WDM-compatible integrated photonic computing core for implementing a neural network

Zhenyu Zhao[1, 2, 3, 4, †] Shuang Zheng[1, 2, 3, 4, †], Weifeng Zhang[1, 2, 3, 4, *]

1. Radar Research Lab, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
2. Key Laboratory of Electronic and Information Technology in Satellite Navigation (Beijing Institute of Technology), Ministry of Education, Beijing 1000 81, China
3. Beijing Institute of Technology Chongqing Innovation Center, Chongqing, 401120, China
4. Chongqing Key Laboratory of Novel Civilian Radar, Chongqing 401120, China
† These authors contributed equally to this work.
**e-mail address: weifeng.zhang@bit.edu.cn*

Abstract: We propose and design an energy-efficient integrated photonic computing core by using a microdisk resonator-assisted Mach-Zehnder interferometer. With the use of the fabricated silicon photonic chip, an optical convolutional neural network for image classification is experimentally demonstrated. © 2022 The Author(s)

1. Introduction

Artificial neural networks (ANNs) have achieved great success in a wide range of cognitive tasks, such as speech recognition and image processing. Limited by the transistor size and the "memory wall" caused by the von Neumann architecture, it is difficult for conventional electronic computing hardware to meet the increased computational demands of the ANNs [1]. In recent years, optical neural networks (ONNs) have been demonstrated to be an emerging neuromorphic platform with ultra-low latency, high bandwidth, and high energy efficiency [2], since computationally-intensive operations in neural networks, e.g., matrix multiplication, can be efficiently realized by optics at the speed of light. In 2017, Shen et al. demonstrated an integrated ONN to realize vowel recognition based on a programmable integrated Mach-Zehnder interferometer (MZI) array [3]. This approach is capable of implementing reconfigurable and multi-layer ONN via complex-valued matrix computation [4]. However, the key problem of the thermo-optic MZI modulator is its high power consumption.

To reduce the power consumption, in this paper, we propose and demonstrate a WDM-compatible integrated photonic computing core that is capable to perform parallel vector-matrix multiplication with a high energy efficiency. Thanks to the microdisk resonator (MDR) assisted MZI configuration, by tuning the resonant wavelength of each MDR independently, the computing core can support multi-wavelength operation, enabling parallel matrix computation with a low power consumption. An experiment is performed in which an optical convolutional neural network (CNN) for image classification is realized with the use of the fabricated silicon photonic chip.

2. Principle

Figure 1(a) shows the schematic of the proposed WDM-compatible integrated photonic computing core. The key component of this computing core is an MDR-assisted MZI structure. Each arm of the MZI is composed of three MDRs with an identical radius of 6.4 μm, the gap between the MDR and bus waveguide is designed as 200 nm to ensure that it works in over-coupling condition. The resonant wavelengths of MDRs are independently tuned by the corresponding thermo-optic electrodes, and so are their phase response. Thus, each MDR can be considered as a tunable phase shifter. With the MZI interferometer, by tuning the resonant wavelengths, at the output different splitting ratios can be achieved at different wavelengths. The input data could be represented by different vectors with 2 components that are encoded in the optical power of different wavelengths via spectral shaping by a waveshaper. And different 2×2 weight matrices could be also implemented by tuning the MDRs. Thus, the input vector and parallel matrix-vector multiplication between a matrix 2×2 and a vector 2×1 could be realized at different wavelengths. The inset of Fig. 1(a) shows the measured optical image of the fabricated silicon photonic computing core. As can be seen, four vertical grating couplers spaced at 127 μm are used for input and output fiber-chip coupling. The thermo-optic microdisk modulators on each arm of the device are separated by insulating trenches.

The transmission spectrum of the WDM-compatible integrated photonic computing core is shown in Fig. 1(b). The FSR of the MDR is about 16 nm, and the average extinction ratio exceeds 16 dB by precisely tuning the resonant wavelengths of the three sets of MDRs. By further tuning, the wavelengths of the three sets of MRRs can be arbitrarily adjusted, thus giving the WDM-compatible photonic computing core the ability to compute in parallel.

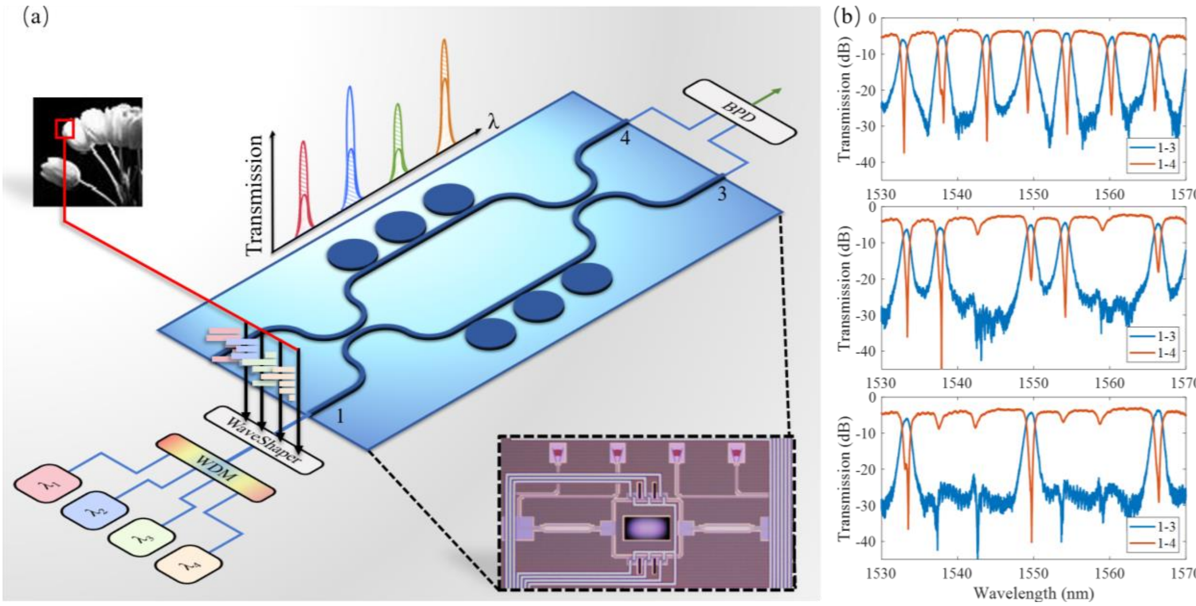

Fig. 1. (a) Schematic of the proposed WDM-compatible photonic computing core. The inset shows the optical image of the fabricated chip. (b) The measured transmission spectrum of the photonic computing core.

3. Experimental results

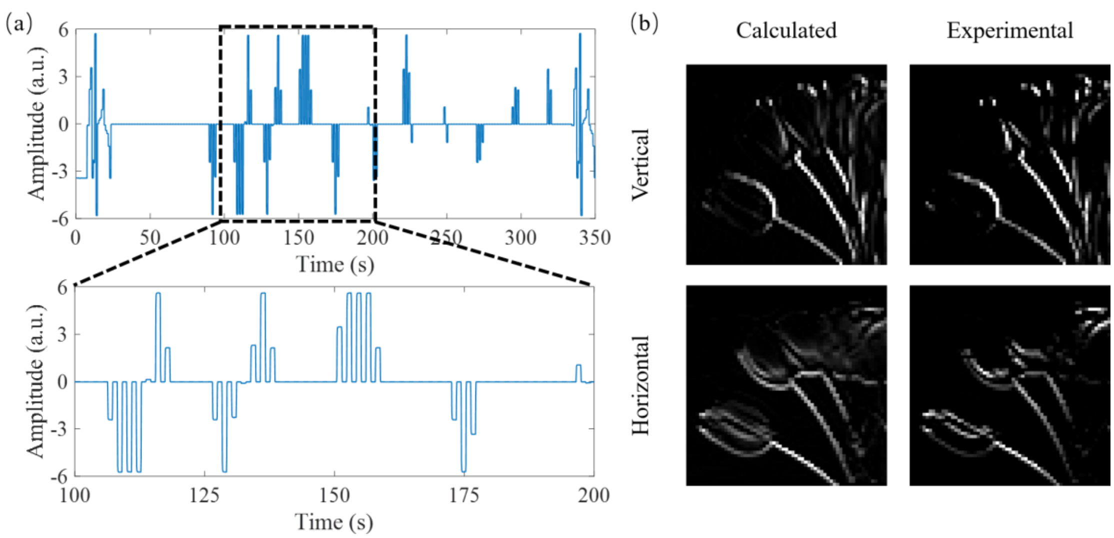

Fig. 2. Convolutional image processing using the fabricated photonic computing core. (a) Part of the measured temporal waveform for convolutional image processing. (b) The theoretically calculated edge-detected images and the experimentally measured edge-detected images.

In the experiment, we build a photoelectric system to conduct convolutional image processing with the proposed photonic computing core, as shown in Fig. 1(a). As a proof-of-concept demonstration, we perform the convolution process of a 64×64 8-bit image of a flower picture to detect its edge feature. The flower image in the inset of Fig. 1(a) is used as the input grayscale image. Two 2×2 kernels $\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$ $\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ are used in the experiment for horizontal and vertical edge detection, respectively. According to the order in which the convolution kernel moves in the convolution calculation, the grayscale image is divided into a series of 2×2 data groups. Each data group contains 4 pixels. Then, the data group and kernel weight are flattened into vectors X, Y respectively. After the flattening operation, vector X is loaded into the WaveShaper by setting different amplitude attenuation on different wavelengths, and vector Y is loaded into the photonic computing core by adjusting the resonant wavelengths of the three sets of MDRs. When the input optical sequences of four different wavelengths propagate through the chip, the vector X represents the image information, and the vector Y represents the kernel weights are multiplied. The multiplication result for each data group is then accumulated through a balanced photodetector (BPD) with a bandwidth of 150 MHz. The processed image can be recovered by sampling and rearranging the output temporal signal from the BPD. As can be seen in Fig. 2(b), the experimentally measured edge-detected image agrees very well with the theoretically calculated one, and the mean square error (MSE) are 0.004 (horizontal) and 0.008 (vertical), respectively. This result confirms the effectiveness of the use of the photonic computing core to perform convolutional image processing, which is a prerequisite for building an optical CNN.

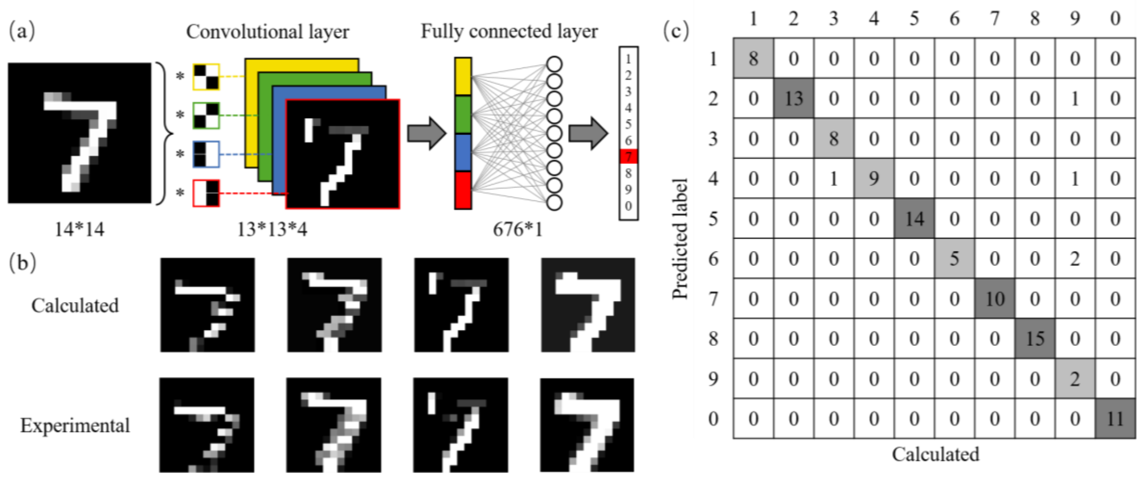

Fig 3. An optical CNN for image classification. (a) Schematic of optical CNN-based image classification. (b) The theoretically calculated feature maps and the experimentally measured feature maps from the convolutional layer. (c) Confusion matrix obtained from 100 repeated experiments.

Here, we further construct an ONN for classifying handwritten digit images in the MNIST dataset based on the proposed photonic computing core. Figure. 3(a) gives the structure of the built CNN with a convolutional layer and a fully connected layer. Four 2×2 convolution kernels are used to extract the feature maps of the 14×14 handwritten digit image in the convolutional layer. By sampling the output temporal waveform of the photonic computing core and using the activation function (ReLu) to process the sampling result, four vectors each containing 169 elements are obtained. Then, arranging these elements into four 13×13 matrices, the feature maps of the input handwritten digit image are obtained. Figure. 3(b) displays the experimental results of the convolutional layer. It can be found that the experimentally reconstructed feature maps are consistent with that of a digital computer. This convolutional layer extracts features that contribute to image recognition. By finishing the following fully connected layer in an auxiliary computing device, the recognition result is obtained. We use the trained convolution kernel to process 100 handwritten images to get their feature maps, and the corresponding classification result gives a classification accuracy of 95%. The diffusion matrix with 10 categories is shown in Fig. 3(c).

In conclusion, we propose a WDM-compatible integrated photonic computing core, which is capable to process parallel matrix computation with a high energy efficiency. Based on the fabricated silicon photonic chip, an optical convolutional neural network is experimentally realized. The demonstrated structure holds great potential for improving the parallel computational capability and reducing the power consumption of conventional MZI structure.

4. Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) (62105028 and 62071042), and the National Key R&D Program of China (2018YFE0201800).

5. References

[1] Tallents, G., Wagenaars, E., & Pert, G. Lithography at EUV wavelengths. Nat. Photonics, 4(12), 809-811 (2010).

[2] Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. Nature, 589(7840), 44-51 (2021).

[3] Shen, Y. et al. Deep learning with coherent nanophotonic circuits. Nat. Photonics 11, 441 (2017)

[4] Zhang, H., Gu, M., Jiang, X.D. et al. An optical neural chip for implementing complex-valued neural network. Nat. Commun. 12, 457 (2021).