Regular Articles

Programming the scalable optical learning operator with spatial-spectral optimization

Yi Zhou [a,b,c,1], Jih-Liang Hsieh [a,b,1,*], Ilker Oguz [a,b], Mustafa Yildirim [a,b], Niyazi Ulas Dinc [a,b], Carlo Gigli [b], Kenneth K.Y. Wong [c], Christophe Moser [a], Demetri Psaltis [b]

[a] *Laboratory of Applied Photonics Devices, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[b] *Optics Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[c] *Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong*

A R T I C L E   I N F O

A B S T R A C T

Electronic computers have evolved drastically over the past years with an ever-growing demand for improved performance. However, the transfer of information from memory and high energy consumption have emerged as issues that require solutions. Optical techniques are considered promising solutions to these problems with higher speed than their electronic counterparts and with reduced energy consumption. Here, we use the optical reservoir computing framework we have previously described (Scalable Optical Learning Operator or SOLO [1]) to program the spatial-spectral output of the light after nonlinear propagation in a multimode fiber. The novelty in the current paper is that the system is programmed through an output sampling scheme, similar to that used in hyperspectral imaging in astronomy. Linear and nonlinear computations are performed by light in the multimode fiber and the high dimensional spatial-spectral information at the fiber output is optically programmed before it reaches the camera. We then used a digital computer to classify the programmed output of the multi-mode fiber using a simple, single layer network. When combining front-end programming and the proposed spatial-spectral programming, we were able to achieve 89.9 % classification accuracy on the dataset consisting of chest X-ray images from COVID-19 patients. At the same time, we obtained a decrease of 99 % in the number of tunable parameters compared to an equivalently performing digital neural network. These results show that the performance of programmed SOLO is comparable with cutting-edge electronic computing platforms, albeit with a much-reduced number of electronic operations.

1. Introduction

Growing demand for artificial intelligence (AI) has evolved dramatically in recent years. The AI revolution is fueled by the immense parallel computing power of electronic hardware such as field-programmable gate arrays and graphic and tensor processing units [2–5]. However, the performance is inherently limited by the fundamental tradeoff between energy efficiency and computing power in electronic computing [6]. The "bigger is better" mentality has dominated, but the energy required to operate large networks becomes a limiting factor. Moreover, as the scale of an electronic transistor approaches its physical limit, it is necessary to investigate and develop new computing processors during the post-Moore's law era [7,8]. Optical computing, using photons instead of electrons as the information carrier,

has the potential to provide high power efficiency, and high-speed processing [9,10].

The development of optical computing platforms has recently been the focus of intense research and commercial interest. Various optical neural network (ONN) architectures have been proposed, including diffractive optical neural networks [11,12], photonic reservoir computing [13,14], photonic spiking neural networks [15,16], and optical convolutional and recurrent neural networks [17,18]. An ONN with a computational speed of more than 10 TOPS has been demonstrated with today's advanced optical technologies [19]. The low energy consumption of ONNs is the key advantage of the technology. In fact, less than one photon per operation has been demonstrated [20], which can lead to orders of magnitude lower energy consumption than digital computation. Adaptive training of the ONNs with an error

backpropagation algorithm is the most common method to guarantee the model's reliable and accurate network inference. Examples of trained ONNs include holographic ONNs [21], physics-aware training of an optoelectronic network [22], a reconfigurable diffractive optoelectronic neural network with three layers [23], and hybrid training of optical neural networks [24].

Scalable Optical Learning Operator (SOLO) was previously proposed by our group to utilize the optical nonlinear propagation of a graded-index multimode fiber (GRIN MMF) to transform the input dataset into a less demanding representation space [1]. This architecture is inspired by the Extreme Learning Machine algorithm [25], in which a fixed nonlinear mapping of the input data is combined with a simple single trainable layer. In SOLO, the data is optically encoded via a spatial light modulator (SLM) and the nonlinear propagation inside an MMF realizes the nonlinear mapping. The advantage of performing the nonlinear mapping optically is that it can reduce the number of trainable parameters in a neural network and provides reasonable performance on many different datasets. The nonlinearity of the MMF affects the output beam shape and broadens the spectrum. As the information bandwidth increases, valuable information may be spatially or spectrally hidden in the output beam. The SOLO system collects the output beam simply with a camera, for which the number of pixels is much smaller than the spatial-spectral information contained in the beam, thus omitting the potential to further increase the classification performance. It can be expected that the system's classification accuracy depends strongly on the measuring approach. Therefore, we expect that optimizing the sampling of the MMF output beam could improve the performance of the system.

In this paper, we implemented the back-end programmed-SOLO (P-SOLO) system and tackled the image classification task on COVID-19 dataset. A digital micromirror device (DMD) placed at the output of the MMF is used to select spatial features by turning DMD pixels on or off. A chromatic dispersion grating is placed after the DMD to disperse the sampled light to achieve an optimized spatial–temporal measurement by the camera. We ran two sampling approaches. In the first approach (multi-line), we sampled the output beam of the MMF as a line, which is perpendicular to the orientation of the dispersion grating vector. The sampling line is swept across the DMD and capturing the sampled beam on the camera simultaneously. This generates multiple sampled images per input. In the second approach (single-shot), we optimized the DMD sampling pixels that generated only one sampled image per input. In the final experiment, we combined the front-end programming described in [26] with the proposed single-shot approach, and we were able to generate the best classification performance (89.9 %) of the COVID-19 dataset on an optical network.

2. Materials and methods

2.1. Experimental setup

The P-SOLO system (Fig. 1) includes an ultrafast laser source (Amplitude Laser Satsuma), whose pulses could be adjusted to have pulse widths from 700 fs to 10 ps and 125 kHz repetition rate. The pulse has a center wavelength of 1034 nm and spectral width of 7 nm. The linearly polarized laser output beam with a Gaussian beam shape encodes the dataset via a phase-only Spatial Light Modulator (SLM, Meadowlark). The SLM is a reflective liquid crystal on silicon, with 1920×1152 pixels, placed with a pitch of 9.2 µm and an 8-bit dynamic range operated at a frame rate of 50 Hz. The laser beam was expanded to cover 520×520 pixels on the SLM, which was used to input the dataset images. The pulses then propagate in a 5 m GRIN 50/125 MMF with a numerical aperture (NA) of 0.2 that supports 240 modes at the operation wavelength. The phase-modulated light from the SLM is coupled to the MMF through a lens with a focal length of 36 mm. The coupled beam covers the whole MMF core area. The output beam pattern from the MMF was collimated and then directed to a DMD. The DMD (Ajile AJD-4500) consists of 912×1140 micromirrors with a pitch of 7.6 µm. The light sampled and reflected by the DMD is then dispersed by a diffraction grating with a 1200 line mm^{-1} period. The dispersed beam after the grating is recorded by a camera with a 6.9 µm pixel pitch and a maximum frame rate of 552 Hz (BFS-U3-04S2M-CS). We program the DMD mask through a surrogate optimization algorithm [27] to find the optimal spatial sampling pattern. The beam is attenuated with an adjustable neutral density filter before camera detection. The input pulse peak power was optimized to maximize the nonlinear interaction of multimodes in MMF as in the SOLO system [1].

2.2. Physical model of SOLO [1]

In the computational framework of the SOLO system, nonlinear mapping is utilized as a physical realization of reservoir computing that takes place when the input pattern propagates in an MMF with high pulse intensity. During the pattern propagation in the MMF, the initial complex modal coefficients evolve according to spatiotemporal linear and nonlinear effects. The nonlinear transformation of information is achieved by nonlinear energy redistribution among the fiber modes. Further, the nonlinear mapping performed by the MMF can be described by the time-dependent beam propagation method (TBPM) involving the
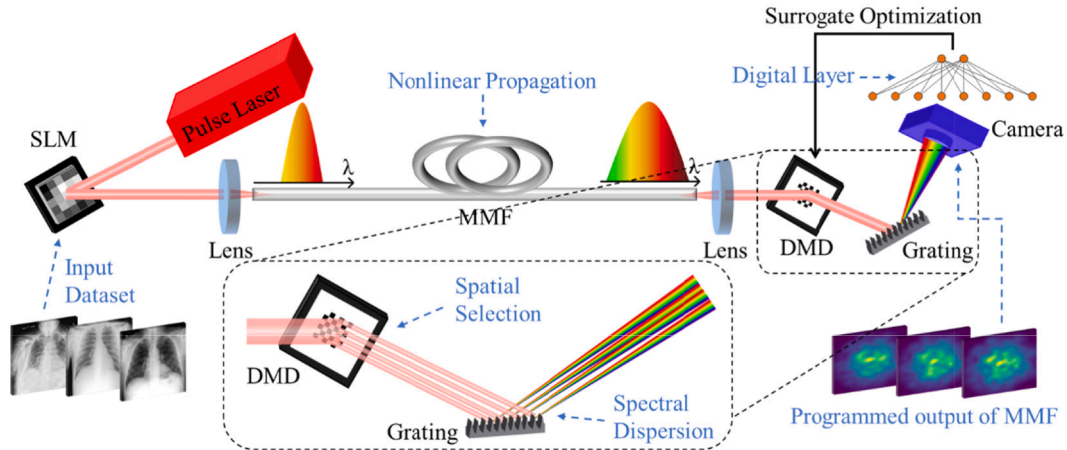

Fig. 1. Schematic of P-SOLO experimental setup. The experimental setup comprised an SLM for encoding dataset images onto laser pulses, an MMF for nonlinear transformation with spatial–temporal optical nonlinearities, a DMD as a space mask for programming combined with a surrogate optimization algorithm to extract spatial feature map, and a grating as the spectral frequency-resolved beam profile measurement technique to further improve classification accuracy.

fiber mode amplitudes (A_p, where p is the considered mode index). In an ideal situation with low power pulse or continuous-wave light input, only the phases of modes coefficients change at different rates due to modal and chromatic dispersion, which does not exist in intermodal power redistribution. This behavior corresponds to a linear transformation of the field when it propagates through the fiber, as shown in the first term of equation (1). Mode-coupling caused by perturbations due to fiber bending or impurities also acts as liner mixing. When the pulse peak power is sufficient to induce nonlinear effects in MMF, nonlinear mode coupling will lead to a nonlinear operation on the information spatially encoded in the input pulse throughout the fiber (the third term in equation (1)). Then, by simply iterating equation (1) with a small propagation step (Δz), nonlinear propagation can be implemented for each step, and the fiber modes will couple to each other according to the linear coupling coefficients and the nonlinear coupling tensor (η). This nonlinear operation can also be numerically modeled at each propagation step by multiplying each three-element combination of mode coefficients with the related entry of the nonlinear mode coupling tensor. The specific definitions of symbols in equation (1) can be found in [1].

$$\frac{\partial A_p}{\partial z} = \underbrace{i\partial\beta_0^p A_p - \partial\beta_1^p \frac{\partial A_p}{\partial t} - i\frac{\beta_2}{2}\frac{\partial^2 A_p}{\partial t^2}}_{\text{Dispersion}} + \underbrace{i\sum_n C_{p,n} A_n}_{\text{Linear mode coupling}} + \underbrace{i\gamma \sum_{l,m,n} \eta_{p,l,m,n} A_l A_m A_n^*}_{\text{Nonlinear mode coupling}}$$

(1)

2.3. Multi-line mask sampling approach

In the experiment, we first performed a sequential line sweep method across the output beam after the MMF. The COVID-19 dataset [28] was served as comparing metrics with 3000 X-ray samples (2400 X-ray samples for training and 600 X-ray samples for testing). The accuracy previously reported in SOLO was 83 % [1]. The X-ray sample is encoded on the phase of the input beam through the SLM while the corresponding fiber output intensity patterns are spatially sampled by the DMD. The distance of the DMD, grating, and camera is 54 cm and 17 cm, respectively (Fig. 2a). The line width of DMD is a parameter that needs to be optimized. In the first experiment, the line mask of DMD is designed to be perpendicular to the dispersion grating vector, meaning that the spatial sampling and spectral sampling are perpendicular to each other. The vertical DMD line masks, found to be optimized at 48 pixel width (see Supplementary), are horizontally swept at 4 sequential locations to spatially sample the MMF output (4H configuration, Fig. 2b). One sample in the dataset will generate 4 intensity images recorded by the camera. After concatenating these 4 images, the digital layer obtains a test accuracy of 86.2 % for 4H configuration.

In the second experiment, the DMD line masks are rotated. The horizontal DMD line masks, found to be optimized at 58 pixel width (see Supplementary), are vertically swept at 5 sequential locations and the camera recorded 5 corresponding intensity images (5 V configuration, Fig. 2c). The test accuracy for 5 V configuration is 85.9 %. In 5 V configuration, spatial sampling and spectral sampling are parallel to each other, thus, feature information hidden in the beam is not extracted as well as 4H configuration. However, with the help of free-space diffraction and grating dispersion, this accuracy is still higher than the 83 % reported previously from the SOLO system [1]. If we further combine the 9 images from 4H and 5 V configuration (4H + 5 V configuration, Fig. 2d), the classification accuracy of P-SOLO can go up to 88 %, surpassing the 87.3 % accuracy obtained by LeNet-5 (82826 training parameter) [29] digital neural networks. The improvement is summarized in Fig. 3 together with the accuracy previously reported in SOLO [1].

The distance between the DMD, grating, and camera, as indicated in Fig. 2a plays an important role since it determines the degree to which the nonlinear spectral components of the field propagate in free-space and are separately recorded on the camera. In the third experiment, we manually optimized the distances between optics components, and
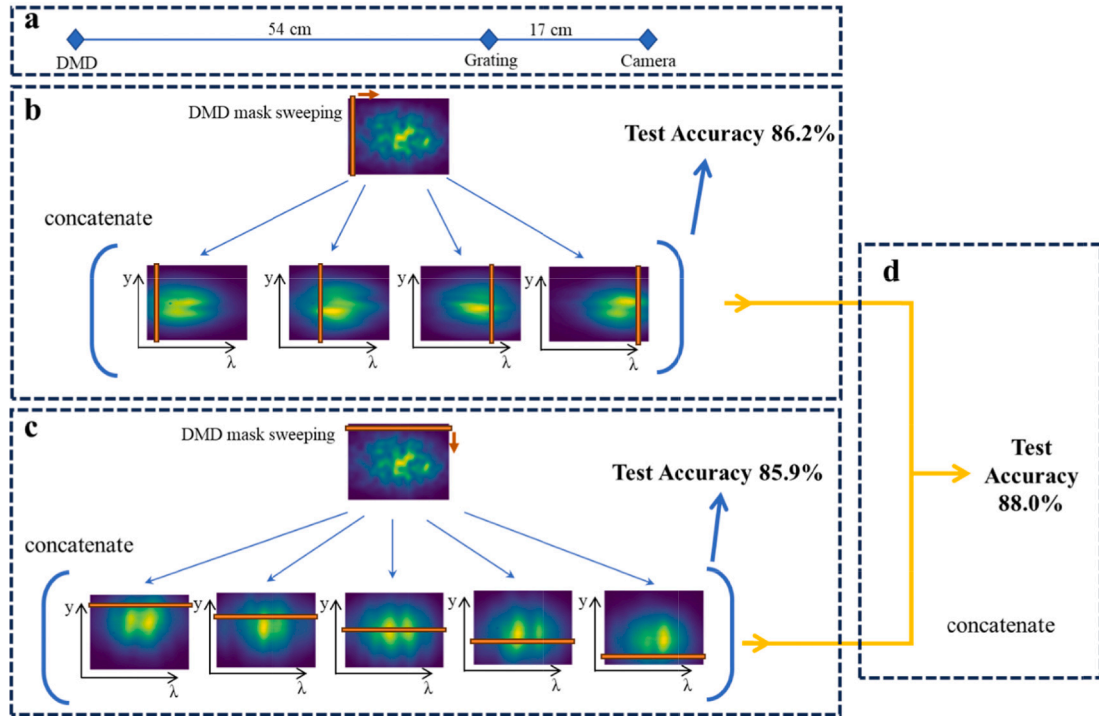

Fig. 2. Multi-line mask sampling approach with different experimental configurations. (a) Distance between the DMD, the grating and the camera. (b) 4 horizontally sweeping vertical DMD lines (4H configuration). (c) 5 vertically sweeping horizontal DMD lines (5 V configuration). (d) Combing both DMD line orientation (4H + 5 V configuration).
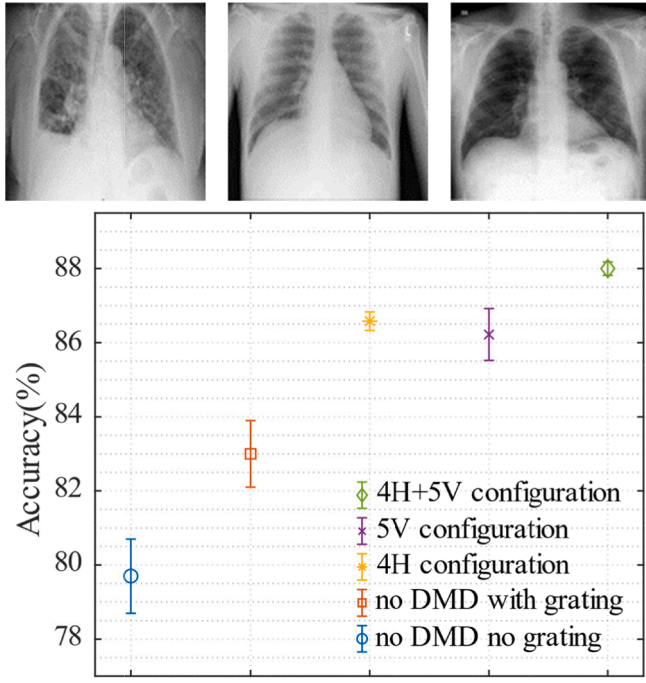
Fig. 3. Comparison of classification accuracy for the COVID-19 X-ray dataset under different multi-line mask sampling. 4H, 5 V, and 4H + 5 V are the methods proposed in this study. No DMD no grating and no DMD with grating are previously reported in SOLO [1].

we were able to improve the classification accuracy for 4H configuration from 86.2 % to 88 % (4H + optimized distance configuration, Fig. 4). This accuracy is as high as the 4H + 5 V configuration while it only required concatenating 4 images instead of 9 images. This could reduce the number of training parameters in the digital decision layer of P-SOLO and shorten the experiment time. This result also demonstrates that free-space diffraction also plays a vital role in the spatial-spectral sampling in the P-SOLO system.

2.4. Single-shot mask approach

Next, we optimized the sampling operation by programming a single DMD mask. The DMD is divided into several superpixels and the MMF output beam covers 80 superpixels. To avoid overlapping between the superpixels at the camera plane, a spacing parameter between the superpixel (zero-padding) was first introduced. A smaller COVID-19 dataset was used (1200 X-ray samples for training and 300 X-ray samples for testing) to efficiently determine the spacing parameter of DMD superpixel. A total of 80 (4×20) superpixels are encoded on the DMD, each superpixel consisting of multiple DMD binary pixels. The height of the superpixel contains 30 DMD pixels. The width of the superpixel is a parameter that needs to be optimized in a range from 16 to 48 DMD pixels and it is determined through surrogate optimization. Specifically, 80 parameters in the optimization algorithm control the on/off of the DMD superpixels, and another 5 binary parameters control the width of the superpixel (a variable between 16 and 48). During surrogate optimization, classification accuracy is generated by the digital decision layer and acts as the objective function for optimization.

The result of the optimized mask indicates that the best width for each superpixel is 48 pixels, which means there is no zero-padding between each superpixels. We next applied the optimized DMD mask on full-size COVID-19 dataset (3000 X-ray samples of COVID-19 dataset with 2400 training samples and 600 testing samples). The iteration of the surrogate optimization is shown in Fig. 5. The accuracy gradually increases with the number of iterations and the best test accuracy is found to be 86.9 % after 400 iterations. The scatter in Fig. 5 indicates the testing accuracy as a function of each iteration number and the blue dashed line represents the best accuracy obtained up to this point. The initial and optimized DMD masks (black superpixel corresponds to the on state and white superpixel corresponds to the off state) are also shown in Fig. 5.

2.5. Combining front-end programming and back-end sampling programming

We then performed the joint programming of the front-end (wave-front shaping of the input data that manipulates the nonlinear
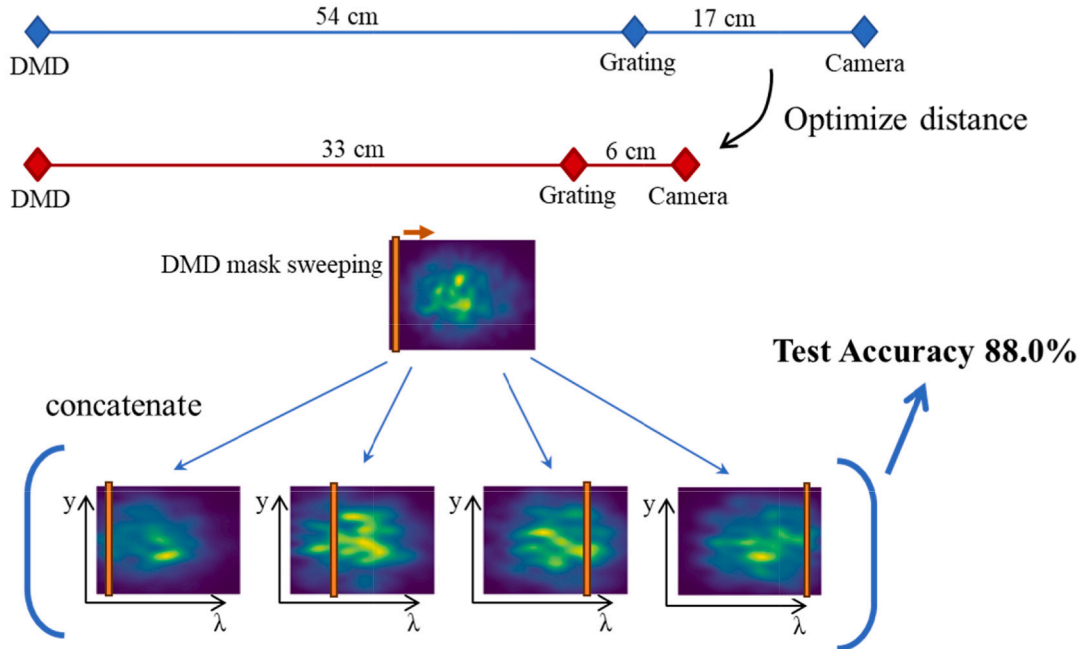

Fig. 4. The classification accuracy improvement after optimizing the distance between the DMD, the grating, and the camera. The optimized distance is shown at the top of the figure. The accuracy obtained is as high as 4H + 5 V configuration while only requiring 4 images instead of 9 images.
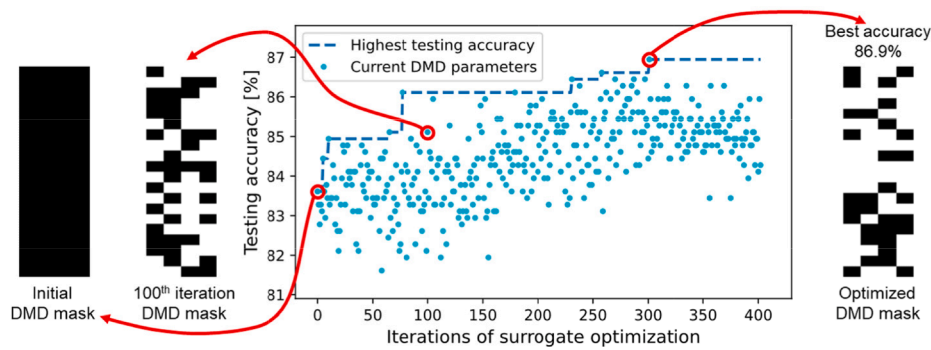
Fig. 5. Classification accuracy evolution of 3000 samples of COVID-19 X-ray dataset by programming single-shot DMD mask with surrogate optimization. Single-shot DMD mask divides into four lines including 80 superpixels with fixed height (30 DMD pixels) and width (48 DMD pixels). On the left and right shows the initial, the 100th iteration and the optimized mask displayed on the DMD.

Table 1
Summary of performance of optical and digital implementations of networks for the classification of the COVID-19 x-ray images database. The brackets style indicates the type of programmable parameters: in round bracket () is digital decision layer, in curly bracket {} is DMD superpixel, and in angle bracket $\langle \rangle$ is front-end programming.

Configuration	# of Training Parameter	Test Accuracy
SOLO [1]	$(22 \times 22 + 1)$	83%
4 horizontally sweeping multi-line DMD (4H)	$(4 \times 22 \times 22 + 1) = 1{,}937$	86.2%
5 vertically sweeping multi-line DMD (5V)	$(5 \times 22 \times 22 + 1) = 2{,}421$	85.9%
LeNet-5 [29]	82,826	87.3%
4H + 5V	$(9 \times 22 \times 22 + 1) = 4{,}357$	88.0%
4H + optimized distance	$(4 \times 22 \times 22 + 1) = 1{,}937$	88.0%
Single-shot DMD	$(30 \times 30 + 1) + \{80\} = 981$	86.9%
EfficientNet-B6 [30]	43,322,513	89.9%
Single shot DMD + front-end programming [26]	$(30 \times 30 + 1) + \{52\} + \langle 50 \rangle = 1{,}003$	89.9%

interaction of MMF [26]) and the single-shot mask optimization of the DMD at the MMF output. We obtained 89.9 % classification accuracy, up from 86.9 % with only the back-end single-shot DMD programming. This accuracy is remarkably the same as a pretrained EfficientNet-B6 architecture with 43,322,513 parameters and followed by a digital decision layer [30]. However, in our case, only 1003 training parameters were used, a reduction of 99 % in training parameters. In the table below, we summarize the performance achieved by the different approaches, digital network and optical network, in terms of classification accuracy and the number of training parameters (Table 1).

3. Conclusion

The P-SOLO system in this work was achieved by spatial-spectral optimization based on DMD sampling and grating dispersion placed at the MMF output. We optimized the binary mask on the DMD to extract important features for classification tasks. The advantage of our P-SOLO system was compared and demonstrated with respect to digital neural networks. The performance of the P-SOLO system might be further improved with mechanically perturbed MMF through polarization control. By combining the wavefront shaping method in [26] and our proposed sampling method at the MMF output, we achieved a remarkable classification accuracy of 89.9 % while reducing training parameters by 99 % compared to an equally performing digital neural network.

In summary, we have experimentally demonstrated a P-SOLO system that can be used for a range of learning task classifications. The programming single-shot spatial mask or multiple-line sweep method combined with spectral dispersion has achieved accuracy improvement for classifying COVID-19 X-ray images. The DMD mask only achieves amplitude modulation for the output beam. We anticipate that using amplitude and phase modulation simultaneously might further improve the performance.

CRediT authorship contribution statement

Yi Zhou: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jih-Liang Hsieh:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ilker Oguz:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mustafa Yildirim:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Niyazi Ulas Dinc:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Carlo Gigli:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Kenneth K.Y. Wong:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Christophe Moser:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Demetri Psaltis:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yofte.2024.103864.

References

[1] U. Teğin, M. Yıldırım, İ. Oğuz, C. Moser, D. Psaltis, Scalable optical learning operator, Nat. Comput. Sci. 1 (2021) 542–549, https://doi.org/10.1038/s43588-021-00112-0.

[2] C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, J. Cong, Optimizing FPGA-based Accelerator Design for Deep Convolutional Neural Networks, in: Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, Association for Computing Machinery, New York, NY, USA, 2015: pp. 161–170. doi: 10.1145/2684746.2689060.

[3] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2012. https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (accessed February 20, 2023).

[4] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.

[5] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016: pp. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html (accessed February 20, 2023).

[6] B. Marr, B. Degnan, P. Hasler, D. Anderson, Scaling energy per operation via an asynchronous pipeline, IEEE Trans. Very Large Scale Integr. VLSI Syst. 21 (2013) 147–151, https://doi.org/10.1109/TVLSI.2011.2178126.

[7] J.M. Shainline, S.M. Buckley, R.P. Mirin, S.W. Nam, Superconducting optoelectronic circuits for neuromorphic computing, Phys. Rev. Appl. 7 (2017) 034013, https://doi.org/10.1103/PhysRevApplied.7.034013.

[8] P.R. Prucnal, B.J. Shastri (Eds.), Neuromorphic Photonics, CRC Press, Boca Raton, 2017, https://doi.org/10.1201/9781315370590.

[9] D. Woods, T.J. Naughton, Photonic neural networks, Nat. Phys. 8 (2012) 257–259, https://doi.org/10.1038/nphys2283.

[10] D.R. Solli, B. Jalali, Analog optical computing, Nat. Photon 9 (2015) 704–706, https://doi.org/10.1038/nphoton.2015.208.

[11] X. Lin, Y. Rivenson, N.T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, A. Ozcan, All-optical machine learning using diffractive deep neural networks, Science 361 (2018) 1004–1008, https://doi.org/10.1126/science.aat8084.

[12] T. Yan, J. Wu, T. Zhou, H. Xie, F. Xu, J. Fan, L. Fang, X. Lin, Q. Dai, Fourier-space diffractive deep neural network, Phys. Rev. Lett. 123 (2019) 023901, https://doi.org/10.1103/PhysRevLett.123.023901.

[13] G.V. der Sande, D. Brunner, M.C. Soriano, Advances in photonic reservoir computing, Nanophotonics 6 (2017) 561–576, https://doi.org/10.1515/nanoph-2016-0132.

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

[15] J. Feldmann, N. Youngblood, C.D. Wright, H. Bhaskaran, W.H.P. Pernice, All-optical spiking neurosynaptic networks with self-learning capabilities, Nature 569 (2019) 208–214, https://doi.org/10.1038/s41586-019-1157-8.

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

[17] M. Miscuglio, Z. Hu, S. Li, J.K. George, R. Capanna, H. Dalir, P.M. Bardet, P. Gupta, V.J. Sorger, Massively parallel amplitude-only Fourier neural network, Optica 7 (2020) 1812–1819, https://doi.org/10.1364/OPTICA.408659.

[18] J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, D. Brunner, Reinforcement learning in a large-scale photonic recurrent neural network, Optica 5 (2018) 756–760, https://doi.org/10.1364/OPTICA.5.000756.

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

[20] T. Wang, S.-Y. Ma, L.G. Wright, T. Onodera, B.C. Richard, P.L. McMahon, An optical neural network using less than 1 photon per multiplication, Nat. Commun. 13 (2022) 123, https://doi.org/10.1038/s41467-021-27774-8.

[21] D. Psaltis, D. Brady, K. Wagner, Adaptive optical networks using photorefractive crystals, Appl. Opt., AO 27 (1988) 1752–1759, https://doi.org/10.1364/AO.27.001752.

[22] L.G. Wright, T. Onodera, M.M. Stein, T. Wang, D.T. Schachter, Z. Hu, P.L. McMahon, Deep physical neural networks trained with backpropagation, Nature 601 (2022) 549–555, https://doi.org/10.1038/s41586-021-04223-6.

[23] T. Zhou, X. Lin, J. Wu, Y. Chen, H. Xie, Y. Li, J. Fan, H. Wu, L. Fang, Q. Dai, Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit, Nat. Photonics 15 (2021) 367–373, https://doi.org/10.1038/s41566-021-00796-w.

[24] J. Spall, X. Guo, A.I. Lvovsky, Hybrid training of optical neural networks, Optica 9 (2022) 803–811, https://doi.org/10.1364/OPTICA.456108.

[25] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70 (2006) 489–501, https://doi.org/10.1016/j.neucom.2005.12.126.

[26] I. Oguz, J.-L. Hsieh, N.U. Dinc, U. Teğin, M. Yildirim, C. Gigli, C. Moser, D. Psaltis, Programming nonlinear propagation for efficient optical learning machines, AP 6 (2024) 016002, https://doi.org/10.1117/1.AP.6.1.016002.

[27] D. Eriksson, D. Bindel, C.A. Shoemaker, pySOT and POAP: An event-driven asynchronous framework for surrogate optimization, (2019). doi: 10.48550/arXiv.1908.00420.

[28] COVID-19 Radiography Database, (n.d.). https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database (accessed March 4, 2024).

[29] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324, https://doi.org/10.1109/5.726791.

[30] M. Tan, Q. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, in: Proceedings of the 36th International Conference on Machine Learning, PMLR, 2019: pp. 6105–6114. https://proceedings.mlr.press/v97/tan19a.html (accessed February 20, 2023).