International Journal of Reconfigurable and Embedded Systems (IJRES)
Vol. 14, No. 2, July 2025, pp. 311~319
ISSN: 2089-4864, DOI: 10.11591/ijres.v14.i2.pp311-319

Building a photonic neural network based on multi-operand multimode interference ring resonators

Thanh Tien Do, Hai Yen Pham, Trung Thanh Le
Faculty of Engineering and Technology, International School (VNU-IS), Vietnam National University (VNU), Hanoi, Vietnam

Article Info

Article history:

Received Aug 19, 2024
Revised Dec 30, 2024
Accepted Mar 29, 2025

Keywords:

4×4 multimode interference
Convolutional neural networks
High-speed computing
Multi-operand ring resonators
Photonic neural networks

ABSTRACT

Photonic neural networks (PNNs) offer significant potential for enhancing deep learning networks, providing high-speed processing and low energy consumption. In this paper, we present a novel PNN architecture that employs nonlinear optical neurons using multi-operand 4×4 multimode interference (MMI) multi-operand ring resonators (MORRs) to efficiently perform vector dot-product calculations. This design is integrated into a photonic convolutional neural network (PCNN) with two convolutional layers and one fully connected layer. Simulation experiments, conducted using Lumerical and Ansys tools, demonstrated that the model achieved a high test accuracy of 98.26% on the MNIST dataset, with test losses stabilizing at approximately 0.04%. The proposed model was evaluated, demonstrating high computation speed, improved accuracy, low signal loss, and scalability. These findings highlight the model's potential for advancing deep learning applications with more efficient hardware implementations.

Corresponding Author:

Trung Thanh Le
Faculty of Engineering and Technology, International School (VNU-IS)
Vietnam National University (VNU)
144 Xuan Thuy Road, Cau Giay, Hanoi 100000, Vietnam
Email: thanh.le@vnu.edu.vn

1. INTRODUCTION

Convolutional neural networks (CNNs), influenced by the principles of biological nervous systems, have developed into a strong form of artificial neural network [1]. They are utilized in image recognition tasks, allowing for a notable reduction in the complexity of the network while still providing highly accurate predictions. Further, there are many applications, consisting of image classification, object detection, facial recognition, real-time language translation, and other related fields [2], [3]. The growing complexity of applications like self-driving vehicles and cloud-based artificial intelligence (AI) services is driving the need for faster, more energy-efficient neuromorphic hardware.

However, current methods, which largely rely on the von Neumann computing architecture, face a fundamental challenge involving a trade-off between the speed of data transfer and the amount of energy consumed. This limitation arises mainly from the fact that, in these systems, memory, and processing units are physically separated, leading to inefficiencies in performance and energy consumption [4]-[7]. This separation creates bottlenecks when handling large volumes of data, making it difficult to simultaneously optimize both speed and energy efficiency.

Furthermore, convolution operational layers make up about 90% of a CNN's computations [8]. While there is strong parallelism within layers, data dependencies across layers challenge any effort of inter-layer parallelization. This leads to scalability issues with power and performance, which are worsened by the inherent power and speed limits of electronics.

Photonic neural networks (PNNs) are considered highly promising for the next generation of hardware processors designed for neuromorphic computing. Photonic devices have very low loss, helping to greatly minimize signal attenuation. Furthermore, they can overcome the bandwidth bottlenecks found in electrical systems, achieving high computing speeds [9]-[12]. Moreover, in PNNs, light transmission directly facilitates data processing, effectively eliminating the need for data shuttling, which is a key inefficiency in the von Neumann computing model. Over recent years, PNNs have garnered significant attention for enabling high-speed, large-scale, and highly parallel optical neuromorphic hardware. Various photonic CNN implementations have been classified into four groups: optical CNNs based on light diffraction [13]-[17], optical CNNs using light interference [18]-[21], optical CNNs utilizing wavelength division multiplexing [22]-[24], and optical CNNs relying on tunable optical attenuation [25], [26].

However, previous traditional PNNs are constrained by high area costs and the limitation of one multiply-accumulate (MAC) operation per photonic device. Using matrix singular value decomposition (SVD) and unitary matrix parametrization as described in Reck *et al.* [27] and Ribeiro *et al.* [28]. Shen *et al.* [29] developed and implemented a fully PNN, achieving a multilayer perceptron (MLP) architecture through arrays of Mach-Zehnder interferometers (MZIs). However, its area cost improvement is limited. These scalability limitations are a key challenge that PNNs seek to overcome. A hardware-software co-design of slimmed PNN based on MZIs has been proposed [30] to achieve a reduction of 15% to 38% in the number of MZIs required for different network sizes. Moreover, an FFT-based architecture proposed by Gu *et al.* [31] achieves 2.2 3.7x area cost improvement compared with other PNNs.

Furthermore, many PNNs only achieve accuracy below 95%, which is significantly lower than the average 99% of electrical CNNs [32]-[34]. Therefore, there is a strong need to design a new PNN architecture that can achieve high accuracy and low loss comparable to that of electrical CNNs.

To address the area cost limitations and achieve high accuracy with low loss in PNNs, we propose a novel PNN architecture using a new multi-operand ring resonator (MORR) design, which relies on a single 4×4 multimode interference (MMI) coupler on silicon waveguides. The key advantages of this new design include low loss, compatibility with complementary metal oxide semiconductor (CMOS) technology, high bandwidth, relaxed fabrication tolerances, and reduced sensitivity to wavelength or polarization variations due to the use of the MMI coupler. The main contributions are as follows:
- Scalability: we introduce a scalable PNN architecture that surpasses previous PNNs in terms of footprint, offering a more compact and efficient design.
- Efficiency: the proposed PNN architecture supports high levels of parallel computation, achieving both high accuracy and low loss. This addresses key challenges found in traditional electrical CNNs, offering improved performance in PNN systems.

2. METHOD

In this section, we first design a novel MORR functioning as a neuron in a PNN. It is constructed using a 4×4 MMI coupler with silicon-based waveguides. Following this, a complete PNN architecture is proposed, demonstrating how data can be transformed and processed within photonic components. The output is generated by converting optical signals into digital data, which is then used to classify images on the MNIST dataset. The experiment to validate the proposed PNN and its performance on the dataset is conducted via simulation using Lumerical and Interconnect (Ansys) tools, integrated with Python (PyTorch) to execute a fully operational PNN.

2.1. Proposed multi-operand ring resonator profile

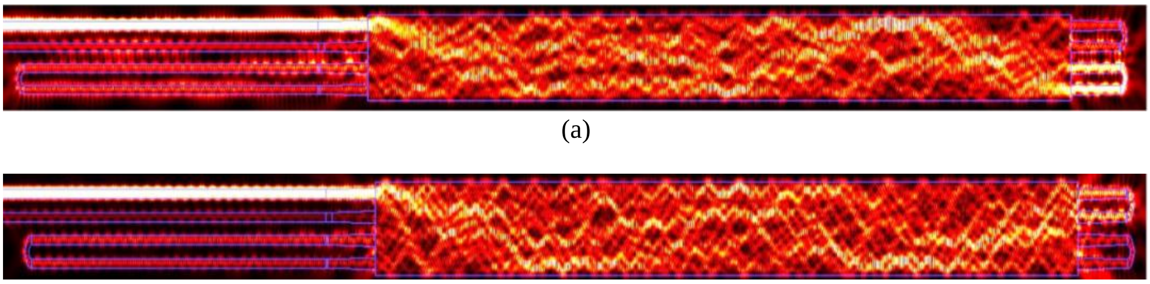
We present a novel MORR controlled by n electrical signals $x_{1,} x_{2,} \dots x_n$ (see Figure 1). The MORR is integrated with a 4×4 MMI coupler on silicon waveguides. The waveguide are made of a silicon on insulator with dimensions of 500 nm in width and 250 nm in height for both input and output paths. The selected length (L_{MMI}) and width (W_{MMI}) of the 4×4 MMI are $L_{MMI} = \frac{3L_\pi}{2} = 214\,\mu m$ and $W_{MMI} = 6\,\mu m$, respectively [35]. Where $L_\pi = \frac{\pi}{\beta_0 - \beta_1}$ is the beat length of the MMI; β_0 and β_1 are the propagation constants of the fundamental and first-order modes supported by the multimode waveguide with the width of W_{MMI}. Each input signal x_i induces a phase shift $\phi_i(x_i)$, with the total accumulated phase shift given by $\phi = \sum_{i=1}^{n} \phi_i\,(x_i)$. We employ them as neurons in PNN. The in/tensity buildup function of ϕ described by the following [36]:

$$\frac{I_{out}}{I_{in}} = f(\phi) = \left|\frac{a^2 - 2arcos\phi + r^2}{1 - 2arcos\phi + a^2r^2}\right|,\ \phi = \sum_{i=1}^{n} \phi_i\,(x_i) \propto w_i x_i$$

In this equation, I_{out} and I_{in} represent the output and input light intensity at the input and through ports, respectively. Both of them are normalized within a range between 0 and 1 ($I_{out}, I_{in} \in [0,1]$), a is self-coupling coefficient and r is the single-pass amplitude transmission factor. The variable x_i denotes the input voltage in the electric domain, and w_i is the weight associated with that input. The weight w_i can be adjusted through various mechanisms such as altering the size of active regions or modifying driving voltages. The $f(\phi)$ transfer function is non-linear and can be leveraged in neural network computation.

Figure 1. A multi-operand micro-ring resonator

For the proposed MORR structure, the free spectral range (FSR) is very high. As a result, a high bandwidth and high bit rate can be achieved. The proposed design is suitable for high-speed systems. The finite-difference time-domain (FDTD) simulation for ON and OFF resonances is shown in Figure 2. The Figure 2(a) shows the ON resonance state. In this state, the signals within the waveguides constructively interfere, resulting in the propagation of the signal through the waveguide system. This state demonstrates strong signal transmission as the microring is resonant at the operating wavelength. The Figure 2(b) shows the OFF resonance state. In this state, destructive interference occurs, resulting in weak signal transmission or almost no signal propagation. This indicates that the microring is not resonant at the operating wavelength, effectively blocking the signal.



(a)

(b)

Figure 2. Simulations of the signal propagation via the MORR using a 4×4 MMI coupler on silicon waveguides (a) ON and (b) OFF resonances

2.2. Proposed photonic convolutional neural network

Building on the proposed MORR, we introduce a novel PNN, where the intensity of input power is created by modulating many different light sources, such as laser diodes that operate at different wavelengths. Initially, these laser diodes are multiplexed at first and then split onto multiple rows in an array of microring resonators (see Figure 3). By utilizing photodiodes at the output to capture the photonic signal from each row, the accumulated power of all wavelengths is measured as the output power of light.

Assuming we have an $X \times Y$ matrix of weights mapping onto an array of MORR, it is split into a grid $M \times N$ sub-matrices of blocks of size $n \times n$ (n - the number of operands in the ring resonators, and $M = X \div n$, $N = Y \div n$). Each block $n \times n$ applies block-circulant weight matrix technique [37]. This structure allows us to deploy an array of MORRs for nonlinear computations in PNNs. In the m_{j-th} row, the MORRs resonate at the wavelength λ_j and the output $I_{j_{out}}(y_j) = f(\sum_{i=1}^{n} w_{ji} x_{ji}).I_{j_{in}}$, as measured by a photodetector. Finally, the light intensity y_m in row m is detected and measured as $I_{m_{out}}(y_m) = \sum_{j=1}^{N} I_{j_{out}}$.

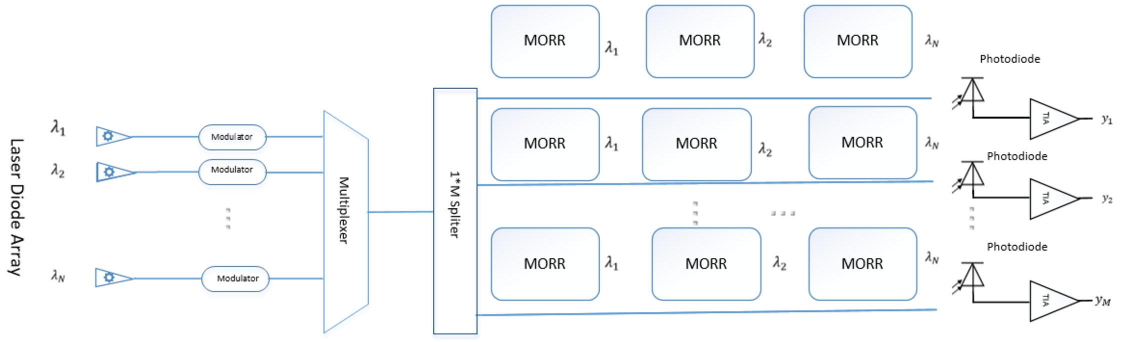


Figure 3. The proposed model

3. RESULTS AND DISCUSSION

3.1. Simulation of the proposed photonic neural network

We simulate a MORR neuron with three operands (n=3) at a resonance wavelength of 1,554.28 nm. A photonic simulation is conducted to test the functionality of the proposed model within a CNN. The CNN configuration includes the first two convolutional layers, followed by a flattened layer for classification. Each convolutional layer includes 32 kernels with a size of 5×5, a stride of 2, and a padding of 1. After each convolutional layer, we apply batch normalization (BatchNorm) using trans-impedance amplifiers (TIA) and offset voltage signals to minimize latency overhead. By adjusting the gain, BatchNorm effectively smooths the amplitude of activation signals. The flattened layer is used for the final classification, producing 10 outputs.

The PNN is trained for 100 epochs using the Adam optimizer. The dataset is MNIST. Input data, kernels, and activation functions are represented in float32 format. The proposed photonic CNN model is implemented using an open-source library based on Pytorch and runs on a machine equipped with an Intel Core i5-9700 CPU and NVIDIA RTX 3600 GPU.

3.2. Performance of proposed photonic neural network

The results indicate that the highest test accuracy of the model is 98.26% at epoch 100. The three highest test accuracies are 98.72% (epoch 79), 98.7% (epoch 82), and 98.68% (epoch 91) (see Figure 4). The test loss stabilizes at approximately 0.04% from epoch 80 (see Figure 5). The test accuracy of our model nearly reaches the 99% level, similar to traditional CNN models like VGG-16, ResNet-50, and AlexNet [38]-[41]. However, the speed of our model based on light-based computation, represents a significant improvement over the digital computations employed by these traditional CNNs [42], [43].

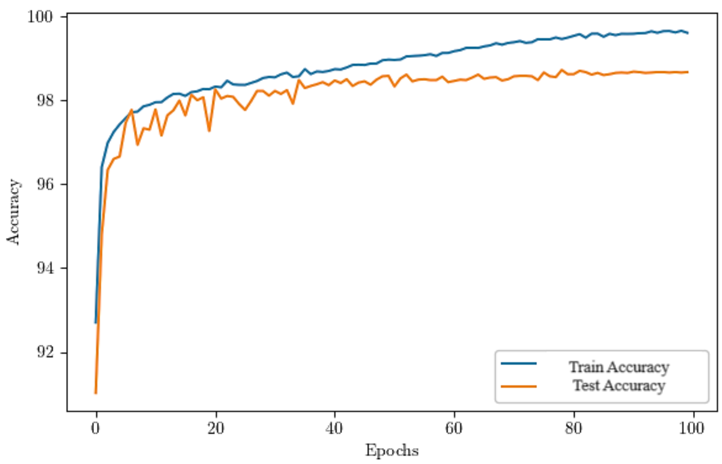


Figure 4. Train accuracy and test accuracy

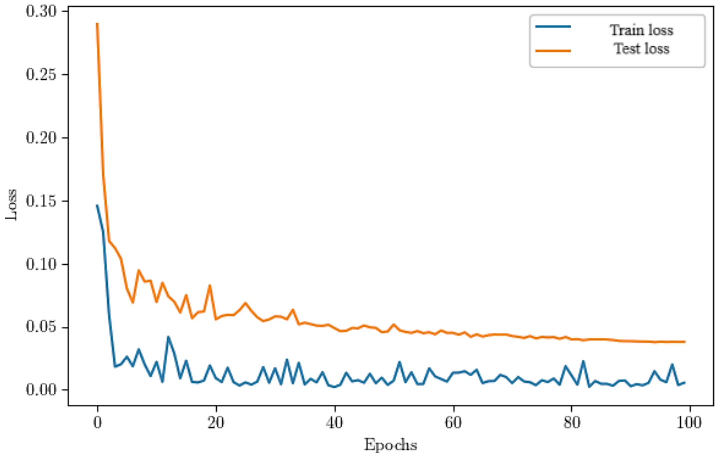


Figure 5. Train loss and test loss

We adjust the number of convolutional layers to observe the changes in accuracy and loss. By evaluating different configurations, we assess how the number of convolutional layers impacts the model's performance in terms of both accuracy and loss. The results show that as the number of convolutional layers increases, accuracy improves while loss decreases (see Table 1). The optimal setup includes three convolutional layers and one fully connected layer, achieving the highest accuracy (99.07%) and the lowest loss (0.03%). However, adding a fourth convolutional layer may lead to overfitting and increased computational complexity. Additionally, increasing the number of layers raises the computational load and resource requirements, which should be carefully considered, particularly when deploying models on devices with limited hardware capabilities.

Table 1. Our models on different number of layers

Our models	Accuracy (%)	Loss (%)
1. Conv+1 FC	97.39	0.08
2. Conv+1FC	98.7	0.06
3. Con+1FC	99.07	0.03

3.3. Comparison with other photonic neural network

We run our model using a precision of 7 bits for both inputs and weights. The model achieves an accuracy of approximately 98.1% (see Figure 6), with a loss of around 0.04%. When compared to other PNN (see Table 2), the results indicate that our model outperforms them in terms of accuracy on the same MNIST dataset.

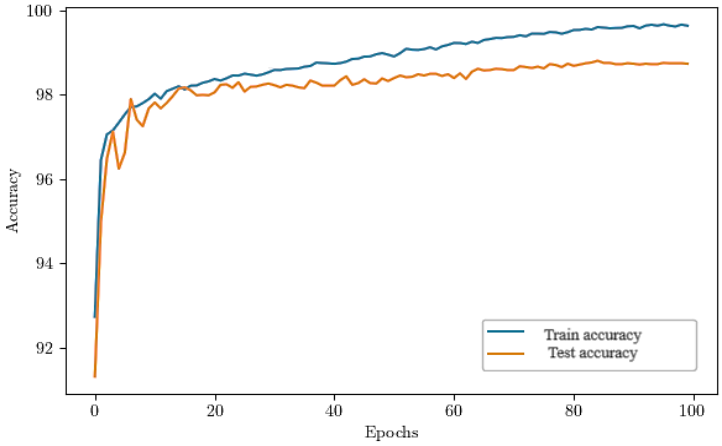


Figure 6. Test accuracy on 7-bit for parameters

Table 2. Comparison between our model with other previous PNN models

Model	Layers	Quantization	Accuracy (%)
Our model	2 Conv+1 FC	7 bit	98.1
IDNN-VMM [16]	2FC	/	92.6
MRR-VMM [17]	3FC	4 bit	97.41
MZI-VMM [18]	2FC	5 bit	76.70
PCM-VMM [19]	1 Conv+1FC	7 bit	95.30
TOPS-CA [20]	Conv+1FC	7 bit	88

3.4. Area reduction and system efficiency

By taking advantage of using the block-circulant weight matrix technique presented by Ding *et al.* [37], our model can reduce the number of weight matrices for better efficiency and area reduction. Figure 7 illustrates the transformation from an unstructured weight matrix with 18 parameters to a block-circulant weight matrix with only 6 parameters in a neural network. In the block-circulant matrix, rows are generated by cyclically shifting the elements of the first row, significantly reducing the number of stored parameters.

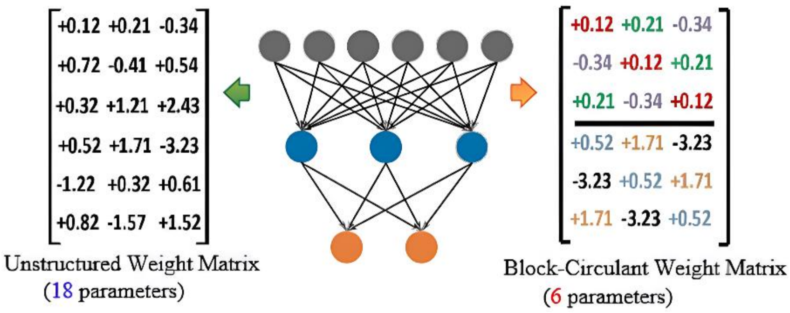


Figure 7. Block-circulant weight technique [37]

By utilising the MORR device, we can compress a vector dot-product operation into a single micro-ring. Specifically, a $X \times Y$ matrix in our model contains $M \times N$ sub-matrices of block of size $n \times n$ (n - the number of operands in the ring resonator). Each sub-matrix of size $n \times n$ is used in the form of a circulant matrix. This means that the sub-matrix has a special structure in which each row (or column) is a cyclic shift of the first row (or column). This circulant structure allows for reducing the number of weights required and making matrix multiplication more efficient. Further, the circulant matrix structure allows for area reduction in device and wavelength usage. As a result, it leads to $O(n^2)$ times area reduction and n times wavelength usage savings. In comparison, other PNNs, such as the one designed and fabricated by Shen *et al.* [29], require $O(n^2)$ MZIs for implementation. Besides this, wavelength savings make optical systems more efficient in terms of cost, energy, and performance while reducing complexity and optimizing processing capabilities.

4. CONCLUSION

In this study, we introduced a novel photonic convolutional neural network (PCNN) based on multi-operand 4×4 MMI micro-ring resonators, demonstrating its effectiveness in deep learning tasks like image recognition. Our model achieved high accuracy, low signal loss, rapid light-based computation, showing great potential compared to traditional digital CNN architectures. The flexibility of this model allows it to be adapted to various CNN structures, including AlexNet, ResNet, and VGG16, offering promising results for future AI and deep learning advancements.

This study shows that PNN can significantly improve speed and efficiency in large-scale data processing and autonomous systems. Future research should focus on integrating this model into more complex architectures and testing its real-time performance. Overcoming challenges in large-scale implementation and adapting it to various environments will further enhance its impact. These findings suggest that PNN could play a key role in the future of AI, offering faster and more efficient solutions for practical applications.

FUNDING INFORMATION

The authors thank the Ministry of Science and Technology of Vietnam for funding this research under project number ĐTĐL.CN-92/21.

AUTHOR CONTRIBUTIONS STATEMENT

This work follows the Contributor Roles Taxonomy (CRediT). The specific contributions of each author are detailed below:

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Thanh Tien Do	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Trung Thanh Le	✓		✓			✓		✓	✓	✓	✓	✓		
Hai Yen Pham				✓			✓			✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The authors confirm that all data supporting the findings of this study are available within the article and its supplementary materials

REFERENCES

[1] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996, doi: 10.1109/2.485891.
[2] N. Shabairou, E. Cohen, O. Wagner, D. Malka, and Z. Zalevsky, "Color image identification and reconstruction using artificial neural networks on multimode fiber images: towards an all-optical design," *Optics Letters*, vol. 43, no. 22, pp. 5603–5606, Nov. 2018, doi: 10.1364/ol.43.005603.
[3] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*, pp. 1–13, 2018, doi: 10.1155/2018/7068349.
[4] M. Naylor and C. Runciman, "The Reduceron: Widening the von Neumann bottleneck for graph reduction using an FPGA," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5083, pp. 129–146, 2008, doi: 10.1007/978-3-540-85373-2_8.
[5] D. A. B. Miller, "Attojoule Optoelectronics for Low-Energy Information Processing and Communications," *Journal of Lightwave Technology*, vol. 35, no. 3, pp. 346–396, 2017, doi: 10.1109/JLT.2017.2647779.
[6] T. N. Theis and H. S. P. Wong, "The End of Moore's Law: A New Beginning for Information Technology," *Computing in Science and Engineering*, vol. 19, no. 2, pp. 41–50, 2017, doi: 10.1109/MCSE.2017.29.
[7] H. Markram *et al.*, "Reconstruction and Simulation of Neocortical Microcircuitry," *Cell*, vol. 163, no. 2, pp. 456–492, 2015, doi: 10.1016/j.cell.2015.09.029.
[8] J. Cong and B. Xiao, "Minimizing computation in convolutional neural networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8681 LNCS, pp. 281–290, 2014, doi: 10.1007/978-3-319-11179-7_36.
[9] M. Nazirzadeh, M. Shamsabardeh, and S. J. B. Yoo, "Energy-Efficient and High-Throughput Nanophotonic Neuromorphic Computing," in *Conference on Lasers and Electro-Optics*, Washington, D.C.: OSA, 2018, doi: 10.1364/CLEO_AT.2018.ATh3Q.2.
[10] Y. Fei *et al.*, "Design of the low-loss waveguide coil for interferometric integrated optic gyroscopes," *Journal of Semiconductors*, vol. 38, no. 4, p. 044009, Apr. 2017, doi: 10.1088/1674-4926/38/4/044009.
[11] R. Slavík, Y. Park, M. Kulishov, R. Morandotti, and J. Azaña, "Ultrafast all-optical differentiators," *Optics Express*, vol. 14, no. 22, pp. 10699–10707, 2006, doi: 10.1364/oe.14.010699.
[12] J. Huang, C. Li, R. Lu, L. Li, and Z. Cao, "Beyond the 100 Gbaud directly modulated laser for short reach applications," *Journal of Semiconductors*, vol. 42, no. 4, p. 041306, Apr. 2021, doi: 10.1088/1674-4926/42/4/041306.
[13] J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Scientific Reports*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-30619-y.
[14] M. Miscuglio *et al.*, "Massively parallel amplitude-only Fourier neural network," *Optica*, vol. 7, no. 12, pp. 1812–1819, 2020, doi: 10.1364/optica.408659.
[15] Z. Hu *et al.*, "High-Throughput Multichannel Parallelized Diffraction Convolutional Neural Network Accelerator," *Laser and Photonics Reviews*, vol. 16, no. 12, pp. 1–9, Dec. 2022, doi: 10.1002/lpor.202200213.
[16] W. Shi *et al.*, "LOEN: Lensless opto-electronic neural network empowered machine vision," *Light: Science and Applications*, vol.

11, no. 1, p. 121, 2022, doi: 10.1038/s41377-022-00809-5.
[17] H. H. Zhu *et al.*, "Space-efficient optical computing with an integrated chip diffractive neural network," *Nature Communications*, vol. 13, no. 1, p. 1044, 2022, doi: 10.1038/s41467-022-28702-0.
[18] X. Xu, L. Zhu, W. Zhuang, L. Lu, and P. Yuan, "A Convolution Neural Network Implemented by Three 3×3 Photonic Integrated Reconfigurable Linear Processors," *Photonics*, vol. 9, no. 2, p. 80, 2022, doi: 10.3390/photonics9020080.
[19] S. Xu *et al.*, "Optical coherent dot-product chip for sophisticated deep learning regression," *Light: Science and Applications*, vol. 10, no. 1, p. 221, 2021, doi: 10.1038/s41377-021-00666-8.
[20] C. Feng *et al.*, "A Compact Butterfly-Style Silicon Photonic-Electronic Neural Chip for Hardware-Efficient Deep Learning," *ACS Photonics*, vol. 9, no. 12, pp. 3906–3916, 2022, doi: 10.1021/acsphotonics.2c01188.
[21] X. Meng *et al.*, "Compact optical convolution processing unit based on multimode interference," *Nature Communications*, vol. 14, no. 1, p. 3000, 2023, doi: 10.1038/s41467-023-38786-x.
[22] V. Bangari *et al.*, "Digital Electronics and Analog Photonics for Convolutional Neural Networks (DEAP-CNNs)," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, 2020, doi: 10.1109/JSTQE.2019.2945540.
[23] S. Xu, J. Wang, S. Yi, and W. Zou, "High-order tensor flow processing using integrated photonic circuits," *Nature Communications*, vol. 13, no. 1, p. 7970, 2022, doi: 10.1038/s41467-022-35723-2.
[24] J. Feldmann *et al.*, "Parallel convolutional processing using an integrated photonic tensor core," *Nature*, vol. 589, no. 7840, pp. 52–58, 2021, doi: 10.1038/s41586-020-03070-1.
[25] S. Xu, J. Wang, R. Wang, J. Chen, and W. Zou, "High-accuracy optical convolution unit architecture for convolutional neural networks by cascaded acousto-optical modulator arrays," *Optics Express*, vol. 27, no. 14, pp. 19778–19787, 2019, doi: 10.1364/oe.27.019778.
[26] F. Ashtiani, A. J. Geers, and F. Aflatouni, "An on-chip photonic deep neural network for image classification," *Nature*, vol. 606, no. 7914, pp. 501–506, 2022, doi: 10.1038/s41586-022-04714-0.
[27] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental Realization of Any Discrete Unitary Operator," *Physical Review Letters*, vol. 73, no. 20, pp. 58–61, 1994, doi: 10.1103/PhysRevLett.73.58.
[28] A. Ribeiro, A. Ruocco, L. Vanacker, and W. Bogaerts, "Demonstration of a 4 × 4-port universal linear circuit," *Optica*, vol. 3, no. 12, pp. 1348–1357, Dec. 2016, doi: 10.1364/OPTICA.3.001348.
[29] Y. Shen, N. C. Harris, D. Englund, and M. Soljacic, "Deep learning with coherent nanophotonic circuits," *Nature photonics*, pp. 441–446, 2017, doi: 10.1109/E3S.2017.8246190.
[30] Z. Zhao *et al.*, "Hardware-software Co-design of slimmed optical neural networks," in *Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC*, New York, NY, USA: ACM, Jan. 2019, pp. 705–710, doi: 10.1145/3287624.3287720.
[31] J. Gu, Z. Zhao, C. Feng, M. Liu, R. T. Chen, and D. Z. Pan, "Towards Area-Efficient Optical Neural Networks: An FFT-based Architecture," in *Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC*, 2020, pp. 476–481, doi: 10.1109/ASP-DAC47756.2020.9045156.
[32] A. Baldominos, Y. Saez, and P. Isasi, "A survey of handwritten character recognition with MNIST and EMNIST," *Applied Sciences*, vol. 9, no. 15, p. 3169, 2019, doi: 10.3390/app9153169.
[33] A. El Sawy, H. El-Bakry, and M. Loey, "CNN for handwritten arabic digits recognition based on LeNet-5," in *Advances in Intelligent Systems and Computing*, 2017, pp. 565–575, doi: 10.1007/978-3-319-48308-5_54.
[34] S. Namasudra, R. Chakraborty, S. Kadry, G. Manogaran, and B. S. Rawal, "FAST: Fast Accessing Scheme for data Transmission in cloud computing," *Peer-to-Peer Networking and Applications*, vol. 14, no. 4, pp. 2430–2442, 2021, doi: 10.1007/s12083-020-00959-6.
[35] T.-T. Le and D.-T. Le, "High FSR and Critical Coupling Control of Microring Resonator Based on Graphene-Silicon Multimode Waveguides," *Waveguide Technologies in Photonics and Microwave Engineering [Working Title]*, 2020, doi: 10.5772/intechopen.92210.
[36] P. R. Prucnal and Bhavin J. Shastri, "*Neuromorphic photonics*," CRC Press, 2017. doi: 10.1201/9781315373234.
[37] C. Ding *et al.*, "CIRCNN: Accelerating and compressing deep neural networks using block-circulant weight matrices," in *Proceedings of the Annual International Symposium on Microarchitecture, MICRO*, 2017, pp. 395–408, doi: 10.1145/3123939.3124552.
[38] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," Dec. 2018, doi: 10.1109/ICSCCC.2018.8703316.
[39] K. Auliasari, M. Wasef, and M. Kertaningtyas, "Leveraging VGG16 for fish classification in a large-scale dataset," *Brilliance*, vol. 3, no. 2, 2023, doi: 10.47709/brilliance.v3i2.3270
[40] A. Saini, K. Guleria, and S. Sharma, "A deep learning-based fine-tuned ResNet50 model for soybean seeds multiclass classification," in *Proceedings of the IEEE Global Conference on Information Technologies and Communication (GCITC)*, pp. 1–6, 2023, doi: 10.1109/gcitc60406.2023.10426079.
[41] K. Banaszewska and M. Plechawska-Wójcik, "Comparative analysis of CNN models for handwritten digit recognition," *Journal of Computer Sciences Institute*, vol. 32, pp. 179–185, 2024.
[42] F. P. Sunny, E. Taheri, M. Nikdast, and S. Pasricha, "A Survey on Silicon Photonics for Deep Learning," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 17, no. 4, pp. 1–57, 2021, doi: 10.1145/3459009.
[43] L. De Marinis, M. Cococcioni, P. Castoldi, and N. Andriolli, "Photonic Neural Networks: A Survey," *IEEE Access*, vol. 7, pp. 175827–175841, 2019, doi: 10.1109/ACCESS.2019.2957245.

BIOGRAPHIES OF AUTHORS

Thanh Tien Do has been awarded the M.A. degree from Swinburne University of Technology, Australia, since 2018. He is currently working toward a Ph.D. degree at the International School–Vietnam National University Hanoi. His research interests include IT architectures, e-government models, machine learning and optical communication. He can be contacted at email: thanhdo@vnuis.edu.vn.

Hai Yen Pham received the M.A. degree from the Military Technology Academy, Vietnam, in 2009. She is currently working toward a Ph.D. degree at the International School – Vietnam National University Hanoi. Her research interests include machine learning and optical communication. She can be contacted at email: yenph@vnu.edu.vn.

Trung Thanh Le is a professor and International School-VNU principal. He received a Ph.D. degree in electrical engineering from La Trobe, Australia, in 2009. With over 100 refereed publications, he made seminal contributions to photonic integrated circuits and quantum optics including laser and fiber-optical communication systems, all-optical computing and optical interconnects, sensors and wireless sensor networks, and integrated circuit systems. He can be contacted at email: thanh.le@vnu.edu.vn.