Optica Applicata, Vol. LIII, No. 3, 2023
DOI: 10.37190/oa230311

Noise quantization simulation analysis of optical convolutional networks

Ye Zhang[1, *], Saining Zhang[2], Danni Zhang[1, 3], Yanmei Su[3], Junkai Yi[1], Pengfei Wang[3, 4], Ruiting Wang[3, 4], Guangzhen Luo[3, 4], Xuliang Zhou[3, 4], Jiaoqing Pan[3, 4, **]

[1] School of Automation, Beijing Information Science and Technology University, Beijing, China

[2] School of Computer Science Technology, Beijing Institute of Technology, Beijing, China

[3] Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China

[4] College of Materials Science and Opto-Electronic Technology,
University of Chinese Academy of Sciences, Beijing, China

[*] Corresponding author: zhangyethu@163.com

[**] Corresponding author: jqpan@semi.ac.cn

Optical neural network (ONN) has been regarded as one of the most prospective techniques in the future, due to its high-speed and low power cost. However, the realization of optical convolutional neural network (CNN) in non-ideal cases still remains a big challenge. In this paper, we propose an optical convolutional networks system for classification problems by applying general matrix multiply (GEMM) technology. The results show that under the influence of noise, this system still has good performance with low TOP-1 and TOP-5 error rates of 44.26% and 14.51% for ImageNet. We also propose a quantization model of CNN. The noise quantization model reaches a sufficient prediction accuracy of about 96% for MNIST handwritten dataset.

Keywords: optical neural network, convolutional neural network, noise, quantization.

1. Introduction

Deep neural networks have already been used in a wide range of applications, from computer vision to natural language processing. Convolutional neural networks (CNNs) are good at using the spatial invariance of various image properties, which leads to the popular applications in solving computer vision problems [1-3]. As the performance of various tasks grows up to significant levels, the number of parameters and connections in these networks also increase dramatically, resulting in the demand of power reduction and storage capacity increment. In order to deal with more data with efficiency, many techniques which have been used to compress neural networks are no longer sufficient. Thus fast, low-power, and small chips are needed urgently. Specialized machine learning processing devices have been investigated, while other applications for embedded vision still focus on inference [4,5], trying to incorporate some image processing functions into the sensors to eliminate or reduce the hassle of transmitting complete image data to the processor. Due to the strict limitation of power and bandwidth, the configuration of CNNs for embedded systems such as mobile vision, vehicles, robotics, wireless smart sensors is still difficult [6,7]. Optical computing is receiving increasing attention due to its high bandwidth, high interconnectivity, and inherently parallel processing. Transmission and computation may be realized at the speed of light [8-10]. Retaining these advantages, optimized and scalable optical configurations for building optical CNN frameworks will be of research interest in the fields of computer vision, robotics, machine learning and optics.

The initial researches on optical neural networks (ONNs) focus on the performance of matrix multiplication for fully connected layers [11]. So far, most research on optical CNNs has focused on the design of convolutional layers based on optical components [12,13]. There are few reports on the application of all-optical CNNs. Therefore, more research is still needed to improve the optical CNNs. For real-life computer vision problems, processing large -scale data remains a big challenge for all-optical CNNs, since optical methods are mainly applied with fully connected structures, and their input scale is determined by hardware parallelism [14].

However, for ONNs in real cases, there would inevitably be some problems under non-ideal conditions that will affect their performance. In practice, the following two conditions which can cause errors could occur in optical devices: (1) Mach–Zehnder interferometer (MZI) devices could generate device-level noise [15-17]. Each MZI contains a configurable thermal-optical phase shifter to encode ONN weights. This phase shift can be affected by the device size, manufacturing defects, voltage control, and environmental changes, resulting in incorrect weight coding. Due to the cascaded architecture of ONN, phase errors caused by limited control resolution and phase shifter variations will propagate and accumulate throughout the system, ultimately resulting in reducing the whole inference accuracy. (2) The electronic control of optical devices has only limited resolution, and the phase shift generated by MZI cannot physically achieve arbitrary accuracy [14,18,19]. As a result, there would be weight encoding errors when mapping high-precision models to physical optical devices. We investigated the noise effect on the optical fully-connect neural network (FCNN) [20]. However, the noise issue in the optical convolutional networks remains a severe problem, which needs to be investigated clearly.

In this paper, an optical CNN system is designed based on an integrated ONN chip through the collaborative design of optics and algorithms, which integrates image acquisition and computation. This system helps to classify the input images. Our aim is to design a system with optimized optical convolution layers for specific classification problems, which demonstrates the capability of photoelectric CNNs. A noise quantization model for convolution operation is presented, and the effect of quantization on the accuracy of the optical CNN is analyzed.

2. Optical convolutional network model

The classic integrated ONN architecture utilizes MZI arrays for multi-layer perception (MLP) inference [21]. In our previous work, we designed an image classification recognition model based on a fully connected neural network (FCNN), and mapped it into a silicon-based integrated optical path [20]. The preliminary simulation experiments show that the ONN chip can classify handwritten digits quickly and accurately, with an accuracy rate of more than 97% [22]. However, the realization of CNN is different from FCNN. In order to realize convolution operations, we design an optical convolutional network based on the same ONN chip, where the detailed structure with five parts has been reported in our previous literatures [20,22].

In neural networks, complete connection layers and convolutional layers can be achieved through general matrix multiply (GEMM) [23]. The implementation of GEMM can achieve high-speed operations by making full use of the system's multi-level memory structure and program execution locality, which is a key function of basic linear algebra subprograms (BLA) [24]. The optical unit performs the matrix-vector multiplication, and executes the optical GEMM in parallel with multiple units with the same set of weights. In addition to fully connected layers, convolutional layers can be implemented on optical GEMM units by employing "Patching" technology [25]. Figure 1 shows a schematic diagram of the convolution operation based on the optical GEMM unit.

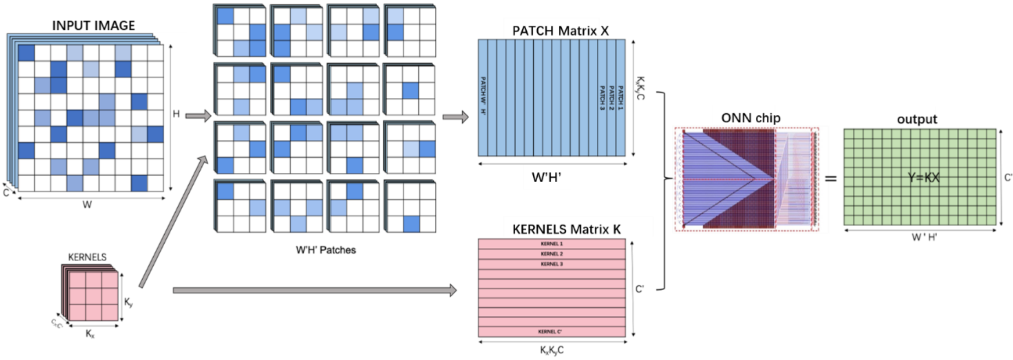

Fig. 1. Schematic diagram of optical GEMM implementing convolution operation.

As shown in Fig. 1, patching technique reconstructs convolution into matrix-matrix multiplication. In the convolutional layer shown in Fig. 1, input $x_{ij;k}$ is an image with dimension $W \times H$ and C channels. The convolution operation output $y_{ij;k}$ is with the dimension of $W' \times H'$, and the number of channels is C'. The relationship between input and output is expressed as

$$y_{ij;k} = \sum_{i'j',\,l} K_{i'j',\,kl}\, x_{(s_x i + i')(s_y j + j');l} \tag{1}$$

where $K_{i'j',kl}$ is the convolution filter, which is a four-bit tensor of size $K_x \times K_y \times C \times C'$, and s_x, s_y is the step size of the convolution. In the convolutional layer shown in Fig. 1, $K_x = K_y = 3$ and $s_x = s_y = 2$. Patching stretches the image into a matrix X of dimensions $K_xK_yC \times W'H'$, with each column corresponding to the vector $K_x \times K_y$ of the image. Rearrange the elements of the filter to form a dense matrix K of size $C' \times K_xK_yC$. Then the formula can be calculated by matrix multiplication of $Y = KX$. It can be concluded that the size of matrix Y is $C' \times W'H'$. In almost any microprocessor, GEMM is a highly optimized function with a very regular memory access pattern. The benefit of rewriting convolution to GEMM is to highlight the redundancy of data storage generated from overlapping patching [26,27]. The time required to rearrange images into a patch matrix is usually very small compared to that required to calculate GEMM. Therefore, by accelerating GEMM, the optical matrix multiplier will significantly improve the speed and energy efficiency of the convolutional layers. Since this algorithm performs convention operation in a matrix-matrix operation, energy cost can also be reduced even if neural networks are not run on large batches of data. Due to the limitation of optical devices, convolution operation can be converted into matrix multiplication through optical GEMM, which is convenient for calculation. Through optical GEMM, CNN can be well mapped into the chip structure.

Figure 2 shows the schematic diagram of the photoelectric system. The signal is input through the input port and output through the balance detector after multiplication and addition operations through the ONN chip. Shot noise will be generated in the balance detector. Through AD/DA conversion, the balance detector can be connected to the circuit. A computer is used to process the nonlinear data. Data quantization would be needed in this process. Both analog noise and quantization will be discussed in detail in the following sections. According to the previous research, it can be concluded that this chip is scalable. Therefore, for large-scale data such as ImageNet, we can arrange

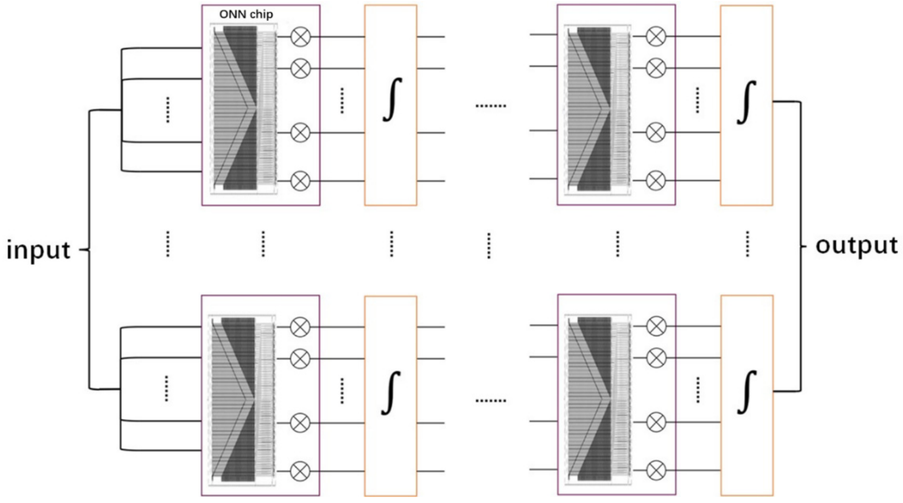

Fig. 2. Photoelectric system diagram.

the chips orderly. Final output can be obtained following the timing input and information processing.

The power consumption of the ONN chip mainly comes from the modulators. The electro-optic phase modulator used in the ONN chip has a low power consumption of less than 1 mW, with the speed of 100 MHz. The chip is made by CMOS technology, which can be mass produced.

3. Noise quantization simulation analysis

3.1. Noise

In a neural network, the output x_{i+1} of a specific layer can be obtained through multiplying and accumulating (MAC) operations of input x_i and weighted signal A_{ij}. The expression is shown as follows:

$$x_{i+1} = f\left(\sum\nolimits_j A_{ij} x_i\right) \tag{2}$$

The ONN chip could operate MAC calculations optically with weighted signal and input, which has been encoded as pulses. The output current follows Poisson distribution $Q/\mathrm{e} \sim \mathrm{Poisson}(|u|^2)$, which will lead to the Gaussian random variable:

$$\frac{Q_i^{(\pm)}}{\mathrm{e}} = \sum\nolimits_j \frac{1}{2}(\overline{A_{ij}} \pm \overline{x_{ij}})^2 + w_i^{(\pm)}\left[\sum\nolimits_j \frac{1}{2}(\overline{A_{ij}} \pm \overline{x_{ij}})^2\right]^{1/2} \tag{3}$$

where $w_i(k) \sim N(0, 1)$ are Gaussian random variables.

Thus the output x_{i+1} can be described as [28]

$$x_{i+1} = f\left(\sum\nolimits_j A_{ij} x_i + w_i \frac{\|A\|\,\|x\|}{\sqrt{N^2 N'}} \frac{\sqrt{N}}{\sqrt{n_{\mathrm{MAC}}}}\right) \tag{4}$$

where $\|\cdot\|$ is the 2-norm, n_{MAC} is the number of photons per MAC, N is the number of input neurons and N' is the number of output neurons.

We adopt the optical GEMM in the simulation, where the optical unit performs the matrix-vector multiplication and runs multiple elements in parallel with the same set of weights to implement the general GEMM, which is a key function in the basic linear algebra subprograms. A typical convolution neural network which is competitive in the large-scale visual recognition challenge of ImageNet is AlexNet [1]. Thus we consider to use this typical network as a benchmark issue. AlexNet is an 8-layer structure, of which the first 5 layers are convolutional layers and the last 3 layers are fully connected layers. Figure 3 shows a schematic diagram of AlexNet network structure, and parameters of the network are shown in the Table.

To be more realistic, we add noise in Eq. (4) to the network. Two error rates, Top-1 and Top-5, are usually reported in ImageNet. In prediction, the meaning of Top-1 error

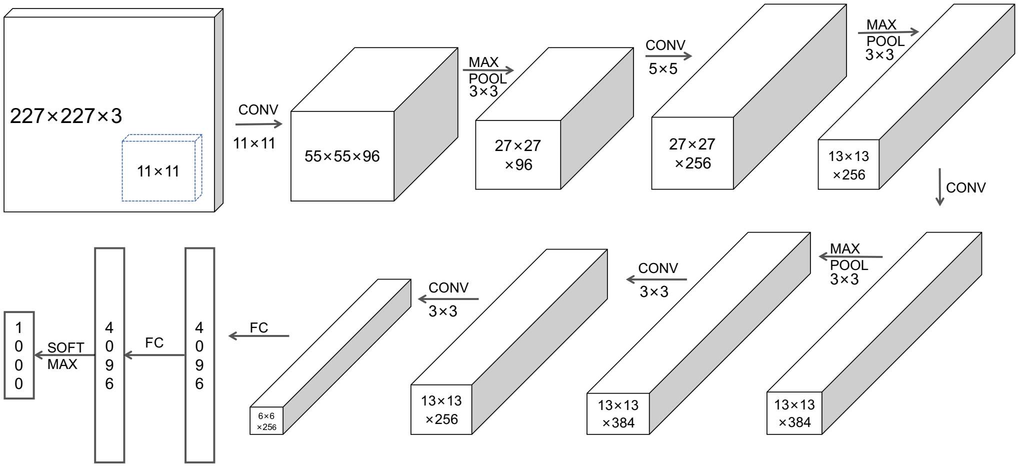

Fig. 3. AlexNet network structure diagram.

Table. Parameters of AlexNet.

Layer	Output	Kernel	Stride	MACs
Conv1	55 × 55 × 96	11 × 11	4	105M
Pool	27 × 27 × 96	–	2	–
Conv2	27 × 27 × 256	5 × 5	1	448M
Pool	13 × 13 × 256	–	2	–
Conv3	13 × 13 × 384	3 × 3	1	150M
Conv4	13 × 13 × 384	3 × 3	1	224M
Conv5	13 × 13 × 256	3 × 3	1	150M
Pool	6 × 6 × 256	–	2	–
FC1	4096	–	–	38M
FC2	4096	–	–	16M
FC3	1000	–	–	4M

rate is to check whether the class with the highest probability is the same as the target label. And the Top-5 error rate is the proportion of the test image where the correct tag is not among the five tags that the model considers to be the most likely. In both cases, the highest score is obtained by dividing the number of matches between the predicted and target labels by the number of evaluated data points. The experimental results are shown in Fig. 4.

As we can see from Fig. 4, both Top-1 and Top-5 error rates decrease as the photons/(number of MACs) increase. When the photons/(number of MACs) are large enough, the effect of noise can be ignored. In this experiment, when the photons/(number of MACs) are large enough, the error rates of Top-1 and Top-5 are 44.26% and

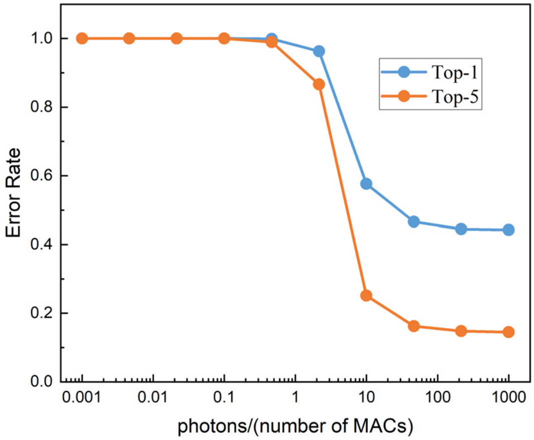

Fig. 4. Error rate of Top-1 and Top-5.

14.51%, respectively. Typical Top-1 and Top-5 error rates were reported to be 37.5% and 17.0% [1]. Compared with the classical data, we achieve comparable Top-1 error rate and lower Top-5 error rate.

3.2. Quantization

Since the electronic control of the optical devices has only limited resolution, the optical neural unit cannot physically achieve arbitrary precision. When a high-precision model is mapped to a physical optical device, a weight encoding error could occur. Due to AD/DA conversion, floating-point numbers (FPs) should be quantized into fixed-point numbers (INT) in the model. This results in smaller models and improved inference speed. Extensive work has shown that more efficient deep neural networks can be achieved through low-bit parameter quantization [29,30]. However, quantization errors after training will lead to performance degradation. The results of some experiments using low-precision numerical representations seem to indicate that the experiment requires precision higher than eight bits to deal with backward propagation and gradients [31]. This will make the implementation of the training more complicated. Therefore, after training the model, it is reasonable to use the quantified weights for inference.

The basic components of CONV and FC layers are MAC operations, which can be easily parallelized. In order to achieve high performance, highly parallelized computing paradigm, including time and space architecture, is widely used. In the above experiment, we found that when the MAC reaches a certain size, the noise has little effect on accuracy. We will quantize the AlexNet model mentioned above and add noise for analysis to form a new noise quantization model. The model before quantification is shown in Fig. 3, and the structure diagram after quantification is shown in Fig. 5.

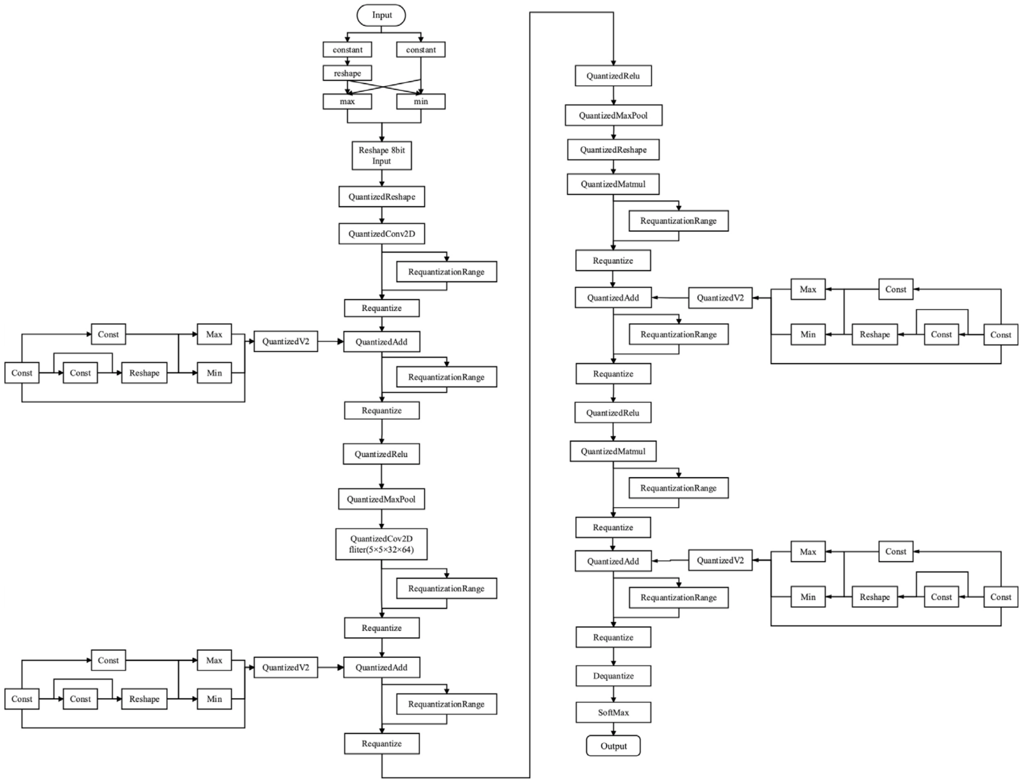

Fig. 5. Flow chart after quantization of noise model.

Accuracy is the most common evaluation criteria for classification problems, which can be described as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} \tag{5}$$

where TP is true positives, TN is true negatives, P and N are positives and negatives, respectively. (TP + TN) presents all cases that have been correctly recognized, while (P + N) presents all cases that have been obtained in the dataset.

The accuracy of the quantified model is analyzed. In the case of FP32, AlexNet is used to predict the accuracy of the same data set. Figure 6 shows the accuracy of using AlexNet to predict MNIST dataset under FP32. It can be seen that the accuracy increases as the step increases. The highest accuracy of the trained network is 96.81%. While the accuracy rate obtained by reasoning after quantifying the model is 96.68%, the classification accuracy drop could achieve as low as 0.13%. The slight loss of accuracy after quantization compared with that before quantization is too small to be neglected. It can be concluded that this quantization model is effective.

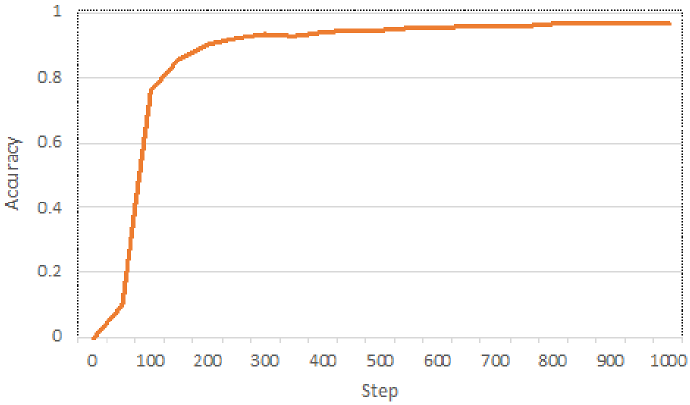

Fig. 6. Accuracy *vs*. step.

4. Conclusion

In this paper, in combination with the GEMM algorithm, we have proposed an optical convolutional networks system optimized for specific classification problems, which enables ONN to classify images. Experimental results on ImageNet dataset show that the system has good expression ability, where the Top-1 and Top-5 error rates are as low as 44.26% and 14.51%, respectively. In addition, we have also proposed a quantization model of CNN for image classification and recognition to adapt ONN to non-ideal environments with low-bit transmission. Through the comparison before and after quantization, the optimized quantization model in this paper is effective with sufficient prediction accuracy which can reach about 96% for MNIST handwritten dataset. Experimental results show that the proposed quantification method can effectively solve the non-ideal ONN problem. We believe that the noise quantization model established in this paper could provide theoretical guide to optical neural network chips in the near future.

Acknowledgements
This work is supported by the Beijing Natural Science Foundation (Z200006), the National Natural Science Foundation of China (62090053, 61934007, 61974141), the Beijing Municipal Science and Technology Project (Z191100004819011), the National Key R&D Program of China (2018YFE0203103), and the Supplementary and Supportive Project for Teachers at the Beijing Information Science and Technology University (5029011103).

References

[1] Krizhevsky A., Sutskever I., Hinton G.E., *ImageNet classification with deep convolutional neural networks*, Communications of the ACM **60**(6), 2017: 84-90. https://doi.org/10.1145/3065386

[2] Acharya U.R., Oh S.L., Hagiwara Y., Tan J.H., Adam M., Gertych A., Tan R.S., *A deep convolutional neural network model to classify heartbeats*, Computers in Biology and Medicine **89**, 2017: 389-396. https://doi.org/10.1016/j.compbiomed.2017.08.022

[3] Shelhamer E., Long J., Darrell T., *Fully convolutional networks for semantic segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4), 2017: 640-651. https://doi.org/10.1109/TPAMI.2016.2572683

[4] Sokolov A.S., Abbas H., Abbas Y., Choi C., *Towards engineering in memristors for emerging memory and neuromorphic computing: A review*, Journal of Semiconductors **42**(1), 2021: 013101. https://doi.org/10.1088/1674-4926/42/1/013101

[5] Park S., Noh J., Choo M.-I., Sheri A. M., Chang M., Kim Y.-B., Kim C. J., Jeon M., Lee B.-G., Lee B.H., Hwang H., *Nanoscale RRAM-based synaptic electronics: toward a neuromorphic computing device*, Nanotechnology **24**(38), 2013: 384009. https://doi.org/10.1088/0957-4484/24/38/384009

[6] Yao P., Wu H., Gao B., Tang J., Zhang Q., Zhang W., Yang J.J., Qian H., *Fully hardware-implemented memristor convolutional neural network*, Nature **577**(7792), 2020: 641-646. https://doi.org/10.1038/s41586-020-1942-4

[7] Park S., Hong I., Park J., Yoo H.-J., *An energy-efficient embedded deep neural network processor for high speed visual attention in mobile vision recognition SoC*, IEEE Journal of Solid-State Circuits **51**(10), 2016: 2380-2388. https://doi.org/10.1109/JSSC.2016.2582864

[8] Larger L., Baylón-Fuentes A., Martinenghi R., Udaltsov V. S., Chembo Y.K., Jacquot M., *High -speed photonic reservoir computing using a time-delay-based architecture: million words per second classification*, Physical Review X **7**(1), 2017: 011015. https://doi.org/10.1103/PhysRevX.7.011015

[9] Peng H.-T., Nahmias M.A., de Lima T.F., Tait A.N., Shastri B.J., *Neuromorphic photonic integrated circuits*, IEEE Journal of Selected Topics in Quantum Electronics **24**(6), 2018: 6101715. https://doi.org/10.1109/JSTQE.2018.2840448

[10] 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 **361**(6406), 2018: 1004-1008. https://doi.org/10.1126/science.aat8084

[11] Shen Y., Harris N. C., Skirlo S., Prabhu M., Baehr-Jones T., Hochberg M., Sun X., Zhao S., Larochelle H., Englund D., Soljačić M., *Deep learning with coherent nanophotonic circuits*, Nature Photonics **11**(7), 2017: 441-446. https://doi.org/10.1038/nphoton.2017.93

[12] Xu X., Tan M., Corcoran B., Wu J., Boes A., Nguyen T.G., Chu S.T., Little B.E., Hicks D.G., Morandotti R., Mitchell A., Moss D.J., *11 TOPS photonic convolutional accelerator for optical neural networks*, Nature **589**(7840), 2021: 44-51. https://doi.org/10.1038/s41586-020-03063-0

[13] Chang J., Sitzmann V., Dun X., Heidrich W., Wetzstein G., *Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification*, Scientific Reports **8**(1), 2018: 12324. https://doi.org/10.1038/s41598-018-30619-y

[14] Gu J., Zhao Z., Feng C., Zhu H., Chen R.T., Pan D.Z., *ROQ: A noise-aware quantization scheme towards robust optical neural networks with low-bit controls*, [In] *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Grenoble, France, 2020: 1586-1589. https://doi.org/10.23919/DATE48585.2020.9116521

[15] 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 Journal of Selected Topics in Quantum Electronics **26**(1), 2020: 7700412. https://doi.org/10.1109/JSTQE.2019.2930455

[16] Fang M.Y.-S., Manipatruni S., Wierzynski C., Khosrowshahi A., DeWeese M.R., *Design of optical neural networks with component imprecisions*, Optics Express **27**(10), 2019: 14009-14029. https://doi.org/10.1364/OE.27.014009

[17] Slussarenko S., Weston M.M., Chrzanowski H.M., Shalm L.K., Verma V.B., Nam S.W., Pryde G.J., *Unconditional violation of the shot-noise limit in photonic quantum metrology*, Nature Photonics **11**(11), 2017: 700-703. https://doi.org/10.1038/s41566-017-0011-5

[18] Harris N.C., Ma Y., Mower J., Baehr-Jones T., Englund D., Hochberg M., Galland C., *Efficient, compact and low loss thermo-optic phase shifter in silicon*, Optics Express **22**(9), 2014: 10487-10493. https://doi.org/10.1364/OE.22.010487

[19] Tait A.N., Nahmias M.A., Shastri B.J., Prucnal P.R., *Broadcast and weight: An integrated network for scalable photonic spike processing*, Journal of Lightwave Technology **32**(21), 2014: 3427-3439.

[20] Zhang D., Zhang Y., Zhang Y., Su Y., Yi J., Wang P., Wang R., Luo G., Zhou X., Pan J., *Training and inference of optical neural networks with noise and low-bits control*, Applied Sciences **11**(8), 2021: 3692. https://doi.org/10.3390/app11083692

[21] Kim J.-Y., Kang J.-M., Kim T.-Y., Han S.-K., *10 Gbit/s all-optical composite logic gates with XOR, NOR, OR and NAND functions using SOA-MZI structures*, Electronics Letters **42**(5), 2006: 303-304. https://doi.org/10.1049/el:20063501

[22] Zhang D., Wang P., Luo G., Bi Y., Zhang Y., Yi J., Su Y., Zhang Y., Pan J., *Design of a silicon -based optical neural network*, Proceedings of the 2019 2nd International Conference on Mathematics, Modeling and Simulation Technologies and Applications (MMSTA 2019), Advances in Computer Science Research, Vol. 93, 2019: 184-186.

[23] Springer P., Bientinesi P., *Design of a high-performance GEMM-like tensor–tensor multiplication*, ACM Transactions on Mathematical Software **44**(3), 2018: 28. https://doi.org/10.1145/3157733

[24] Lawson C.L., Hanson R.J., Kincaid D.R., Krogh F.T., *Basic linear algebra subprograms for Fortran usage*, ACM Transactions on Mathematical Software **5**(3), 1979: 308-323. https://doi.org/10.1145/355841.355847

[25] Vasudevan A., Anderson A., Gregg D., *Parallel multi channel convolution using general matrix multiplication*, [In] *2017 IEEE 28th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, Seattle, WA, USA, 2017: 19-24. https://doi.org/10.1109/ASAP.2017.7995254

[26] Kurzak J., Tomov S., Dongarra J., *Autotuning GEMM kernels for the Fermi GPU*, IEEE Transactions on Parallel and Distributed Systems **23**(11), 2012: 2045-2057. https://doi.org/10.1109/TPDS.2011.311

[27] Barrachina S., Dolz M.F., San Juan P., Quintana-Ortí E.S., *Efficient and portable GEMM-based convolution operators for deep neural network training on multicore processors*, Journal of Parallel and Distributed Computing **167**, 2022: 240-254. https://doi.org/10.1016/j.jpdc.2022.05.009

[28] Hamerly R., Bernstein L., Sludds A., Soljačić M., Englund D., *Large-scale optical neural networks based on photoelectric multiplication*, Physical Review X **9**(2), 2019: 021032. https://doi.org/10.1103/PhysRevX.9.021032

[29] Moren K., Göhringer D., *A framework for accelerating local feature extraction with OpenCL on multi-core CPUs and co-processors*, Journal of Real-Time Image Processing **16**(4), 2019: 901-918. https://doi.org/10.1007/s11554-016-0576-0

[30] Wu J., Leng C., Wang Y., Hu Q., Cheng J., *Quantized convolutional neural networks for mobile devices*, [In] *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016: 4820-4828. https://doi.org/10.1109/CVPR.2016.521

[31] Gupta S., Agrawal A., Gopalakrishnan K., Narayanan P., *Deep learning with limited numerical precision*, Proceedings of the 32nd International Conference on Machine Learning, 2015: 1737 -1746.

Received January 16, 2023
in revised form March 1, 2023