Article

An Efficient Frequency Encoding Scheme for Optical Convolution Accelerator

Gongyu Xia [1] , Jiacheng Liu [1], Qilin Hong [1], Pingyu Zhu [2], Ping Xu [2,* and Zhihong Zhu [1,*

[1] College of Advanced Interdisciplinary Studies & Hunan Provincial Key Laboratory of Novel Nano Optoelectronic Information Materials and Devices, National University of Defense Technology, Changsha 410073, China; zxc274261@163.com (G.X.); 18464268894@163.com (J.L.); qlhong_95@163.com (Q.H.)

[2] Institute for Quantum Information and State Key Laboratory of High-Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China; pingyuzhu@yeah.net

[*] Correspondence: pingxu520@nju.edu.cn (P.X.); zzhwcx@163.com (Z.Z.)

Abstract: In today's era where the demand for computational resources by large models is increasingly high, optical computing offers an alternative physical platform for computation. With its high parallelism and the maturation of integrated photonic technologies, optical computing is expected to further support the computational resources required by large models. For one-dimensional optical convolution accelerators, existing methods can fully utilize the working bandwidth of electro-optic modulators and the frequency resources of light sources. However, most convolution computations currently require the execution of two-dimensional matrix convolutions, and existing encoding schemes suffer a drop in terms of effective computations per second when performing two-dimensional matrix convolutions. In response to this, we propose a new encoding scheme that can fully utilize the computational resources of optical convolution accelerators. For convolution operations with a kernel size of $M \times M$, compared to existing encoding schemes, it can achieve an M-fold increase in effective computation rate. This implies that under the constraints of essentially the same hardware physical performance, adopting our newly proposed encoding scheme can achieve a significant improvement in computational performance. We construct an optical convolution acceleration system to demonstrate the principles of the new encoding.

Keywords: convolution neural network; optical computing; matrix convolution

1. Introduction

The concept of the artificial neural network (ANN) originates from the study of the mechanisms underlying the biological neural processing of signals. With the increasing power of electronic integrated circuits, artificial neural networks and large language models have demonstrated remarkable capabilities across an expanding array of fields, including image recognition, natural language processing, and autonomous driving. However, the evolution of traditional electronic computing chips has encountered two significant limitations: the power consumption wall and the computation rate wall. These constraints have impeded further enhancements in computational performance, urgently necessitating the development of new physical computing platforms to surmount these obstacles.

In response to these challenges, a variety of processing chips aimed at intelligent computing have emerged, such as IBM TrueNorth [1] and Google TPU [2]. These chips have been designed to accelerate specific types of computations, particularly those involved

in deep learning and neural network processing. However, they still operate within the constraints of electronic computing, which limits their scalability and energy efficiency. For instance, electronic chips face challenges in handling the massive quantity of data and complex computations required for training and inference in large-scale neural networks, leading to increased power consumption and longer processing times.

Optical computing, leveraging the high-speed transmission and high parallelism characteristics of light, stands out as a promising candidate for a high-performance computation platform. The inherent advantages of optical computing include its ability to perform computations in parallel, its low power consumption, and its potential for high-speed data processing. These features make it particularly suitable for handling the computationally intensive tasks required by large-scale artificial neural networks and other advanced AI applications. For instance, optical computing can process multiple data streams simultaneously, reducing the need for sequential processing and enabling faster computation of complex operations such as matrix multiplications and convolutions.

Currently, research is being conducted into several areas of optical computing, including coherent optical computing [3], reservoir computing [4–6], and diffractive network computing [7–9]. Coherent optical computing utilizes the phase and amplitude of light waves to perform computations, offering high precision and efficiency. Reservoir computing, on the other hand, uses the natural dynamics of optical systems to process information, providing a novel approach to machine learning and signal processing. Diffractive network computing employs diffractive optical elements to create complex neural networks, enabling efficient and compact optical processing systems.

Notable advancements have been made in multi-wavelength computation, as evidenced by the significant progress achieved in this domain [10–13]. Multi-wavelength computation allows for the simultaneous processing of multiple wavelengths of light, increasing the computational capacity and efficiency of optical systems. This is particularly useful in applications such as optical communication and data processing, where large quantities of data need to be transmitted and processed quickly.

Among these advancements, the employment of high-speed electro-optical modulators facilitates the multiplexing of frequency- and time-domain signals, thereby enabling optically accelerated matrix convolution operations. Electro-optical modulators are crucial components in optical computing systems, as they convert electrical signals into optical signals, allowing for the manipulation and processing of light. The latest work utilizes a Bragg grating array to optimize electro-optic modulators for the realization of convolution accelerators [14]. Bragg grating arrays can precisely control the phase and amplitude of light, enhancing the performance of electro-optical modulators and improving the overall efficiency of optical convolution accelerators.

For one-dimensional convolution operations, existing computational acceleration schemes [13] can fully utilize the frequency resources of optical signals and the bandwidth of electro-optic modulators for the optical acceleration of computations. However, considering practical applications, most convolution operations in artificial intelligence computations are two-dimensional matrix convolutions. Existing encoding schemes may compromise actual computational efficiency when dealing with two-dimensional matrix convolutions. When the size of a single convolution kernel is $M \times M$, the number of effective multiplication and addition operations per second is only $\frac{1}{M}$ of that for one-dimensional vector convolutions. This limitation significantly impacts the performance of optical convolution accelerators in handling complex AI tasks that involve large-scale two-dimensional data, such as image and video processing.

To address these limitations, we propose a novel matrix–vector mapping encoding scheme. This scheme can fully utilize the frequency resources of optical signals and the

bandwidth of electro-optic modulators, achieving the same number of effective multiplication and addition operations per second as one-dimensional vector convolutions. Compared to existing two-dimensional matrix encoding schemes, it offers an *M*-fold increase in computational operations. This enhancement not only improves the efficiency of optical convolution accelerators but also paves the way for more complex and larger-scale neural network applications to be efficiently processed using optical computing platforms. By overcoming the challenges associated with two-dimensional matrix convolutions, our proposed scheme enables optical computing to better meet the demands of modern AI applications, offering a significant step forward in the development of high-performance computing systems.

Moreover, the proposed encoding scheme has the potential to significantly impact various domains that rely heavily on convolutional neural networks (CNNs). In the field of medical imaging, for example, CNNs are extensively used for tasks such as disease diagnosis, anomaly detection, and image segmentation. By accelerating the convolution operations in CNNs, our encoding scheme can enable faster and more accurate analysis of medical images, leading to improved diagnostic outcomes and more efficient healthcare workflows. In the realm of autonomous vehicles, CNNs play a crucial role in perception tasks, such as object detection, scene understanding, and navigation. The enhanced computational efficiency provided by our scheme can contribute to real-time processing of sensor data, enabling autonomous vehicles to make quicker and more reliable decisions, thereby enhancing their overall performance and safety.

Furthermore, the scalability of our encoding scheme makes it suitable for integration with emerging technologies such as quantum computing and neuromorphic computing. Quantum computing holds the promise of solving complex problems that are intractable for classical computers, while neuromorphic computing aims to emulate the human brain's neural structure and functionality. By combining the strengths of optical computing with these advanced technologies, we can potentially unlock new levels of computational power and efficiency, paving the way for breakthroughs in fields such as cryptography, optimization, and artificial intelligence research.

The remainder of this paper is organized as follows. In Section 2, we present the details of the more efficient frequency encoding scheme for optical convolution accelerators. Section 3 describes the experimental setup and results, demonstrating the effectiveness of our proposed scheme. Finally, Section 4 discusses the implications of our findings and suggests directions for future research.

2. More Efficient Frequency Encoding Scheme for Optical Convolution Accelerator

The optical vector convolution accelerator leverages the high bandwidth characteristics of high-speed electro-optic modulators to perform simultaneous multiplication and addition operations on optical signals of multiple frequencies [15–17]. The input vector $X \in \mathbb{R}^{1 \times n}$ is encoded in the form of a time series within a one-dimensional voltage signal. When the electro-optic intensity modulator is biased at $\pi/4$, its response to the input optical signal can be approximated as the product of the radio-frequency electrical signal and the input optical power. The one-dimensional convolution kernel $W \in \mathbb{R}^{1 \times m}$, which is the reversed, W^r, where $W_i^r = W_{m-i+1}$, is encoded into the frequency domain of the optical signal. At timestamp i, the vector W^r composed of optical signals of different frequencies will undergo the same multiplicative factor X_i. The total power of the current optical signal

is given by the sum $\sum_{j=1}^{m} X_i W_j^r$. The dispersion of the optical fiber is utilized to interleave the frequencies of the optical signal. The following results are obtained.

$$\sum_{j=1}^{m} X_i W_j^r \xrightarrow[\text{Interleaving}]{\text{Frequency}} \sum_{j=1}^{m} X_{i+j-1} W_j^r = \sum_{j=1}^{m} X_{i+j-1} W_{m-j+1} = Y_i \tag{1}$$

Considering the sampling rate of the electro-optic modulator is Ω, which means the time interval between adjacent elements of vector X is $\frac{1}{\Omega}$, and the number of available frequency points of the optical signal is M, then, within a unit time, the vector convolution accelerator can perform $2 \times \Omega \times M$ multiplication and addition operations. With a sampling rate of 172 GHz [18] and input of an optical signal with 1400 frequency components [19], the optical vector convolution accelerator can perform multiplication and addition operations at a rate of $2 \times 172 \times 1400 = 481$ T per second.

Convolution neural networks predominantly require the execution of convolution operations on two-dimensional matrix inputs. For one-dimensional optical vector convolution accelerators, performing two-dimensional matrix convolution necessitates an encoding scheme that transforms the convolution operations between two-dimensional matrices into computations between one-dimensional vectors.

Taking a 3×3 convolution kernel as an example, let the input matrix be $X \in \mathbb{R}^{N \times N}$, which is mapped to $X' \in \mathbb{R}^{1 \times N^2}$. The mapping is denoted as $f : \mathbb{R}^{N \times N} \to \mathbb{R}^{1 \times N^2}$. An existing mapping scheme, proposed by Xu et al. [13] in 2021, is denoted as $f_0 : \mathbb{R}^{N \times N} \to \mathbb{R}^{1 \times N^2}$.

$$f_0(\mathbf{X}) = \mathbf{X}', \ X_k' = X_{i,j}, \ k = ((i-1) \bmod 3) + 1 + (j-1) * 3 + \lfloor (i-1)/3 \rfloor * 3N \tag{2}$$

As shown in Figure 1a, the first 3 rows of the input two-dimensional matrix are taken and concatenated into a one-dimensional vector by columns, and then this operation is repeated by taking 3 rows at a time to complete the mapping from a two-dimensional matrix to a one-dimensional vector. After completing the dimensionality transformation, the one-dimensional vector is placed into the optical vector convolution accelerator. As shown in Figure 1b, after frequency interleaving, every three timestamps, one is the value required for the actual convolution computation. This means that compared to one-dimensional vector convolution operations, the effective number of computations for two-dimensional vector convolution after the encoding scheme f_0 is only $1/3$ of that of one-dimensional vectors.

We propose another encoding scheme $f_1 : \mathbb{R}^{N \times N} \to \mathbb{R}^{1 \times N^2}$.

$$f_1(\mathbf{X}) = \mathbf{X}', \ X_k' = X_{i,j}, \ k = i + (j-1)N \tag{3}$$

The matrix at the bottom of Figure 1a concatenates the entire input two-dimensional matrix into a one-dimensional vector by rows. In the computation, as shown in Figure 1c, we take the power sum of every three values in the frequency domain $W_{i,1} \sim W_{i,3}$, and then sum $\sum_j Y_{k,j}$. It is not difficult to find that compared to the original one-dimensional vector convolution operation, after adopting the encoding scheme f_1, the number of effective computations is essentially the same as the original, and compared to f_0, for a 3×3 convolution kernel, there is a three-fold increase. As showned in Table 1, given a sampling rate of 172 GHz and an input optical signal with 1400 frequency components, the actual number of effective multiply–accumulate operations per second required for the two-dimensional matrix convolution of a 9×9 convolution kernel is increased from $2 \times 172 \times 1400/9 = 53.5$ T to $2 \times 172 \times 1400 = 481.6$ T.

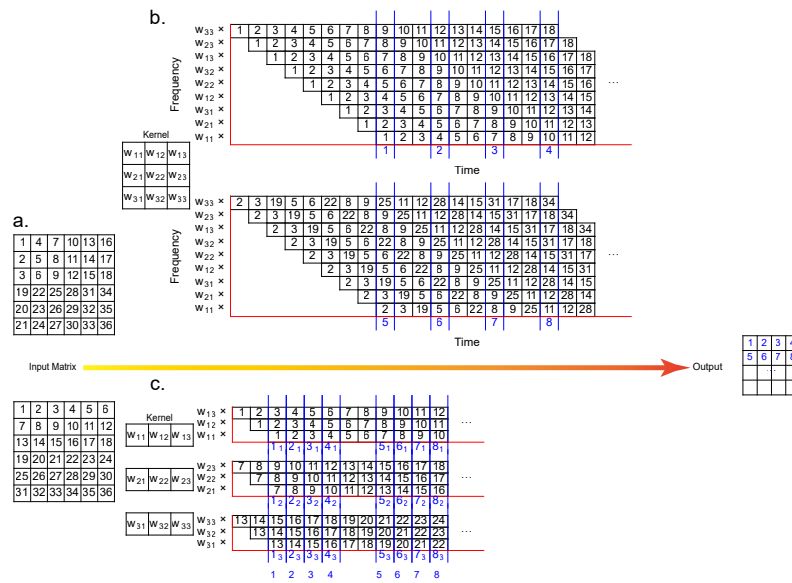**Figure 1.** (**a**) Image signals converted to one-dimensional signal coding scheme: original scheme (top), new scheme (bottom). (**b**) Calculation process of optical convolution for existing scheme [13]. (**c**) New optical convolution scheme proposed in this work.

Table 1. The number of multiply–add operations performed per second of original scheme and our scheme for different kernel sizes.

Kernel Size	3×3	5×5	7×7	9×9
Original scheme	160.53 T	96.32 T	68.8 T	53.51 T
Our scheme	481.6 T	481.6 T	481.6 T	481.6 T

We constructed an optical convolutional neural network computing system to demonstrate the actual working principle of the new encoding scheme. Figure 2a shows the structure of the convolutional network, which consists of five layers. From front to back, they are the input layer of size 28×28, a convolutional layer of size $8 \times 26 \times 26$ with 8 convolution kernels each of size 3×3, a pooling layer of size $8 \times 13 \times 13$, followed by a flattening operation, and finally a fully connected layer with 10 nodes and the activation function SoftMax for recognizing the handwritten digit dataset of the Mixed National Institute of Standards and Technology database (MNIST), which comprises a total of 50,000 training data entries and 10,000 testing data entries. Out of the 50,000 training data entries, 10,000 were separated as evaluation data during the training process. The overall network was trained by a computer, with the convolutional calculations performed by the optical system shown in Figure 2e, and the subsequent fully connected layers and activation functions completed by an electronic computer.

Eight convolution kernels of size 3×3 were transformed into one-dimensional vectors and encoded onto the frequency domain of the input optical signal. In Figure 2b, for a single convolution kernel, row-by-row concatenation was performed, and the encoding operation was completed by a spectral shaper, ensuring that each element corresponded to the intensity of an optical signal at a specific frequency, with adjacent frequency intervals of $\Delta\omega$.

For the input matrix, we took the handwritten digit "2" as an example. In Figure 2c, according to the encoding scheme f_1 (3), the two-dimensional input matrix was concatenated row-by-row to obtain a one-dimensional vector, which was then encoded into the radio-frequency voltage input of the electro-optic intensity modulator, with adjacent element

time intervals of $1/\Omega$, where Ω is the sampling rate of the arbitrary waveform generator. The bias of the electro-optic intensity modulator was set to $\pi/4$.

$$T(U) = \sin^2\left(\frac{\pi}{2}\frac{U}{U_\pi} + \phi\right) \approx \frac{1}{2} + \frac{\pi}{2}\frac{U}{U_\pi} \quad \left(\phi = \frac{\pi}{4}, U \ll U_\pi\right) \tag{4}$$

The transmittance T of the electro-optic intensity modulator as a function of the external field voltage U, when the bias ϕ was set to $\frac{\pi}{4}$, could be approximated as a linear response, and subtracting the DC term completed the multiplication of the input optical intensity X_i and the modulator radio-frequency voltage W_j.
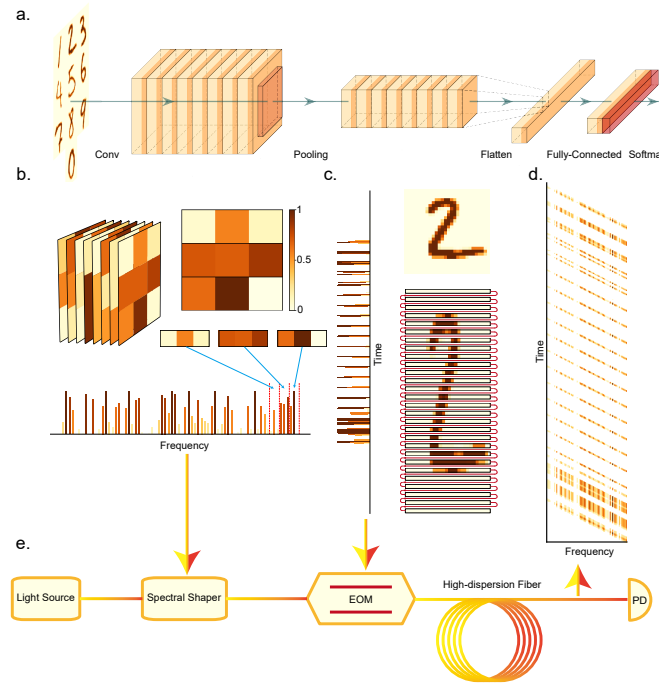

Figure 2. (**a**) Convolutional network model framework. (**b**) Convolutional kernel decomposition process. (**c**) Image decomposition process. (**d**) Calculation results after dispersive fiber. (**e**) Experimental framework diagram.

To achieve frequency interleaving, the optical convolution system also required a certain dispersion module to create a time-domain displacement of optical signals at different frequencies. When the sampling rate of the arbitrary waveform generator is Ω, and the frequency interval between adjacent elements of the convolution kernel is $\Delta\omega$, in a system with a dispersion coefficient of D (ps/(km · nm)), such as an optical fiber, the required total length is $L = \frac{1}{\Omega}\frac{\omega}{D\lambda\Delta\omega}$. After passing through the dispersion system, the result shown in Figure 2d was obtained, and the required computational result was obtained using the calculation method of Figure 1c and a power meter. For the multi-frequency computation process, taking a 3×3 convolution kernel as an example, after the optical signal passed through the dispersion fiber, a wavelength division multiplexer (WDM) was used to group every three adjacent frequencies together, and the output was directed to an optical power meter to measure the optical power sum corresponding to the current step.

For the experimental system we constructed to demonstrate the new encoding scheme of the optical convolutional neural network, we could calculate the actual number of effective operations per second and compare it with the existing encoding scheme. As shown in Figure 3a, we built a light source with 9 frequency teeth, and the electro-optic modulator operated at a sampling rate of 1.429 GHZ. As shown in Figure 3d, the current system response had a standard deviation of 0.03987, corresponding to a signal-to-noise ratio of 27.98 dB. We conducted simulation calculations using this numerical precision and

obtained the performance of the current demonstration system for the MNIST handwritten digit recognition task, as depicted in Figure 3f. For one-dimensional vector convolution, the actual number of computations per second was $2 \times 9 \times 1.429 = 25.7$ G multiplication and addition operations. After using the encoding scheme f_1, the actual number of computations per second for two-dimensional convolution operations was the same as for one-dimensional vector convolution. For the encoding scheme f_0, the actual number of computations per second was reduced to $2 \times 9 \times 1.429/3 = 8.57$ G multiplication and addition operations. It can be seen that when the convolution kernel was a 3×3 matrix, using the new encoding scheme f_1 for computing matrix convolution had three times the improvement compared to the existing scheme. For larger convolution kernels, such as 5×5, 7×7, $N \times N$, the improvement would increase with the increase in N, and it w3would have N times the improvement compared to the existing scheme.
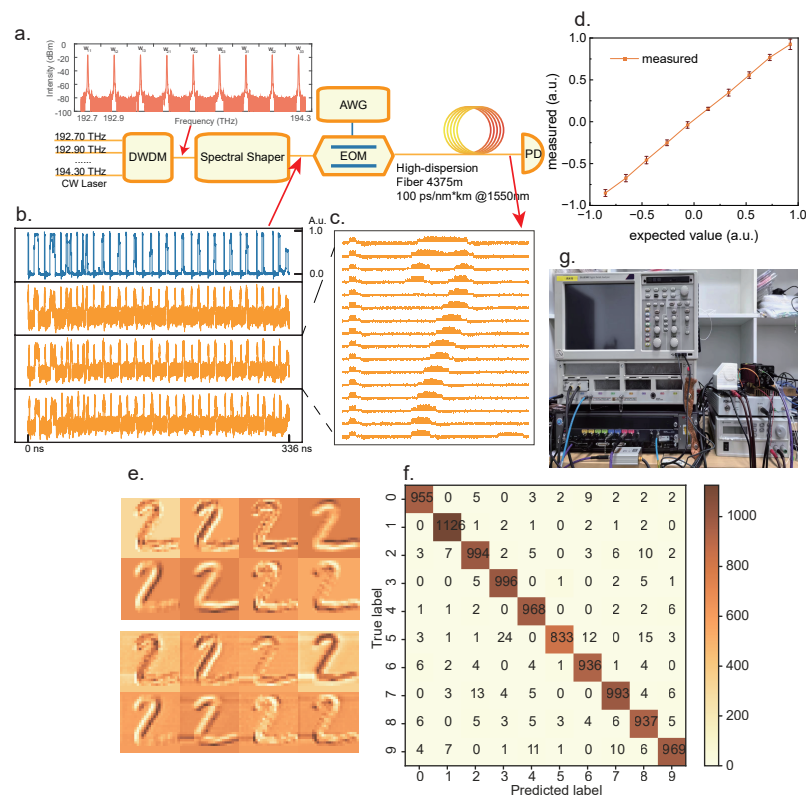

Figure 3. (**a**) System experimental diagram, Dense Wavelength Division Multiplexer (DWDM), Arbitrary Waveform Generators (AWG), electro-optic modulator (EOM), Photodetector (PD). (**b**) Time-domain electrical signals loaded onto the modulator and finally detected by the power meter. (**c**) Results of single, small-kernel calculations. (**d**) Error analysis of the system. (**e**) Features obtained by computer calculations of the eight convolution kernels (top) versus features obtained by convolution of the optical system (bottom). (**f**) Confusion table. (**g**) Experimental setup.

3. Experimental Setup

The experimental procedure is meticulously illustrated in Figure 3a. The light source employed in this study was a continuous-wave laser, which was characterized by a series of frequency teeth ranging from 192.7 to 194.3 THz, with intervals of 0.2 THz, thereby comprising a total of nine distinct frequency teeth. Prior to the introduction of the light through the spectrum shaper, the initial spectral distribution is depicted in Figure 3a.

The spectrum shaper, specifically the Finisar 16000A model, played a crucial role in encoding the convolution kernel information. That information, initially in a two-dimensional format, was transformed into a one-dimensional vector and subsequently

embedded into the spectral profile of the light source. This encoding process was essential for the subsequent optical signal processing.

The input data for the experiment were derived from the MNIST dataset, which consists of handwritten digit images. Each image is composed of a 28×28 pixel grid, with grayscale values ranging from 0 to 255. To facilitate the encoding process, these grayscale values were normalized to the interval $[0, 1]$. The normalized values were then encoded onto the radio-frequency input of the lithium niobate electro-optic intensity modulator (iXblue, MX-LN-20, Saint-germain-en-laye, Ile-de-France, France) using an arbitrary waveform generator (Keysight, M8196A). The arbitrary waveform generator operated at a sampling rate of 1.429 GHz, as shown in Figure 3b. Under the limitation of the current electro-optic modulator's 3 dB bandwidth, we found that at the current modulation frequency of 1.429 GHz, the system could ensure a high signal-to-noise ratio of 27.98 dB. Upon increasing the modulation frequency, it became challenging to maintain a high level of signal-to-noise ratio; hence, we chose the current modulation frequency as a compromise.

Due to the anisotropy of lithium niobate crystals, their electro-optic modulators are polarization-dependent. Before the optical signal entered the electro-optic modulator, it was necessary to use a polarization controller to rotate the polarization of the incident light to the appropriate direction. The electro-optic modulator, which had an approximate half-wave voltage of 5 V, was biased at $\pi/4$ to achieve a more accurate linear approximation of the input signal. The output voltage of the arbitrary waveform generator was modulated to 100 mV to ensure that the response of the optical intensity to the voltage signal could be approximated linearly.

Following the modulation process, the optical signal passed through an optical fiber with a dispersion coefficient of 100 ps/(km · nm) and a length of 4375 m. The dispersion introduced by the fiber was critical for achieving the desired frequency interleaving, which was essential for the accurate computation of the convolution operations. After traversing the fiber, the optical signals were measured using an optical power meter (Thorlabs, RXM40AF, Newton, NJ, USA), as indicated by the orange line in Figure 3b. The measured signals were then converted into a two-dimensional matrix to facilitate the observation of the computational results at that stage, as shown in Figure 3c.

For the 8 3×3 convolution kernels used in this experiment, their positive and negative elements were separated to form 16 convolution kernel matrices, each with uniform sign. The optical system computed the results, which are displayed at the bottom of Figure 3d. These results were compared with the convolution results obtained directly from an electronic computer, as shown at the top of Figure 3d. The current computational framework, as depicted in Figure 2a, enabled the system to achieve an approximate recognition accuracy of 97%, as illustrated in Figure 3e.

4. Discussion

It is a misconception to equate the product of frequency resources and modulation bandwidth with the actual computational power in existing optical convolution systems. As illustrated in Figure 1, for a single convolution kernel of size 3×3, the actual computational power is merely one-third of the aforementioned product. This efficiency is likely to decrease with an increasing convolution kernel size. In our computational framework, the aforementioned product can be exploited to a high degree of efficiency, albeit at the cost of additional summing operations at the backend. As a trade-off for improving the efficiency of matrix convolution, we require additional Dense Wavelength Division Multiplexing (DWDM) to perform the block summation depicted in Figure 1c. Fortunately, the currently mature DWDM technology allows us not to worry about this additional cost.

Currently, in the case of multiple physical computing systems with limited frequency resources, these can be computed collectively within our framework and subsequently integrated into a larger system with enhanced computational capabilities. In the demonstration experiments, only multiple independent lasers were employed to illustrate that the subsystems can be integrated into a larger system. In actual deployment, thin-film lithium niobate on-chip optical frequency combs coupled with on-chip electro-optical intensity modulators can be utilized to achieve higher integration and more frequency resources, which is anticipated to facilitate the deployment of optical accelerated computing into practical applications in a more expedient manner.

Furthermore, our research reveals the potential of optical computing in handling large-scale datasets. For instance, in image recognition tasks, traditional electronic computing methods require substantial computational resources and time, whereas our optical convolution accelerator can significantly enhance computational speed, thus offering a distinct advantage in real-time image processing and video analysis applications. This advantage is not only reflected in speed but also in energy efficiency. The energy consumption of optical computing systems is far lower than that of traditional electronic computing systems, which is crucial for applications that require long-term operation and processing of large quantities of data.

However, we also recognize that there are challenges in practical applications. First, the stability and reliability of optical computing systems need to be further improved. For example, optical signals may be affected by environmental factors during transmission, such as temperature changes and vibrations, leading to errors in computational results. Therefore, it is necessary to develop more advanced optical signal processing technologies and error correction algorithms to ensure the stability and reliability of the systems. Second, the integration and scalability of optical computing systems also need to be optimized. Although current technologies for on-chip optical frequency combs and electro-optical modulators have made some progress, there are still technical challenges in achieving large-scale integration and diverse functionalities. Future research can explore new materials and manufacturing processes to improve the integration and scalability of optical computing systems.

Moreover, the application scope of optical computing systems needs to be expanded. Currently, optical computing is mainly applied to specific types of computational tasks, such as convolutional neural networks, and the potential of optical computing in other types of tasks, such as graph computing and optimization problems, has not been fully explored. Future research can investigate the possibilities of optical computing in a broader range of application fields, such as bioinformatics, financial analysis, and weather forecasting, developing optical computing algorithms and system architectures suitable for these areas.

In summary, our research provides new insights and methods for the development of the field of optical computing. By proposing a novel frequency encoding scheme and optimizing system architecture, we have not only improved the computational performance of optical convolution accelerators but also laid the foundation for the deployment of optical computing in practical applications. In the future, with the continuous advancement and innovation of related technologies, optical computing is expected to play a significant role in more fields, providing more efficient and energy-saving solutions for solving complex computational problems.

Author Contributions: Conceptualization, G.X.; methodology, G.X. and J.L.; investigation, software, G.X. and P.Z.; writing—original draft preparation, G.X.; writing—review and editing, G.X. and Q.H.; supervision, Z.Z.; project administration, G.X.; funding acquisition, P.X. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

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

References

1. Esser, S.K.; Merolla, P.A.; Arthur, J.V.; Cassidy, A.S.; Appuswamy, R.; Andreopoulos, A.; Berg, D.J.; McKinstry, J.L.; Melano, T.; Barch, D.R.; et al. Convolutional networks for fast, energy-efficient neuromorphic computing. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11441–11446. [CrossRef]
2. Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwińska, A.; Colmenarejo, S.G.; Grefenstette, E.; Ramalho, T.; Agapiou, J.; et al. Hybrid computing using a neural network with dynamic external memory. *Nature* **2016**, *538*, 471–476. [CrossRef]
3. 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]
4. Larger, L.; Soriano, M.C.; Brunner, D.; Appeltant, L.; Gutierrez, J.M.; Pesquera, L.; Mirasso, C.R.; Fischer, I. Photonic information processing beyond Turing: An optoelectronic implementation of reservoir computing. *Opt. Express* **2012**, *20*, 3241. [CrossRef] [PubMed]
5. Vandoorne, K.; Mechet, P.; Van Vaerenbergh, T.; Fiers, M.; Morthier, G.; Verstraeten, D.; Schrauwen, B.; Dambre, J.; Bienstman, P. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **2014**, *5*, 3541. [CrossRef]
6. Van der Sande, G.; Brunner, D.; Soriano, M.C. Advances in photonic reservoir computing. *Nanophotonics* **2017**, *6*, 561–576. [CrossRef]
7. Lin, X.; Rivenson, Y.; Yardimci, N.T.; Veli, M.; Luo, Y.; Jarrahi, M.; Ozcan, A. All-optical machine learning using diffractive deep neural networks. *Science* **2018**, *361*, 1004–1008. [CrossRef] [PubMed]
8. Mengu, D.; Zhao, Y.; Yardimci, N.T.; Rivenson, Y.; Jarrahi, M.; Ozcan, A. Misalignment resilient diffractive optical networks. *Nanophotonics* **2020**, *9*, 4207–4219. [CrossRef]
9. Fu, T.; Zang, Y.; Huang, Y.; Du, Z.; Huang, H.; Hu, C.; Chen, M.; Yang, S.; Chen, H. Photonic machine learning with on-chip diffractive optics. *Nat. Commun.* **2023**, *14*, 70. [CrossRef] [PubMed]
10. Tait, A.N.; Nahmias, M.A.; Shastri, B.J.; Prucnal, P.R. Broadcast and Weight: An Integrated Network For Scalable Photonic Spike Processing. *J. Light. Technol.* **2014**, *32*, 4029–4041. [CrossRef]
11. 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]
12. 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] [PubMed]
13. 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]
14. Cai, K.; Chen, L.; Zhang, Y.; Wang, J.; Lin, W.; Duan, S.; Liu, B. On-chip photoelectric hybrid convolutional accelerator based on Bragg grating array. *Results Phys.* **2024**, *65*, 107968. [CrossRef]
15. Rizzo, A.; Novick, A.; Gopal, V.; Kim, B.Y.; Ji, X.; Daudlin, S.; Okawachi, Y.; Cheng, Q.; Lipson, M.; Gaeta, A.L.; et al. Massively scalable Kerr comb-driven silicon photonic link. *Nat. Photonics* **2023**, *17*, 781–790. [CrossRef]
16. Fu, T.; Zhang, J.; Sun, R.; Huang, Y.; Xu, W.; Yang, S.; Zhu, Z.; Chen, H. Optical neural networks: Progress and challenges. *Light. Sci. Appl.* **2024**, *13*, 263. [CrossRef]
17. McMahon, P.L. The physics of optical computing. *Nat. Rev. Phys.* **2023**, *5*, 717–734. [CrossRef]
18. Mao, J.; Uemura, F.; Yazdani, S.A.; Yin, Y.; Sato, H.; Lu, G.W.; Yokoyama, S. Ultra-fast perovskite electro-optic modulator and multi-band transmission up to 300 Gbit/s. *Commun. Mater.* **2024**, *5*, 114. [CrossRef]
19. Cheng, R.; Yu, M.; Shams-Ansari, A.; Hu, Y.; Reimer, C.; Zhang, M.; Lončar, M. Frequency comb generation via synchronous pumped $\chi(3)$ resonator on thin-film lithium niobate. *Nat. Commun.* **2024**, *15*, 3921. [CrossRef]