Optical convolutional neural network with atomic nonlinearity

MINGWEI YANG,[1,2,*] ELIZABETH ROBERTSON,[1,2]
LUISA ESGUERRA,[1,2] KURT BUSCH,[3,4] (iD)
AND JANIK WOLTERS[1,2]

[1]*Deutsches Zentrum für Luft- und Raumfahrt, Institute of Optical Sensor Systems, Rutherfordstraße 2, 12489 Berlin, Germany*
[2]*Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany*
[3]*Humboldt-Universität zu Berlin, Institut für Physik, AG Theoretische Optik & Photonik, 12489 Berlin, Germany*
[4]*Max-Born-Institut, 12489 Berlin, Germany*
[*]*mingwei.yang@dlr.de*

Abstract: Due to their high degree of parallelism, fast processing speeds and low power consumption, analog optical functional elements offer interesting routes for realizing neuromorphic computer hardware. For instance, convolutional neural networks lend themselves to analog optical implementations by exploiting the Fourier-transform characteristics of suitable designed optical setups. However, the efficient implementation of optical nonlinearities for such neural networks still represents challenges. In this work, we report on the realization and characterization of a three-layer optical convolutional neural network where the linear part is based on a 4f-imaging system and the optical nonlinearity is realized via the absorption profile of a cesium atomic vapor cell. This system classifies the handwritten digital dataset MNIST with 83.96% accuracy, which agrees well with corresponding simulations. Our results thus demonstrate the viability of utilizing atomic nonlinearities in neural network architectures with low power consumption.

1. Introduction

In recent years, convolutional neural networks (CNNs) have established themselves as a key method in computer vision tasks, with applications that range from fundamental studies in condensed-matter physics [1] and particle physics [2] all the way to autonomous driving [3] and cancer detection [4]. With the rising popularity of CNNs, significant concerns regarding their energy consumption relative to simpler network architectures have emerged. Specifically, about $\sim 80\%$ of the inference time required by CNNs is utilized for carrying out the convolution [5] so that energy-efficient computing paradigms for computing convolutions have hence become an active field of research. Due to their inherent parallelism, potential for GHz modulation speeds and low energy consumption (when using only passive elements), free-space-optics implementations have been identified as an attractive possibility for analog computations of convolutions [6,7]. In fact, within the broader context of artificial neural networks, linear-optics implementations have been demonstrated based on diffractive materials [8], spatial light modulators (SLMs) [9–13], ring resonators [14], arrays of Mach-Zender interferometers [15,16] and wavelength-division multiplexing techniques [17]. For further information, we would like to refer to recent reviews [18–20] and the convolutional layer design that has recently been demonstrated by Miscuglio et al. [21]. However, the most efficient optical implementation of the nonlinearities required by neural networks remains an open question. Recently, Zuo et al. have demonstrated an optical nonlinearity by utilizing electromagnetically induced transparency in a gas of ultra-cold ^{87}Rb

atoms [22]. Ryou et al. introduced the idea of avoiding the overhead associated with ultra-cold atoms by realizing a nonlinearity through saturable absorption in a thermal vapor of Rb atoms [23]. Other mechanisms for realizing optical nonlinearities for neuromorphic applications include phase-change materials (PCMs) [17] and the Kerr effect combined with two-photon absorption [24].

In this work, we introduce a nonlinearity in the form of a saturated absorption profile of cesium atomic vapor into an optical convolution setup based on free-space SLMs. As we shall demonstrate below, this system classifies the handwritten digit dataset MNIST [25] with 83.96% accuracy. As an essential part of the experimental development of multi-layer optical neural networks, the optical nonlinearity is proven to be effectively provided by the cesium atomic vapor. In fact, we expect a potential benefit of optical neural networks with atomic nonlinearity for data processing on board of satellites. The standard procedure for processing complex sensor data is the use of artificial neural networks, which is served on the ground by specialized digital hardware such as graphics cards, tensor flow processors, etc. The availability of these options is limited for data processing in orbit due to the extreme requirements for energy consumption, thermal management and radiation hardness. However, data transfer and processing on the ground can also be challenging due to the immense amount of data. Therefore, high-performance computers under space conditions would be desirable and optical computers have a high potential to fill the gap, enabling energy-efficient machine learning in orbit. Digital micro mirror devices (DMDs), as main components of the setup, are proven to be extremely robust and promising for orbit conditions, given their resistance to radiation, vibration, vacuum, mechanical shock loads, and cryogenic temperatures [26]. Thus, optical convolutional neural networks (OCNN) are feasible as powerful optical co-processors for specific tasks, with performance exceeding space-qualified conventional digital hardware.

2. Methods

Here we demonstrate an optical convolutional neural network in which both linear operations and nonlinearity are realized optically. We implement the OCNN with one input layer, one optical convolutional layer, one fully connected layer followed by one output layer. An optical nonlinearity is applied after the convolutional layer (Fig. 1). The convolution of the input and
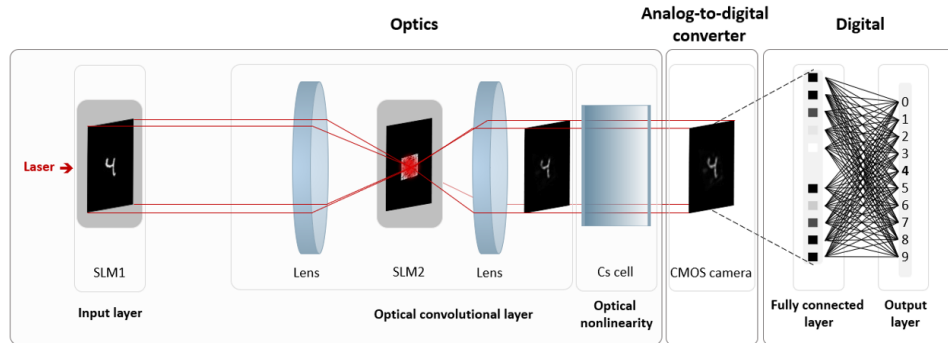
Fig. 1. Layout of the optical convolutional network architecture used in this work.

the kernel of the OCNN is performed optically by pointwise multiplications in the Fourier plane, based on a 4f-imaging system and the convolution theorem. A CMOS camera acts as an analog-to-digital converter. A digital fully connected layer is implemented in the computer to connect the nonlinear activated feature maps and the output layer. We simulate the system based on the experimental setup, which is used to train both the kernel of the OCNN and the digital fully

connected layer. The trained kernel is then displayed in the experiment and the inference process is executed for both simulation and experiment, for the same kernel and fully connected layer.

We train and test the system with handwritten digits from the MNIST dataset. The input data is encoded as a two-dimensional intensity profile by an amplitude-only SLM, which is based on a digital micromirror device (DMD). This electro-optic conversion is implemented by controlling the individual micromirrors of the DMD. These mirrors can rapidly tilt between an on-and-off position and selectively reflect the incident light to the optical path of the OCNN. One micromirror and its corresponding memory unit constitute one pixel of the DMD. The properties of the large-scale display (1920×1080 pixels) and the high update rate up to 10^4 Hz for binary images of the micromirror array allow, in principle, complex multi-channel data processing at high speed.

The conventional convolution process is formed of the pixel-wise multiplication and summation of a subsection of an image with a kernel. The kernel is scanned across the whole image, where this multiplication and summation repeats, resulting in an image convolved with a kernel. By contrast, the OCNN performs the convolution process by employing the convolution theorem $f(x) * h(x) = \mathcal{F}^{-1}\{\mathcal{F}[f(x)] \cdot \mathcal{F}[h(x)]\}$ and the Fourier transform properties of lenses [27]. Specifically, the convolution ($*$) of the input $f(x)$ and the kernel $h(x)$ is the inverse Fourier transform ($\mathcal{F}^{-1}$) of the pointwise product ($\cdot$) of their Fourier transforms ($\mathcal{F}[f(x)], \mathcal{F}[h(x)]$). By constructing a 4f-system with two SLMs and two lenses, the Fourier transform of the input, the dot product in Fourier space and the inverse Fourier transform of the dot product can be performed optically and passively. Thus, the convolved images can be observed in the front focal plane of the second lens.

In the experiment, we place the cesium vapor cell after the convolutional layer, so that the convolved pattern is nonlinearly absorbed, thereby introducing an optical nonlinearity into the system. A CMOS camera (CS165MU- 1.6 MP Monochrome CMOS Camera, Thorlabs, pixel size $3.45 \ \mu m \times 3.45 \ \mu m$) is positioned in the front focal plane of the second lens. The images are captured by the camera and saved on a computer. Figure 2 depicts the sketch of the experimental setup. The light source is a distributed feedback diode laser assembly with a Doppler-free
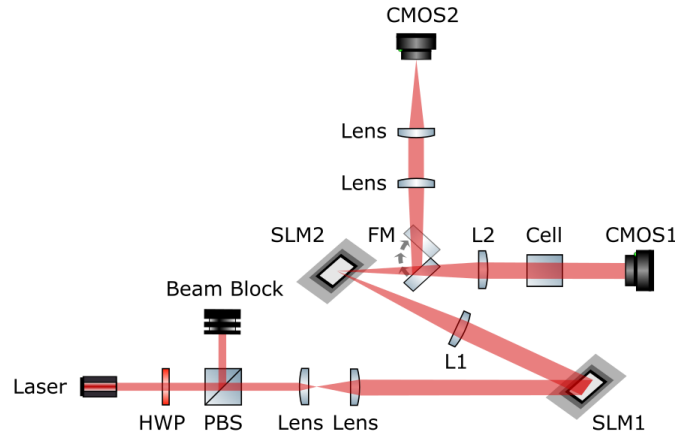
Fig. 2. Sketch of the experimental setup. M: mirror, HWP: half-wave plate, PBS: polarizing beam splitter, L: lens, FM: flip mirror. The intensity of light is controlled by the combination of the half-wave plate and the polarizing beam splitter. A telescope is composed of a 50 mm and a 100 mm focal length lens, expanding the beam to 8 mm diameter.

spectroscopy setup. The wavelength is fine-tuned and actively stabilized to the cesium D1 line transition ($6^2S_{1/2} \rightarrow 6^2P_{1/2}$) at about 894 nm. The laser is further coupled into and emerges from an optical fiber. A half-wave plate and a polarizing beam splitter are used to adjust the light

intensity. Thereafter, two lenses act as a telescope to enlarge the beam diameter to approximately 8 mm. This large collimated beam is sufficient to cover the pattern surface of SLM1 (DLP6500, Texas Instruments, pixel size 7.6 μm × 7.6 μm). A computer connects and loads the patterns to SLM1, therefore, the incoming beams are specifically modulated to the image shapes of the binarized MNIST dataset and reflected to the optical path. For the used SLM model, a theoretical maximum pattern rate of 10 kHz can be reached with binary data. A lens (f = 250 mm, L1) is located at a distance of 250 mm from SLM1. The second identical SLM is then placed in its back focal plane, multiplying the corresponding Fourier-transformed input with the displayed Fourier-transformed kernel. The amplitude modulated dot products of the Fourier patterns are then inverse Fourier transformed by a second lens (f = 250 mm, L2), and subsequently captured by CMOS1 after being nonlinearly activated by the Cs vapor cell. Consequently, the images captured by the camera are the nonlinearly activated convolutions of the input images and the kernel. The computer collects all such processed images and converts them into arrays of gray scale values. Each array is connected to a trained fully connected layer, which outputs the prediction of the neural network via matrix multiplication. In addition, a flip mirror can be used to reflect the beam to CMOS2, which is used to image the Fourier plane. Practically, it is indispensable for optimizing the SLM2 alignment. We verify the precise placement of SLM2 in the Fourier plane by means of observing the Fourier plane image of the input. Furthermore, according to the imaging of the Fourier plane, we fine tune the position and the angle of SLM2 by pixel matching the Fourier patterns of the input and the kernel displayed on SLM2.

As discussed above, much attention has been drawn to optical nonlinearities. Atomic vapor cells are competitive options due to their non-vacuum, room-temperature experimental environment. Moreover, as a result of the passive pass of laser light, there is no need for additional energy. The relation between the output intensity I and the input intensity I_0 is [28]

$$I = I_0 \exp\left(\frac{-\mathrm{OD}_0}{1 + I_0/I_s}\right), \tag{1}$$

where I, I_0, I_s represent the output intensity, input intensity, and saturation intensity, respectively. The optical depth at $I_0 = 0$ is given by $\mathrm{OD}_0 = N\sigma z$, where N is the number of cesium atoms per unit volume, σ denotes the corresponding cross-section, and z is the cell length. When the input intensity I_0 reaches the saturation intensity I_s, the absorption is reduced by 50%. The input-output relation thus follows a nonlinear shape. In the experiment, the light intensities are converted to 8-bit arrays of integer values by the camera. To ensure that the corresponding simulation is consistent with the experiment, we experimentally determine the input-output relation. To this end, we display a square pattern on SLM1 representing the average number of bright pixels in the MNIST dataset. While varying the incident laser power, we measure the resulting pixel intensity detected by the CMOS camera with and without the Cs cell, respectively. The measured and fitted input-output relation of the nonlinearity is shown in Fig. 3.

As a programming language, we have employed Python (version 3.8.5.) in combination with the Pytorch framework (version 1.7.1) [29] for training. This combination features an autograd function that nicely supports our custom nonlinear activation function during the backpropagation process. Contrary to a conventional CNN training process, different methods and additional procedures must be considered in the simulation in order to fit the optical setup. The OCNN model has a three-layer structure including the atomic nonlinearity. The original image size of the MNIST is 28 × 28, however, for experimental convenience, the input size is enlarged to 256 × 256. Moreover, according to the convolution theorem, the kernel size must be same as the input size, so the kernel size is also set to 256 × 256. The input and kernel are binarized for the purpose of effectively interacting with the cesium atoms. In the OCNN model, the convolution stems from the inverse Fourier transform of the multiplication of the Fourier-transformed input and the kernel. Thus, we have implemented a custom convolutional layer into the Pytorch
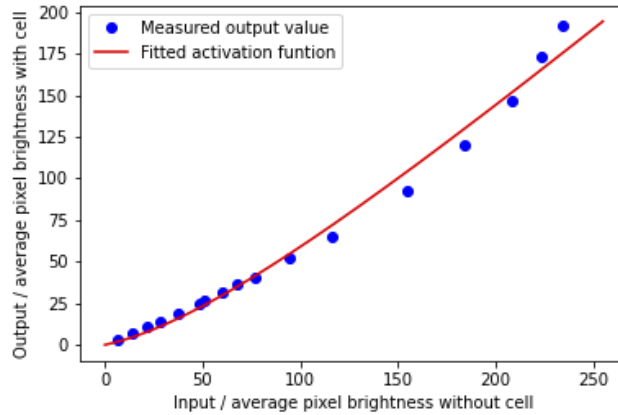
Fig. 3. Measured input-output average pixel brightness and the corresponding fit curve. Fitting the experimental data to Eq. (1) gives $OD_0 = 1.38, I_s = 62.31$ pixel brightness. The axes represent the average grayscale value of the picture without (input) and with (output) the vapor cell, respectively.

machine-learning framework [29]. Due to the fact that we initialize the learnable kernel values directly in complex Fourier space [30], no Fourier transform is required for the kernel during the training process. The real-part is composed of random positive values. As amplitude-only SLMs do not deal with the phase information, all the imaginary values are set to zero. The first step is to perform a two-dimensional Fourier transform of the input images and multiply them with the kernel in Fourier space. Although the real-part values of the kernel are initialized to be positive, they may contain negative values during training. As a result, we apply a rectified linear unit (ReLU) function to the kernel to replace negative values with zero. Furthermore, we ensure that the pixel position of the Fourier pattern of the input light has a one-to-one match to the pixel position of the kernel structure on the SLM2. However, exact matching is experimentally difficult for the SLM hardware that has micron-scale pixels. Accordingly, the 256×256 kernel is max-pooled to 32×32, followed by a process of enlarging back to 256×256. Next, the element-wise multiplication is performed in Fourier space, and the following inverse Fourier transform gives the feature map of the OCNN in real space, the size maintains 256×256. It is worth mentioning that, in order for the optical patterns to be activated in a nonlinear manner, the light intensity is adjusted to that the input intensity of the vapor cell falls within the nonlinear region around the saturation intensity. The feature maps are max-pooled to 28×28 and flattened to one-dimensional arrays, then a fully connected layer with a size of 784×10 further processes the feature maps and outputs a 10×1 array representing the predicted digits. The model is trained for 5 epochs, using a gradient descent optimization algorithm with the Softmax classifier as loss function, Adam [31] as the optimizer, and 0.001 learning rate.

3. Results

We have performed the training of our OCNN with the $60,000$ training images from MNIST and have determined the corresponding accuracy for $10,000$ test data. For the MNIST test set, the simulation of the OCNN achieve 92.6% and 93.2% accuracy with the cesium absorption function and the ReLU function as nonlinearity, respectively. The accuracy is relatively low for the MNIST dataset comparing to other SLM-based optical network systems [11,21], since the dataset is binarized in the OCNN and there is only one convolutional layer with a single kernel for the first demonstration and better reproducibility of the training. Without any nonlinearity in the entire network, the accuracy goes down to 90.8%. For comparison, a conventional feed

forward model is trained (5 epochs, 0.001 learning rate). This model which consists of one fully connected layer of size 784×10 gives an accuracy of 80.5%. This result is equivalent to using a blank kernel in OCNN.

The experimental classification accuracy for both cases, with and without the vapor cell are tested directly using the individually simulated kernel and fully connected layer. Figure 4(a) shows examples of simulated and experimentally captured feature maps. For the absence of the vapor cell, the OCNN setup achieves 70.72% accuracy for test data. With the cesium vapor cell in the setup, the performance is improved to 71.84%. The confusion matrix indicates that the experimental results have particular errors in predicting the digits "1", "5", and "8" (Fig. 4(b)). The comparable low accuracy is caused by experimental imperfections. In order to compensate for these imperfections, we train a mapping matrix using the full $10,000$ test data being processed by the optical system. This linear mapping is used to modify the fully connected layer, i.e. to retrain it to the experimental data. As shown in the confusion matrix, the trained mapping matrix significantly reduces mispredictions (Fig. 4(c)). With the help of this training, the accuracy increases to 83.96% and 91.9% as a result of with and without the vapor cell, respectively. For comparison, the accuracy of the OCNN without any additional nonlinearity gives an accuracy of 89.3%, indicating the effectiveness of the nonlinearity introduced by the camera. In Table 1 we have summarized the results on the accuracy. After retraining the mapping matrix, the experimental classification performance is in good agreement with the simulation. We attribute the slightly lower performance of the optical system to the reasons: 1) As the dynamic range and analog-to-digital conversion accuracy of the camera are limited, the distribution of pixel values does not accurately match the theoretical values. 2) The Fourier-transformed input images and the kernel image displayed on SLM2 do not overlap perfectly. 3) Distortions and aberrations caused by optical imperfections of the optical components also affect the experimental results.

(a)

(b)　　　　　　　　　　　(c)　　　　　　　　　　　(d)
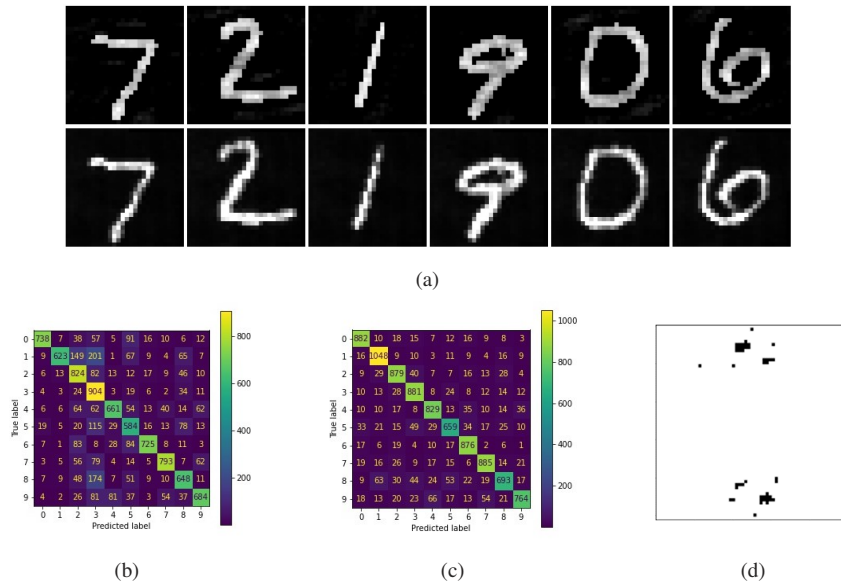
Fig. 4. Visualization of feature maps, confusion matrices, and the kernel. (a) Examples of simulated (first row) and experimental (second row) feature maps. (b) Experimental confusion matrix of the MNIST classification. (c) Confusion matrix of the MNIST classification with the trained mapping matrix. (d) Pretrained binary kernel. Note that although this kernel shows uncomplicated patterns, the resulting improvement in prediction accuracy is significant when compared to that of the blank kernel.

Table 1. Summary of OCNN prediction accuracy.[a]

Non-linearity	Cs Vapor	ReLU	No nonlinearity
Simulation	92.6%	93.2%	90.8%
OCNN (No mapping matrix)	71.84%	70.72%	45.59%
OCNN (With mapping matrix)	83.96%	91.9%	89.3%

[a]Each kernel is individually trained for each case. No nonlinearity means that no additional nonlinearity is added beyond the inherent nonlinearity of the camera.

4.   Conclusion

In summary, we have demonstrated an optical convolutional neural network with atomic nonlinearity. The linear part of the system consists of a 4f-system made from SLMs and lenses, the optical nonlinearity is realized by a cesium vapor cell. The prediction accuracy of the OCNN setup for recognizing the MNIST handwritten digit dataset reaches 83.96%, which is in reasonable agreement with the simulation (92.6%). The effectiveness of the atomic vapor cell suggests that it has great potential to provide optical nonlinear activation functions for neural networks with other topologies. Moreover, it is an attractive candidate for use in multilayer optical neural networks with several SLMs and vapor cells in series, as the light beam passively acquires nonlinear output characteristics without requiring a further source of energy. In addition, the operation complexity is also relatively low and for further integrations, the atomic vapor cell can also be replaced by semiconductor saturable absorbers (SESAM) [32]. Besides, [33–35] show that optical nonlinearities can also be used for mode filtering or data pre-processing, etc. aside from being used as nonlinear activation functions in neural networks.

For the specific task of classifying MNIST, the nonlinearity has only limited impact. However, our results show the great potential of OCNN with nonlinearity for machine learning problems in general. The optical convolution and optical nonlinearity with Rb atoms have been demonstrated in two independent experiments in [21] and [23], respectively. To make the forward pass completely optical, we have substantiated that atomic nonlinearity can be included in a standard machine learning architecture, e.g. CNN, a dominant approach in various computer vision tasks. In contrast, the diffraction method in [23] is not recognized as a standard model architecture. Compared to Rb atoms, Cs atoms have only one stable isotope and a vapor pressure of $\sim 2 \cdot 10^{-6}$ Torr at room temperature, and the Cs D lines show excellent absorption properties, enabling pronounced nonlinearities without isotopic purification or power consuming cell heating. The simplicity in the use of Cs cell is particularly important for realizing optical nonlinearity in field applications. Moreover, an optical nonlinearity is also essential for realizing a multiple layer structure, i.e., making the light propagate and loop in the convolutional layer multiple times. In Table 2 we list some of the differences between both approaches in [21] and [23]. Comparison with [23] is of interest as they benchmark against the same task, employing a different optical neural network architecture. From this work, we can draw several conclusions: 1) The convolutional neural network architecture is better suited to solving this task. We observe higher accuracy in both simulation and experiment for OCNN compared to the system in [23], with or without nonlinearity. First considering the case of no nonlinearity, in simulation, the OCNN achieves a 90.8% accuracy compared to 66.4% in [23]. In experiment, the performance of both systems drops slightly, with the OCNN delivering a higher accuracy; 89.3% to 26.7%. When introducing the atomic nonlinearity, the OCNN achieves 92.6% accuracy in simulation and 83.96% accuracy in experiment, compared to 66.6% and 33% for the system in [23], respectively. 2) [23] distinguishes between idealized and experimental parameters describing the nonlinearity. In our work, we consider only experimentally measured parameters, and our simulations suggest our Cs nonlinearity should improve the performance by about 2%. When comparing this to [23],

we see they expect an improvement of only 0.2%. Although in [23] they obtain an increase by 6% in the experiment with the nonlinearity, they attribute this to the noise filtering instead of the nonlinear property of the vapor cell. Therefore, our system is more responsive to the introduction of atomic nonlinearity for an optical neural network. 3) As they do not use a mapping matrix, we draw comparison to the experiments where we also used no mapping matrix. We see that in both cases the introduction of a nonlinearity improves the performance of the system. In the OCNN system, the accuracy increases from 45.59% to 71.84%, while in [23] it improves from 26.7% to 33%.

Table 2. Comparison of parameters and results among different approaches.

-	Training images	Test images	Kernel(s)	Training epochs	Resolution [*bits*]	Additional nonlinearity	Hardware model accuracy
OCNN [this work]	60,000	10,000	1	5	1	Cs cell	83.96%
Fourier neural network [21]	5,5000 training 5,000 validation	10,000	16	10	8	None	98%
Free space ONN with Rb [23]	10,000	1,000	None	50	8	Rb cell	33%

Previous experiments demonstrated various aspects of SLM-based optical neural networks [21,23]. Inspired by these works, we demonstrate an all-optical implementation of a convolutional layer including the nonlinear activation. By this, we take an important step for realizing optical co-processors for the most important machine learning method, enabling even optical multi-layer CNN architectures typically used in image analysis. The performance of OCNN can theoretically reach $10^6 \times 10^6 \times 10^4 = 10 \times 10^{15}$ operations per second (OPS) with image and kernel size of 10^6 each and binary resolution allowing for a maximum pattern rate of 10^4 Hz. Then, bottlenecks will include the resolution and update rate of hardware, the transmission rate of interfaces, etc. As a comparison, Google's latest release of the state-of-the-art TPU v4 chip, provides up to $\sim 275 \cdot 10^{12}$ OPS. Regarding power consumption of the OCNN, the laser and camera together require only about 2 W, while the two SLMs need on the order of 10 W. The used computer requires about 50 W. However, the atomic nonlinearity enables the possibilities for realizing multilayer ONNs with all-optical training [36], so the computer might be omitted in the future. Moreover, the SLMs mainly consume power for switching. Thus, the numerous quasi-static SLMs that provide connectivity in such a multilayer ONN are expected to have negligible impact on the power consumption being dominated by the input SLM. Hence, we expect a multilayer optical neural network to drastically increase the computational performance without increasing the energy consumption significantly, thereby rendering optical neural networks with atomic nonlinearity an attractive alternative to conventional hardware.

Disclosures. The authors declare no conflicts of interest.

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

References

1. J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," Nat. Phys. **13**(5), 431–434 (2017).
2. A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, "Machine learning at the energy and intensity frontiers of particle physics," Nature **560**(7716), 41–48 (2018).
3. H. Fujiyoshi, T. Hirakawa, and T. Yamashita, "Deep learning-based image recognition for autonomous driving," IATSS Res. **43**(4), 244–252 (2019).

4. R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights Imaging **9**(4), 611–629 (2018).
5. X. Li, G. Zhang, H. H. Huang, Z. Wang, and W. Zheng, "Performance analysis of gpu-based convolutional neural networks," in *2016 45th International Conference on Parallel Processing (ICPP)*, (2016), pp. 67–76.
6. H. J. Caulfield and S. Dolev, "Why future supercomputing requires optics," Nat. Photonics **4**(5), 261–263 (2010).
7. H. J. Caulfield, J. Kinser, and S. K. Rogers, "Optical neural networks," Proc. IEEE **77**(10), 1573–1583 (1989).
8. 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).
9. 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).
10. J. Spall, X. Guo, T. D. Barrett, and A. I. Lvovsky, "Fully reconfigurable coherent optical vector–matrix multiplication," Opt. Lett. **45**(20), 5752–5755 (2020).
11. T. Zhou, X. Lin, J. Wu, Y. Chen, H. Xie, Y. Li, J. Fan, H. Wu, L. Fang, and Q. Dai, "Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit," Nat. Photonics **15**(5), 367–373 (2021).
12. S. A. Goorden, J. Bertolotti, and A. P. Mosk, "Superpixel-based spatial amplitude and phase modulation using a digital micromirror device," Opt. Express **22**(15), 17999–18009 (2014).
13. J. Bueno, S. Maktoobi, 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).
14. A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: an integrated network for scalable photonic spike processing," J. Lightwave Technol. **32**(21), 4029–4041 (2014).
15. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**(7), 441–446 (2017).
16. H. Zhou, Y. Zhao, Y. Wei, F. Li, J. Dong, and X. Zhang, "All-in-one silicon photonic polarization processor," Nanophotonics **8**(12), 2257–2267 (2019).
17. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," Nature **569**(7755), 208–214 (2019).
18. X. Sui, Q. Wu, J. Liu, Q. Chen, and G. Gu, "A review of optical neural networks," IEEE Access **8**, 70773–70783 (2020).
19. J. Cheng, H. Zhou, and J. Dong, "Photonic matrix computing: From fundamentals to applications," Nanomaterials **11**(7), 1683 (2021).
20. L. De Marinis, M. Cococcioni, P. Castoldi, and N. Andriolli, "Photonic neural networks: A survey," IEEE Access **7**, 175827–175841 (2019).
21. M. Miscuglio, Z. Hu, S. Li, J. K. George, R. Capanna, H. Dalir, P. M. Bardet, P. Gupta, and V. J. Sorger, "Massively parallel amplitude-only fourier neural network," Optica **7**(12), 1812–1819 (2020).
22. Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, "All-optical neural network with nonlinear activation functions," Optica **6**(9), 1132–1137 (2019).
23. A. Ryou, J. Whitehead, M. Zhelyeznyakov, P. Anderson, C. Keskin, M. Bajcsy, and A. Majumdar, "Free-space optical neural network based on thermal atomic nonlinearity," Photonics Res. **9**(4), B128–B134 (2021).
24. C. Mesaritakis, A. Kapsalis, and D. Syvridis, "All-optical reservoir computing system based on ingaasp ring resonators for high-speed identification and optical routing in optical networks," Proc. SPIE **9370**, 937033 (2015).
25. Y. LeCun, "The mnist database of handwritten digits," http://yann. lecun. com/exdb/mnist/ (1998).
26. A. Travinsky, D. Vorobiev, Z. Ninkov, A. Raisanen, M. A. Quijada, S. A. Smee, J. A. Pellish, T. Schwartz, M. Robberto, S. Heap, D. Conley, C. Benavides, N. Garcia, Z. Bredl, and S. Yllanes, "Evaluation of digital micromirror devices for use in space-based multiobject spectrometer application," Journal of Astronomical Telescopes, Instruments, and Systems **3**(03), 1–035003 (2017).
27. J. W. Goodman, "Introduction to fourier optics," *Introduction to Fourier optics*, 3rd ed., by JW Goodman. (Englewood, CO: Roberts & Co. Publishers, 2005), 1.
28. C. J. Foot, *Atomic physics*, vol. 7 (OUP Oxford, 2004).
29. https://pytorch.org/.
30. H. Pratt, B. Williams, F. Coenen, and Y. Zheng, "Fcnn: Fourier convolutional neural networks," *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, (Springer, 2017), pp. 786–798.
31. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv, arXiv:1412.6980 (2014).
32. A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of absorption," Opt. Express **22**(9), 10868–10881 (2014).
33. T. Bu, S. Kumar, H. Zhang, I. Huang, and Y.-P. Huang, "Single-pixel pattern recognition with coherent nonlinear optics," Opt. Lett. **45**(24), 6771–6774 (2020).
34. S. Kumar, T. Bu, H. Zhang, I. Huang, and Y. Huang, "Robust and efficient single-pixel image classification with nonlinear optics," Opt. Lett. **46**(8), 1848–1851 (2021).
35. M. Yildirim, I. Oguz, F. Kaufmann, M. R. Escale, R. Grange, D. Psaltis, and C. Moser, "Nonlinear optical data transformer for machine learning," arXiv, arXiv:2208.09398 (2022).
36. X. Guo, T. D. Barrett, Z. M. Wang, and A. Lvovsky, "Backpropagation through nonlinear units for the all-optical training of neural networks," Photonics Res. **9**(3), B71–B80 (2021).