Chinese Physics Letters **40**, 034201 (2023) **Express Letter**

Optical Neural Network Architecture for Deep Learning with Temporal Synthetic Dimension

Bo Peng(彭擘)[1†], Shuo Yan(颜硕)[1†], Dali Cheng(成大立)[2], Danying Yu(俞丹英)[1], Zhanwei Liu(刘展维)[1], Vladislav V. Yakovlev[3], Luqi Yuan(袁璐琦)[1*], and Xianfeng Chen(陈险峰)[1,4,5]

[1]*State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*Ginzton Laboratory and Department of Electrical Engineering, Stanford University, Stanford, CA 49305, USA*
[3]*Texas A&M University, College Station, Texas 77843, USA*
[4]*Shanghai Research Center for Quantum Sciences, Shanghai 201315, China*
[5]*Collaborative Innovation Center of Light Manipulation and Applications, Shandong Normal University, Jinan 250358, China*

(Received 18 November 2022; accepted manuscript online 1 February 2023)

The physical concept of synthetic dimensions has recently been introduced into optics. The fundamental physics and applications are not yet fully understood, and this report explores an approach to optical neural networks using synthetic dimension in time domain, by theoretically proposing to utilize a single resonator network, where the arrival times of optical pulses are interconnected to construct a temporal synthetic dimension. The set of pulses in each roundtrip therefore provides the sites in each layer in the optical neural network, and can be linearly transformed with splitters and delay lines, including the phase modulators, when pulses circulate inside the network. Such linear transformation can be arbitrarily controlled by applied modulation phases, which serve as the building block of the neural network together with a nonlinear component for pulses. We validate the functionality of the proposed optical neural network for the deep learning purpose with examples handwritten digit recognition and optical pulse train distribution classification problems. This proof of principle computational work explores the new concept of developing a photonics-based machine learning in a single ring network using synthetic dimensions, which allows flexibility and easiness of reconfiguration with complex functionality in achieving desired optical tasks.

DOI: 10.1088/0256-307X/40/3/034201

Optical neural networks (ONNs) are under extensive studies recently with an ultimate goal of achieving machining learning in a photonic system.[1–12] Recent advancements have revealed that ONNs exhibit important computation capability with photonic tools[13–17] and training optical fields for some specific optimization purposes.[18] On the other hand, realizations of ONNs on different platforms also attract great interest from theoretical and computational perspectives. For example, training ONNs through *in situ* back propagation[19,20] and quantum ONNs can conduct the non-classical tasks.[21] In addition, the recurrent neural network,[22–26] as an important machine learning model, has been studied with the optical-based technologies.[27] Nevertheless, it has been found that most of the ONN designs depend on the number of photonic devices in each layer as well as the total layer number, which makes an ONN system require N^2 photonic devices with tunable externally controlled components and makes its practical implementation rather complex and lacks the freedom and options for further reconfiguration and miniaturization.[13–15] It is therefore important to investigate alternative photonic ONN design architectures, which can potentially offer enough freedom towards arbitrary functionality. Thus, it is essential to explore novel physical principles, and the approach based on synthetic dimensions offers an intriguing opportunity to overcome some of the existing challenges and limitations.

Synthetic dimension is a rapidly arising concept in photonics which facilitates utilization of different degrees of freedom of light to simplify experimental arrangements and to get the most out of those.[28–32] Recently, it has been suggested that an ONN with synthetic dimensions can potentially provide simpler design of the ONN to achieve a complicated functionality.[33–35] However, the proper implementation of those appeared to be challenging. In this Letter, we investigate the time-multiplexed architecture using temporal information[36–41] that has been demonstrated as a highly promising way for optical computations such as coherent Ising machines,[39] photonic reservoir computing,[42] and ONNs with synthetic nonlinear lattices.[43]

In this work, we introduce and validate through computational experiments a new paradigm to achieve the optical neural network in a single resonator network, with the

†These authors contributed equally to this work.
*Corresponding author. Email: yuanluqi@sjtu.edu.cn

temporal synthetic dimension constructed by connecting different temporal positions of pulses with pairs of delay lines. Different from pioneering works in Refs. [33,43] that proposed ONNs with synthetic lattices in coupled rings, the proposed approach in this study offers an alternative solution to the ONN problem in a single ring. The optical resonator network with reconfigurable couplings between different arrival times (i.e., temporal positions) of optical pulses supports time-multiplexed lattice [39] and creates the temporal synthetic dimension. With controllable splitters and phase modulators used to build desired connections between pulses, we show the way of constructing multiple layers of an ONN in a single resonator [see Fig. 1(a)]. A nonlinear operation is used to perform complex modulations which are being controlled by external signals with the aid of a computer. As validations for the deep-learning functionality, we perform the training of the proposed platform for the ONN with the training data set of MNIST handwritten digit database with appropriate noises considered. [44] The striking feature of our ONN is that it needs only one resonator but gives arbitrary size of layers in the network, which makes our system unlimited in the total layer (roundtrip) number with high reconfigurability. Moreover, this single resonator network is capable of conducting arbitrary optical tasks, after performing the proper training. For example, we conduct a pulse train classification problem, which recognizes different distributions of pulse trains. Our work hence points out a concept for realizing the ONN with synthetic dimensions, which is highly scalable and therefore gives the extra freedom for further simplification of the setup with possible reconfiguration.

Model. We start by considering a resonator composed of the main cavity loop of the waveguide [see Fig. 1(a)]. By neglecting the group velocity dispersion of the waveguide, we assume that there are N optical pulses simultaneously propagating inside the loop, and every two nearby pulses is temporally separated by a fixed time Δt. Each pulse is labelled by its temporal position t_n (or arrival time, with $t_{n+1} - t_n = \Delta t$), [39] and we use $n = 1, \ldots, N$ to denote each pulse at different temporal positions.

To construct the temporal synthetic dimension, we add a pair of delay lines, which are connected with the main loop through splitters and couplers. Each splitter is controlled by parameter $\phi_{1(2)}$, which determines that a portion of the pulse with the amplitude $\cos\phi_{1(2)}$ remains in the main loop while the rest of the pulse with the amplitude $i\sin\phi_{1(2)}$ gets into the delay line. [36,37] Lengths of delay lines are carefully designed. For the pulse at the temporal position n propagating through the shorter delay line, it combines into the main loop at a time Δt ahead of its original arrival time t_n and contributes to the pulse at the time $t_{n-1} = t_n - \Delta t$, i.e., $\Delta n = -1$. On the other hand, for the pulse propagating through the longer delay line, it combines into the main loop at a time Δt behind t_n and contributes to the pulse at $t_{n+1} = t_n + \Delta t$, i.e., $\Delta n = +1$. Such a design constructs the temporal synthetic dimension [see Fig. 1(b)], where the n-th pulse during the m-th roundtrip with the amplitude $A(n, m)$ (in units of a reference amplitude A_0) is connected to its nearest neighbor sites in the temporal synthetic lattice after each roundtrip. The boundary of this lattice can be created by further introducing the intracavity intensity modulator to suppress unwanted pulses in the main loop. [45]

We place phase modulators inside the main loop as well as two delay lines. Each phase modulator is controlled by external voltage and adds a modulation phase θ_i ($i = 1, 2, 3$) for the pulse propagating through it. [39,45] Moreover, we use the complex modulator as the nonlinear component, which can convert the input pulse to an output pulse with a complex nonlinear function. In such an ONN, parameters ϕ_i and θ_i can be precisely controlled at any time, meaning that one can manipulate ϕ_i and θ_i for each pulse n at each roundtrip number m.

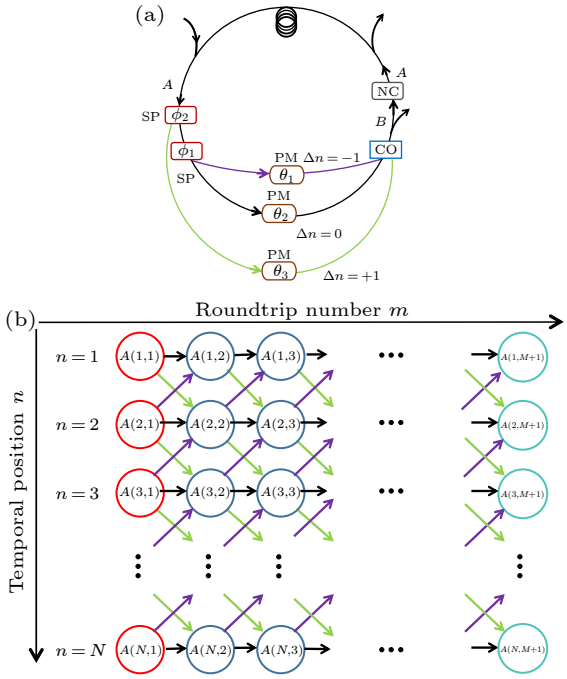

Fig. 1. (a) The schematic of the single resonator network with two delay lines in purple and green respectively. CO: combiner, SP: splitter, PM: phase modulator, NC: nonlinear component. A denotes the field amplitude while B denotes the output amplitude defined in Eq. (1). (b) The connectivity of the synthetic photonic lattice along the temporal dimension (n-axis) implemented in (a) for pulses evolving after roundtrips (m). A number of N pulses in each roundtrip (shown in circles) are considered and the pulses evolve for M roundtrips in total, which therefore construct the ONN with M layers and N neurons sites in each layer. Green, black, and purple arrows correspond to different optical branches of delay lines in (a).

In this temporal synthetic lattice, the propagation process of pulses in each single roundtrip can compose the linear transformation, described by [36,37]

$$\begin{aligned} B(n,m) = {} & A(n,m)\cos\phi_1(n,m)\cos\phi_2(n,m)e^{i\theta_2(n,m)} \\ & - iA(n+1,m)\sin\phi_2(n+1,m)e^{i\theta_3(n+1,m)} \\ & - iA(n-1,m)\cos\phi_2(n-1,m) \\ & \sin\phi_1(n-1,m)e^{i\theta_1(n-1,m)}, \end{aligned} \tag{1}$$

where $B(n, m)$ denotes output amplitudes for the set of pulses after the linear transformation. A very small portion of pulses are dropped out and collected by detectors,

which are stored in the computer for the further analysis. The pulses then pass the nonlinear component where we use a formula similar to a saturable absorber [46,47] but with amplitudes, so a complex nonlinear operation is performed

$$2B(n,m)(1-T_{n,m})/A_0 = \ln(T_{n,m}), \tag{2}$$

$$A(n,m+1) = B(n,m)T_{n,m}. \tag{3}$$

For a given input pulse $B(n,m)$, the nonlinear coefficient $T_{n,m}$ can be calculated in the computer with Eq. (2), and then appropriate external signal is applied to the complex modulator [48] so the output pulse after the nonlinear component follows Eq. (3), which turns out to be the input pulse $A(n,m+1)$ for the next layer (the next roundtrip). We find that this particular choice of the complex nonlinear function works extremely well, compared to regular real nonlinear activation functions such as sigmoid function or hyperbolic tangent function.

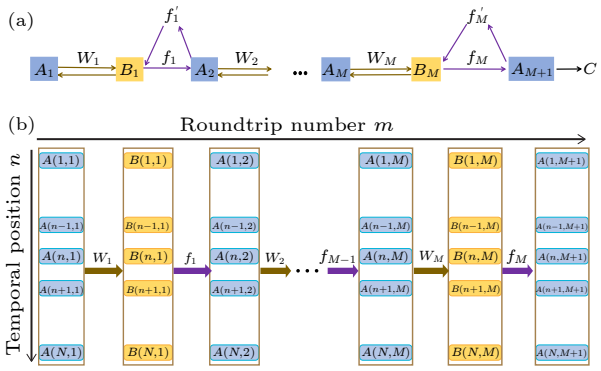

Fig. 2. (a) Schematic of the architecture of an optical neural network. A_1 is the vector of pulses imported in the first layer when training starts. A_m: vector of the output pulses after the $(m-1)$-th roundtrip (layer), which is also the input vector for the m-th roundtrip (layer); W_m: matrix for the linear transformation during the m-th roundtrip (layer); B_m: vector of pulses after the linear transformation during the m-th roundtrip (layer); f_m: nonlinear activation operation; f_m': derivative of f_m during back propagation. C is the cost function for the output signal. (b) Illustration of the signal flow through roundtrips (layers) in the resonator in Fig. 1(a).

Figure 2 summarizes the forward transmissions with linear transformations and nonlinear operations on pulses. Theoretically, the total number of layers, M as well as the total pulse number N, can be arbitrary. In Fig. 2, we use W_m to define the linear transformation in Eq. (1) and f_m to define the nonlinear operation in Eqs. (2) and (3) for the m-th roundtrip. Hence the forward transmission at each layer m follows $B_m = W_m A_m$ and $A_{m+1} = f_m B_m$, where A_m and B_m are vectors of $A(n,m)$ and $B(n,m)$, respectively. Pulse information $A(n,m+1)$ $(B(n,m))$ after (before) the nonlinear operation at the n-th temporal position during the m-th roundtrip is collected by dropping a small portion of pulses out of the resonator network into detectors. Such information of A_m and B_m is stored in the computer for further backward propagation in training the ONN.

Once the forward propagation is finished after M roundtrips in the optical resonator network, the backward propagation can be performed in the computer following the standard procedure to correct control parameters, [19,49] which is briefly summarized here. The backward propagation equations read [19,49]

$$\tilde{B}_m = B_m + f_m'(A_{m+1} - \tilde{A}_{m+1}), \tag{4}$$

$$\tilde{A}_m = W_m^T \tilde{B}, \tag{5}$$

with $\tilde{A}_m$ and $\tilde{B}_m$ being vectors at the m-th layer, calculated through the back propagation from the stored information of A_{m+1} and B_m. Here f_m' is the derivative of the nonlinear operation at the m-th layer in Eq. (4), W_m^T is the inverse of W_m, and $\tilde{A}_{M+1}$ is the target vector A_{target}, which is the expected output vector of the training set. The cost function after the m-th layer can therefore be calculated by

$$C_m = \frac{1}{2N}\sum_{i=1}^{N} |A(i,m+1) - \tilde{A}(i,m+1)|^2. \tag{6}$$

Throughout the backward propagation, optical controlling parameters $\phi_1(n,m)$, $\phi_2(n,m)$, $\theta_1(n,m)$, $\theta_2(n,m)$, and $\theta_3(n,m)$ can be trained by calculating the derivative of C_m with respect to these parameters, i.e.,

$$\frac{\partial C_m}{\partial \phi_{1,2}(n,m)} = [(A_{m+1} - \tilde{A}_{m+1})]^T \odot f_m' \cdot \frac{\partial W^T}{\partial \phi_{1,2}(n,m)} \cdot A_m, \tag{7}$$

$$\frac{\partial C_m}{\partial \theta_{1,2,3}(n,m)} = [(A_{m+1} - \tilde{A}_{m+1})]^T \odot f_m' \cdot \frac{\partial W^T}{\partial \theta_{1,2,3}(n,m)} \cdot A_m, \tag{8}$$

where $\odot$ is the vector multiplication, with $\boldsymbol{c} = \boldsymbol{a} \odot \boldsymbol{b}$ defined as $c_n = a_n b_n$. We can obtain the corrections of parameters as follows: [49]

$$\Delta\phi_{1,2}(n,m) = -a\frac{\partial C_m}{\partial \phi_{1,2}(n,m)}, \tag{9}$$

$$\Delta\theta_{1,2,3}(n,m) = -a\frac{\partial C_m}{\partial \theta_{1,2,3}(n,m)}, \tag{10}$$

with a being the learning rate for this training. Then $\phi_{1,2}(n,m)$ becomes $\phi_{1,2}(n,m) + \Delta\phi_{1,2}(n,m)$ and $\theta_{1,2,3}(n,m)$ becomes $\theta_{1,2,3}(n,m) + \Delta\theta_{1,2,3}(n,m)$. Following the backward propagation procedure summarized above, the parameters for controlling the forward propagation of each pulse at the n-th temporal position for the m-th roundtrip are updated backwardly from the M-th layer to the 1-st layer.

Having the entire procedure in hand, one can train an ONN with a training set of data to prepare the ONN ready for doing the designed all-optical computation with optical pulses in this single resonator network.

Results. To show the validity and reliability of our proposed ONN, we consider an MNIST handwritten digit recognition problem as commonly used for ONNs, [44] with noises included. The MNIST data set is chosen from the classic data set in the field of machine learning. It consists of 60000 training samples and 10000 test samples. Each sample is a 28×28 pixel grayscale handwritten digital picture, representing a number from 0 to 9. Some typical visualization legends are given in Fig. 3.

Fig. 3. Typical visualization legends from the MNIST dataset.

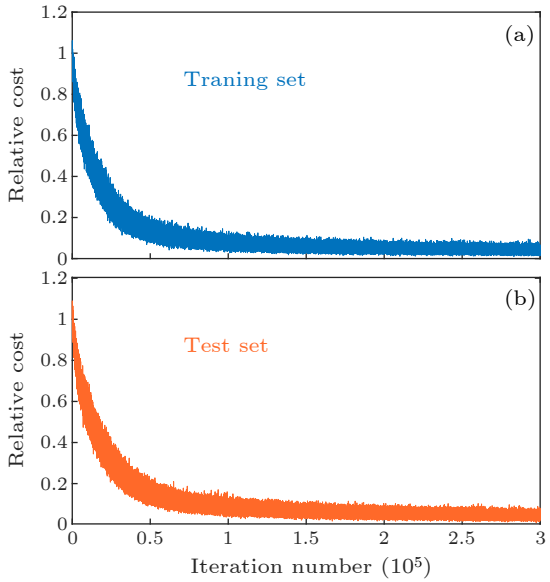

Fig. 4. Relative cost functions defined in Eq. (6) versus the training iteration number during the training process or (a) training set, and (b) test set, respectively, for the Hand written recognition problem.

In simulations, we use 49 pulses ($N = 49$) and 45 roundtrips ($M = 45$), with learning rate $a = 0.001$. For the simplicity purpose, we pre-process the original MNIST handwritten digit database,[44] where each input data supposes to have an array of 784 elements, with the maximum pooling twice,[50] so the input data can be mapped on 49 input pulses in our ONN architecture. Moreover, after the final roundtrip in the single resonator network, we add another full connection layer between collected signals from 49 pulses and 10 additional output sites, which shall be assisted by the computer. In this full connection layer on the computer, $a = 0.02$, and we use the sigmoid nonlinear function as the activation function. In Fig. 4, we plot the normalized cost function, defined in Eq. (6), for training set and test set versus the computation iteration number. Such a cost function based on the mean square error has been used in the literature for classification problems.[51–54] Both cost functions decreases as the iteration number increases. Therefore, the ONN in the temporal synthetic dimension works fine for the handwritten digit recognition problem. We emphasize that the preprocessing and the additional full connection layer make this model less competitive with previous ONNs,[5,7,55] and this simulation is only for the purpose of demonstrating the validity of our ONN and the stability with certain noises. To this end, in the test set, we add random noises on the 49 input pulses with their amplitudes multiplied by $1 + R \cdot \delta/2$, with $R \in (-0.5, 0.5)$ being a random number and δ denoting amplitude of noises, where we choose $\delta = 0$, 2%, 4%, 6%, 8%, 10%. Here, 60000 sets of training data and 10000 sets of test data are used for simulations. After training, noises with δ are appended into the ONN to carry out the test. We list errors of prediction in Table 1. One can see that the error of prediction in our ONN architecture is 21.1% if there is no noise in input pulses from the test set. However, when we add noises into the system, the error increases up to 29.7% for $\delta = 10\%$. Small noises may be tolerated in this proposed ONN architecture. However, large noises could affect the performance of the system, which may need further improvement in the future. Although the effects of noises in our proposed ONN architecture are difficult to be compared with those in other ONN systems due to the very different design associated to synthetic dimensions, typical experiments with time-multiplexed architecture can be carried out with small noises.[39]

Table 1. Errors of prediction for handwritten digit recognition with different noises.

δ in test set (%)	0	2	4	6	8	10
Error of prediction (%)	21.1	24.6	25.8	27.1	28.0	29.7

We have demonstrated the validity of the proposed ONN. One of the key importance of this proposal is to provide a possible trained optical network to act a certain photonic functionality intelligently. As a simple proof-in-principle verification, we perform a home-made optical pulse train distribution classification problem.

Our goal is to train an optical neural network to recognize five different profiles of optical pulse trains composed by 101 pulses, where shapes of five profiles are chosen as sinusoidal functions as $\sin(k\pi t_i/T)$ for pulse at temporal position t_i, with $T = 100\Delta t$ and $k = 1, 2, 3, 4, 5$ labeling five profiles, respectively. For both training and test procedures, each pulse is interrupted with noises. There are 30000 training sets and 5000 test sets constructed in simulations. Similar noise is used so that the pulse is modified in amplitude by $1 + R_{1(2)}\delta_{1(2)}$, where $\delta_{1(2)}$ is amplitude of noises in the training (test) sets and $R_{1(2)} \in (-0.5, 0.5)$ is a random number. The choice of $\delta_{1(2)}$ is listed in Table 2.

Table 2. Errors of prediction for optical pulse train distribution classification problems.

δ_2 (%)	δ_1(%)			
	0	2	4	6
0	1.1	16.6	24.4	29.8
2	43.7	18.7	26.6	30.8
4	68.9	19.9	26.7	30.9
14	75.4	30.7	29.4	32.4

In the simulations, 101 pulses ($N = 101$) and 31 roundtrips ($M = 31$) are chosen for the ONN, and after the final roundtrip, another full connection layer between 101 pulses and 5 output sites is used for predictions. For the training procedure, the learning rate a is 0.001 for $\delta_1 = 0$, 0.0017 for $\delta_1 = 2\%$, 0.021 for $\delta_1 = 4\%$, and 0.011 for $\delta_1 = 6\%$. The choice of $\delta_1 = 0$ results in the invalidation due to lack of data type (all training sets having same labels are identical), while the noise amplitude $\delta_1 = 6\%$ induces the complexity caused by high volatility and instability of our data. Therefore, for these two training

procedures, the test result has relatively high errors of prediction. Nevertheless, one can see from Table 2 that, for noise amplitudes $\delta_1 = 2\%$ and 4% in the training procedures, the errors of prediction in the test procedures show relatively good results (error of prediction $\lesssim 30\%$), even for high noise amplitudes $\delta_2 = 14\%$ in the test procedures.

The training process with zero noise of pulses in the training set is invalidated due to the monotonicity of the data set. However, in the case of low noise amplitudes of pulses in the training set, our ONN system shows a relatively stable prediction for optical pulse train distribution classifications. Furthermore, as another important feature, one can see that, for a larger noise amplitude δ_1 in the training set (for example, comparing $\delta_1 = 2\%$ and 4%), although it gives larger errors for smaller noise amplitudes δ_2 in the test set, one can obtain a smaller error (such as 30.7% and 29.4%) for the relatively large noise $\delta_2 = 14\%$. The example therefore shows the capability of our proposed ONN architecture in performing direct optical processing.

Discussion and Summary. The proposed platform is experimentally feasible with the state-of-the-art photonic technology. The fiber ring resonator with kilometer-long roundtrip length can be constructed with hundreds of temporal separated pulses circulating inside the resonator.[39,45] In particular, in Ref. [45] Leefmans *et al.* showed the capability for constructing an ONN consisting of 64 time-multiplexed optical resonant sites with pulses produced by an input 1550 nm mode-locked laser, separated by 4 ns, which points out an excellent possible experimental platform for realizing our theoretical proposal. Moreover, our proposal for achieving the temporal synthetic dimension can also be realized in a resonator with the free-space optics.[41] In both setups, delay lines (channels) are used to create the nearest-neighbor couplings along the temporal synthetic dimension. Moreover, appropriate delay lines (channels) can also connect pulses at time separations with double, triple, and/or high-order Δt, i.e., providing the long-range couplings. Our ONN therefore holds the possibility for generating more than three connectivities between sites in two layers in Fig. 1(b). It is possible to further increase the accuracy of the ONN. These delay lines may induce small errors, but as one sees in Tables 1 and 2, the synthetic ONN can tolerate small noises. The current nonlinear function in the proposal is performed in the computer. However, it is possible to consider nonlinear component operated by amplitude and phase modulations[48,56,57] or other nonlinear components,[43,58] which can perform alternative different complex nonlinear functions in optics. One notices that, in the proposed approach, the back propagation in the training process is conducted with a computer and then obtained optimal parameters are transferred to the physical system. Such *ex situ* training may bring extra errors, but is currently a reasonable strategy utilized in recent experiments for demonstrating ONN functionality.[3,5,7,14] In Ref. [19], Hughes *et al.* suggested a possible way to realize *in situ* backward propagation in optical systems, which may greatly improve speed in ONNs. The inclusion of such *in situ* backward propagation in our proposed ONN could be of interest for future research.

In summary, we have proposed a novel paradigm to achieve an ONN in a single resonator network. The proposed approach is based on a physical concept of the temporal synthetic dimension. As the proof of principle, we study the MNIST handwritten digit recognition problem to verify the validation of the deep learning functionality of our proposed ONN. Furthermore, we demonstrate the possibility of photonic intelligent features, by showing the performance of a home-made optical pulse train distribution classification problem. Our proposed ONN in the temporal synthetic dimension uses the trade-off between time and space complexity, and therefore does not have the advantages in energy and speed. However, the key achievement here is that we propose an alternative model with relatively high flexibility, which can be re-configurable and scalable on the number of sites (pulses) in each layer as well as the number of layers (roundtrips) for each computation. Distinguished from other relevant studies,[33,43] our proposal focuses one resonator supporting temporal synthetic dimension and shows the opportunity for constructing a flexible ONN that is capable for various optical tasks once getting trained. The construction in Fig. 1(b) can be easily linked to architectures of conventional neural networks with long-range connectivities added via additional delay lines, which can be further generalized to a recurrent neural network.[22–26] Furthermore, one can also prepare the set of pulses with the single-photon state instead,[41] which mat make our proposal with the temporal synthetic dimension possible for constructing the quantum neural network in the future study. Our work therefore shows the opportunity for constructing a flexible ONN in a single resonator, which points to a broad range of potential applications from all-optical computation to intelligent optical information processing[59] and biomedical imaging.[60,61]

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 12122407, 11974245, and 12192252), the Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01-ZX06). V. V. Y. acknowledges partial funding from NSF (Grant Nos. DBI-1455671, ECCS-1509268, and CMMI-1826078), AFOSR (Grant Nos. FA9550-15-1-0517, FA9550-18-1-0141, FA9550-20-1-0366, and FA9550-20-1-0367), DOD Army Medical Research (Grant No. W81XWH2010777), NIH (Grant Nos. 1R01GM127696-01 and 1R21GM142107-01), and the Cancer Prevention and Research Institute of Texas (Grant No. RP180588). L. Y. thanks the sponsorship from Yangyang Development Fund and the support from the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning.

References

[1] Rosenbluth D, Kravtsov K, Fok M P, and Prucnal P R 2009 *Opt. Express* **17** 22767
[2] Tait A N, Nahmias M A, Shastri B J, and Prucnal P R 2014 *J. Lightwave Technol.* **32** 4029
[3] Shen Y C, Harris N C, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S J, Larochelle H, Englund D, and Soljačić M 2017 *Nat. Photon.* **11** 441
[4] Tait A N, de Lima T F, Zhou E, Wu A X, Nahmias M A, Shastri B J, and Prucnal P R 2017 *Sci. Rep.* **7** 7430

[5] Lin X, Rivenson Y, Yardimci N T, Veli M, Luo Y, Jarrahi M, and Ozcan A 2018 *Science* **361** 1004
[6] Ying Z F, Wang Z, Zhao Z, Dhar S, Pan D Z, Soref R, and Chen R T 2018 *Opt. Lett.* **43** 983
[7] Feldmann J, Youngblood N, Wright C D, Bhaskaran H, and Pernice W H P 2019 *Nature* **569** 208
[8] Zuo Y, Li B, Zhao Y, Jiang Y, Chen Y C, Chen P, Jo G B, Liu J, and Du S 2019 *Optica* **6** 1132
[9] Hamerly R, Bernstein L, Sludds A, Soljačić M, and Englund D 2019 *Phys. Rev. X* **9** 021032
[10] Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, and Yu Z 2019 *Photon. Res.* **7** 823
[11] Zhang T, Wang J, Dan Y, Lanqiu Y, Dai J, Han X, Sun X, and Xu K 2019 *Opt. Express* **27** 37150
[12] Zhang H, Thompson J, Gu M, Jiang D, Cai H, Liu P Y, Shi Y, Zhang Y, Karim M F, Lo G Q, Luo X, Dong B, Kwek L C, and Liu A Q 2021 *ACS Photon.* **8** 1662
[13] Wetzstein G, Ozcan A, Gigan S, Fan S, Englund D, Soljačić M, Denz C, Miller D A B, and Psaltis D 2020 *Nature* **588** 39
[14] Nahmias M A, de Lima T F, Tait A N, Peng H T, Shastri B J, and Prucnal P R 2020 *IEEE J. Sel. Top. Quantum Electron.* **26** 7701518
[15] Bogaerts W, Pérez D, Capmany J, Miller D A B, Poon J, Englund D, Morichetti F, and Melloni A 2020 *Nature* **586** 207
[16] Xu X Y, Tan M X, Corcoran B, Wu J Y, Boes A, Nguyen T G, Chu S T, Little B E, Hicks D G, Morandotti R, Mitchell A, and Moss D J 2021 *Nature* **589** 44
[17] Feldmann J, Youngblood N, Karpov M, Gehring H, Li X, Stappers M, Gallo M L, Fu X, Lukashchuk A, Raja A S, Liu J, Wright C D, Sebastian A, Kippenberg T J, Pernice W H P, and Bhaskaran H 2021 *Nature* **589** 52
[18] Jiang J Q, Chen M K, and Fan J A 2021 *Nat. Rev. Mater.* **6** 679
[19] Hughes T W, Minkov M, Shi Y, and Fan S 2018 *Optica* **5** 864
[20] Zhou T K, Fang L, Yan T, Wu J M, Li Y P, Fan J T, Wu H Q, Lin X, and Dai Q H 2020 *Photon. Res.* **8** 940
[21] Steinbrecher G R, Olson J P, Englund D, and Carolan J 2019 *npj Quantum Inf.* **5** 60
[22] Connor J T, Martin R D, and Atlas L E 1994 *IEEE Trans. Neural Networks. Learn. Syst.* **5** 240
[23] Dorffner G 1996 *Neural Netw. World* **6** 447
[24] Hüsken M and Stagge P 2003 *Neurocomputing* **50** 223
[25] Yao K, Zweig G, Hwang M Y, Shi Y, and Yu D 2013 *Proceedings of Interspeech* pp 2524–2528
[26] Goodfellow I, Bengio Y, and Courville A 2016 *Deep Learning* (Cambridge: MIT) vol 1 pp 326–366
[27] Hughes T W, Williamson I A D, Minkov M, and Fan S 2019 *Sci. Adv.* **5** eaay6946
[28] Yuan L Q, Lin Q, Xiao M, and Fan S H 2018 *Optica* **5** 1396
[29] Ozawa T and Price H M 2019 *Nat. Rev. Phys.* **1** 349
[30] Lustig E and Segev M 2021 *Adv. Opt. Photon.* **13** 426
[31] Liu H, Yan Z, Xiao M, and Zhu S 2021 *Chin. Opt. Lett.* **41** 0123002
[32] Yuan L Q, Dutt A, and Fan S H 2021 *APL Photon.* **6** 071102
[33] Pankov A V, Sidelnikov O S, Vatnik I D, Sukhorukov A A, and Churkin D V 2019 *Proc. SPIE* **11192** 111920N
[34] Buddhiraju S, Dutt A, Minkov M, Williamson I A D, and Fan S 2021 *Nat. Commun.* **12** 2401
[35] Lin Z, Sun S, Azana J, Li W, Zhu N, and Li M 2020 arXiv:2009.03213 [eess.SP]
[36] Regensburger A, Bersch C, Hinrichs B, Onishchukov G, Schreiber A, Silberhorn C, and Peschel U 2011 *Phys. Rev. Lett.* **107** 233902
[37] Regensburger A, Bersch C, Miri M A, Onishchukov G, Christodoulides D N, and Peschel U 2012 *Nature* **488** 167
[38] Wimmer M, Regensburger A, Bersch C, Miri M A, Batz S, Onishchukov G, Christodoulides D N, and Peschel U 2013 *Nat. Phys.* **9** 780
[39] Marandi A, Wang Z, Takata K, Byer R L, and Yamamoto Y 2014 *Nat. Photon.* **8** 937
[40] Wimmer M, Price H M, Carusotto I, and Peschel U 2017 *Nat. Phys.* **13** 545
[41] Chen C, Ding X, Qin J, He Y, Luo Y H, Chen M C, Liu C, Wang X L, Zhang W J, Li H, You L X, Wang Z, Wang D W, Sanders B C, Lu C Y, and Pan J W 2018 *Phys. Rev. Lett.* **121** 100502
[42] Larger L, Baylón-Fuentes A, Martinenghi R, Udaltsov V S, Chembo Y K, and Jacquot M 2017 *Phys. Rev. X* **7** 011015
[43] Pankov A V, Vatnik I D, Sukhorukov A A 2022 *Phys. Rev. Appl.* **17** 024011
[44] Lecun Y and Botto L 1998 *Proc. IEEE* **86** 2278
[45] Leefmans C, Dutt A, Williams J, Yuan L, Parto M, Nori F, Fan S, and Marandi A 2022 *Nat. Phys.* **18** 442
[46] Bao Q L, Zhang H, Ni Z H, Wang Y, Polavarapu L, Shen Z X, Xu Q H, Tang D Y, and Loh K P 2011 *Nano Res.* **4** 297
[47] Cheng Z, Tsang H K, Wang X, Xu K, and Xu J B 2014 *IEEE J. Sel. Top. Quantum Electron.* **20** 43
[48] Xie Q J, Zhang H H, and Shu C 2020 *J. Lightwave Technol.* **38** 339
[49] Bengio Y 2009 *Found. Trends Mach. Learn.* **2** 1
[50] Scherer D, Müller A, and Behnke S 2010 *Proc. 20th International Conference on Artificial Neural Networks* **6354** LNCS (PART 3) p 92
[51] Raudys S 1998 *Neural Networks* **11** 283
[52] Lehtokangas M and Saarinen J 1998 *Neurocomputing* **20** 265
[53] Sebastiani F 2002 *ACM Comput. Surv.* **34** 1
[54] Saleem N and Khattak M I 2020 *Appl. Acoust.* **167** 107385
[55] Psaltis D, Brady D, and Wagner K 1988 *Appl. Opt.* **27** 1752
[56] Tainta S, Erro M J, Amaya W, Garde M J, Sales S, and Muriel M A 2012 *IEEE J. Sel. Top. Quantum Electron.* **18** 377
[57] Malacarne A and Azaña J 2013 *Opt. Express* **21** 4139
[58] Willianmson I A D, Hughes T W, Minkov M, Bartlett B, Pai S, and Fan S 2020 *IEEE J. Sel. Top. Quantum Electron.* **26** 7700412
[59] Chen Z G and Segev M 2021 *ELight* **1** 2
[60] Duran-Sierra E, Cheng S, Cuenca R, Ahmed B, Ji J, Yakovlev V V, Martinez M, Al-Khalil M, Al-Enazi H, Cheng Y S L, Wright J, Busso C, Jo J A 2021 *Cancers* **13** 4751
[61] Shirshin E A, Gayerm A V, Nikonova E E, Lukina M M, Yakimov B P, Budylin G S, Dudenkova V V, Ignatova N I, Komarov D V, Zagaynova E V, Yakovlev V V, Becker W, Shcheslavskiy V I, Shirmanova M, and Scully M O 2022 *Proc. Natl. Acad. Sci. USA* **119** e2118241119