i\varphi} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \\
 &= \frac{1}{2} \begin{bmatrix} e^{i\varphi}(e^{i\theta} - 1) & ie^{i\varphi}(e^{i\theta} + 1) \\ ie^{i\theta} + 1 & 1 - e^{i\theta} \end{bmatrix} \\
 &= \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix},
 \end{aligned} \tag{5}$$

Sequentially, according to Equation (7), the entire transfer matrix of the unitary MZI mesh can be inferred as

$$T = T_{2,1}^{10} T_{3,2}^9 T_{4,3}^8 T_{5,4}^7 T_{2,1}^6 T_{3,2}^5 T_{4,3}^4 T_{2,1}^3 T_{3,2}^2 T_{2,1}^1. \quad (9)$$

Since the MZI mesh can represent any unitary matrix, it is worth using the MZI mesh to represent any real-valued matrix through singular value decomposition (SVD) [92]. Any real-valued matrix M with dimensions of $p \times q$ can be decomposed by SVD as

$$M = U\Sigma V^\dagger, \quad (10)$$

where U is a $p \times p$ unitary matrix, Σ is a $p \times q$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and V^\dagger is the complex conjugate of the $q \times q$ unitary matrix V . From Equation (7), unitary matrices U and V^\dagger can be implemented with the MZI mesh. The rectangular diagonal matrix Σ can be implemented with optical attenuators or optical amplifiers. Therefore, any real-valued matrix can be realized with the dimension-matched MZI mesh. Since any real-valued matrix M is realized, parallel convolution of p real-valued kernels with q elements in each kernel can be performed with the MZI mesh.

The most representative optical interference-based CNN was proposed in 2018 [85]. The rectangular architecture MZI mesh in Figure 5a was used as the kernel matrices, and the optical delay lines introduced different time delays to implement multiple convolutional layers. The classification accuracy of the handwritten digits was simulated to 97% in the MNIST database. Then, 4×4 MZI-based linear optical processors were experimentally demonstrated in multiple material platforms, such as silicon-on-insulator (SOI) and silicon nitride (SiN) [86]. The triangular architecture MZI mesh has also been proposed to realize optical CNNs [87]. The fabricated SOI chip of the triangular architecture MZI mesh network is shown in Figure 5b [89] to realize three 3×3 kernels. Recognition accuracies of 86.9% for the MNIST database and 79.3% for Fashion-MNIST were experimentally demonstrated.

Apart from the MZI-based unitary matrix realizing the convolution operation, the optical coherent dot-product chip (OCDC) realized by the interference principle was also an important implementation of optical CNN [88]. The schematic of the chip is shown in Figure 5c, where the light wave is first divided into several branches and then sent into the dot-product chip with one as the reference and the rest to perform the dot-product operation. In each computing branch of Figure 5c, double modulators are cascaded to accomplish the dot-product. The reference branch was utilized to introduce an amplitude bias to perform real-valued computing. By reusing the chip, the AUTOMAP neural network, including matrix multiplication and convolution computing, was experimentally demonstrated and achieved image reconstruction quality comparable to that of a 32 bit digital computer.

In addition to the interference of coherent light with the MZI, the combination of WDM technology and interference was another important realization of optical CNN [37,93,94]. Figure 5d shows the application of WDM technology in a simplified butterfly-style MZI mesh [93]. In this scheme, incoherent light was utilized to avoid additional hardware costs for phase control between input channels and the influence of phase fluctuations of optical signals in off-chip fibers when using coherent light. In their experimental demonstration, a handwritten digit recognition accuracy of 94.16% was realized. As shown in Figure 5e, a compact optical convolution processing unit (OCPU) was realized by combining a multimode interference mechanism and WDM technology [37]. In this work, the process of multiplication was accomplished with multimode interference, and the process of addition was realized with square-law detection of multiple wavelengths with photodetectors. With two 4×4 multimode interference regions and four phase shifters, three correlated 2×2 real-valued kernels were realized with a convolution operation precision of 5 bit. Kernels were reconfigured by adjusting phase shifters, the compute density of 12.74 TMACs/s/mm² was demonstrated, and a classification accuracy of 92.17% was realized on handwritten digits in the MNIST database. The greatest advantage of the proposed OCPU is that the number of reconfigurable units scales linearly with the scale of the kernel size, which translates into solid potential for large-scale integration.

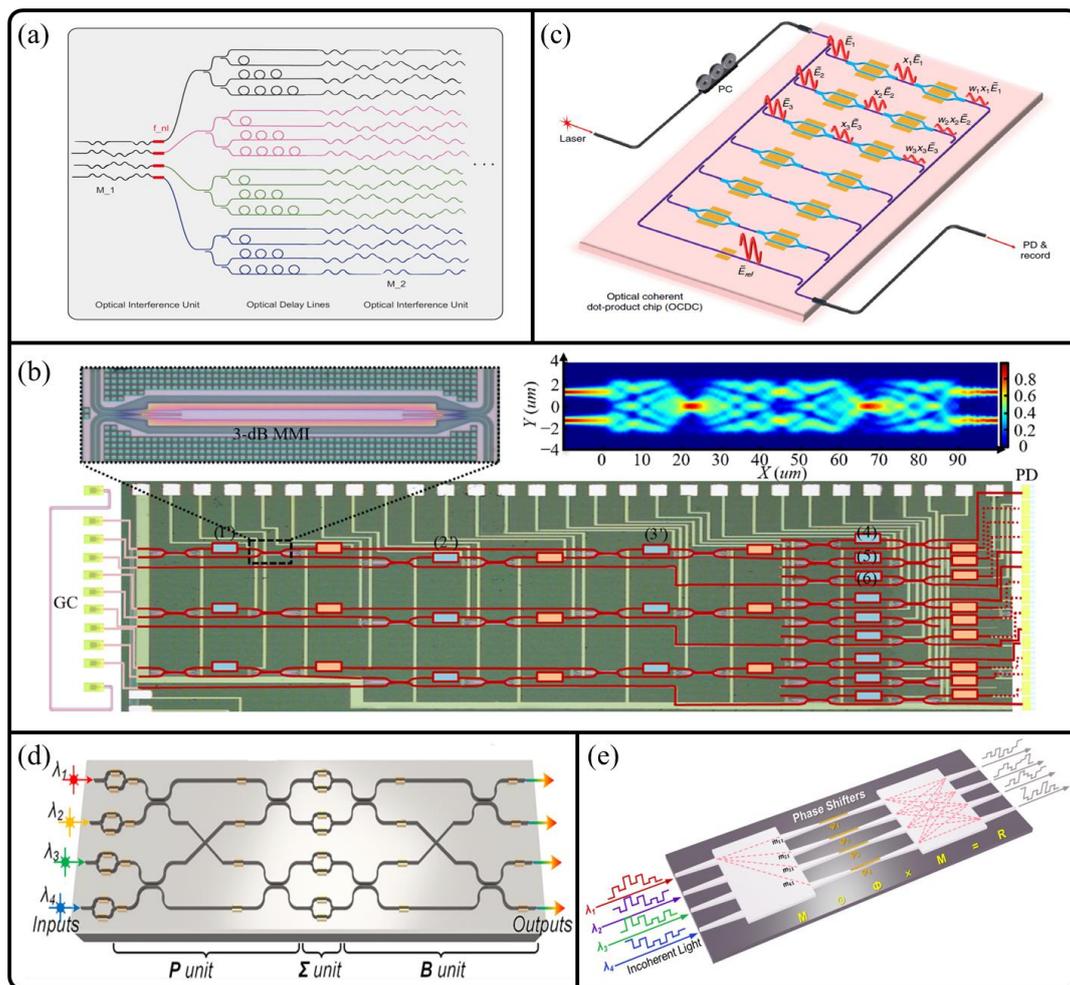


Figure 5. Schemes of optical CNNs based on optical interference. (a) Scheme of optical CNN based on rectangular architecture MZI mesh [85]. (b) Scheme of optical CNN based on triangular architecture MZI mesh [89]. (c) Scheme of optical coherent dot-product chip (OCDC) to realize optical CNN [88]. (d) Scheme of optical CNN based on the combination of WDM and MZI mesh [93]. (e) Scheme of optical CNN based on the combination of WDM and multimode interference [37]. (a) Reprinted from Ref. [85] with permission of the authors. (b,c,e) Reprinted with permission from Refs. [37,88,89] under a CC BY 4.0 license. (d) Reprinted with permission from Ref. [93]. 2022, American Chemical Society.

2.3. Optical CNN Based on Wavelength Division Multiplexing

A WDM-based CNN offers a new solution by comprehensive use of the multidimensional physical information of light waves in terms of the wavelength, amplitude, spatial, temporal, polarization, and so on. By leading into the wavelength dimension in Figure 6, VVM operation is divided into the multiplication process and the addition process, where the multiplication process is accomplished using electrooptic modulations, and the addition process is achieved via wavelength multiplexing and photoelectrical conversion at the photoelectric detector. WDM-based VVMs are familiarly split into two categories: one is the wavelength-independent weight-loaded architecture (shown in Figure 6a), and the other is the wavelength-sensitive weight-loaded architecture (Figure 6b), in which the division criteria is whether wavelength-sensitive variable optical amplitude modulators (VOAMs) are used.

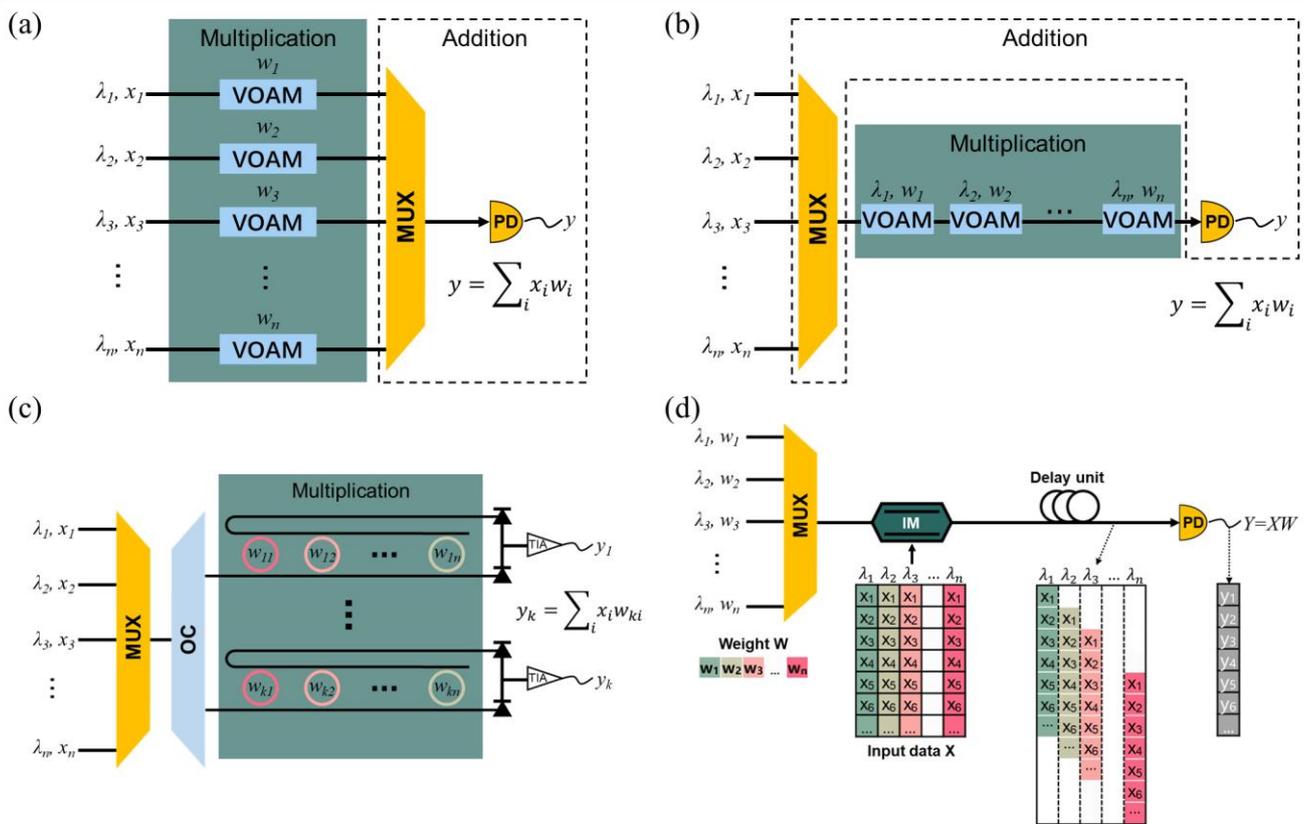


Figure 6. The principle of vector-vector multiplication (VVM) with WDM technology. (a) Wavelength-independent weight-loaded VVM architecture. (b) Wavelength-sensitive weight-loaded VVM architecture. (c) Classical MRR-based real-valued VMM architecture. (d) WDM-based VMM with data input serially by introducing optical delay. VOAM: variable optical amplitude modulator, MUX: multiplexing, PD: photodetector, OC: optical coupler, TIA: transimpedance amplifier, IM: intensity modulator, DCF: dispersion compensation fiber.

First, for the wavelength-independent weight-loaded VVM shown in Figure 6a, n wavelength components loaded with input data vector X were parallelly sent into the computing hardware, where each light wave carried one element of the data vector X . Then, wavelength components underwent amplitude modulation to load the weight vector W , which was the multiplication process of elements in two vectors. As a main optical device in the WDM-based CNN, the variable optical amplitude modulators can be implemented using MZIs [95], microring resonators (MRRs) [96–103], phase change materials (PCMs) [39,104], and waveshapers [35,105–109]. The addition process was conducted using square-law photoelectric detection and the summation of the optical intensity of wavelength components at the photodetector. To obtain the summation signal of different wavelength components, the beat of optical sidebands between adjacent wavelength components should be effectively avoided. Therefore, in the WDM-based CNN system, the wavelength interval of wavelengths fed to the photodetector as well as the bandwidth for both the modulator and photodetector need to satisfy the following condition:

$$f_w - 2f_m > f_p, \tag{11}$$

where f_w is the wavelength interval of adjacent wavelength components, f_m is the modulation bandwidth, and f_p is the bandwidth of the photodetectors.

Figure 6b shows a wavelength-sensitive weight-loaded VVM architecture where the process of multiplication was performed with wavelength-sensitive variable optical amplitude modulators [97–102]. Wavelength components loaded with data vector X were

transmitted through one physical channel by wavelength-division multiplexing. Then, n wavelength-sensitive variable optical amplitude modulators serially loaded n elements of the weight vector W and completed the process of multiplication. Finally, the weighted light waves were fed to a photodetector to complete the addition function.

The classical MRR-based real-valued VMM architecture on the parallel data inputting is shown in Figure 6c, where the MRR weight banks were used as the wavelength-sensitive variable optical amplitude modulator to implement the multiplication process. The wavelength components after wavelength multiplexing were loaded with data X and then divided into k channels using an optical coupler, where each channel contained n wavelength components. In each channel, n add-drop MRRs were used to couple the light waves with n wavelengths to the drop port of the MRR. The proportion of light that each microring couples to the drop port h_{ij} depends on the weight w_{ij} . Then, the transfer matrix for the drop ports and through ports of MRRs can be written as

$$H_{drop} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ & & \vdots & \\ h_{k1} & h_{k2} & \cdots & h_{kn} \end{bmatrix}, \tag{12}$$

$$H_{through} = \begin{bmatrix} 1 - h_{11} & 1 - h_{12} & \cdots & 1 - h_{1n} \\ & & \vdots & \\ 1 - h_{k1} & 1 - h_{k2} & \cdots & 1 - h_{kn} \end{bmatrix}. \tag{13}$$

H_{drop} at the drop port and $H_{through}$ at the through port were sent into the balanced photodetector (BPD) to make both the photoelectric conversion and difference. The transfer matrix H of the MRR weight banks coupled with BPDs can be written as

$$\begin{aligned} H &= H_{through} - H_{drop} \\ &= \begin{bmatrix} 1 - 2h_{11} & 1 - 2h_{12} & \cdots & 1 - 2h_{1n} \\ & & \vdots & \\ 1 - 2h_{k1} & 1 - 2h_{k2} & \cdots & 1 - 2h_{kn} \end{bmatrix}. \end{aligned} \tag{14}$$

As seen from Equation (14), real-valued VMM can be performed by adjusting the coupling coefficient of each MRR.

Additionally, WDM-based CNNs with serial data input approaches have also been reported [35,96,101–103,105–110]. As shown in Figure 6d, the weight vector W was still weighted on the amplitudes of the wavelength components, and the 1D data X were serially modulated to the amplitude of the wavelength components. Owing to the delay between different wavelengths via the delay unit (single mode fiber, dispersion compensating fiber, optical tunable delay lines, etc.), the weighted optical signal generated a time delay among the adjacent wavelength components. Following this method, the input data X were serially loaded into the light waves with a single intensity modulator, and different elements of input data X were fed to the photodetector at the same time with precise delay control. Owing to using a single modulator to load the data in the CNN system, it released the demand for full clock synchronization.

In 2018, Reference [99] proposed a WDM-based photonics convolutional neural network accelerator (PCNNA). Here, the MRR weight banks were the wavelength-sensitive variable optical amplitude modulators to adjust the kernels. The input data were modulated into intensities of different wavelengths. The weights were loaded while passing through the MRR weight banks, and wavelength-based weight components were summed at the photodetectors to generate the convolutional results. In Figure 7a, the analog–digital hybrid CNN computing architecture (e.g., digital electronics and analog photonics (DEAP)), as the most popular architecture, was proposed in 2019 [97]. Two MRR arrays were occupied to load the data matrix and the weight matrix. The simulation recognition accuracy on 500 images from the MNIST test dataset reached 97.6%. After that, optical tunable delay

lines were introduced to serially load the input data [102,103,111] and were experimentally demonstrated in 2022 [101]. As shown in Figure 7b, different from other schemes, one convolution operation only needs a single wavelength component, and the WDM technology enables the multiple wavelength channel to form an on-chip photonic tensor flow processor (PTFP). Using the multi-dimensional multiplexing of light waves, including wavelength, space, and time, high-order tensors can be directly processed in the optical domain. The integrated photonics tensor core on phase change materials is demonstrated in Figure 7c [39]. In this scheme, phase change materials work as wavelength-independent variable optical amplitude modulators to enable efficient calculation without power to maintain the state. A chip-based optical frequency comb was introduced as the light source to perform high parallelism optical convolution processing with a speed of 2 tera-MAC operations per second. The experimental recognition accuracy on 10,000 images in the MNIST test dataset was up to 95.3%. A phase-change metasurface mode converter (PMMC) was also fabricated to form an optical CNN with 6-bit weight precision [104]. The waveguide of two spatial modes was controlled by the phase change material-based phase-gradient metasurface, and 2×2 real-valued kernels were experimentally demonstrated.

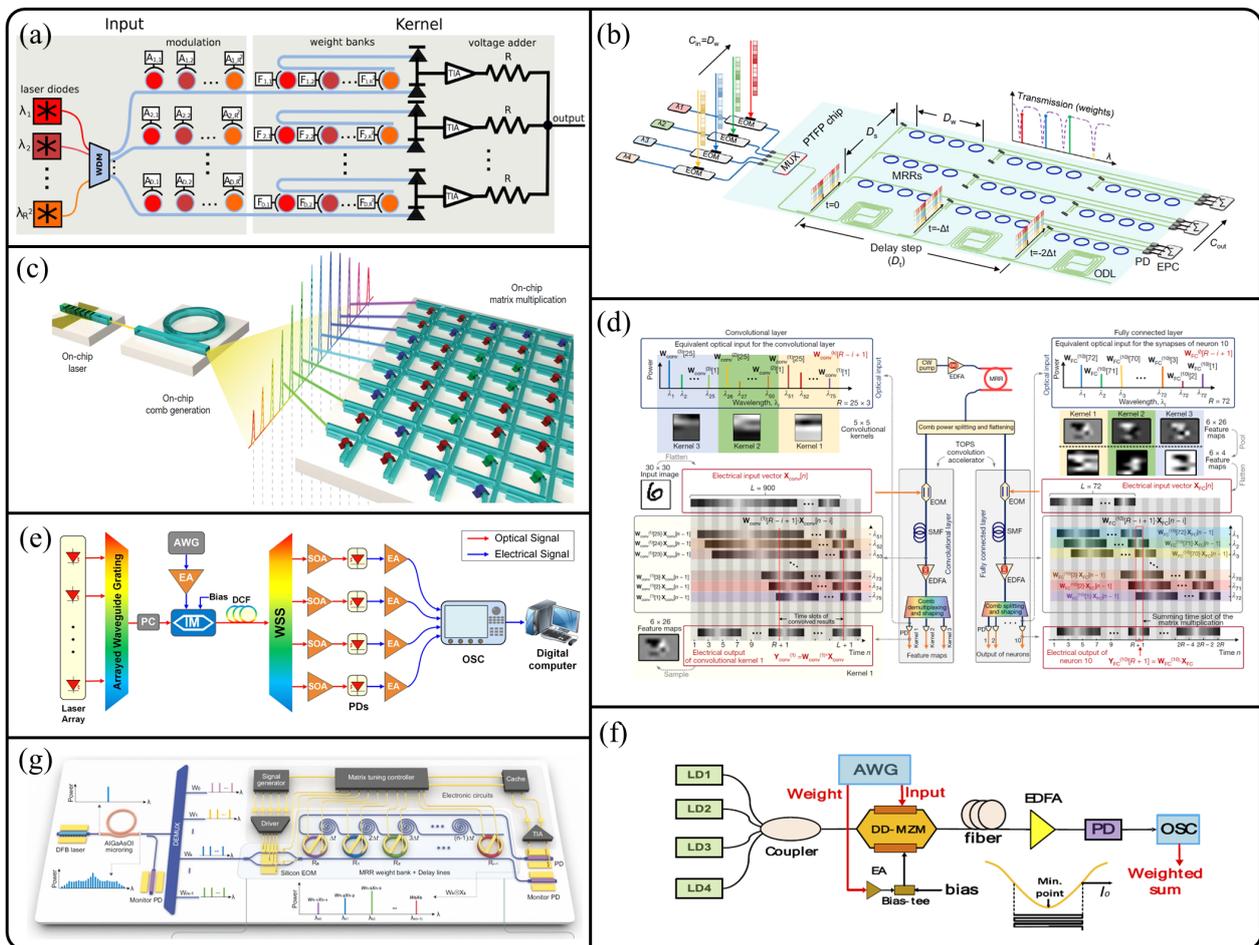


Figure 7. WDM-based optical CNN realizations. (a) The principle of digital electronics and analog photonics (DEAP) architecture CNNs with MRRs as the variable optical amplitude modulators (VOAMs) [97]. (b) Integrated photonic tensor flow processor (PTFP) based on MRRs [101]. (c) The first realization of a photonics tensor core with phase change material (PCM) as the VOAM [39]. (d) Photonic convolution accelerator [35]. (e) On-demand reconfigurable optical matrix operator [106]. (f) Optical realization of binary convolution based on a dual-drive Mach-Zender (DD-MZM) modulator [110].

(g) MRR-based photonic processing unit (PPU) [36]. (a) Reprinted with permission from Ref. [97]. 2020, IEEE. (b,f,g) Reprinted with permission from Refs. [36,101,110] under a CC BY 4.0 license. (c) and (d) Reprinted with permission from Refs. [35,39]. 2021, Springer Nature. (e) Reprinted with permission from Ref. [106]. 2023, IEEE.

Otherwise, a parallel photonics convolution accelerator with ten 3×3 kernels in Figure 7d is illustrated using a soliton optical frequency comb as the light source to operate with a speed of 11 TOPS [35]. Herein, the 2D image data are flattened into a 1D vector and modulated to multiple wavelengths to achieve electro-optical conversion in the single intensity modulator with the serial data input approach. The weight matrix is represented by the amplitude of optical combs and adjusted with a waveshaper. The proposed photonic convolution accelerator cooperates with a fully connected optical layer to classify handwritten digits with an experimental accuracy of 88% on 50 images. With the homologous function (Figure 7e), laser arrays are used to replace the optical frequency comb to improve the robustness of the computing system and obtain an experimental recognition accuracy of 96.01% [106]. The on-demand real-time reconfiguration between the optical perceptron and optical convolution operator is also experimentally demonstrated. As illustrated in Figure 7f, the binary optical convolution operation [110] is also realized by controlling the dual-drive Mach–Zehnder modulators to work at two quadrature points, where the input data and the weights simultaneously drive two arms of the modulator and an optical fiber is occupied for dispersive delay. In 2023, a compact MRR-based photonic processing unit was reported (Figure 7g) with all the essential photonic components integrated, and a photonic-core compute density of over 1 TOPS/mm² was achieved [36]. Using the through port of the MRR as the monitor port, a weight precision of 9 bit was realized. Negative elements in kernels were accomplished via subtraction between two channels, and a handwritten digit recognition accuracy of 96.6% was experimentally illustrated, which is comparable with a digital computer.

2.4. Optical CNN Based on Tunable Optical Attenuation of Coherent Light

The last category of optical CNNs is generally based on the tunable attenuation of coherent light. In this architecture, the data matrix and the weight matrix are represented by the intensities of light waves, and two cascaded intensity modulations are used to load the input data and weights to accomplish the dot-product operation in the optical domain. In addition, optical-electrical conversion is achieved on each optical path, and the sum operation is conducted in the electrical domain.

As shown in Figure 8a [112], two acousto-optical modulator (AOM) arrays were utilized, with one for data loading and the other for kernel weighing. A single laser was employed as the coherent light and divided into several equal power beams. Then, the light waves passed through two acousto-optical modulator arrays and converted the optical power into electrical voltage at the PD array. Subsequently, a switching array decided whether the voltages were accumulated positively or negatively to realize real-valued computing. The proof-of-concept experiment was verified with a recognition accuracy of 98.9% on the MNIST dataset and 91.5% on the Fashion-MNIST dataset. P-doped–intrinsic–N-doped (PIN) current-controlled attenuators have also been reported to adjust the optical power for the weight matrix (Figure 8b) [113]. The PIN junction works in the forward bias state, and the optical power is tuned by changing the injected carriers. The input data are also represented by the optical power, and the linear dot-product process is realized with tunable optical attenuation. The addition operation is conducted in the electrical domain on the PD, and nonlinear activation is accomplished with an on-chip microring modulator. An integrated end-to-end photonic deep neural network (PDNN) with one convolutional layer and two fully connected layers was experimentally demonstrated. In this architecture, through the design of the physical connection mode of the photoelectric link, convolution and full connection were realized by the combination of photoelectricity, and nonlinearity was realized by using a microring modulator. The classification results can be directly generated at the output port of the integrated chip. The printed letters were demonstrated

with a two-class classification accuracy of 93.8% and a four-class classification accuracy of 89.8%. In addition, serial dot-product coupled with integrated balanced homodyne detection was also proposed [114]. The VMM was completed in each optical channel and the parallel multiple channels enabled the VMM operation. In addition, other elements, such as MRR and MZI, have also been reported as tunable optical attenuators to realize optical CNNs [88,101].

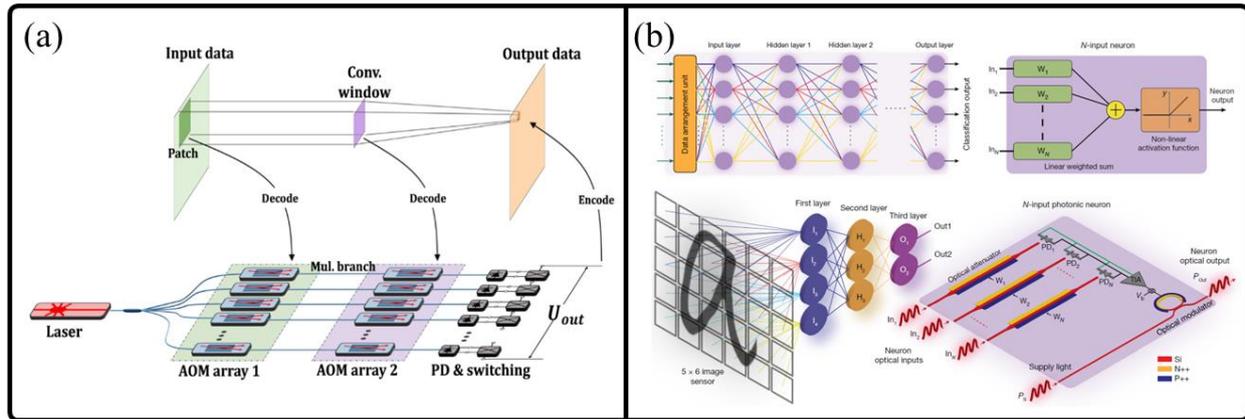


Figure 8. Optical CNN schemes based on tunable optical attenuation. (a) Optical CNN realization with acousto-optic modulator arrays [112]. (b) Integrated end-to-end photonic deep neural network (PDNN) [113]. (a) Reprinted with permission from Ref. [112] under a CC BY 4.0 license. (b) Reprinted with permission from Ref. [113]. 2022, Springer Nature.

3. Discussion and Conclusions

As an emerging research direction, optical CNNs have garnered significant attention since their inception. Table 1 presents a comparison of four optical CNN schemes, highlighting their parallelism, computing speed, integration density, and reconfigurability.

Table 1. Comparison of optical CNN schemes.

Type	Parallelism	Computing Speed	Integration Density	Reconfigurability
Diffraction	high	high	low	low
Interference	low	medium	medium	high
WDM	high	high	high	high

Diffraction-based optical CNNs exhibit advantages in terms of parallelism, computing speed, and scalability due to their utilization of spatial light. The presence of spatial light allows for a large number of neurons in each layer, facilitating the expansion of multiple channels and kernels in the spatial domain, thus enabling high parallelism. The abundance of neurons and high parallelism contributes to achieving high computing speeds. However, diffraction-based optical CNNs also suffer from notable disadvantages. The discrete components used make the system bulky, and attempts at integration result in performance degradation. Moreover, kernel-loading devices such as DOE and metamaterials are nearly impossible to reconfigure, while SLM and DMD devices have low data rates (typically ~kHz).

An interference-based optical CNN excels in reconfigurability. This scheme often utilizes MZI for kernel matrix loading, enabling rapid refresh rates in the range of tens of GHz. Despite the advantages of high-speed reconstruction offered by MZIs, their relatively large volume limits the integration density of interference-based schemes. Furthermore, the use of coherent light restricts the transmission of only one light at a time in the optical waveguide, thereby constraining parallelism and computing speed.

WDM-based optical CNN represents a promising and extensively researched solution. This scheme fully exploits the optical wavelength dimension, leading to high parallelism. The use of MRR as a wavelength-sensitive optical attenuator, with its radius as small as several micrometers and modulation rate reaching tens of GHz, further enhances computing speed, integration density, and reconfigurability.

The optical CNN based on tunable optical attenuation depends on the specific characteristics of the optical attenuator used. Due to the absence of a unified conclusion, it is not listed in the table. Additionally, apart from the four types of optical CNNs discussed above, there are other noteworthy optical CNN solutions, such as photon frequency synthesis [115] and photodetectors with adjustable responsivity [116], which warrant further research.

Presently, although optical CNNs exhibit advantages in terms of bandwidth, latency, and computational speed compared with electrical architectures, optical CNNs face challenges in surpassing the limitations of small realizable matrix sizes, a limited range of realizable functions, and low computing precision. Consequently, extensive efforts are required for optical CNNs to gain widespread usage.

First, on-chip large-scale integration needs to be broken through. The reported on-chip integrated optical accelerated computing architectures only realize the integration of tens of thousands of devices at present, which is far less than their electrical counterparts. In optical computing, the power required for calculations such as electro-optical conversion, photoelectric conversion, and analog-to-digital conversion remains basically unchanged. By integrating more photonic devices, the common power required for computing will be averaged, thereby improving the energy efficiency of optical computing and giving full play to the advantages of optical computing.

Second, more functions should be realized with optical methods. Most optical CNN solutions primarily focus on the optical implementation of convolutions. Although there are related studies on the optical implementation of nonlinear activation and full connection [35,38,91,113,117–120], these studies remain relatively limited and warrant further exploration. In particular, limited by the small matrix size, it is still a challenge to achieve large-scale full connections using optical methods. Implementing more functions using optical methods will be conducive to expanding the application field of optical computing and promoting the practical application of optical computing.

Finally, the development of an in situ trainable arbitrary reconfigurable computing architecture is essential. At present, most optical CNN implementations adopt the offline training method, and the weight matrix is pretrained in the neural network simulation model. As a result, a deviation between the simulation model and the experimental system inevitably appears. In situ training, which updates the weight directly and performs the computation at the original place, offers a new form to accelerate the reconfiguration performance of the neural network and improve its precision. Several recently proposed in situ training schemes, such as physics-aware training [121], adaptive training [122], and other relevant methods [123,124], have been successfully introduced to optical computing. These approaches can be effectively incorporated into the optical CNN framework. By incorporating in situ training, it is more convenient to reconstruct the network structure (such as changing the size and number of convolution kernels, changing the number of convolution layers, etc.), and the influence of errors can be considered during the training process. In this way, the refactoring of more different, more complex classification tasks on the same hardware will become a reality, rather than the simple tasks that are fixed today (such as classifying handwritten digits).

In summary, various optical methods have been used to conduct convolution operations with the benefit of high computing speed, low power consumption, and robust parallelism. Here, the review primarily introduces the recent advancements in optical CNNs, classifying them into four categories and summarizing their respective strengths and weaknesses. The challenges faced by optical CNNs are also summarized, and the future prospects of this technology are discussed. Next, the untapped potential of opti-

cal CNNs warrants further exploration, as they hold promise in shaping next-generation artificial neural network platforms.

Funding: This research was funded in part by the National Natural Science Foundation of China, grant numbers 62235011, 61925505, and 62075212, and in part by the Youth Innovation Promotion Association of Chinese Academy of Sciences, grant number 2022111.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fukushima, K.; Miyake, S. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition. *Comput. Coop. Neural Nets* **1982**, *36*, 267–285.
2. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [\[CrossRef\]](#)
3. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
4. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
5. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the Computer Vision—ECCV 2014, Cham, Switzerland, 6–12 September 2014; pp. 818–833.
6. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2015**, arXiv:1409.1556.
7. Szegedy, C.; Wei, L.; Yangqing, J.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
9. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. [\[CrossRef\]](#) [\[PubMed\]](#)
10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
11. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
12. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [\[CrossRef\]](#)
13. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
14. Lee, H.; Pham, P.; Largman, Y.; Ng, A. Unsupervised feature learning for audio classification using convolutional deep belief networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 7–10 December 2009; pp. 1–9.
15. Abdel-Hamid, O.; Mohamed, A.R.; Jiang, H.; Penn, G. Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 4277–4280.
16. Sainath, T.N.; Mohamed, A.; Kingsbury, B.; Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 8614–8618.
17. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [\[CrossRef\]](#)
18. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [\[CrossRef\]](#)
19. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [\[CrossRef\]](#)
20. Lakhani, P.; Sundaram, B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* **2017**, *284*, 574–582. [\[CrossRef\]](#)

21. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Rob.* **2020**, *37*, 362–386. [CrossRef]
22. Cui, Y.; Chen, R.; Chu, W.; Chen, L.; Tian, D.; Li, Y.; Cao, D. Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 722–739. [CrossRef]
23. Data Is Giving Rise to a New Economy. Available online: <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy> (accessed on 6 May 2017).
24. Kawatsu, C.; Koss, F.; Gillies, A.; Zhao, A.; Crossman, J.; Purman, B.; Stone, D.; Dahn, D. Gesture recognition for robotic control using deep learning. In Proceedings of the NDIA Ground Vehicle Systems Engineering and Technology Symposium, Detroit, MI, USA, 15–17 August 2017; pp. 1–7.
25. Schaller, R.R. Moore’s law: Past, present and future. *IEEE Spectr.* **1997**, *34*, 52–59. [CrossRef]
26. Moore, G.E. Cramming more components onto integrated circuits. *Electronics* **1965**, *38*, 82–85. [CrossRef]
27. Theis, T.N.; Wong, H.S.P. The End of Moore’s Law: A New Beginning for Information Technology. *Comput. Sci. Eng.* **2017**, *19*, 41–50. [CrossRef]
28. Leiserson, C.E.; Thompson, N.C.; Emer, J.S.; Kuszmaul, B.C.; Lamson, B.W.; Sanchez, D.; Schardl, T.B. There’s plenty of room at the Top: What will drive computer performance after Moore’s law? *Science* **2020**, *368*, eaam9744. [CrossRef]
29. Stone, H.S. A Logic-in-Memory Computer. *IEEE Trans. Comput.* **1970**, *100*, 73–78. [CrossRef]
30. Patterson, D.; Anderson, T.; Cardwell, N.; Fromm, R.; Keeton, K.; Kozyrakis, C.; Thomas, R.; Yelick, K. Intelligent RAM (IRAM): Chips that remember and compute. In Proceedings of the IEEE International Solids-State Circuits Conference (ISSCC), San Francisco, CA, USA, 8 February 1997; pp. 224–225.
31. Sengupta, B.; Stemmler, M.B. Power Consumption During Neuronal Computation. *Proc. IEEE* **2014**, *102*, 738–750. [CrossRef]
32. Miller, D.A.B. Attojoule Optoelectronics for Low-Energy Information Processing and Communications. *J. Light. Technol.* **2017**, *35*, 346–396. [CrossRef]
33. Kitayama, K.-I.; Notomi, M.; Naruse, M.; Inoue, K.; Kawakami, S.; Uchida, A. Novel frontier of photonics for data processing—Photonic accelerator. *APL Photonics* **2019**, *4*, 090901. [CrossRef]
34. Nahmias, M.A.; de Lima, T.F.; Tait, A.N.; Peng, H.T.; Shastri, B.J.; Prucnal, P.R. Photonic Multiply-Accumulate Operations for Neural Networks. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 7701518. [CrossRef]
35. Xu, X.; Tan, M.; Corcoran, B.; Wu, J.; Boes, A.; Nguyen, T.G.; Chu, S.T.; Little, B.E.; Hicks, D.G.; Morandotti, R.; et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **2021**, *589*, 44–51. [CrossRef] [PubMed]
36. Bai, B.; Yang, Q.; Shu, H.; Chang, L.; Yang, F.; Shen, B.; Tao, Z.; Wang, J.; Xu, S.; Xie, W.; et al. Microcomb-based integrated photonic processing unit. *Nat. Commun.* **2023**, *14*, 66. [CrossRef]
37. Meng, X.; Zhang, G.; Shi, N.; Li, G.; Azaña, J.; Capmany, J.; Yao, J.; Shen, Y.; Li, W.; Zhu, N.; et al. Compact optical convolution processing unit based on multimode interference. *Nat. Commun.* **2023**, *14*, 3000. [CrossRef]
38. Sludds, A.; Bandyopadhyay, S.; Chen, Z.; Zhong, Z.; Cochrane, J.; Bernstein, L.; Bunandar, D.; Dixon, P.B.; Hamilton, S.A.; Streshinsky, M.; et al. Delocalized Photonic Deep Learning on the Internet’s Edge. *Science* **2022**, *378*, 270–276. [CrossRef]
39. Feldmann, J.; Youngblood, N.; Karpov, M.; Gehring, H.; Li, X.; Stappers, M.; Le Gallo, M.; Fu, X.; Lukashchuk, A.; Raja, A.S.; et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **2021**, *589*, 52–58. [CrossRef]
40. Cheng, J.; Zhou, H.; Dong, J. Photonic Matrix Computing: From Fundamentals to Applications. *Nanomaterials* **2021**, *11*, 1683. [CrossRef]
41. Zhou, H.; Dong, J.; Cheng, J.; Dong, W.; Huang, C.; Shen, Y.; Zhang, Q.; Gu, M.; Qian, C.; Chen, H.; et al. Photonic matrix multiplication lights up photonic accelerator and beyond. *Light Sci. Appl.* **2022**, *11*, 30. [CrossRef]
42. Abdollahramezani, S.; Hemmatyar, O.; Adibi, A. Meta-optics for spatial optical analog computing. *Nanophotonics* **2020**, *9*, 4075–4095. [CrossRef]
43. Bogaerts, W.; Perez, D.; Capmany, J.; Miller, D.A.B.; Poon, J.; Englund, D.; Morichetti, F.; Melloni, A. Programmable photonic circuits. *Nature* **2020**, *586*, 207–216. [CrossRef]
44. Gupta, P.; Li, S. *4F Optical Neural Network Acceleration: An Architecture Perspective*; SPIE: Bellingham, WA, USA, 2022; Volume 12019.
45. Lima, T.F.d.; Peng, H.T.; Tait, A.N.; Nahmias, M.A.; Miller, H.B.; Shastri, B.J.; Prucnal, P.R. Machine Learning With Neuromorphic Photonics. *J. Light. Technol.* **2019**, *37*, 1515–1534. [CrossRef]
46. Pérez, D.; Gasulla, I.; Capmany, J. Programmable multifunctional integrated nanophotonics. *Nanophotonics* **2018**, *7*, 1351–1371. [CrossRef]
47. Pérez, D.; Gasulla, I.; Das Mahapatra, P.; Capmany, J. Principles, fundamentals, and applications of programmable integrated photonics. *Adv. Opt. Photonics* **2020**, *12*, 709–786. [CrossRef]
48. Xu, X.; Han, W.; Tan, M.; Sun, Y.; Li, Y.; Wu, J.; Morandotti, R.; Mitchell, A.; Xu, K.; Moss, D.J. Neuromorphic computing based on wavelength-division multiplexing. *IEEE J. Sel. Top. Quantum Electron.* **2022**, *29*, 7400112. [CrossRef]
49. Zhang, Q.; Yu, H.; Barbiero, M.; Wang, B.; Gu, M. Artificial neural networks enabled by nanophotonics. *Light Sci. Appl.* **2019**, *8*, 42. [CrossRef] [PubMed]
50. Bai, Y.; Xu, X.; Tan, M.; Sun, Y.; Li, Y.; Wu, J.; Morandotti, R.; Mitchell, A.; Xu, K.; Moss, D.J. Photonic multiplexing techniques for neuromorphic computing. *Nanophotonics* **2023**, *12*, 795–817. [CrossRef]
51. Huang, C.; Sorger, V.J.; Miscuglio, M.; Al-Qadasi, M.; Mukherjee, A.; Lampe, L.; Nichols, M.; Tait, A.N.; Ferreira de Lima, T.; Marquez, B.A.; et al. Prospects and applications of photonic neural networks. *Adv. Phys. X* **2022**, *7*, 1981155. [CrossRef]

52. Liu, J.; Wu, Q.; Sui, X.; Chen, Q.; Gu, G.; Wang, L.; Li, S. Research progress in optical neural networks: Theory, applications and developments. *Photonix* **2021**, *2*, 5. [[CrossRef](#)]
53. Marinis, L.D.; Cococcioni, M.; Castoldi, P.; Andriolli, N. Photonic Neural Networks: A Survey. *IEEE Access* **2019**, *7*, 175827–175841. [[CrossRef](#)]
54. Shastri, B.J.; Tait, A.N.; Ferreira de Lima, T.; Pernice, W.H.P.; Bhaskaran, H.; Wright, C.D.; Prucnal, P.R. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **2021**, *15*, 102–114. [[CrossRef](#)]
55. Sui, X.; Wu, Q.; Liu, J.; Chen, Q.; Gu, G. A Review of Optical Neural Networks. *IEEE Access* **2020**, *8*, 70773–70783. [[CrossRef](#)]
56. Wetzstein, G.; Ozcan, A.; Gigan, S.; Fan, S.; Englund, D.; Soljačić, M.; Denz, C.; Miller, D.A.B.; Psaltis, D. Inference in artificial intelligence with deep optics and photonics. *Nature* **2020**, *588*, 39–47. [[CrossRef](#)]
57. Li, X.; Zhang, G.; Huang, H.H.; Wang, Z.; Zheng, W. Performance Analysis of GPU-Based Convolutional Neural Networks. In Proceedings of the International Conference on Parallel Processing (ICPP), Philadelphia, PA, USA, 16–19 August 2016; pp. 67–76.
58. Vasudevan, A.; Anderson, A.; Gregg, D. Parallel Multi Channel convolution using General Matrix Multiplication. In Proceedings of the IEEE International Conference on Application-Specific Systems, Architectures and Processors (ASAP), Seattle, WA, USA, 10–12 July 2017; pp. 19–24.
59. Abushagur, M.A.G.; Caulfield, H.J. 7—Optical Matrix Computations. In *Optical Processing and Computing*; Arsenault, H.H., Szoplík, T., Macukow, B., Eds.; Academic Press: Cambridge, MA, USA, 1989; pp. 223–249.
60. Cutrona, L.; Leith, E.; Palermo, C.; Porcello, L. Optical data processing and filtering systems. *IRE Trans. Inf. Theory* **1960**, *6*, 386–400. [[CrossRef](#)]
61. Ambs, P. Optical Computing: A 60-Year Adventure. *Adv. Opt. Technol.* **2010**, *2010*, 372652. [[CrossRef](#)]
62. Ozaktas, H.M.; Mendlovic, D. Fractional Fourier optics. *J. Opt. Soc. Am. A* **1995**, *12*, 743–751. [[CrossRef](#)]
63. Goodman, J.W. *Introduction to Fourier Optics*; W. H. Freeman: New York, NY, USA, 2005.
64. Chang, J.; Sitzmann, V.; Dun, X.; Heidrich, W.; Wetzstein, G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* **2018**, *8*, 12324. [[CrossRef](#)]
65. Wang, H.; Jin, W.; Guo, C.; Zhao, N.; Rodrigues, S.P.; Fan, S. Design of Compact Meta-Crystal Slab for General Optical Convolution. *ACS Photonics* **2022**, *9*, 1358–1365. [[CrossRef](#)]
66. Zheng, H.; Liu, Q.; Zhou, Y.; Kravchenko, I.I.; Huo, Y.; Valentine, J. Meta-optic accelerators for object classifiers. *Sci Adv* **2022**, *8*, eabo6410. [[CrossRef](#)]
67. Burgos, C.M.V.; Yang, T.; Zhu, Y.; Vamivakas, A.N. Design framework for metasurface optics-based convolutional neural networks. *Appl. Opt.* **2021**, *60*, 4356–4365. [[CrossRef](#)]
68. Colburn, S.; Chu, Y.; Shilzerman, E.; Majumdar, A. Optical frontend for a convolutional neural network. *Appl. Opt.* **2019**, *58*, 3179–3186. [[CrossRef](#)]
69. Miscuglio, M.; Hu, Z.; Li, S.; George, J.K.; Capanna, R.; Dalir, H.; Bardet, P.M.; Gupta, P.; Sorger, V.J. Massively parallel amplitude-only Fourier neural network. *Optica* **2020**, *7*, 1812–1819. [[CrossRef](#)]
70. Hu, Z.; Li, S.; Schwartz, R.L.T.; Solyanik-Gorgone, M.; Miscuglio, M.; Gupta, P.; Sorger, V.J. High-Throughput Multichannel Parallelized Diffraction Convolutional Neural Network Accelerator. *Laser Photonics Rev.* **2022**, *16*, 2200213. [[CrossRef](#)]
71. Ma, G.; Yu, J.; Zhu, R.; Zhou, C. Optical multi-imaging-casting accelerator for fully parallel universal convolution computing. *Photonics Res.* **2023**, *11*, 299–312. [[CrossRef](#)]
72. Shi, W.; Huang, Z.; Huang, H.; Hu, C.; Chen, M.; Yang, S.; Chen, H. LOEN: Lensless opto-electronic neural network empowered machine vision. *Light Sci. Appl.* **2022**, *11*, 121. [[CrossRef](#)]
73. Zhu, H.H.; Zou, J.; Zhang, H.; Shi, Y.Z.; Luo, S.B.; Wang, N.; Cai, H.; Wan, L.X.; Wang, B.; Jiang, X.D.; et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **2022**, *13*, 1044. [[CrossRef](#)]
74. Ong, J.R.; Ooi, C.C.; Ang, T.Y.L.; Lim, S.T.; Png, C.E. Photonic Convolutional Neural Networks Using Integrated Diffractive Optics. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 7702108. [[CrossRef](#)]
75. Li, S.; Yang, H.; Wong, C.W.; Sorger, V.J.; Gupta, P. PhotoFourier: A Photonic Joint Transform Correlator-Based Neural Network Accelerator. In Proceedings of the 2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA), Montreal, QC, Canada, 25 February–1 March 2023. [[CrossRef](#)]
76. Peserico, N.; Schwartz, R.; Yang, H.; Ma, X.; Hosseini, M.; Gupta, P.; Dalir, H.; Sorger, V.J. FFT-based Convolution Neural Network on Silicon Photonics Platform. In Proceedings of the IEEE Photonics Conference (IPC), Vancouver, BC, Canada, 13–17 November 2022; pp. 1–2.
77. Gu, Z.; Huang, Z.; Gao, Y.; Liu, X. Training optronic convolutional neural networks on an optical system through backpropagation algorithms. *Opt. Express* **2022**, *30*, 19416–19440. [[CrossRef](#)]
78. Gu, Z.; Gao, Y.; Liu, X. Position-robust optronic convolutional neural networks dealing with images position variation. *Opt. Commun.* **2022**, *505*, 127505. [[CrossRef](#)]
79. Fei, Y.; Sui, X.; Gu, G.; Chen, Q. Zero-power optical convolutional neural network using incoherent light. *Opt. Lasers Eng.* **2023**, *162*, 107410. [[CrossRef](#)]
80. Ibadulla, R.; Chen, T.M.; Reyes-Aldasoro, C.C. FatNet: High-Resolution Kernels for Classification Using Fully Convolutional Optical Neural Networks. *AI* **2023**, *4*, 361–374. [[CrossRef](#)]
81. Sadeghzadeh, H.; Koohi, S. High-Speed Multi-Layer Convolutional Neural Network Based on Free-Space Optics. *IEEE Photonics J.* **2022**, *14*, 8835312. [[CrossRef](#)]

82. Clements, W.R.; Humphreys, P.C.; Metcalf, B.J.; Kolthammer, W.S.; Walmsley, I.A. Optimal design for universal multiport interferometers. *Optica* **2016**, *3*, 1460–1465. [[CrossRef](#)]
83. Reck, M.; Zeilinger, A.; Bernstein, H.J.; Bertani, P. Experimental realization of any discrete unitary operator. *Phys Rev Lett* **1994**, *73*, 58–61. [[CrossRef](#)]
84. Shen, Y.; Harris, N.C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.; Laroche, H.; Englund, D.; et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **2017**, *11*, 441–446. [[CrossRef](#)]
85. Bagherian, H.; Skirlo, S.; Shen, Y.; Meng, H.; Ceperic, V.; Soljacic, M. On-Chip Optical Convolutional Neural Networks. *arXiv* **2018**, arXiv:1808.03303.
86. De Marinis, L.; Cococcioni, M.; Liboiron-Ladouceur, O.; Contestabile, G.; Castoldi, P.; Andriolli, N. Photonic Integrated Reconfigurable Linear Processors as Neural Network Accelerators. *Appl. Sci.* **2021**, *11*, 6232. [[CrossRef](#)]
87. Xiaofeng, X.; Lianqing, Z.; Wei, Z. Convolutional neural networks with coherent nanophotonic circuits. In Proceedings of the 10th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Intelligent Sensing Technologies and Applications, Chengdu, China, 15–17 June 2021; pp. 1–7.
88. Xu, S.; Wang, J.; Shu, H.; Zhang, Z.; Yi, S.; Bai, B.; Wang, X.; Liu, J.; Zou, W. Optical coherent dot-product chip for sophisticated deep learning regression. *Light Sci. Appl.* **2021**, *10*, 221. [[CrossRef](#)]
89. Xu, X.; Zhu, L.; Zhuang, W.; Lu, L.; Yuan, P. A Convolution Neural Network Implemented by Three 3×3 Photonic Integrated Reconfigurable Linear Processors. *Photonics* **2022**, *9*, 80. [[CrossRef](#)]
90. Xu, X.; Zhu, L.; Zhuang, W.; Zhang, D.; Yuan, P.; Lu, L. Photoelectric hybrid convolution neural network with coherent nanophotonic circuits. *Opt. Eng.* **2021**, *60*, 117106. [[CrossRef](#)]
91. Yang, Z.; Tan, W.M.; Zhang, T.J.; Chen, C.D.; Wang, Z.X.; Mao, Y.; Ma, C.X.; Lin, Q.; Bi, W.J.; Yu, F.; et al. MXene-Based Broadband Ultrafast Nonlinear Activator for Optical Computing. *Adv. Opt. Mater.* **2022**, *10*, 2200714. [[CrossRef](#)]
92. Lawson, C.L.; Richard, J.H. *Solving Least Squares Problems*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1995.
93. Feng, C.; Gu, J.; Zhu, H.; Ying, Z.; Zhao, Z.; Pan, D.Z.; Chen, R.T. A Compact Butterfly-Style Silicon Photonic–Electronic Neural Chip for Hardware-Efficient Deep Learning. *ACS Photonics* **2022**, *9*, 3906–3916. [[CrossRef](#)]
94. Huang, Y.; Yue, H.; Ma, W.; Zhang, Y.; Xiao, Y.; Tang, Y.; Tang, H.; Chu, T. A highly parallel photonic acceleration processor for matrix-matrix multiplication. *Opt. Lett.* **2023**, *48*, 3231–3234. [[CrossRef](#)]
95. Totovic, A.; Giamougiannis, G.; Tsakyridis, A.; Lazovsky, D.; Pleros, N. Programmable photonic neural networks combining WDM with coherent linear optics. *Sci. Rep.* **2022**, *12*, 5605. [[CrossRef](#)]
96. Jiang, Y.; Zhang, W.; Yang, F.; He, Z. Photonic Convolution Neural Network Based on Interleaved Time-Wavelength Modulation. *J. Light. Technol.* **2021**, *39*, 4592–4600. [[CrossRef](#)]
97. Bangari, V.; Marquez, B.A.; Miller, H.; Tait, A.N.; Nahmias, M.A.; Lima, T.F.d.; Peng, H.T.; Prucnal, P.R.; Shastri, B.J. Digital Electronics and Analog Photonics for Convolutional Neural Networks (DEAP-CNNs). *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 7701213. [[CrossRef](#)]
98. Cheng, J.; Zhao, Y.; Zhang, W.; Zhou, H.; Huang, D.; Zhu, Q.; Guo, Y.; Xu, B.; Dong, J.; Zhang, X. A small microring array that performs large complex-valued matrix-vector multiplication. *Front. Optoelectron.* **2022**, *15*, 15. [[CrossRef](#)]
99. Mehrabian, A.; Al-Kabani, Y.; Sorger, V.J.; El-Ghazawi, T. PCNNA: A Photonic Convolutional Neural Network Accelerator. In Proceedings of the IEEE International System-on-Chip Conference (SOCC), Arlington, VA, USA, 4–7 September 2018; pp. 169–173.
100. Mehrabian, A.; Miscuglio, M.; Alkabani, Y.; Sorger, V.J.; El-Ghazawi, T. A Winograd-Based Integrated Photonics Accelerator for Convolutional Neural Networks. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 6100312. [[CrossRef](#)]
101. Xu, S.; Wang, J.; Yi, S.; Zou, W. High-order tensor flow processing using integrated photonic circuits. *Nat. Commun.* **2022**, *13*, 7970. [[CrossRef](#)]
102. Xu, S.; Wang, J.; Zou, W. Optical Convolutional Neural Network With WDM-Based Optical Patching and Microring Weighting Banks. *IEEE Photonics Technol. Lett.* **2021**, *33*, 89–92. [[CrossRef](#)]
103. Xu, S.; Wang, J.; Zou, W. Optical patching scheme for optical convolutional neural networks based on wavelength-division multiplexing and optical delay lines. *Opt. Lett.* **2020**, *45*, 3689–3692. [[CrossRef](#)]
104. Wu, C.; Yu, H.; Lee, S.; Peng, R.; Takeuchi, I.; Li, M. Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nat. Commun.* **2021**, *12*, 96. [[CrossRef](#)]
105. Huang, Y.; Zhang, W.; Yang, F.; Du, J.; He, Z. Programmable matrix operation with reconfigurable time-wavelength plane manipulation and dispersed time delay. *Opt Express* **2019**, *27*, 20456–20467. [[CrossRef](#)] [[PubMed](#)]
106. Meng, X.Y.; Shi, N.N.; Li, G.Y.; Zhang, G.J.; Li, W.; Zhu, N.H.; Li, M. On-Demand Reconfigurable Incoherent Optical Matrix Operator for Real-Time Video Image Display. *J. Light. Technol.* **2023**, *41*, 1637–1648. [[CrossRef](#)]
107. Meng, X.; Shi, N.; Shi, D.; Li, W.; Li, M. Photonics-enabled spiking timing-dependent convolutional neural network for real-time image classification. *Opt. Express* **2022**, *30*, 16217–16228. [[CrossRef](#)] [[PubMed](#)]
108. Xu, Z.; Tang, K.; Ji, X.; Sun, Z.; Wang, Y.; Hong, Z.; Dai, P.; Xiao, R.; Shi, Y.; Chen, X. Experimental demonstration of a photonic convolutional accelerator based on a monolithically integrated multi-wavelength distributed feedback laser. *Opt. Lett.* **2022**, *47*, 5977–5980. [[CrossRef](#)]
109. Zang, Y.; Chen, M.; Yang, S.; Chen, H. Optoelectronic convolutional neural networks based on time-stretch method. *Sci. China Inf. Sci.* **2021**, *64*, 122401. [[CrossRef](#)]

110. Huang, L.; Yao, J. Optical processor for a binarized neural network. *Opt. Lett.* **2022**, *47*, 3892–3895. [[CrossRef](#)]
111. Xu, S.; Wang, J.; Zou, W. High-energy-efficiency integrated photonic convolutional neural networks. *arXiv* **2019**, arXiv:1910.12635.
112. Xu, S.; Wang, J.; Wang, R.; Chen, J.; Zou, W. High-accuracy optical convolution unit architecture for convolutional neural networks by cascaded acousto-optical modulator arrays. *Opt. Express* **2019**, *27*, 19778–19787. [[CrossRef](#)]
113. Ashtiani, F.; Geers, A.J.; Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **2022**, *606*, 501–506. [[CrossRef](#)]
114. Hamerly, R.; Bernstein, L.; Sludds, A.; Soljačić, M.; Englund, D. Large-Scale Optical Neural Networks Based on Photoelectric Multiplication. *Phys. Rev. X* **2019**, *9*, 021032. [[CrossRef](#)]
115. Fan, L.; Zhao, Z.; Wang, K.; Dutt, A.; Wang, J.; Buddhiraju, S.; Wojcik, C.C.; Fan, S. Multidimensional Convolution Operation with Synthetic Frequency Dimensions in Photonics. *Phys. Rev. Appl.* **2022**, *18*, 034088. [[CrossRef](#)]
116. Zhen, W.; Zhou, X.; Weng, S.; Zhu, W.; Zhang, C. Ultrasensitive, Ultrafast, and Gate-Tunable Two-Dimensional Photodetectors in Ternary Rhombohedral ZnIn₂S₄ for Optical Neural Networks. *ACS Appl. Mater. Interfaces* **2022**, *14*, 12571–12582. [[CrossRef](#)] [[PubMed](#)]
117. Williamson, I.A.D.; Hughes, T.W.; Minkov, M.; Bartlett, B.; Pai, S.; Fan, S. Reprogrammable Electro-optic Nonlinear Activation Functions for Optical Neural Networks. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 1–12. [[CrossRef](#)]
118. Zuo, Y.; Li, B.H.; Zhao, Y.J.; Jiang, Y.; Chen, Y.C.; Chen, P.; Jo, G.B.; Liu, J.W.; Du, S.W. All-Optical neural network with nonlinear activation functions. *Optica* **2019**, *6*, 1132–1137. [[CrossRef](#)]
119. Guo, X.; Barrett, T.D.; Wang, Z.M.; Lvovsky, A.I. Backpropagation through nonlinear units for the all-optical training of neural networks. *Photonics Res.* **2021**, *9*, B71–B80. [[CrossRef](#)]
120. Filipovich, M.J.; Guo, Z.; Al-Qadasi, M.; Marquez, B.A.; Morison, H.D.; Sorger, V.J.; Prucnal, P.R.; Shekhar, S.; Shastri, B.J. Silicon photonic architecture for training deep neural networks with direct feedback alignment. *Optica* **2022**, *9*, 1323–1332. [[CrossRef](#)]
121. Wright, L.G.; Onodera, T.; Stein, M.M.; Wang, T.; Schachter, D.T.; Hu, Z.; McMahon, P.L. Deep physical neural networks trained with backpropagation. *Nature* **2022**, *601*, 549–555. [[CrossRef](#)]
122. Zhou, T.; Lin, X.; Wu, J.; Chen, Y.; Xie, H.; Li, Y.; Fan, J.; Wu, H.; Fang, L.; Dai, Q. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **2021**, *15*, 367–373. [[CrossRef](#)]
123. Zhou, T.; Fang, L.; Yan, T.; Wu, J.; Li, Y.; Fan, J.; Wu, H.; Lin, X.; Dai, Q. In situ optical backpropagation training of diffractive optical neural networks. *Photonics Res.* **2020**, *8*, 940–953. [[CrossRef](#)]
124. Hughes, T.W.; Minkov, M.; Shi, Y.; Fan, S.H. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* **2018**, *5*, 864–871. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.