electronics

MDPI

Article

Integrated Photonic Processor Implementing Digital Image Convolution

Chensheng Wang [1,2], Wenhao Wu [2], Zhenhua Wang [2], Zhijie Zhang [2], Wei Xiong [1] and Leimin Deng [1,*]

[1] Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China; d201980829@hust.edu.cn (C.W.); weixiong@hust.edu.cn (W.X.)

[2] Huazhong Institute of Electro-Optics, Wuhan 430299, China; wuwenhao@cssc717.com (W.W.); wangzhenhua@cssc717.com (Z.W.); zhangzhijie@cssc717.com (Z.Z.)

* Correspondence: dlm@hust.edu.cn

Abstract: Upon the advent of the big data era, information processing hardware platforms have undergone explosive development, facilitating unprecedented computational capabilities while significantly reducing energy consumption. However, conventional electronic computing hardware, despite significant upgrades in architecture optimization and chip scaling, still faces fundamental limitations in speed and energy efficiency due to Joule heating, electromagnetic crosstalk, and capacitance. A new type of information processing hardware is urgently needed for emerging data-intensive applications such as face identification, target tracking, and autonomous driving. Recently, integrated photonics computing architecture, which possesses remarkable compactness, wide bandwidth, low latency, and inherent parallelism, has harvested great attention due to its enormous potential to accelerate parallel data processing, such as digital image convolution. In this study, an integrated photonic processor based on a Mach-Zehnder interferometer (MZI) network is proposed and demonstrated. The processor, being scalable and compatible with complementary metal oxide semiconductors, facilitates mass production and seamless integration with other silicon-based optoelectronic devices. An experimental verification for digital image convolution is also performed, and the result deviations between our processor and a commercial 64-bit computer are less than 2.3%.

Keywords: photonics computing; optoelectronic integration; artificial intelligence; optical neural networks; integrated photonics; digital image convolution

check for updates

Received: 8 January 2025
Revised: 31 January 2025
Accepted: 11 February 2025
Published: 12 February 2025

Citation: Wang, C.; Wu, W.; Wang, Z.; Zhang, Z.; Xiong, W.; Deng, L. Integrated Photonic Processor Implementing Digital Image Convolution. *Electronics* **2025**, *14*, 709. https://doi.org/10.3390/electronics14040709

1. Introduction

Image processing techniques have been used extensively in many different applications, such as medical diagnosis [1–4], intelligent driving [5–8], and target identification [9–11]. Digital image convolution is one of the most fundamental processing techniques for extracting image features [12–15]. This convolution process involves multiplying the image pixel matrix with a kernel, encompassing numerous parallel multiply-accumulate (MAC) operations. The increasing demand for real-time and high-quality image processing has sparked a rapid expansion in custom hardware designed to accelerate MAC operations. Diverse electronic computing hardware, encompassing field-programmable gate arrays (FPGAs) and graphics processing units (GPUs), have been developed to enhance computational capabilities. Meanwhile, various fast and efficient computing technologies have also been developed, such as reversible computing [16] and neuromorphic computing [17]. However, against the backdrop of gradual failure of Moore's law [18], these electronic schemes still encounter bottlenecks in terms of speed and energy efficiency. The enhancement of computing speed heavily depends on the scaling up of hardware, resulting in an annual surge in data center costs and power consumption. However, even this expansion fails to keep pace with the explosive development of technologies such as artificial intelligence, cloud computing, and big data. Furthermore, a multitude of small- and medium-sized unmanned platforms, including drones and robots, hindered by their restricted payload capacities, cannot accommodate bulky computing hardware, thereby generating an imperative need for solutions that offer high computing power with low power consumption. In recent years, photonic computation, which employs photons as the information carrier instead of electrons, has been rapidly developed due to its numerous advantages, including intrinsically large bandwidth, low latency, and high parallelism. Although a prototype made up of discrete devices was demonstrated decades ago [19], it remained excessively bulky and unstable. To address these limitations, photonic integration technology [20–22] was introduced, offering compactness, scalability, and cost-effectiveness. In 2007, Shen et al. pioneered the concept of a coherent nanophotonic chip using Mach-Zehnder interferometer (MZI) meshes for vowel recognition [23], which opened a new era of integrated photonic computation. Since then, various integrated photonic computing chips have been extensively reported, primarily categorized into two groups: those utilizing MZI meshes [24–27] and those employing micro-ring (MRR) meshes [28–30]. The former typically utilizes a single coherent light source and carries out MAC operations through light interference within the MZI meshes, while the latter employs multiple light sources with varying wavelengths, modulated by MRRs operating in distinct states, for loading weights during MAC operations. Both types of integrated photonic computing chips possess their own advantages and are widely studied. In this paper, taking into account the consumption of light sources, we opted for the cascaded MZIs configuration and presented a scalable silicon-based photonic computing processor capable of executing digital image convolution. The proposed chip, measuring 1.5 mm × 6 mm, is comprised of 20 MZIs and capable of executing arbitrary matrix transformations with a dimension of 4 × 4. Along with the off-chip laser source, photodetector (PD) arrays, and upper computer, a digital image convolution experiment platform is constructed based on the packaged photonic computing chip. A self-configuring algorithm based on gradient descent method is utilized for weight training to load convolution kernel. The proposed chip is characterized by comparing with a 64-bit computer in performing convolution for a digital image with a resolution of 320 × 256, and the relative computation error is less than 2.3%. Under plausible assumptions, notably the integration of cutting-edge photonic I/O technology and the realization of substantially larger chip dimensions, the proposed processor promises remarkable enhancements in both computing speed and energy efficiency, potentially achieving improvements spanning one to two orders of magnitude when compared to current top-tier electronic computing devices, such as NVIDIA's AI computing cards.

2. Device Design and Experimental Setup

The schematic of the photonic computing chip demonstrated in this work is shown in Figure 1. The chip consists of five parts. Part (1) is a 1-to-4 power splitter, and Part (2) is composed of four parallel MZIs, which connect to the respective outputs of the power splitter. These MZIs are used to load the input signals by modulating the light intensity. Parts (3), (4), and (5) are all MZI arrays, which can perform an arbitrary matrix transformation as a whole according to singular value decomposition [31]. Specifically, Parts (3) and (5) are the same and are composed of six MZIs, respectively, which are arranged as a rectangular mesh, as reported in Ref. [29]. These two parts can perform arbitrary unitary matrix transformation. Part (4) has 4 MZIs used for intensity attenuation, achieving arbitrary diagonal matrix transformation. The rectangular scheme, rather than the triangular scheme [32] designed by Reck, is chosen in order to halve the optical depth, which is important for minimizing transmission loss and reducing chip size. Besides, the rectangular scheme has a natural symmetry that makes it significantly more robust to fabrication errors [31]. In general, the chip contains 20 MZIs and 40 phase shifters in total.

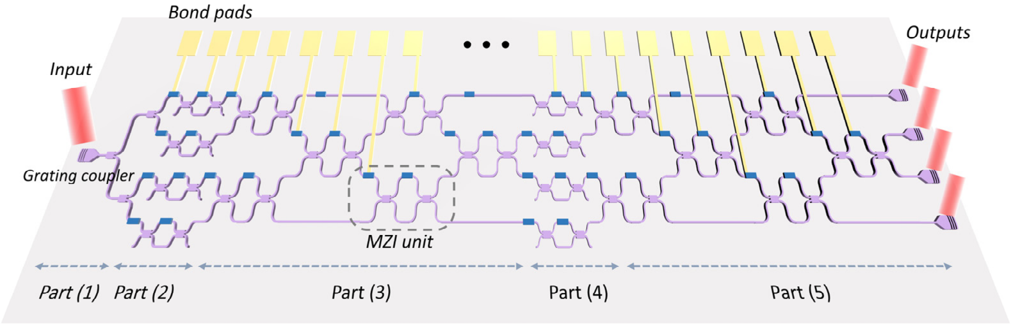

Figure 1. The schematic of the photonic computing chip. The red components signify the input/output optical signals, whereas the purple section represents the silicon waveguide. Positioned on the waveguide, the blue squares denote the thermal phase shifters. Furthermore, the gold squares indicate the metal bond pads, which are interconnected to the thermal phase shifters on-chip and external driving circuit through a wire-bonding process (not illustrated in the figure).

The external light is firstly coupled into the photonic computing chip through a grating coupler, subsequently divided into four equal beams by Part (1). Following transmission through Part (2), all of them undergo modulation with corresponding electrical signals, subsequently being mixed within the MZI network encompassing Parts (3), (4), and (5), which performs MAC operations in the optical domain through splitting and interference. Eventually, the four mixed light beams are coupled out of the chip via grating couplers and captured by four commercial photodetectors. The aforementioned process is capable of executing matrix-vector multiplication, represented by the equation $\boldsymbol{X}\cdot\boldsymbol{B} = \boldsymbol{A}$. In this equation, $\boldsymbol{B}$ denotes a four-dimensional vector determined by the input electrical signals transmitted to Part (2), $\boldsymbol{X}$ represents the matrix transformation carried out by the MZI network, and $\boldsymbol{A}$ signifies another four-dimensional vector that is determined by the output signals collected by the photodetectors.

The photonic computing chip has been crafted on the silicon-on-insulator (SOI) platform, featuring a top Si layer of 220 nm and SiO_2 cladding of 2 μm. The grating coupler employed is of the focused type [31], with a shallow etch depth of 150 nm. The grating's period and duty cycle are designed at 650 nm and 0.5, respectively, optimized for a center wavelength of 1550 nm. The power splitter in Part (1) is comprised of three cascaded 1 × 2 multimode interference (MMI) couplers, with detailed structures depicted in Figure 2a. The first MMI stage divides the incident light into two equal parts, and, subsequently, the second MMI stage further divides this light into four equal components. The dimensions of the multimode waveguide are specified as L_1 = 7.9 μm in length and W_1 = 3.2 μm in width. To mitigate abrupt width transitions and minimize reflection losses at the junctions between single-mode and multimode waveguides, tapered structures with a length of L_t = 10 μm are introduced. We employ a light source with a wavelength of 1550 nm and a power of 10.6 dBm as the input for the beam splitter, resulting in optical powers of approximately 4.3 dBm, 4.2 dBm, 4.2 dBm, and 4.2 dBm at its four output terminals, respectively. Consequently, the beam splitter exhibits an excess loss of approximately 0.4 dB.

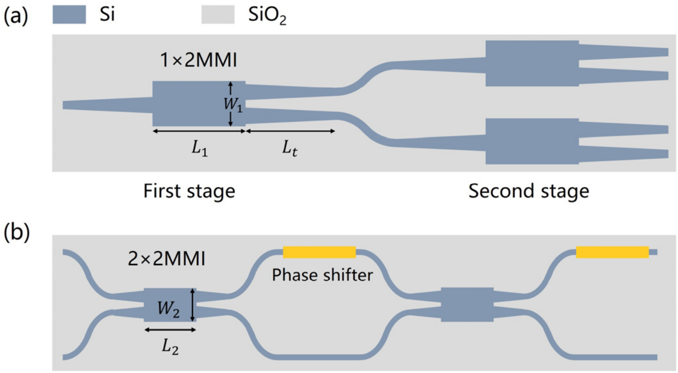

Figure 2. (**a**) Schematics of the 1-to-4 power splitter. There is one MMI in the first stage and two in the second. (**b**) Schematic of an MZI unit.

The MZI serves as the fundamental building block of the photonic computing chip, consisting of two 3-dB couplers and two phase shifters, as depicted in Figure 2b. The 3-dB coupler employs a 2 × 2 MMI configuration, featuring dimensions of L_2 = 31 µm in length and W_2 = 5.2 µm in width. The phase shifter, utilizing the thermal-optic effect, is achieved by depositing a 0.1 µm-thick layer of TiN film above the waveguide, serving as a thermal resistance. The TiN film is designed to exhibit a resistance of 480 Ω, with dimensions of 100 µm × 2.5 µm.

To drive the proposed photonic computing processor, a custom-designed control circuit has been developed, utilizing six 8-channel digital-to-analog converters (DACs, AD5592R, Analog Devices, Wilmington, MA, USA) and an FPGA (XC7Z020-2CLG484I, Xilinx, San Jose, CA, USA). Notably, the AD5592R converters possess dual functionality, serving as both DACs and analogue-to-digital converters (ADCs).

The light source of the system is a laser (SFL1550P, Thorlabs, Newton, NJ, USA) emitting at 1550 nm with an output power of 10.6 dBm. To enhance the coupling efficiency between the light source and the chip, a polarization controller (CPC900, Thorlabs, Newton, NJ, USA) is utilized. The MZI network is pre-configured to perform matrix multiplication with the control circuit by tuning the output voltage of the DACs. After transmission, the outputs of the chip are obtained by four photodetectors (DXM20AF, Thorlabs, Newton, NJ, USA), converted to four photocurrents, and then acquired by the ADCs.

3. Results

The microscope images depicting the photonic computing chip and its crucial component, the Mach-Zehnder interferometer, are presented in Figure 3. The chip, with dimensions of 6 mm in length and 1.5 mm in width, is fabricated using a mature CMOS process. The detailed fabrication procedure is as follows: First, the SOI wafer is cleaned and spin-coated with a photoresist. Using ultraviolet lithography, the waveguide pattern is formed on the photoresist. Nest, the waveguide structures are etched using Reactive Ion Etching (RIE), followed by the deposition of a silicon dioxide cladding via Plasma-Enhanced Chemical Vapor Deposition (PECVD). A TiN film is then formed through magnetron sputtering, and metal electrical contacts and interconnects are formed using electron beam evaporation. Another layer of silicon dioxide is deposited as a passivation layer, followed by the final steps of etching pad opening. In Figure 3b, TiN heaters are fabricated on both arms of the MZI switch to reduce the loss difference and enhance the extinction ratio of the MZI switch. Figure 3c offers a magnified perspective of the thermal phase shifter, where the two dark squares represent deep silicon-etched grooves positioned on both sides of the heater. Their purpose is to minimize thermal crosstalk among phase shifters.

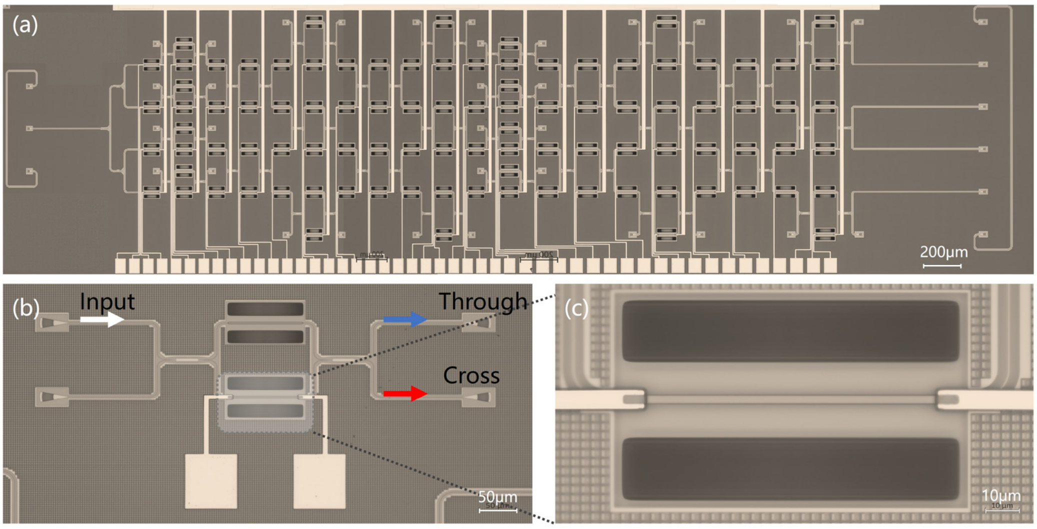

Figure 3. (**a**) Microscope images of the fabricated photonic computing chip using the CMOS process. (**b**) The fabricated reference MZI switch. When light is introduced through the upper port, the output port situated above is termed the "Through" port, while the port positioned below is designated as the "Cross" port. (**c**) Magnified perspective of the thermal phase shifter.

Prior to testing the entire device, the modulation efficiency and speed of the MZI unit are initially characterized. Figure 4a,b depicts the measured transmission spectrum of the MZI unit functioning as an optical switch. Regardless of whether the MZI switch is in the "ON" or "OFF" state, its excess loss, a metric representing the dB loss of the total optical power at all output ports compared to the input optical power, remains under 1 dB. Additionally, at the operating wavelength of 1550 nm, the extinction ratio of the MZI switch surpasses 30 dB, demonstrating excellent performance. Note that 18 mW electrical power is needed to change the MZI state between "ON" and "OFF". Figure 4c illustrates the optical response of the MZI unit when driven by a 10 kHz square wave electrical signal. The ascending phase of the optical response, which encompasses a transition from 10% to 90% of its normalized maximum, endures approximately 11µs. This indicates that the modulation speed of the MZI reaches approximately 90 kHz.

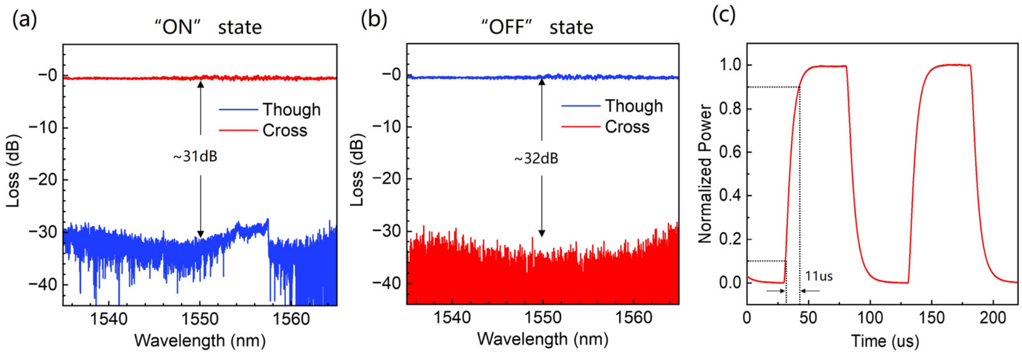

Figure 4. The measured transmission spectra of the MZI switch in (**a**) the "ON" state and (**b**) the "OFF" state. (**c**) The optical response of the MZI switch when driven by a 10 kHz square wave electrical signal.

As previously stated, the MZIs in Part (2) function as intensity modulators and require pre-calibration to establish a relationship model between optical output and electrical input. Figure 5 illustrates the normalized optical output power P_i (i = 1, 2, 3, 4) plotted against the electric power applied to the MZI_i (W_i) in Part (2). This relationship can be theoretically described by the equation provided below, with W_{min} and W_{max} representing the electric power corresponding to the minimal and maximal optical output, respectively:

$$P_i = \frac{1}{2}\left[1 - \cos\left(\frac{W_i - W_{min}}{W_{max} - W_{min}}\right)\right] \tag{1}$$

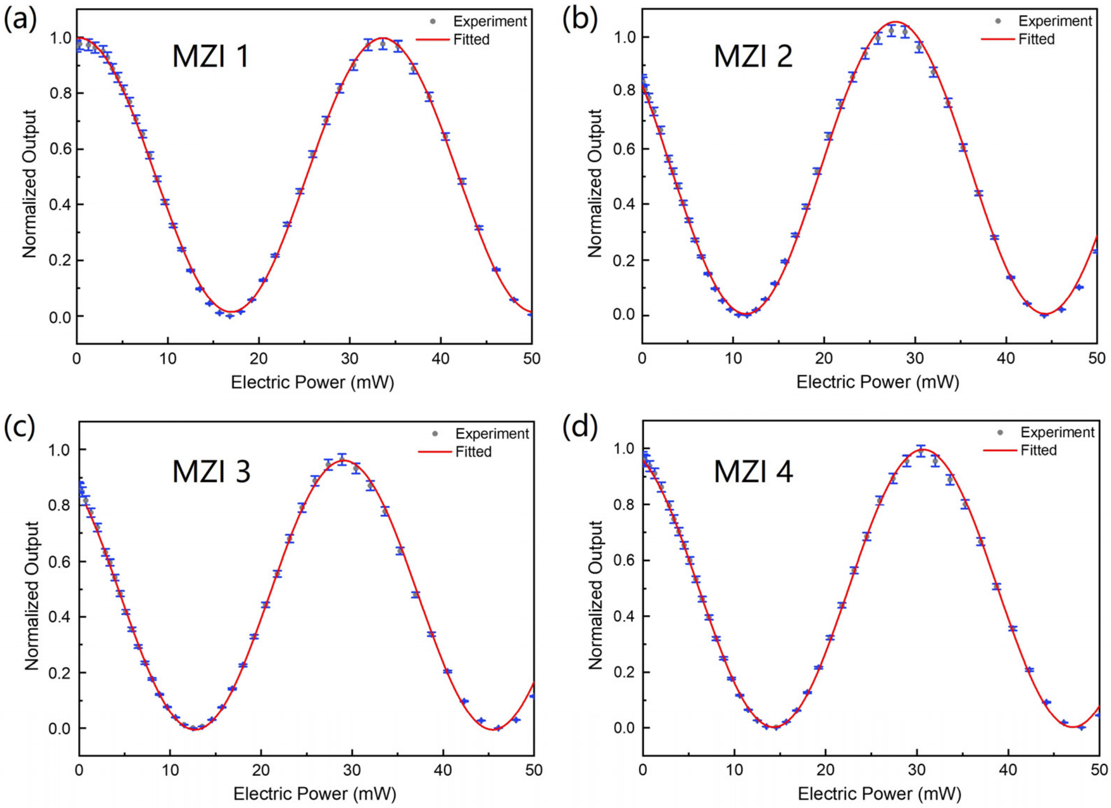

Figure 5. (**a**–**d**) The relationship between the normalized output of the MZI_i (i = 1, 2, 3, 4) in Part (2) and the electric power applied, respectively.

The red lines depicted in Figure 5 represent the fitting curves utilizing the sine function, with a correlation ratio exceeding 0.999, thereby indicating an excellent agreement between the theoretical predictions and experimental observations.

Next, the convolution kernels should be loaded onto Part (3). For this study, we have selected four different 2 × 2 kernels, designated as $\boldsymbol{K}_i$ (i = 1, 2, 3, 4), and integrated them into a 4 × 4 matrix $\boldsymbol{X}$, as shown below:

$$\boldsymbol{X} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ 1 & 0 & 0 & -1 \end{bmatrix} \tag{2}$$

Each row of the matrix X represents a convolution kernel. Specifically, the first kernel can blur the input image, whereas the second and third kernels are designed to extract vertical and horizontal edges, respectively. The fourth kennel can be regarded as a fusion of the second and third kennels, which can highlight the oblique outlines. According to the matrix decomposition principle demonstrated in reference [31], the theoretical retrieval of every phase delay within the phase shifter in Part (3) is feasible, provided that the objective matrix is given. Nevertheless, due to the unknown fabrication deviation, the MZI-based computation network typically remains an enigmatic network, resembling a black box that necessitates training. The training process can be denoted as finding solutions for equation of $\boldsymbol{XB} = \boldsymbol{A}$ when $\boldsymbol{A}$ and $\boldsymbol{B}$ are given. Here, $\boldsymbol{X}$ is the 4 × 4 dimensional matrix needed to be trained, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are 4 × n dimensional matrices. The equation is rewritten in the format of column vectors as:

$$\boldsymbol{X}\begin{bmatrix}\boldsymbol{B}_1 & \boldsymbol{B}_2 & \ldots & \boldsymbol{B_n}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots & \boldsymbol{A_n}\end{bmatrix} \tag{3}$$

During the training process, the phase shifters in Part (3) are tuned using a self-configuring algorithm to manipulate the transmission matrix ($\boldsymbol{X_{part2}}$) towards achieving $\boldsymbol{X_{part2}B} = \boldsymbol{A}$. $\boldsymbol{B_1}, \boldsymbol{B_2}, \boldsymbol{B_3}, \ldots, \boldsymbol{B_n}$, which are defined via a random vector generator and loaded by Part (2). The corresponding outputs are measured and recorded as $\boldsymbol{A_{expi}}$ (i = 1, 2, …, n). In comparison, objective results of $\boldsymbol{X_{object}B_i}$ (i = 1, 2, …, n), where $\boldsymbol{X_{object}}$ represents the objective matrix, are recorded as $\boldsymbol{A_i}$ (i = 1, 2, …, n). Obviously, when $\boldsymbol{A_{expi}} = \boldsymbol{A_i}$, the trained matrix $\boldsymbol{X_{part2}}$ will be equal to the objective matrix $\boldsymbol{X_{object}}$.

The detailed training process is explained in detail, step by step, as follows.

(a) To characterize the training effect, a cost function (*CF*) should be initially established. In this paper, the similarity between the provided matrix $\boldsymbol{A}$ and the experimentally derived matrix $\boldsymbol{A_{exp}}$ is defined and can be expressed by the equation below:

$$CF = \frac{\left|\boldsymbol{A}\cdot\boldsymbol{A_{exp}}\right|}{\|\boldsymbol{A}\|\left\|\boldsymbol{A_{exp}}\right\|} \tag{4}$$

The operation "·" in the numerator denotes the scalar product of two vectors, and "|| ||" in the denominator represents the Frobenius norm of a vector or matrix. Evidently, the *CF* ranges inclusively between 0 and 1, with *CF* = 0 or 1 indicating either irrelevance or consistency between the experimental and theoretical matrices.

(b) To initiate the process, randomly apply voltages to all the phase shifters in Part (2) and subsequently compute the initial *CF*.

(c) Tune the first phase shifter to change its phase delay from θ1 to θ1 + Δθ.

If *CF*(θ1 + Δθ) ≥ *CF*(θ1), replace θ1 with θ1 + Δθ, refresh *CF* with *CF*(θ1 + Δθ), and turn to step (d).
If *CF*(θ1 + Δθ) < *CF*(θ1), first replace θ1 with θ1 − Δθ and calculate *CF*(θ1 − Δθ), then compare the value of *CF*(θ1 − Δθ) and *CF*(θ1). If *CF*(θ1 − Δθ) ≥ *CF*(θ1), replace θ1 with θ1 − Δθ and refresh the *CF* as *CF*(θ1 − Δθ), else, remain the phase delay to θ1 and turn to step (d).

(d) Repeat step (c) for all phase shifters in Part (2) sequentially. This is called a round of iteration.

(e) Repeat step (c) and (d) until the *CF* is converged or reaches target value. Record voltage values loaded on all phase shifters.

During the training process, it is quite significant to choose a proper phase delay step Δθ. Too great a step makes the *CF* difficult to converge, while too small a step could be time-consuming and fall into local convergence. In this paper, first we choose a slightly larger Δθ to accelerate iteration speed, then gradually reduce Δθ until the *CF* is converged. Before the 100th round of iteration, Δθ is set as 0.08 V, and then is reduced by half every 50 rounds of iteration to 0.01 V. The *CF* is converged over 0.999 after 200 rounds of iteration, which indicates a strong correlation between $\boldsymbol{A}$ and $\boldsymbol{A_{exp}}$.

The digital image convolution function is demonstrated using the experimental setup shown in Figure 6. The continuous-wave laser at 1550 nm is sent into the photonic computer chip after being manipulated by a polarization controller. The original digital image is transmitted from the upper computer to the control circuit and loaded, pixel by pixel, in a certain order on Part (2) of the photonic computing chip through DACs. To simplify the experiment, an 8-bit grayscale image with 320 × 256 pixel resolution is selected, with every pixel value between 0 and 255, where 0 means black and 255 means white. Part (3) of the chip is pre-trained to perform four kennels and remain stable during the convolution process. The four outputs of the chip are obtained by photodetectors and sent to the control circuits for analog to digital conversion. The four converted digital signals are then reconstructed to four convolution images and displayed on the upper computer. In order to prove the convolution effects, we introduce a control group, which is the convolution result from the 64-bit upper computer. Figure 7 demonstrates the original image and convolution images from two different kinds of computing system. It is easy to see that both kinds of computing system can perform image convolution and extract outlines correctly. However, the difference between two results is hard to distinguish by human eye. To further quantitatively characterize the convolution process, the relative error RE is defined using the equation below:

$$RE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(\frac{b_{expi} - b_{comi}}{b_{comi}}\right)^2} \tag{5}$$

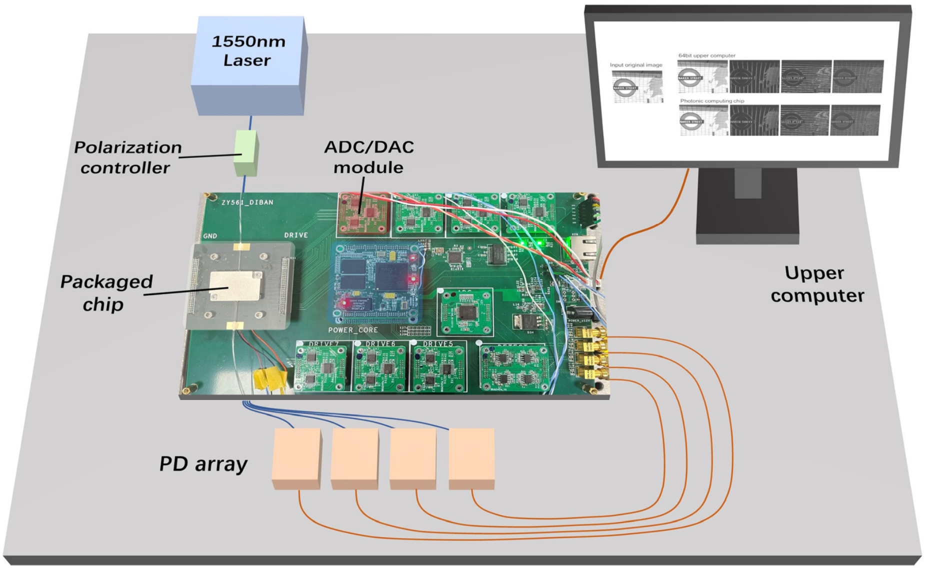

Figure 6. Experimental setup of the digital image convolution function verification platform. PD: photodetector; ADC/DAC: analog to digital converter/digital to analog converter.

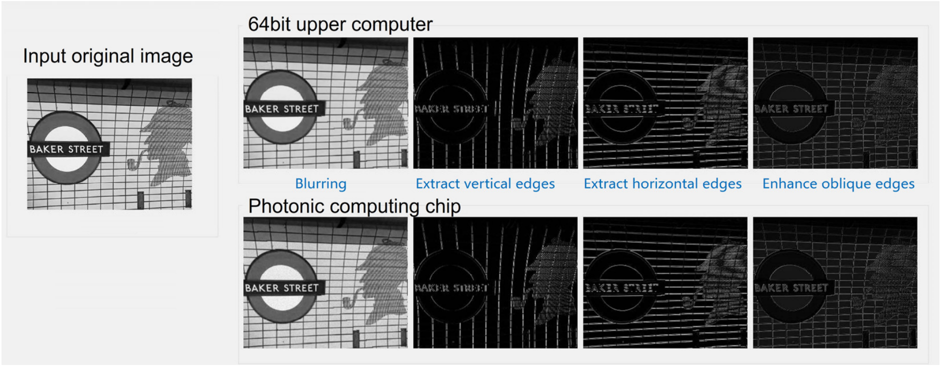

Figure 7. The image convolution results of the 64-bit upper computer and photonic computing chip. The operations in the figure, from left to right, are blurring, extracting vertical edges, extracting horizontal edges, and enhancing oblique edges.

Here, N is the pixel number of the output image. b_{expi} and b_{comi} are the ith pixel value of the output image from photonic computing chip and 64-bit upper computer, respectively. The calculated RE is less than 2.3%, indicating a good validity of the proposed photonic computing chip.

Due to the relatively low computational complexity in the experiment, the computing time for both cases is less than one second, making it difficult to measure. Instead of focusing on computation time, it is more insightful to examine the computation capabilities of both systems. The commercial computer, equipped with a single Intel CPU (i5 12400), offers a computing capability of 240 GFLOPS (FLOPS, floating point operations per second), as outlined in Intel's official documentation (APP Metrics for Intel® Microprocessors—Intel® Core™ Processor). As for the proposed photonic processor, its computing capability can be expressed by 2 × 4 × 4 × BW, where BW represents the lower value between the modulator's and detector's bandwidth [23]. In order to reduce fabrication costs, four thermal-optic modulators with constrained modulation bandwidth were employed, which subsequently limits the computation capability (~2.88MFLOPS). If the latest photonic I/O technology were adopted, the modulation/detection bandwidth could potentially soar to 100 GHz [33,34], thereby elevating the computing capability to 3.2TFLOPS. Furthermore, with the expansion of the photonic integration scale, the computation capability of the photonic processor could witness substantial enhancements.

4. Discussion

The deviations in the demonstration were mainly caused by calibration errors of Part (2) while configuring the MZIs to pure intensity modulator since the redundant phase modulation would impact the matrix building in Part (3). In the digital image convolution demonstration, the outputs of the photonic chip were actually squared because the photodetector array can only acquire light intensity, which equals the square of the light field. Although we had extracted the square root in the final results presented in Figure 7, the sign signal was missing. A feasible method to retrieve sign signal of convolution results is the adoption of coherent detection with balanced PDs, which was reported in reference [27].

Once the computation mode of the chip is configured, the calculation process is executed through passive optical transmission. Notably, each thermo-optic phase shifter requires an average power of only ~9 mW to stabilize its state. Given the chip's small scale, which encompasses just 40 phase shifters, the total power consumption is 360 mW. This is significantly lower than the power consumption of other off-chip devices and circuits, which typically range in the tens of watts, primarily due to the light source's power requirements. In this study, we focus solely on the chip's power consumption. When integrated with photonic I/O technology boasting a bandwidth of 100 GHz, the chip attains a computing power of 3.2TFLOPS. Consequently, the energy efficiency ratio is calculated as 3.2TFLOPS/360 mW = 8.9TFLOPS/W. For comparison, NVIDIA's Tesla T4 GPU has an energy efficiency ratio of 0.87TFLOPS/W, which is an order of magnitude lower than that of the photonic computing approach presented in this paper. Furthermore, the utilization of non-volatile phase-change materials allows the phase to be stabilized without consuming energy, further enhancing the chip's energy efficiency.

The proposed solution, while facing process cost constraints, undeniably presents certain limitations, including a relatively slow data loading speed for the thermo-optic modulator, a restricted chip size, and the reliance on off-chip lasers and detectors. However, these challenges can be tackled by incorporating ultra-high-speed electro-optic modulators [33], on-chip silicon-germanium detectors [35], heterogeneously integrated lasers [36], and employing low-loss waveguides to augment the chip's dimensions.

5. Conclusions

In conclusion, we have proposed a CMOS-compatible photonic computing chip for accelerating MAC operations and using it to perform digital image convolution. The working principle of the device is first elucidated, and then a proof-of-concept device is fabricated on an SOI wafer. Afterwards, the chip is packaged and applied in a convolution demonstration platform, along with the commercial laser source, photodetector array, and home-build drive circuits. A self-configuration algorithm is introduced to train the fabricated chip to perform convolution kernel. Experimental results show a good convolution effect by comparing it with a conventional 64-bit computer. The proposed CMOS-compatible photonic computing chip is scalable and can be integrated with other silicon-based devices, showing enormous potential for large-scale photonic computing.

Our proposed solution directly simulates the computational process using the passive propagation of light beams, harnessing the vast bandwidth of optics to attain high computational frequencies. Additionally, it efficiently executes large-scale matrix operations, capitalizing on the parallel nature of light propagation. If the latest photonic I/O technology can be incorporated, it could achieve a computing capability of 3.2TFLOPS with an energy efficiency of 8.9TFLOPS/W. As the chip size scales up, it is anticipated to reach computational power in the hundreds of TFLOPS range, potentially even higher. This would enable it to rival the computation capabilities of advanced GPUs while surpassing them in energy efficiency by a factor of one to two orders of magnitude. Furthermore, the fabrication of photonic chips does not necessitate the most cutting-edge lithography technology, such as extreme ultraviolet (EUV) lithography. This reduction in technical complexity and associated costs paves the way for future large-scale industrialization, making it more feasible and accessible.

Author Contributions: Conceptualization, C.W. and W.W.; methodology, C.W.; software, C.W. and Z.Z.; validation, C.W., W.W. and Z.W.; formal analysis, C.W. and Z.W.; investigation, C.W. and L.D.; resources, W.X.; data curation, C.W. and Z.Z.; writing—original draft preparation, C.W.; writing—review and editing, C.W., L.D. and W.X.; visualization, C.W. and W.W.; supervision, L.D. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China (Nos. 52175405, 52105568). China Postdoctoral Science Foundation (2024M750993), Postdoctoral Fellowship Program (Grade B) of China Postdoctoral Science Foundation (No. GZB20240241).

Data Availability Statement: Data are contained within the article.

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

References

1. Dilsizian, S.E.; Siegel, E.L. Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. *Curr. Cardiol. Rep.* **2013**, *16*, 441. [CrossRef]
2. Park, S.H.; Han, K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology* **2018**, *286*, 800–809. [CrossRef]
3. Arena, P.; Basile, A.; Bucolo, M.; Fortuna, L. Image processing for medical diagnosis using CNN. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* **2003**, *497*, 174–178. [CrossRef]
4. Doi, K. Diagnostic imaging over the last 50 years: Research and development in medical imaging science and technology. *Phys. Med. Biol.* **2006**, *51*, R5–R27. [CrossRef] [PubMed]
5. Meiring, G.A.M.; Myburgh, H.C. A Review of Intelligent Driving Style Analysis Systems and Related Artificial Intelligence Algorithms. *Sensors* **2015**, *15*, 30653–30682. [CrossRef]
6. Feng, S.; Yan, X.; Sun, H.; Feng, Y.; Liu, H.X. Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment. *Nat. Commun.* **2021**, *12*, 748. [CrossRef]
7. Fujiyoshi, H.; Hirakawa, T.; Yamashita, T. Deep learning-based image recognition for autonomous driving. *IATSS Res.* **2019**, *43*, 244–252. [CrossRef]

8. Gruyer, D.; Magnier, V.; Hamdi, K.; Claussmann, L.; Orfila, O.; Rakotonirainy, A. Perception, information processing and modeling: Critical stages for autonomous driving applications. *Annu. Rev. Control* **2017**, *44*, 323–341. [CrossRef]
9. Tatem, A.J.; Lewis, H.G.; Atkinson, P.M.; Nixon, M.S. Super-resolution target identification from remotely sensed images using a Hopfield neural network. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 781–796. [CrossRef]
10. Fetz, V.; Prochnow, H.; Brönstrup, M.; Sasse, F. Target identification by image analysis. *Nat. Prod. Rep.* **2016**, *33*, 655–667. [CrossRef] [PubMed]
11. Kim, K.-T.; Seo, D.-K.; Kim, H.-T. Radar target identification using one-dimensional scattering centres. In *IEE Proceedings-Radar, Sonar and Navigation*; The Institution of Engineering and Technology (IET): London, UK, 2001; Volume 148, pp. 285–296.
12. Al-Saffar, A.A.M.; Tao, H.; Talab, M.A. Review of deep convolution neural network in image classification. In Proceedings of the 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Jakarta, Indonesia, 23–24 October 2017; pp. 26–31.
13. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
14. Liu, Y.H. Feature Extraction and Image Recognition with Convolutional Neural Networks. *J. Phys. Conf. Ser.* **2018**, *1087*, 062032. [CrossRef]
15. Yang, A.; Yang, X.; Wu, W.; Liu, H.; Zhuansun, Y. Research on Feature Extraction of Tumor Image Based on Convolutional Neural Network. *IEEE Access* **2019**, *7*, 24204–24213. [CrossRef]
16. De Vos, A. Reversible computing. *Prog. Quantum Electron.* **1999**, *23*, 1–49. [CrossRef]
17. Marković, D.; Mizrahi, A.; Querlioz, D.; Grollier, J. Physics for neuromorphic computing. *Nat. Rev. Phys.* **2020**, *2*, 499–510. [CrossRef]
18. Khan, H.N.; Hounshell, D.A.; Fuchs, E.R.H. Science and research policy at the end of Moore's law. *Nat. Electron.* **2018**, *1*, 14–21. [CrossRef]
19. Goodman, J.W.; Dias, A.R.; Woody, L.M. Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms. *Opt. Lett.* **1978**, *2*, 1–3. [CrossRef]
20. Bowers, J.; Komljenovic, T.; Davenport, M.; Hulme, J.; Liu, A.; Santis, C.; Spott, A.; Srinivasan, S.; Stanton, E.; Zhang, C. *Recent Advances in Silicon Photonic Integrated Circuits*; SPIE: Bellingham, WA, USA, 2016; Volume 9774.
21. Debrégeas-Sillard, H.; Kazmierski, C. Challenges and advances of photonic integrated circuits. *C. R. Phys.* **2008**, *9*, 1055–1066. [CrossRef]
22. Ang, K.W.; Liow, T.Y.; Fang, Q.; Yu, M.B.; Ren, F.F.; Zhu, S.Y.; Zhang, J.; Ng, J.W.; Song, J.F.; Xiong, Y.Z.; et al. Silicon photonics technologies for monolithic electronic-photonic integrated circuit (EPIC) applications: Current progress and future outlook. In Proceedings of the 2009 IEEE International Electron Devices Meeting (IEDM), Baltimore, MD, USA, 7–9 December 2009; pp. 1–4.
23. Shen, Y.; Harris, N.C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.; Larochelle, H.; Englund, D.; et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **2017**, *11*, 441–446. [CrossRef]
24. Zhang, H.; Gu, M.; Jiang, X.D.; Thompson, J.; Cai, H.; Paesani, S.; Santagati, R.; Laing, A.; Zhang, Y.; Yung, M.H.; et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **2021**, *12*, 457. [CrossRef]
25. 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*, 7700412. [CrossRef]
26. Zhang, T.; Wang, J.; Dan, Y.; Lanqiu, Y.; Dai, J.; Han, X.; Sun, X.; Xu, K. Efficient training and design of photonic neural network through neuroevolution. *Opt. Express* **2019**, *27*, 37150–37163. [CrossRef] [PubMed]
27. Tian, Y.; Zhao, Y.; Liu, S.; Li, Q.; Wang, W.; Feng, J.; Guo, J. Scalable and compact photonic neural chip with low learning-capability-loss. *Nanophotonics* **2021**, *11*, 329–344. [CrossRef]
28. Tait, A.N.; de Lima, T.F.; Zhou, E.; Wu, A.X.; Nahmias, M.A.; Shastri, B.J.; Prucnal, P.R. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **2017**, *7*, 7430. [CrossRef] [PubMed]
29. Feldmann, J.; Youngblood, N.; Wright, C.D.; Bhaskaran, H.; Pernice, W.H.P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **2019**, *569*, 208–214. [CrossRef]
30. 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]
31. 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]
32. Reck, M.; Zeilinger, A.; Bernstein, H.J.; Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **1994**, *73*, 58–61. [CrossRef]
33. Shen, J.; Zhang, Y.; Zhang, L.; Li, J.; Feng, C.; Jiang, Y.; Wang, H.; Li, X.; He, Y.; Ji, X.; et al. Highly Efficient Slow-Light Mach–Zehnder Modulator Achieving 0.21V·cm Efficiency with Bandwidth Surpassing 110 GHz. *Laser Photonics Rev.* **2024**, *19*, 2401092. [CrossRef]

34. Ding, Y.; Cheng, Z.; Zhu, X.; Yvind, K.; Dong, J.; Galili, M.; Hu, H.; Mortensen, N.A.; Xiao, S.; Oxenløwe, L.K. Ultra-compact integrated graphene plasmonic photodetector with bandwidth above 110 GHz. *Nanophotonics* **2020**, *9*, 317–325. [CrossRef]
35. Benedikovič, D. Advances in chip-integrated silicon-germanium photodetectors. In *Photodetectors*, 2nd ed.; Nabet, B., Ed.; Woodhead Publishing: Cambridge, UK, 2023; pp. 233–266. [CrossRef]
36. Guo, X.; He, A.; Su, Y. Recent advances of heterogeneously integrated III–V laser on Si. *J. Semicond.* **2019**, *40*, 101304. [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.