Highly efficient convolution computing architecture based on silicon photonic Fano resonance devices*

NI Jiarong[1], LU Wenda[2], LAI Xiaohan[2], LU Lidan[1], OU Jianzhen[1], and ZHU Lianqing[1]****

1. Beijing Information Science and Technology University, Beijing 100192, China

2. State Grid Zhejiang Electric Power Corporation Information & Telecommunication Branch, Hangzhou 310007, China

(Received 16 March 2023; Revised 23 March 2023)

Convolutional neural networks (CNNs) require a lot of multiplication and addition operations completed by traditional electrical multipliers, leading to high power consumption and limited speed. Here, a silicon waveguide-based wavelength division multiplexing (WDM) architecture for CNN is optimized with high energy efficiency Fano resonator. Coupling of T-waveguide and micro-ring resonator generates Fano resonance with small half-width, which can significantly reduce the modulator power consumption. Insulator dataset from state grid is used to test Fano resonance modulator-based CNNs. The results show that accuracy for insulator defect recognition reaches 99.27% with much lower power consumption. Obviously, our optimized photonic integration architecture for CNNs has broad potential for the artificial intelligence hardware platform.

Document code: A **Article ID:** 1673-1905(2023)11-0646-7

DOI https://doi.org/10.1007/s11801-023-3047-4

The development of large-scale silicon photonic integration technique provides broader prospect and opportunity for optical computing compared with other photonic integration platforms. At the same time, silicon photonic devices become more compact and sensitive for tuning thanks to the benefit of the continuous advancement of manufacturing technique. It has been used in some large-scale reconfigurable linear systems[1-3]. Combining convolutional neural networks (CNNs) with photonic devices provides a brand-new way to process optical signals. With the help of flexible simulation system, controllable physical method, and the equation deviation, the neural network model is isomorphic[4-6]. The training and prediction of CNNs and other neural pseudo-morphic networks require a lot of computing resources. Although traditional computers are mature in the design and operation principles of the architecture, they are gradually facing the bottleneck of Moore's law and urgently need new architecture and technical regulations as substitutes. The large-scale silicon photonic integration has undeniable advantages in this field with ultra-fast speed and neural pseudo morphic structure.

There are two types of implementations of weights or photonic synapses, mainly including Mach-Zehnder interferometer (MZI)[7] and micro ring resonator (MRR)-based wavelength division multiplexing (WDM) architecture[8,9]. Herein, the MRR-based WDM architecture performs higher integration, computing energy efficiency ratio, and computing density.

One kind of photonic convolution neural network accelerator (PCNNA) is proposed to accelerate the convolution operation[10]. The accelerator utilizes the synergy among WDM, parallelism of light, the sparsity of input feature mapping, and connection between cores in CNN.

The compatibility of silicon photonic architecture and CNN is reflected in the parallelism of the micro-ring resonance array, the transmissibility of the physical resonance equation, and the mathematical equation of convolution calculation. The silicon photonics-based CNN's structure is similar to the biological neural network's synaptic structure. When the neural signal below a certain threshold is transmitted to the synapse, it is blocked. Otherwise, it would pass and multiply the weight value. The weight corresponds to the different degrees of response in the biological neural network. Moreover, the micro-ring array can repeat multiple columns to achieve a deeper neural network structure[11].

To further improve the MRR-based WDM architecture performance, the researchers optimize the performance of neuron nodes (modulators). Some researchers use photonic crystal resonators as neuron modulation units, but the process cannot be realized on the standard multi-project wafer process line[12]. Based on our research group in MRR optimization, the Fano resonance

* This work has been supported by the Science and Technology Project of the State Grid Zhejiang Electric Power Company Limited (No.B311XT21004G).

** E-mails: lldan_dido@bistu.edu.cn; zhulianqing2020@126.com

effect is used to reduce the system's operation power consumption[13]. In this paper, the principle of CNN and the transfer matrix of the MRR-based Fano resonator are investigated. The optoelectronic hybrid architecture and the generation of the Fano lineshape are introduced. The insulator dataset is used to test the energy efficiency ratio of the architecture, and its performance is compared with that of the current architecture.

Furthermore, this study primarily relies on simulation experiments conducted using the transfer matrix method. The analysis results are derived mainly from theoretical calculations and numerical simulations.

With the increase in the depth and breadth of CNNs, the amount of computation required for its training and inference is also increasing rapidly, especially for convolution computation. One of the challenges with convolutions is that they are computationally intensive operations, taking up from 86% to 94% of execution time for CNNs.

CNNs are a type of artificial neural network commonly used in deep learning and image processing tasks. CNNs require a significant amount of computational power and energy to process large amounts of data with convolution operation, making them difficult to deploy on low-power and mobile devices. The power consumption of CNNs can be very high due to the large number of computations required for training and inference. High power consumption can limit the practical deployment of CNNs, particularly in battery-powered devices such as smartphones or wearables. To meet the demands of computing equipment progress, researchers are exploring ways to increase efficiency, such as optical computing. Optical computing using silicon photonics is a promising approach for achieving ultrafast information processing. It involves using light instead of electricity to perform computations and transmit data.

As shown in Fig.1, convolution is integral that expresses the amount of overlap of one function as it is shifted over another function. Each convolution kernel has three dimensions, namely length, width, and depth. The depth of the convolution core is the same as the depth of the current image (the number of feather maps). For example, in the original image layer (input layer), if the image is a gray image and the number of feather maps is 1, the depth of the convolution core is 1. If the image is color and the number of feather maps is 3, the depth of the convolution kernel is 3.

The size of an image I is $H\times W\times C$. H and W are the length and width of the image, and C is the number of channels of the picture. The convolution kernel is R in size, and the number of channels is Z. Use the convolution check image for convolution calculation. R represents the size of kernel F. Image A transformed into a matrix A of dimensionality. The stride size is S. The value of a particular output pixel is defined by

$$O_{i,j}=\sum_{k=1}^{R}\sum_{l=1}^{R}F_{k,l}A_{i+k,j+l}. \qquad (1)$$

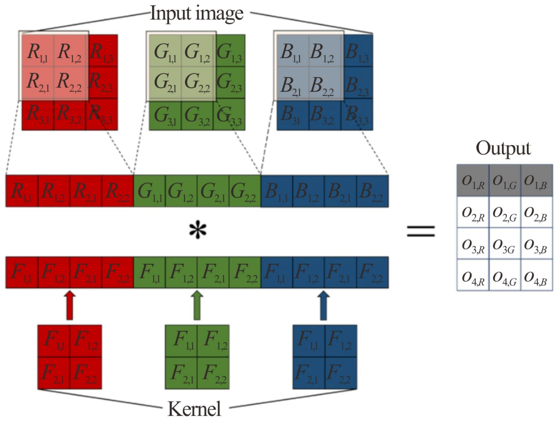

Fig.1 Convolution computation scheme ($R_{i,j}$, $G_{i,j}$, and $B_{i,j}$ represent different channels of input images; The input image's value represents a pixel's intensity at that spatial location)

Convolution calculation cannot be directly supported on graphics processing unit (GPU), so when performing convolution calculation on GPU, it is necessary to first preprocess and expand the convolution core and characteristic graph according to specific rules. Fig.1 shows how to calculate convolution in an RGB image using GPU.

As shown in Fig.2, the system for convolution calculation contains a light source with multiple wavelengths, an input image module, a weight bank, and photodetectors. Input image module and weight bank compose of multiple microrings with Fano resonance. Through WDM technology, lasers with different wavelengths are transmitted in the photonic chip. The pixels of the insulator image are loaded into the throughput Fano resonance modulator with high bandwidth. The weight bank composes of multiple add-drop Fano resonance modulators with low bandwidth. Because the Fano resonance modulator is sensitive to heat, electrical heating is used to modulate it, which shifts the resonance peak, thereby controlling the output light intensity to achieve multiplication operations. The photodiode converts the optical signal into an electrical signal, realizing the function of the adder. Then, transimpedance amplifier (TIA) amplifies the signal strength with a magnified multiply-add calculation result.

The source of insulator dataset is work result of TAO el al[14]. The insulator dataset is used as the input and divided into normal insulators and defective insulators. The dataset is generated by manually labeling the images with Labeling software. Due to the small number of defective images, the dataset of defective insulators is expanded by data augmentation (DA).

First, input image $A_{i,i}$ can be reshaped to a column of pixels. This step transforms the input image into a series of input streams, which is more natural to silicon photonics devices and promotes the convolution's efficiency. Pixels flow into the corresponding channel and convolution kernel for multiplication and addition. The kernel is a matrix with dimensionality of $R\times R\times Z$. The kernel slides over the image using the stride S. After the convolution kernel sliding, the size of the output graph is $(H-1)\times(W-1)\times C$, filling the edge with pixels with a value of 0.

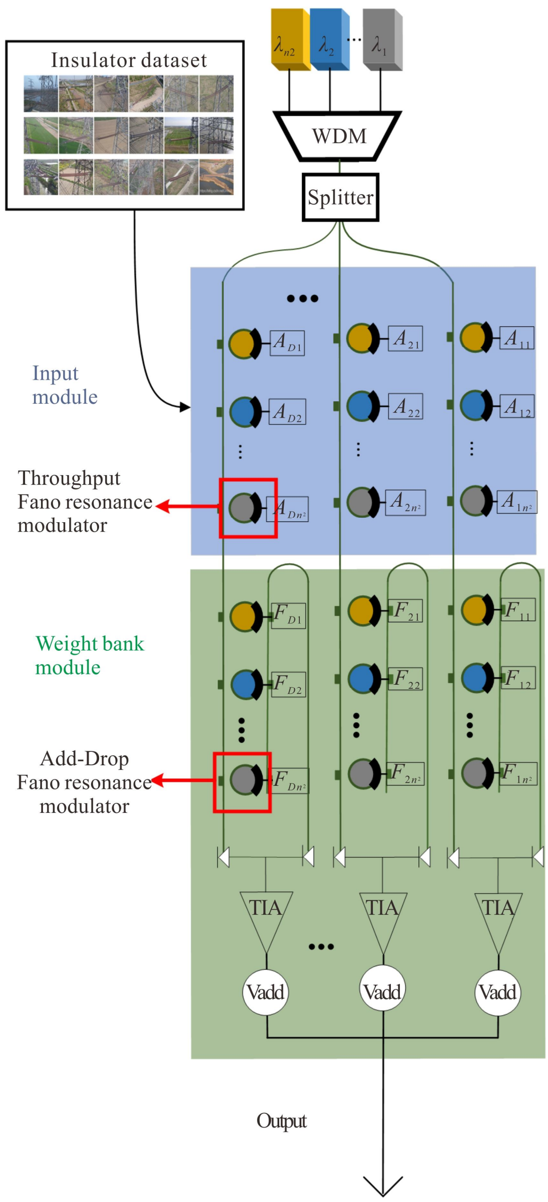

Fig.2 Photonic architecture for convoluting

Photonic product units are highly parallel. The dimension of the input image is $n\times n\times d$, where n represents the size of input pixels, and d represents the number of multiplex channels. Modulators modulate the light intensity. The input pixel value is transferred to the modulator with proper light intensity. It can input several pixels simultaneously, such as 4 pixels. Here, the WDM is used to merge different lasers with a single wavelength, realizing multiple wavelengths loads simultaneously. Multiple convolution calculations need to be performed. There must be a concurrence of several convolution kernels at the same time. This function can be realized by adding a beam splitter, followed by WDM.

Fano resonance is an interference phenomenon leading to a unique asymmetric spectral response caused by interference between continuous and discrete states. Fano resonance is commonly observed in various physical systems, including atoms, molecules, photonic structures, and plasmonic materials. The relationship between wavelength and transmission in a silicon waveguide depends on the waveguide's dispersion properties, which determine how the waveguide responds to different wavelengths of light. Generally, the transmission of light decreases as the wavelength increases due to the waveguide's dispersion properties.

Two methods to describe Fano resonance are coupled-mode theory (CMT) and the transfer matrix method (TMM). CMT is a more straightforward and intuitive method that can provide a qualitative understanding of the Fano resonance phenomenon. TMM is a more rigorous and detailed approach that can handle more complex structures and give a quantitative description of the spectral response. To analyze structure directly, we choose the TMM method to interpret Fano resonance.

In the transfer matrix method, the Fano resonance can be modeled by considering the transmission of light through a layered structure with a resonant mode coupled to a continuum of states. The resonance mode can be described by a Lorentzian lineshape with a central frequency and a linewidth, while a constant transmission coefficient describes the continuum of states. The interference between the resonant mode and the continuum brings on the Fano resonance. The Fano resonance in a silicon waveguide can be expressed using the following formula

$$T(\omega)=|t(\omega)|^2=\frac{|t_{\mathrm{bg}}(\omega)+t_{\mathrm{cv}}(\omega)|^2}{1+q^2}, \tag{2}$$

where $T(\omega)$ is the transmission spectrum, and $t(\omega)$ is the complex transmission amplitude, which describes both the amplitude and the phase of the transmitted wave. $t_{\mathrm{bg}}(\omega)$ is the transmission amplitude of the background continuum of the guided modes, $t_{\mathrm{cv}}(\omega)$ is the transmission amplitude of the resonant mode in the cavity, and q is the dimensionless Fano parameter that describes the relative strengths of the resonance and continuum modes. The parameter q depends on the shape of the Fano resonance.

The structures of the cavity or the properties of the waveguide can change Fano resonance. Resonance is symmetric when q is small, similar to the Lorentz lineshape. The resonance becomes highly asymmetric when q is large, and a sharp peak occurs on the particular position.

In this work, we use the T-waveguide coupled MRR to generate Fano resonance based on a standard silicon-on-insulator (SOI) with top silicon of 220 nm. The width and thickness of the waveguide are 500 nm and 220 nm, respectively. The structure of T-waveguide coupled MRR is shown in Fig.3(c).

Fig.3(a) shows the structure of add-drop MRR. The input light passes through the coupling region, and a part of it continues to transmit in the upper straight waveguide, forming a continuous propagation field. Moreover, the rest of the light is coupled into the MRR. This part of the light forms a discrete resonant wave. The light coupled into the micro-ring propagates one circle in the micro-ring, resulting in phase delay. When the light in the micro-ring enters the other coupling region, part light would be coupled to the straight waveguide.

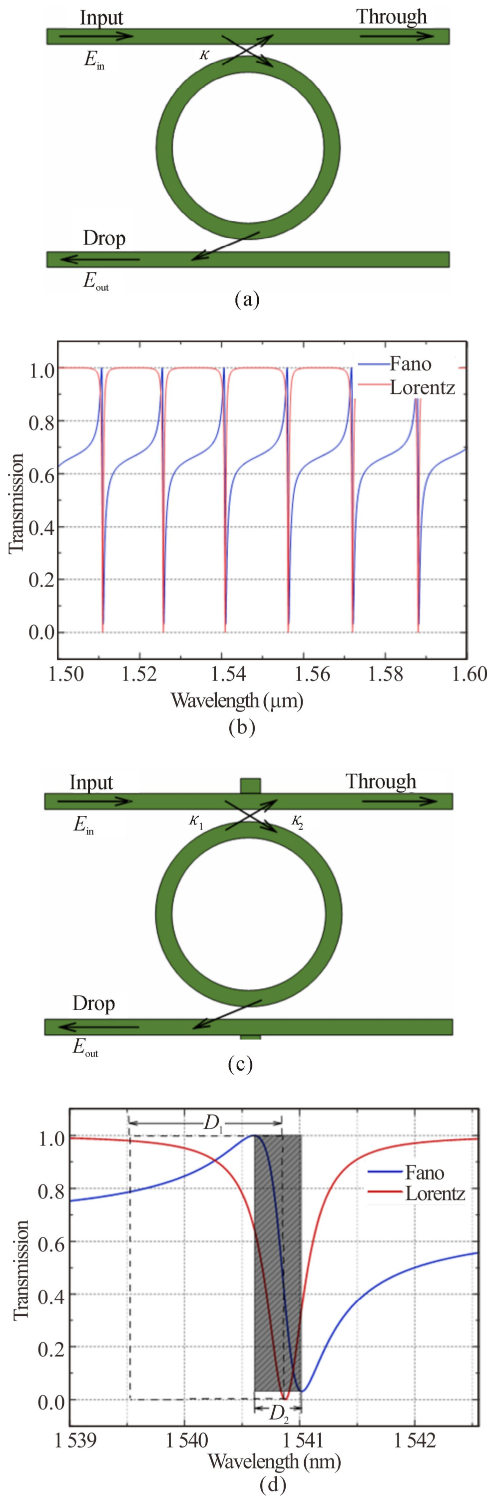

Fig.3 (a) Parallel channel MRR structure; (b) Fano curve generated by a T-shape structure and the Lorentz curve generated by the add-drop MRR; (c) T-shape waveguide MRR structure; (d) Comparison between Lorentz and Fano curves (The wavelength range for tuning the transmission from 0 to 1 is much narrower in Fano lineshapes, which reduces the cost of modulation, and strengthens the merits of resolution of optical spectrum and extinction ratio)

According to Ref.[15], to generate Fano resonance lineshapes, the phase difference between the discrete (microring) and continuous states (T-shaped waveguide) must be changed, but not an integer multiple of 2π. Therefore, the waveguide is designed as a T-shaped waveguide, which makes the bidirectional coupling coefficient in the coupling area different. κ_1 and κ_2 represent the forward and reverse coupling coefficients of the coupling region, t represents the amplitude transmission coefficient of the coupling area between the waveguide and the input/output fiber, and $\kappa^2=\kappa_1\times\kappa_2=1-t^2$. This structure can make the continuous propagation mode in the straight waveguide produce a phase shift factor, thus the electrical filed of through port can be expressed as

$$\begin{aligned} E_{\mathrm{out}} &= tE_0\cdot \mathrm{e}^{-\mathrm{i}\Delta\phi} + \mathrm{i}\kappa E_1 + \mathrm{i}\kappa E_2 + \cdots = \\ &tE_0\mathrm{e}^{-\mathrm{i}\Delta\phi} + \mathrm{i}\kappa_1(\mathrm{i}\kappa_2)\alpha\mathrm{e}^{\mathrm{i}\delta}E_0 + \mathrm{i}\kappa_1(\mathrm{i}\kappa_2)t\alpha\mathrm{e}^{\mathrm{i}2\delta}E_0 + \cdots = \\ &tE_0\mathrm{e}^{-\mathrm{i}\Delta\phi} + \mathrm{i}\kappa_1(\mathrm{i}\kappa_2)\frac{a\mathrm{e}^{\mathrm{i}\delta} - t^n a^{n+1}\mathrm{e}^{\mathrm{i}(n+1)\delta}}{1-ta\mathrm{e}^{\mathrm{i}\delta}}E_0(t<, n\to\infty) = \\ &\left(t\mathrm{e}^{-\mathrm{i}\Delta\phi} - \frac{\kappa_1\kappa_2\alpha\mathrm{e}^{\mathrm{i}2\pi^2 nR/\lambda}}{1-t\alpha\mathrm{e}^{\mathrm{i}2\pi nL/\lambda}}\right)E_0. \end{aligned} \tag{3}$$

The following formula gives the transmission spectrum of output of the coupling system as

$$T(\lambda) = \left|\frac{E_{\mathrm{out}}}{E_{\mathrm{in}}}\right|^2 = \left|t\mathrm{e}^{-\mathrm{i}\Delta\phi} - \frac{\kappa_1\kappa_2 a\mathrm{e}^{\mathrm{i}2\pi nL_R/\lambda}}{1-ta\mathrm{e}^{\mathrm{i}2\pi nL_R/\lambda}}\right|^2. \tag{4}$$

According to Fig.3(c), the Fano resonance arises from the interference between a discrete resonance mode and a broad continuum of modes, resulting in a sharp rise in the transmission spectrum. The interference between the resonant and non-resonant modes in a Fano resonance can result in a high-Q factor. Q factor can be used to enhance the light-matter interaction and increase the efficiency of the devices. In conclusion, Fano resonance has advantages of compatibility with silicon-based technology, low power consumption, high efficiency, and small footprint.

The temperature of the waveguide would change by current heating. The refractive index of the waveguide and the phase of the light can be modified with a changing temperature. The effective refractive index in silicon can change by temperature for the thermo-optic effect. Thermo-optic coefficient of silicon is 1.84×10^{-4}/K. A slight temperature variation would change the transmission spectrum of MRR. It has been verified that the Fano resonance devices can obtain high-temperature sensitivity[16-19].

As shown in Fig.4(a), heating the micro-ring causes the phase shift of Fano resonance. The original spectrum's peak wavelength corresponds to the phase-shifted spectrum's trough, which acurrately modulates transmission to the range of [0, 1].

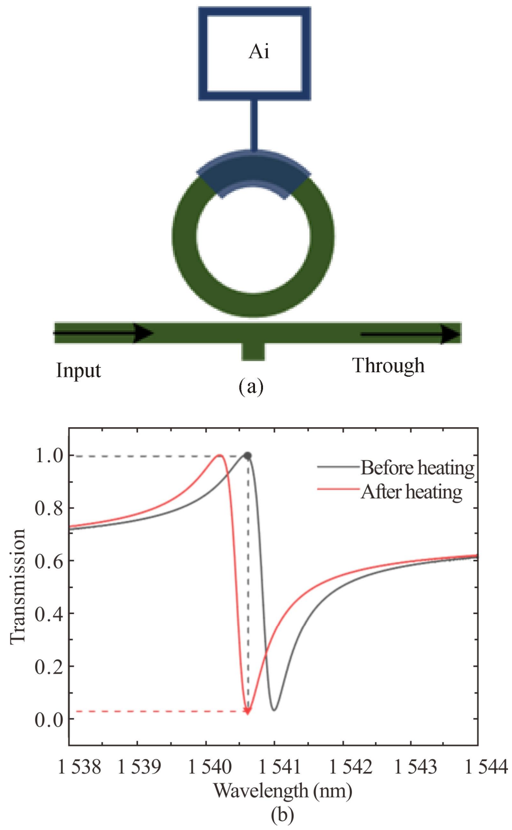

Fig.4 (a) Add-drop waveguide; (b) Phase shift of Fano line caused by heating waveguide

To extend the transmission of the output to [−1, 1], the T-waveguide structure in Fig.5(a) is applied here, and the light intensity at the added end and the drop end is subtracted. Since $T_{add}+T_{through}=1$ (without considering loss), subtracting T_{add} and $T_{through}$ shown in Fig.5(b) would achieve the weight value of [−1, 1] shown in Fig.5(c). Photodiode converts the optical signal in waveguide into an electrical signal. Two balanced photodiodes with opposite directions are connected at the drop and through ports of the T-shaped coupled MRR for transmission subtraction. The basic principle of a TIA is to convert the current signal from the photodiode into a voltage signal by using a feedback resistor. The output voltage of the TIA is proportional to the input current, and the feedback resistor determines the proportionality constant. After TIA amplification, the kernel value is not limited to the range of [−1, 1]. This electronic gain can be performed using a TIA, which can be manufactured in a standard CMOS process and packaged or integrated with the photonic chip[20].

The network comprises of an image loading module and a kernel module. The image loading module is based on an electrical modulator with a modulation speed of giga hertz. On the other hand, the kernel module uses a thermal modulator with a modulation speed of kHz level. However, the kernel is relatively fixed and does not require a high loading speed.

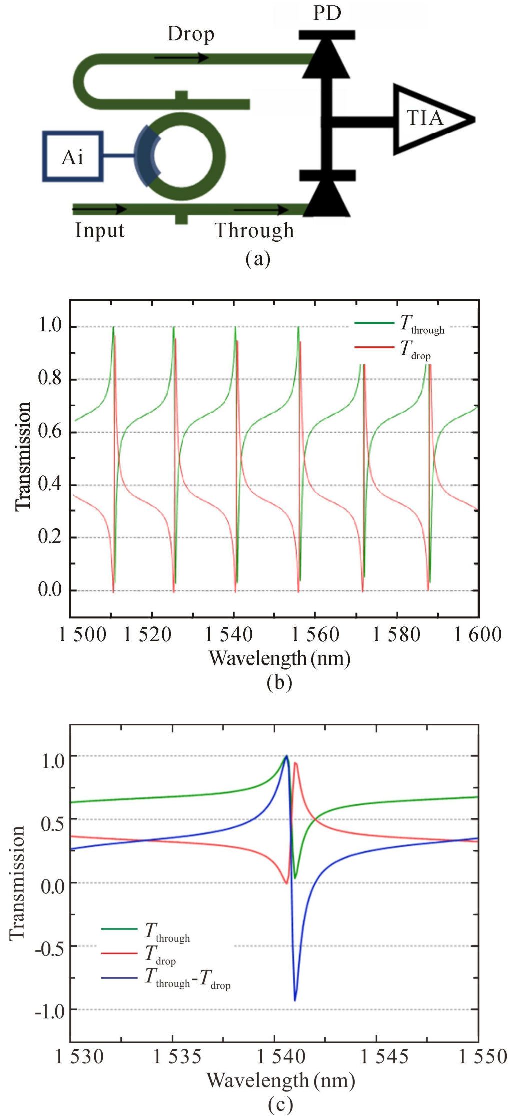

Fig.5 (a) Add-drop MRR modulator; (b) Output of through and drop; (c) $T_{through}$, T_{drop} and $T_{through}$-T_{drop} (The convolution kernel of [-1, 1] or broader can be realized using this product unit)

In a previous study, RAHIM et al[21] investigated the relationship between the power of the heater and the shift in resonance wavelength. They found that the amount of wavelength shift is directly proportional to the heater's power. These results suggest that Fano resonance can significantly enhance the efficiency of thermal tuning.

It is necessary to derive the relationship between electrical modulation power and wavelength drift.

In 2011, NEDELJKOVIC et al[22] updated the relationship between the refractive index and absorption coefficient of silicon, which is caused by changes in free carrier concentration. The relationship for 1 550 nm is as follows

$$\begin{cases}\Delta n=-5.4\times10^{-22}\Delta N^{1.011}-1.53\times10^{-18}\Delta P^{0.838} \\ \Delta\alpha=8.88\times10^{-21}\Delta N^{1.167}+5.84\times10^{-20}\Delta P^{1.109}\end{cases}, \quad (5)$$

where Δn represents changes in the real part of the refractive index, $\Delta\alpha$ represents changes in the absorption coefficient, and ΔN and ΔP represent changes in the concentration of electrons and holes. The calculation formula for junction capacitance is given as follows[23]

$$C_j=q\frac{\Delta n_e}{\Delta V}=q\frac{\Delta n_h}{\Delta V}. \quad (6)$$

The modulation power consumption of the MRR modulator can be estimated based on the junction capacitance

$$P_{av}=\frac{C_j V_{pp}^2\sum_{i=1}^{N-1}(N-i)\left(\frac{i}{N-1}\right)^2}{N^2\log_2 N}, \quad (7)$$

where N represents the amplitude level, while V_{pp} represents the peak value of the driving voltage. As modulation power consumption is proportional to the square of the driving voltage, improving the modulation efficiency can significantly reduce modulation power consumption by decreasing the drive voltage. The shift in resonant wavelength under different bias voltages is denoted as $\Delta\lambda$, and the refractive index is given as

$$\Delta n_{eff}=\frac{n_g\Delta\lambda}{F\lambda}, \quad (8)$$

where λ represents the resonant wavelength of MRR resonator, and F is the ratio of the p-n junction length to the length of the microring resonator. The relationship between the wavelength shift and modulation power can be deduced from the above formula as follows

$$P_{av}=\frac{qn_g\Delta\lambda V_{pp}^2\sum_{i=1}^{N-1}(N-i)(\frac{i}{N-1})^2}{F\lambda\Delta VN^2\log_2 N}. \quad (9)$$

It has been proven that for a given electrically tunable microring resonator, the relationship between wavelength shift and modulation power is linear. When the wavelength shift decreases to 0.07 times, the electrically tunable power also decreases to 0.07 times.

We train the CNN network architecture on the computer, then upload the network parameters to the insulator defect detection architecture and detect the insulator defects based on the architecture. Finally, the performance of the CNN architecture is analyzed. A lightweight algorithm model can reduce the consumption of hardware resources. However, the lightweight model can easily lead to information loss and precision reduction to match the ability of algorithms and hardware devices. How to reduce the precision loss is a complex problem. The neural network's performance can be optimized by extensively optimizing network architecture, minimizing quantization error, and other methods.

As shown in Fig.6, we simulate the network with two convolution layers and two full connection layers. The size of the first and second convolution layer are (2, 3, 2, 2) and (4, 3, 2, 2), repectively. The size of full connect layers are (1 936, 50) and (50, 2). To adapt to the device, the convolution network has been appropriately quantized. For example, mapping pixels in the range from [0, 255] to [0, 1], which would result in accuracy loss. Here, the quantized accuracy of 5 bits is adopted. Input pixel of insulator image is 1 152×864. Finally, the predicted accuracy of defected insulator is 99.27% as shown in Fig.7.

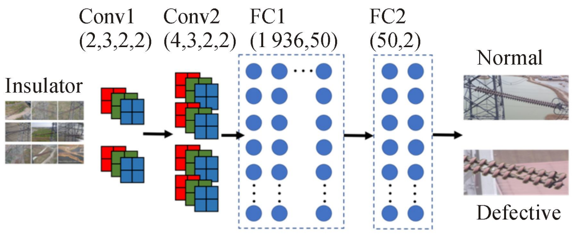

Fig.6 An illustrative block diagram of the two-layer CNN solving insolators

Here, T-shaped waveguide coupled MRR-based CNN architecture is adoped. Compared with MRR-based CNN, our designed CNN can maintain the same accuracy. However, the power consumption is lower for a lager sprectrum resolution for T-shaped waveguide coupled MRR Fano resonator. Our team's previous research shows that under the condition of achieving the same light intensity change, Fano resonance only needs to shift the wavelength by 0.07 times compared with MRR. Moreover, it has same computing density with normal resonance.

Fig.7 presents the results of functional verification based on an architecture that includes four multiplexed wavelengths, eight MRR resonators, and two channels. Each convolution extracts four pixel values from the image, which are then transferred to light intensity using modulators. The 2×2×2 kernel is loaded onto eight thermally modulated MRRs, and the modulated light is multiplied with the eight thermally modulated MRRs. The calculation results are outputted at the end. The results of the insulator defect recognition task showed that overall accuracy reached 99.27%.

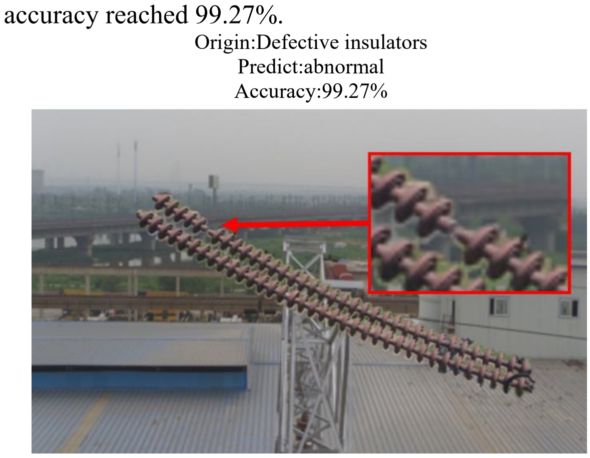

Fig.7 Insulator recognition

In conclusion, a highly efficient convolutional computing architecture based on silicon photonic Fano resonant devices is proposed. Using Fano resonance, we optimize the performance of the MRRs. The negative weight bank values can be applied by subtracting the optical and electrical signals at the drop and through ports of the MRR. The predicted accuracy of defected insulator is 99.27%. Overall, the photonic integration architecture with Fano resonance MRR has broad prospect and application potential in the future field of neural network photonic chips, which can significantly improve the energy efficiency of optical computing.

Ethics declarations

Conflicts of interest

The authors declare no conflict of interest.

References

[1] PÉREZ D, GASULLA I, CRUDGINGTON L, et al. Multipurpose silicon photonics signal processor core[J]. Nature communications, 2017, 8(1)：636.

[2] LIU W, LI M, GUZZON R S, et al. A fully reconfigurable photonic integrated signal processor[J]. Nature photonics, 2016, 10(3)：190-195.

[3] HARRIS N C, BUNANDAR D, PANT M, et al. Large-scale quantum photonic circuits in silicon[J]. Nanophotonics, 2016, 5(3)：456-468.

[4] TAIT A N, DE LIMA T F, ZHOU E, et al. Neuromorphic photonic networks using silicon photonic weight banks[J]. Scientific reports, 2017, 7(1)：1-10.

[5] SHEN Y, HARRIS N C, SKIRLO S, et al. Deep learning with coherent nanophotonic circuits[J]. Nature photonics, 2017, 11(7)：441-446.

[6] SHAINLINE J M, BUCKLEY S M, MIRIN R P, et al. Superconducting optoelectronic circuits for neuromorphic computing[J]. Physical review applied, 2017, 7(3)：034013.

[7] ZHANG H, GU M, JIANG X D, et al. An optical neural chip for implementing complex-valued neural network[J]. Nature communications, 2021, 12(1)：457.

[8] XU S, WANG J, YI S, et al. High-order tensor flow processing using integrated photonic circuits[J]. Nature communications, 2022, 13(1)：7970.

[9] MISCUGLIO M, SORGER V J. Photonic tensor cores for machine learning[J]. Applied physics reviews, 2020, 7(3)：031404.

[10] MEHRABIAN A, AL-KABANI Y, SORGER V J, et al. A photonic convolutional neural network accelerator[C]//2018 31st IEEE International System-on-Chip Conference (SOCC). New York：IEEE, 2018：169-173.

[11] ZHOU H, QIU C, JIANG X, et al. Compact, submilliwatt, 2×2 silicon thermo-optic switch based on photonic crystal nanobeam cavities[J]. Photonics research, 2017, 5(2)：108-112.

[12] LU L, ZHU L, ZENG Z, et al. Fano resonance ion sensor enabled by 2D plasmonic sub-nanopores-material[J]. IEEE sensors journal, 2021, 21(13)：14776-14783.

[13] XU Y, LU L, CHEN G, et al. T-shaped silicon waveguide coupled with a micro-ring resonator-based Fano resonance modulator[J]. Applied optics, 2022, 61(31)：9217-9224.

[14] TAO X, ZHANG D, WANG Z, et al. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks[J]. IEEE transactions on systems, man, and cybernetics：systems, 2018, 50(4)：1486-1498.

[15] GU L, FANG L, FANG H, et al. Fano resonance lineshapes in a waveguide-microring structure enabled by an air-hole[J]. APL photonics, 2020, 5(1)：016108.

[16] ZHANG Y, ZOU J, HE J J. Temperature sensor with enhanced sensitivity based on silicon Mach-Zehnder interferometer with waveguide group index engineering[J]. Optics express, 2018, 26(20)：26057-26064.

[17] DING Z, LIU P, CHEN J, et al. On-chip simultaneous sensing of humidity and temperature with a dual-polarization silicon microring resonator[J]. Optics express, 2019, 27(20)：28649-28659.

[18] CHEN F, ZHANG H, SUN L, et al. Temperature tunable Fano resonance based on ring resonator side coupled with a MIM waveguide[J]. Optics & laser technology, 2019, 116：293-299.

[19] XU Y, OU Z, CHEN J, et al. High sensitivity refractive index and temperature sensors with tunable multiple fano resonances[C]//2021 IEEE 6th Optoelectronics Global Conference (OGC), August 31-September 3, 2021, Shenzhen, China. New York：IEEE, 2021：239-242.

[20] ZHENG H, MA R, ZHU Z. A linear and wide dynamic range transimpedance amplifier with adaptive gain control technique[J]. Analog integrated circuits and signal processing, 2017, 90：217-226.

[21] RAHIM A, SPUESENS T, BAETS R, et al. Open-access silicon photonics：current status and emerging initiatives[J]. Proceedings of the IEEE, 2018, 106(12)：2313-2330.

[22] NEDELJKOVIC M, SOREF R, MASHANOVICH G Z. Free-carrier electrorefraction and electroabsorption modulation predictions for silicon over the 1-14 μm infrared wavelength range[J]. IEEE photonics journal, 2011, 3(6)：1171-1180.

[23] VAN MIEGHEM P, MERTENS R P, VAN OVERSTRAETEN R J. Theory of the junction capacitance of an abrupt diode[J]. Journal of applied physics, 1990, 67(9)：4203-4211.