



# High-accuracy optical convolution unit architecture for convolutional neural networks by cascaded acousto-optical modulator arrays

SHAOFU XU, JING WANG, RUI WANG, JIANGPING CHEN, AND WEIWEN ZOU\*

State Key Laboratory of Advanced Optical Communication Systems and Networks, Intelligent Microwave Lightwave Integration Innovation Center (iMLic), Department of Electronic Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

\*wzou@sjtu.edu.cn

**Abstract:** Optical neural networks (ONNs) have become competitive candidates for the next generation of high-performance neural network accelerators because of their low power consumption and high-speed nature. Beyond fully-connected neural networks demonstrated in pioneer works, optical computing hardware can also conduct convolutional neural networks (CNNs) by hardware reusing. Following this concept, we propose an optical convolution unit (OCU) architecture. By reusing the OCU architecture with different inputs and weights, convolutions with arbitrary input sizes can be done. A proof-of-concept experiment is carried out by cascaded acousto-optical modulator arrays. When the neural network parameters are *ex-situ* trained, the OCU conducts convolutions with SDR up to 28.22 dBc and performs well on inferences of typical CNN tasks. Furthermore, we conduct *in situ* training and get higher SDR at 36.27 dBc, verifying the OCU could be further refined by *in situ* training. Besides the effectiveness and high accuracy, the simplified OCU architecture served as a building block could be easily duplicated and integrated to future chip-scale optical CNNs.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

With the development of machine learning technologies since recent years, deep neural networks exhibit revolutionary performance enhancement in various emerging applications [1]. Particularly, deep convolutional neural networks (CNNs) have made a profound impact in fields like computer vision [2,3], image processing [4–6], speech processing [7,8], medical diagnosis [9], games [10,11], and signal processing [12], becoming the cornerstone of modern artificial intelligence. In spite of the advanced performances introduced by deep neural networks, their complicated architectures and lots of parameters consume massive computing resources at training and inference procedures. Therefore, neural network accelerators with high-speed and low power consumption are of urgent requirement.

Optical methods are potential for the next generation of neural network accelerators since optical components and technologies have appealing features of ultra-broad bandwidth and low power consumption [13,14]. Optical technologies including spatial light diffraction [15–18], on-chip coherent interference [19], wavelength division multiplexing [20,21] were utilized to demonstrate the feasibility of optical neural networks (ONNs). And the high-speed and low-power performances are convincingly inferred from the numerical and experimental results. In these pioneer works on ONNs, fully-connected neural networks are majorly considered and thus these architectures are designed to be vector-matrix multipliers. When it comes to convolutional neural networks (CNNs), these architectures could face heavy challenges because an immense optical circuit is necessary to transform convolutional layers to vector-matrix multiplications. The number of embedded parameters of that optical circuits is at the scale of  $N^4$  if the size of input image is  $N \times N$ . A viable way to conquer this hindrance

is to transform convolutional layers to matrix-matrix multiplications by reusing optical hardwares. Consequently, the number of embedded parameters is significant reduced (to around several tens), and the full calculations are done within  $N^2$  time cycles [22].

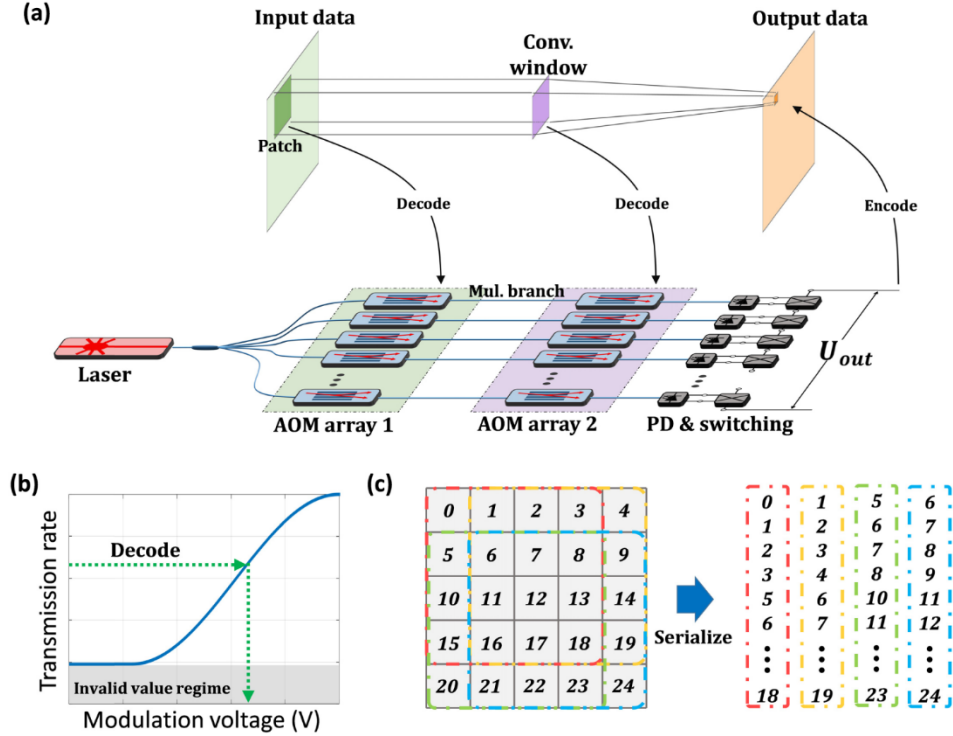


Fig. 1. (a) Schematic of proposed OCU architecture. The OCU majorly comprise 2-layer AOM arrays. Values of an input patch is decoded to the modulation voltages to the AOM array 1, and values of convolution window (Conv. window) are decoded to modulate the AOM array 2. Two cascaded AOMs form a multiplier branch (Mul. branch) to execute optical power multiplication. The optical power is provided by a laser and equally split to the multiplier branches. After PDs transforming optical powers to voltages, the switching array is controlled to give a positive or a negative copy of the voltages. Output voltage  $U_{out}$  is the sum of all voltage copies. Output voltages are encoded to grey scale values to get the output data. (b) Decoding method based on the modulation curve of AOM. A non-negative value is represented by the transmission rate of modulators, so it can be mapped to a modulation voltage. If the extinction ratio of the modulator is low, the invalid value regime could be large, influencing the accuracy of the OCU. (c) An example of serialization method. The numbers are notations of pixels rather than values of pixels. The size of input 2-dimensional image is  $5 \times 5$  and the convolution window is  $2 \times 2$ . Therefore, the number of multiplier branches is 4 and the input image is serialized to 4 input sequences.

Following the hardware reusing concept, here we propose an optical convolution unit (OCU) architecture, which can be reused to execute all the convolutions in arbitrarily complicated CNNs in a single unit. Rather than a matrix multiplier, the proposed architecture is designed to conduct dot-product operations, and it thus mitigates the hardware complexity significantly. Since a matrix multiplication can be equivalently realized by multiple dot-product operations, the OCU can be reused to fulfill the same functionalities of matrix multipliers with released controlling difficulty. In the proof-of-concept experiment, the OCU is implemented with cascaded acousto-optical modulator (AOM) arrays, and reused by simply changing the modulation voltages to the AOMs. Effectiveness of the proposed architecture on typical CNN tasks are demonstrated. Furthermore, we conduct *in situ* training on the

experimental setup, verifying the proposed OCU architecture could be further refined by *in situ* training.

## 2. Architecture of optical convolution unit

As illustrated in Fig. 1(a), the implemented OCU is mainly composed with two cascaded acousto-optical modulator (AOM) arrays, where AOMs are paralleled to form several multiplier branches. In each branch, two cascaded AOMs work as an optical power multiplier. A patch of the input data (i.e., input patch) is used to modulate the AOM array 1 after decoding and the values of convolution window are decoded to the AOM array 2. Besides the AOM arrays, a laser provides optical power and the optical coupler divides the optical power equally into multiplier branches. Photo-detectors (PDs) transform the optical power to electrical signal (voltage) proportionally and the switching array decides whether the voltages are added up positively or negatively.

Equation (1) describes a dot-product operation of a single input patch and the convolution window within the OCU. Note that the input patch can move on the input data, so a flow of the dot-product results constitute the convolution output.

$$y = (x_0, x_1, x_2, \dots, x_{W-1})(w_0, w_1, w_2, \dots, w_{W-1})^T = \sum_{k=0}^{W-1} \frac{\eta P}{W} \text{sign}(w_k) T(x_k) T(|w_k|). \quad (1)$$

The optical power of the laser is assumed to be  $P$ . The  $k$ -th value of inputs,  $x_k$  and  $w_k$ , are multiplied in the  $k$ -th multiplier branch after being decoded to the AOM's transmission rates,  $T(x_k)$  and  $T(|w_k|)$ . The sign of  $w_k$ ,  $\text{sign}(w_k)$ , is maintained by the switches. PDs transform the optical powers to voltages with a photo-electronic efficiency of  $\eta$ .  $W$  represents the size of convolution window. The maximal transmission rate of AOMs represents 1 and minimal transmission rate represents 0. Therefore, if the cascaded AOMs are modulated properly with values from 0 to 1, the output optical powers of the cascaded AOM array represent the multiplied results. In order to control the transmission rates of AOMs with corresponding values, the input data and convolution window are decoded to modulation voltages based on the modulation curve of AOMs (shown in Fig. 1(b)). Typically, the values of input data are non-negative, so the positive transmission rate is adequate to represent them. However, the values of convolution windows are real numbers; therefore, the absolute value of convolution windows are presented by the transmission rate of AOMs and the sign of them are maintained using switches. If a window value is positive, the switch is controlled to give a positive copy of PD voltage output; if not, a negative voltage is given. Consequently, the signs of convolution window values are maintained when all voltages are added up. During image convolution, the input patch moves on the input data but convolution window stays unchanged. We can change the modulation voltages to AOM array 1 to move the input patch over the whole input data.

A serialization method is used to generate sequences of modulation voltages to AOM array 1. Suppose the input data is a 2-dimensional image ( $M \times N$ ) and the size of convolution window is  $W = \sigma \times \sigma$ , the serialization method is described by:

$$x_k(n) = \text{Image} \left( \left\lfloor \frac{k}{\sigma-1} \right\rfloor + \left\lfloor \frac{n}{N-\sigma+1} \right\rfloor, (k \bmod (\sigma-1)) + (n \bmod (N-\sigma+1)) \right), \quad (2)$$

where the input data is a sequence  $x_k(n)$  rather a single value  $x_k$  in Eq. (1).  $\text{Image}(i, j)$  is the pixel value at the location of  $(i, j)$ .  $n = 0, 1, 2, 3, \dots, (M-\sigma+1) \times (N-\sigma+1)$ . A simple example of the serialization method is illustrated in Fig. 1(c). The size of input image is  $5 \times 5$  and the size of convolution window is  $2 \times 2$ , so the size of input patch is  $2 \times 2$  and the number of multiplier branches is 4. Therefore, the input image is serialized to 4 input sequences by Eq. (2).

Since the proposed OCU architecture executes convolutions in analog regime, the extinction ratio between the maximal and minimal transmission rates of modulators turns to be critical for the computing accuracy (see Fig. 1(b)). If the extinction ratio is low, the invalid-value regime is large. Consequently, values cannot be decoded accurately to the modulation voltages, introducing essential distortions to the convolution results. To characterize the achievable accuracy of the OCU architecture, AOMs with extinction ratio up to 50 dB are adopted to implement proof-of-concept experiments.

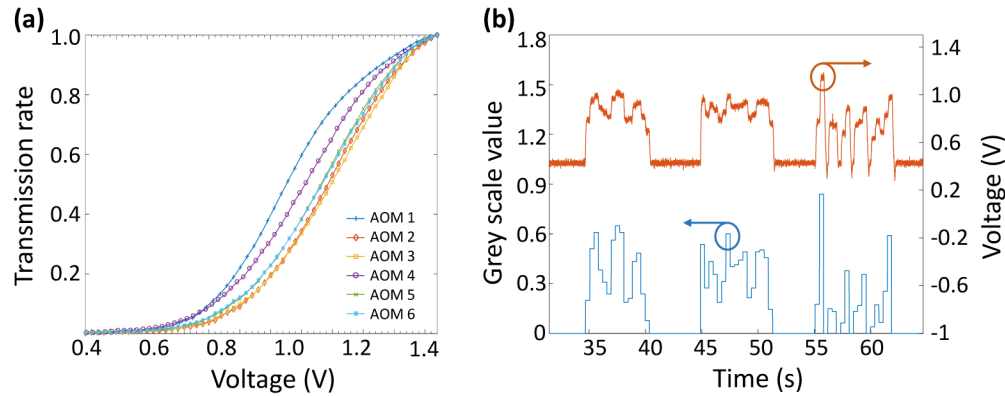


Fig. 2. (a) Measured modulation curves of the adopted AOMs. (b) An example of input sequence (blue curve) and its corresponding modulation voltage (orange curve). The original input image is a “Shirt” in Fashion-MNIST. The image is transformed to an input sequence by serialization method and is decoded to the modulation voltage by the measured modulation curves.

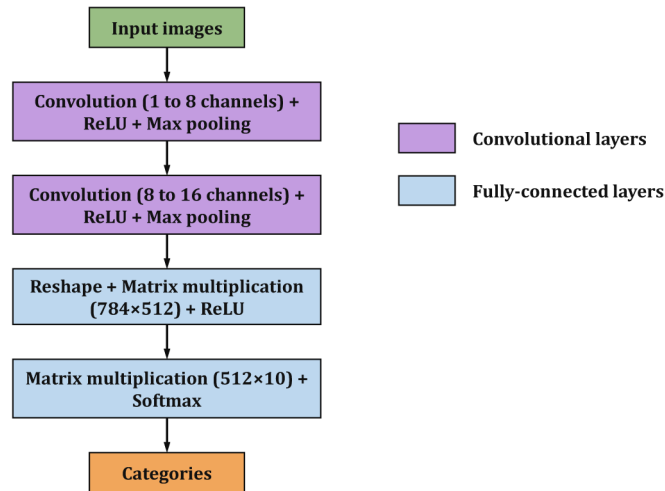


Fig. 3. The CNN model adopted in this work. This model comprises 2 convolutional layers and 2 fully-connected layers. The applied non-linear activation methods and pooling methods are described in the figure. The OCU executes the convolutions in the convolutional layers and other operations are conducted by a computer in the proof-of-concept experiment.

### 3. Experimental demonstration

In the proof-of-concept experiments, we verify the feasibility of the proposed OCU and demonstrate its high accuracy with two CNN classification tasks, that is, MNIST handwritten number classification [23] and Fashion-MNIST attire classification [24]. The size of the convolution windows for demonstration is set to  $3 \times 3$ , so the OCU should comprise 9

multiplier branches. Owing to each multiplier branch works independently, the  $3 \times 3$  convolution can be divided to three  $1 \times 3$  convolutions as follows:

$$\sum_{i=1}^9 x_i \cdot w_i = \sum_{i=1}^3 x_i \cdot w_i + \sum_{i=4}^6 x_i \cdot w_i + \sum_{i=7}^9 x_i \cdot w_i. \quad (3)$$

Therefore, the  $3 \times 3$  convolution can be executed by reusing the  $1 \times 3$  OCU for 3 times and the number of multiplier branches are lessened to 3.

In the experimental setup, a continuous-wave laser diode (Alnair Labs TLG-200) is adopted to serve as the stable optical source for 3 multiplier branches. The measured modulating curves of the adopted AOMs (CETC SGTF100-1550) are illustrated in Fig. 2(a). With these modulation curves, the input data and the convolution window data are decoded to modulation voltages to AOMs. These modulation voltages are generated by two programmable voltage sources (Keithley 2230G-30-1) and loaded on the AOM arrays. Figure 2(b) shows a segment of the generated voltage and the corresponding input sequence. PDs (LightSensing Technology LSIPD-A75) transform optical power to voltages and the PD voltages are added up by the switching array. Finally, the output voltage is recorded by an oscilloscope (Keysight DSO-S 804A) and is encoded to grey scale values.

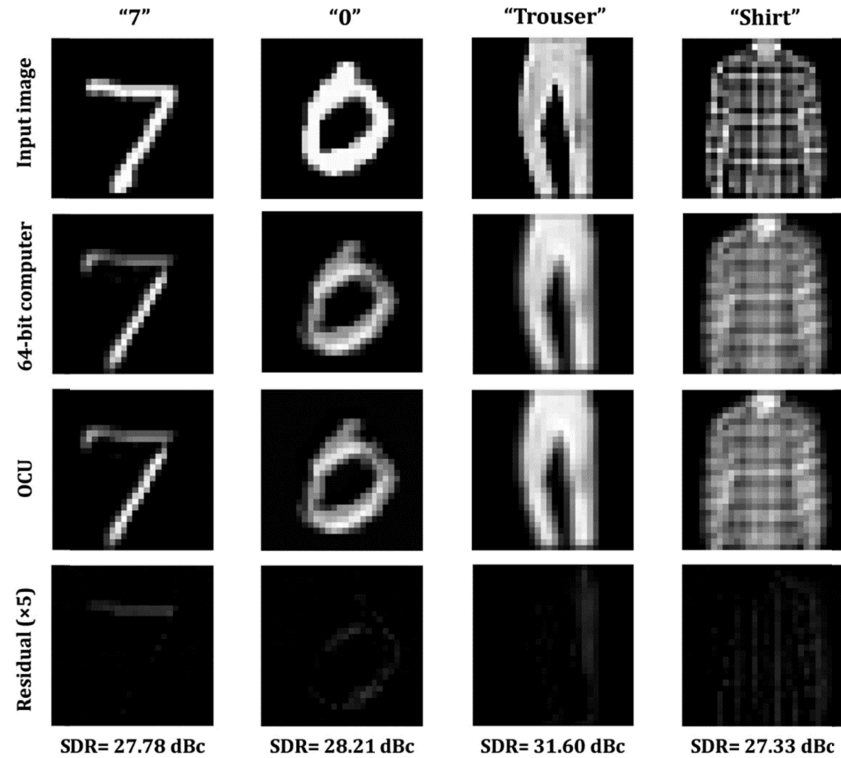


Fig. 4. OCU convolution results of MNIST-handwritten numbers and Fashion-MNIST data sets. The input image is shown in the first row. Through the same convolution window, the results of a 64-bit digital computer and the OCU are depicted in the second and third rows, respectively. Taken the digital computer results as reference, the residual errors of the OCU results are calculated. For better visibility, the residual errors are amplified by 5 times in the fourth row.

As shown in Fig. 3, a classical CNN model comprising two convolutional layers and two fully-connected layers is adopted to finish two classification tasks of MNIST-handwritten numbers and Fashion-MNIST. In the first part of the experiment, The CNN model is *ex-situ* trained and parameters are saved in a 64-bit digital computer. The OCU is used for the

convolutions in inference and the other neural network operations of bias, ReLU activation, max pooling, and matrix multiplications are carried out in the computer.

Figure 4 illustrates some convolution examples calculated by the OCU and the 64-bit digital computer, respectively. The input images are illustrated in the first row. After the same convolution window, the OCU and digital computer yield similar results. Taking the computer results as reference, we can give the residual calculation errors of the OCU. For a better visibility of the residual errors, their values are amplified by 5 times. It can be seen that residual errors of the OCU concentrates on the bright part of the images, meaning that the errors are mainly caused by the system distortions rather than noise. Therefore, we can characterize the accuracy performance of the OCU by the signal-to-distortion ratio (SDR). By averaging the residual errors within 100 image convolutions, the SDR of the OCU is characterized to be 28.22 dBc. To further characterize the prediction accuracy of the OCU in CNN tasks, we simulate the OCU to carry out MNIST-handwritten-number and Fashion-MNIST classifications. By comparing the ideal output and the OCU output in the experiment, we can construct a mapping between ideal results and OCU-distorted results, which is shown in Fig. 5. Using this mapping, ideal convolution results can be transformed to OCU-distorted ones. Altering all ideal convolutions with distorted ones, we can simulate the OCU-distorted CNN and characterize its performances in classification tasks.

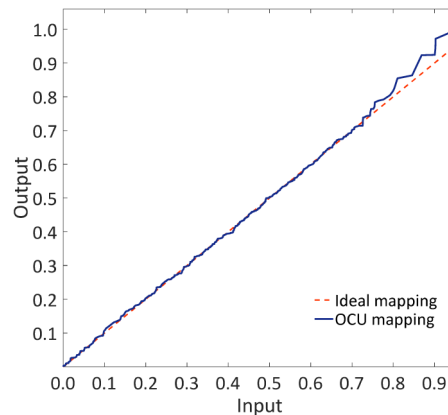


Fig. 5. The distortion mapping of the OCU. By comparing the ideal convolution results (digital computer) and the OCU convolution results, the distortion effect of the OCU can be represented by a mapping (the blue curve). The ideal mapping (red line,  $y = x$ ) is also provided for reference.

Figure 6 gives the prediction distributions of ideal CNN and OCU-distorted CNN. Inputting an image with an original label to the CNN, a predicted label is given. The prediction accuracy is calculated over 1000 samples in the test data sets. Correct predictions concentrate on the diagonal line of the prediction distributions. In the MNIST-handwritten-number classification task, the ideal CNN can reach a prediction accuracy of 99.0% and OCU-distorted turns to be 98.9%. In the Fashion-MNIST classification, the prediction accuracy of ideal CNN is 92.0% and that of OCU-distorted is 91.5%. The prediction accuracy of the OCU closely approaches the ideal results and the prediction distributions of the OCU is similar with that of ideal ones, implying that the OCU distortions introduce minor influences on the CNN tasks.



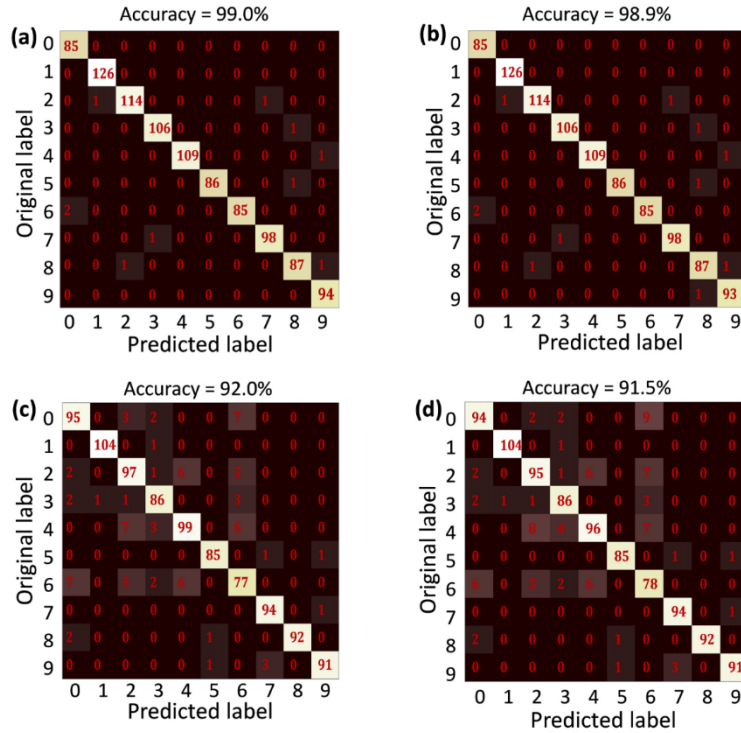


Fig. 6. Prediction distributions of CNN executed by the proposed OCU and digital computer, respectively. Numbers on the figure denote how many times do neural network predictions cast on the coordinate (a), (b) are the results of MNIST-handwritten-number classification. (c), (d) are results of Fashion-MNIST. (a), (c) are yielded by the 64-bit digital computer and (b), (d) are generated by the proposed OCU.

#### 4. *In situ* training for higher accuracy

In the above experiment, the network parameters are *ex-situ* trained in a digital computer and they are not perfectly suitable for the implemented OCU. Imperfections, such as unequal light splitting, unequal insertion loss, and inaccurate decoding among the multiplier branches, could result in deviations and degrade the OCU accuracy. This problem can be solved by *in situ* training [25], where training is carried out directly based on the configured OCU system. We use forward-propagation algorithm to train the network parameters. Instead of calculation of the gradients of all parameters at a time by back-propagation [25], the forward-propagation algorithm updates one parameter every single time as the following formulas [19]:

$$g = \frac{L(\theta + \Delta\theta) - L(\theta)}{\Delta\theta} \quad (4)$$

$$\hat{\theta} = \theta + r \cdot g \quad (5)$$

By shifting a small  $\Delta\theta$  of the parameter  $\theta$ , the loss function  $L$  varies and thus its gradient  $g$  over  $\theta$  is calculated. The parameter  $\theta$  is updated referring the learning rate  $r$  and the gradient  $g$ . In the *in situ* training experiment, we optimize a single convolution window (i.e. voltages to the AOM array 2) rather than all windows of the entire CNN. Therefore, the loss function is calculated by the mean absolute error between the OCU output data and the reference convolution result calculated by the digital computer. The modulation voltages to the AOM array 2 are initialized by the *ex-situ* trained parameters and they are trained once in each epoch. As described above, a  $3 \times 3$  convolution window is separated to three  $1 \times 3$  windows.

Therefore, a complete training of a  $3 \times 3$  convolution window can be done through three rounds of  $1 \times 3$  training. The learning rate is set to be 0.5. Figure 7 depicts the results of the *in situ* training. The loss functions decrease during training and reach the steady limitations. The loss functions could not infinitely drop because of imperfect decoding of the AOM array 1 and system distortion and/or noise. After the *in situ* training, the residual error between the reference (computer) and the OCU result gets lower and the corresponding SDR increases from 27.33 to 36.27 dBc. These results show that *in situ* training provides an effective way to further reduce the influence by the system imperfections and improve the accuracy of the proposed OCU architecture.

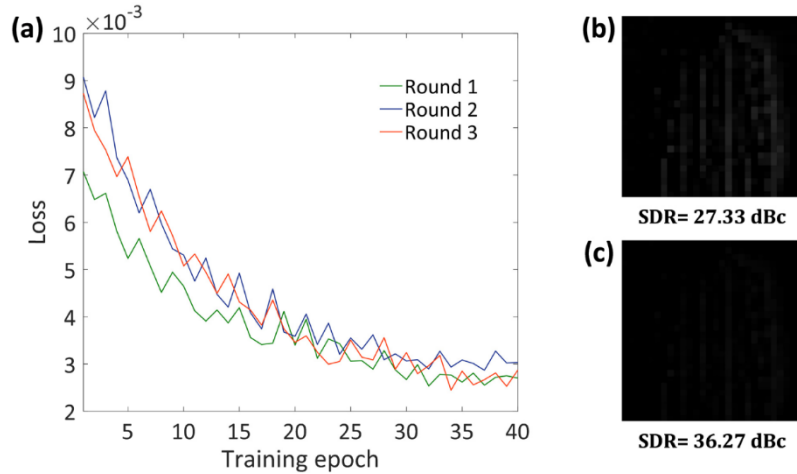


Fig. 7. Result of *in situ* training. (a) Losses descending with training epochs. A  $3 \times 3$  convolution window is trained through 3 rounds of  $1 \times 3$  convolution windows. (b) Residual error of the proposed OCU before *in situ* training and its corresponding SDR of 27.33 dBc. The original input image is a “Shirt” in Fashion-MNIST classification. (c) Residual error of the proposed OCU after *in situ* training, and its corresponding SDR of 36.27 dBc. Residual errors are amplified by 5 times for better visibility.

## 5. Conclusion and discussion

The OCU architecture based on dot-product operation is proposed to realize convolutions in general CNNs. To take the advantage of hardware reusing concept, the OCU is designed to include two cascaded modulator arrays. By changing the modulation voltages on the modulators, the OCU is reused and thus conducts convolutions with arbitrary input sizes. In the experiments, AOM arrays are deployed for their high-extinction ratio so that we can demonstrate the achievable accuracy of the proposed architecture. With *ex-situ* trained parameters, the SDR of the OCU could averagely reach 28.22 dBc. Two typical CNN classification tasks (MNIST handwritten numbers and Fashion MNIST) are then simulated under this accuracy. The prediction accuracies of OCU approach closely to the ideal results yielded from a 64-bit digital computer. Furthermore, by *in situ* training, the SDR of the proposed OCU is enhanced to 36.27 dBc, validating the refinement of accuracy base on the proposed architecture.

It is worth noting that the current demonstration of OCU is a proof-of-concept version based on a power-consuming fiber platform. To realize the full advantages of optical technologies on computing speed and energy consumption, the components should be integrated in chip-scale. Similarly to other ONN paradigms, the proposed OCU also suffers from the latency and power consumption introduced by optical/electrical (O/E) interconversions. However, demonstrated in recent ONN researches [18,26], a large-scale optical computing platform dilutes these margin time/energy costs to ultra-low levels. By



regarding the OCU as building blocks to construct a large-scale integrated convolutional array, the time/energy requirement of each convolution operation will be significantly reduced. Moreover, the integrated convolutional array would also enable parallel computing and thus increases computing speed by multiple times, exploiting the high-speed advantage of ONNs over traditional electronic implementations. Thanks to the recent dramatic development of the chip-scale electro-photon hybrid integration [27], it is promising to manufacture the integrated version of the convolutional array in the near future. And the future adopting of high-speed and low-power integrated PDs [13] and electro-optic modulators [28,29] into the integrated array will boost the convolution speed and reduce power consumption significantly.

## Funding

National Natural Science Foundation of China (NSFC) (grant no. 61822508, 61571292, 61535006).

## References

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
2. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," preprint at arXiv, <https://arxiv.org/abs/1512.03385> (2015).
3. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," preprint at arXiv, <https://arxiv.org/abs/1409.1556> (2015).
4. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.* **26**(7), 3142–3155 (2017).
5. Y. Rivenson, Z. Gorocs, H. Gunaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).
6. B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature* **555**(7697), 487–492 (2018).
7. D. Wang and J. Chen, "Supervised speech separation based on deep learning: and overview," *IEEE Trans. Audio Speech Lang. Process.* **26**(10), 1702–1726 (2018).
8. A. Ephrat, I. Mosseri, O. Lang, T. Dekel, K. Wilson, A. Hassidim, W. T. Freeman, and M. Rubinstein, "Looking to listen at the cocktail party: a speaker-independent audio-visual model for speech separation," *ACM Trans. Graph.* **37**(4), 1–11 (2018).
9. M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung patten classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imaging* **35**(5), 1207–1216 (2016).
10. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature* **529**(7587), 484–489 (2016).
11. D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Science* **362**(6419), 1140–1144 (2018).
12. S. Xu, X. Zou, B. Ma, J. Chen, L. Yu, and W. Zou, "Analog-to-digital conversion revolutionized by deep learning," preprint at arXiv, <https://arxiv.org/abs/1810.08906> (2018).
13. L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J. M. Hartmann, P. Crozat, E. Cassan, C. Kopp, H. Zimmermann, and J. M. Fédéli, "Zero-bias 40Gbit/s germanium waveguide photodetector on silicon," *Opt. Express* **20**(2), 1096–1101 (2012).
14. J. Cardenas, C. B. Poitras, J. T. Robinson, K. Preston, L. Chen, and M. Lipson, "Low loss etchless silicon photonic waveguides," *Opt. Express* **17**(6), 4752–4757 (2009).
15. J. Bueno, S. Matkooobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large-scale photonic recurrent neural network," *Optica* **5**(6), 756–760 (2018).
16. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science* **361**(6406), 1004–1008 (2018).
17. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Sci. Rep.* **8**(1), 12324 (2018).
18. S. Colburn, Y. Chu, E. Shilzerman, and A. Majumdar, "Optical frontend for a convolutional neural network," *Appl. Opt.* **58**(12), 3179–3186 (2019).
19. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljacic, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* **11**(7), 441–446 (2017).

20. L. Yang, R. Ji, L. Zhang, J. Ding, and Q. Xu, "On-chip CMOS-compatible optical signal processor," *Opt. Express* **20**(12), 13560–13565 (2012).
21. A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Rep.* **7**(1), 7430 (2017).
22. H. Banherian, S. Skirlo, Y. Shen, H. Meng, V. Ceperic, and M. Soljacic, "On-chip optical convolutional neural networks," preprint at arXiv, <https://arxiv.org/abs/1808.03303> (2018).
23. Y. LeCun, C. Cortes, and C. Burges, "The MNIST database of handwritten digits," at <http://yann.lecun.com/exdb/mnist/>.
24. H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," preprint at arXiv, <https://arxiv.org/abs/1708.07747>, (2017).
25. T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of photonic neural networks through in situ backpropagation and gradient measurement," *Optica* **5**(7), 864–871 (2018).
26. R. Hamerly, A. Sludds, L. Bernstein, M. Soljacic, and D. Englund, "Large-Scale Optical Neural Networks based on Photoelectric Multiplication," preprint at <https://arxiv.org/abs/1812.07614> (2018).
27. A. H. Atabaki, S. Moazeni, F. Pavanella, H. Gevorgyan, J. Notaros, L. Alloatti, M. T. Wade, C. Sun, S. A. Kruger, H. Meng, K. Al Qubaisi, I. Wang, B. Zhang, A. Khilo, C. V. Baiocco, M. A. Popović, V. M. Stojanović, and R. J. Ram, "Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip," *Nature* **556**(7701), 349–354 (2018).
28. C. Wang, M. Zhang, X. Chen, M. Bertrand, A. Shams-Ansari, S. Chandrasekhar, P. Winzer, and M. Lončar, "Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages," *Nature* **562**(7725), 101–104 (2018).
29. M. He, M. Xu, Y. Ren, J. Jian, Z. Ruan, Y. Xu, S. Gao, S. Sun, X. Wen, L. Zhou, L. Liu, C. Guo, H. Chen, S. Yu, L. Liu and X. Cai, "High-performance hybrid silicon and lithium niobate Mach-Zehnder modulator for 100 Gbit s<sup>-1</sup> and beyond," *Nature Photonics*, online at <https://doi.org/10.1038/s41566-019-0378-6> (2019).