



Multi-Task Learning by Multi-Wave Optical Diffractive Network

By Jing Su, Yafei Yuan, Chunmin Liu & Jing Li

Fudan University

Abstract- Recently, there has been tremendous researches in Optical neural networks that could complete comparatively complex computation by optical characteristic with much more fewer dissipation than electrical networks. Existed neural networks based on the optical circuit are structured with an optical grating platform with different diffractive phase at different diffractive points. In this study, it proposed a multi-wave deep diffractive network with approximately 106 synapses, and it is easy to make hardware implementation of neuromorphic networks. In the optical architecture, it can utilize optical diffractive characteristic and different wavelengths to perform different tasks. Different wavelengths and different tasks inputs are independent of each other. Moreover, we can utilize the characteristic of them to inference several tasks, simultaneously. The results of experiments were demonstrated that the network could get a comparable performance to single-wavelength single-task. Compared to the multi-network, single network can save the cost of fabrication with lithography. We train the network on MNIST and MNIST-FASHION which are two different datasets to perform classification of 32×32 inputs with 10 classes. Our method achieves competitive results across both of them. In particular, on the complex task MNIST-FASION, our framework obtains an excellent accuracy improvement with 3.2%. On the meanwhile, MNSIT also have the improvement with 1.15%.

GJCST-A Classification: C.2.1



Strictly as per the compliance and regulations of:



Multi-Task Learning by Multi-Wave Optical Diffractive Network

Jing Su ^α, Yafei Yuan ^σ, Chunmin Liu ^ρ & Jing Li ^ω

Abstract- Recently, there has been tremendous researches in Optical neural networks that could complete comparatively complex computation by optical characteristic with much more fewer dissipation than electrical networks. Existed neural networks based on the optical circuit are structured with an optical grating platform with different diffractive phase at different diffractive points. In this study, it proposed a multi-wave deep diffractive network with approximately 10^6 synapses, and it is easy to make hardware implementation of neuromorphic networks. In the optical architecture, it can utilize optical diffractive characteristic and different wavelengths to perform different tasks. Different wavelengths and different tasks inputs are independent of each other. Moreover, we can utilize the characteristic of them to inference several tasks, simultaneously. The results of experiments were demonstrated that the network could get a comparable performance to single-wavelength single-task. Compared to the multi-network, single network can save the cost of fabrication with lithography. We train the network on MNIST and MNIST-FASHION which are two different datasets to perform classification of 32×32 inputs with 10 classes. Our method achieves competitive results across both of them. In particular, on the complex task MNIST-FASHION, our framework obtains an excellent accuracy improvement with 3.2%. On the meanwhile, MNSIT also have the improvement with 1.15%.

1. INTRODUCTION

Store and retrieve data units based on the Von-Neumann architecture are far more time-consuming and power-hungry than optical device [1-4]. Different from modern computers, cis integrated of data computation, storage and fetch, which more effective, less power, large storage capacity and higher integration level [5-7]. Besides, Artificial neural network is [8] similar to the way in which human and animal store and process data is successful in a wide range of tasks such as image analysis [9], speech recognition [10] and language translation [11]. Artificial neural network can get comparable or even superior performance than the human with the increasing data volume, problem complexity and structure depth. Most of tasks cannot be migrated well in smart portable devices for its complexity and power. The less-power, more-efficiency and faster speed is becoming more and more critical for deep learning implemented on embedded device.

Author ^α ^ρ ^ω: Dept. of Optical Science and Engineering, Fudan University, 200433, Shanghai China. e-mail: yyf@fudan.edu.cn

Author ^σ: Dept. of Electronic Engineering, Center for Intelligent Medical Electronics, Fudan University, 200433, Shanghai, China.

The neuromorphic computing seeks a brain-like processing, which overcome the limitation from conventional computers. IBM [12] built a 5.4-billion transistor chip with 4096 neurosynaptic cores called True North - a fully functional digital chip. To provide the extreme complexity of the human cerebral cortex, M Prezioso [13] et al combined complementary metal oxide-semiconductors (CMOS) and two-terminal resistive device with electric circuits. Spin-transfer torque magnetic memory [14] (STTMRAM) with non-volatility, high-speed and high endurance, is suitable as a stochastic memristive device, considering the functional implication of synaptic neuronal plasticity. Alexander N. Tait [15] inspired by spiking neural networks integrated laser devices to explore highly interactive information at speeds with optical-electronic systems. This approach promises to incorporate photonic spike processing in the training system. Besides, Carlos Rios dramatically improve storage capacity to implement all-photic nonvolatile [16] multi-level where memory electric-photic interconnect technologies bring not only opportunities but also challenges to the unconventional circuits and systems. To overcome the wastage of optical-electric conversions coupling, all photic device can be performed with fast computational speed and lower power. On-chip nonvolatile photic [17] device would dramatically improve performance in existing brain-like neural networks [18] to eliminate electronic latency and reduce electronic consuming. The on-chip optical architecture is designed for network protocol computational element and waveguide medium to communicate among high-performance spiking neurons.

The architecture of fully-optical network with Mach-Zehuder interferometer [19] promises the reduction of data-movement energy cost. All-optical diffractive deep neural network architecture (D2NN) [20] utilize passive component and optical diffraction. D2NN can be easily scaled up and fabricated by 3D lithography [21] in a power-efficient manner.

In general, optical networks have more trainable parameters with complex-value modulation which provide phase and amplitude of each neuron rather than only amplitude in electric networks. Unfortunately, optical device to forming neural network has some problems. Firstly, all-optical neural network is designed for a single task, but multi-tasks [22] are significant and important. Secondly, learning rates for different tasks are



important to the accuracy. It is non-trivial to balance the tasks and learning rates.

In this paper, to address above two issues, we make most of the optical characteristics to express different tasks with different wavelengths. The one wavelength to use as baseband, the other is used as a carrier frequency. Therefore, the base band wave can be set to a large learning rate and vice versa. Extensive experiments based on MNIST and MNIST-Fasion [23] are conducted to investigate the efficacy and properties of the proposed multi-wavelength diffractive network (MWDN). In both two tasks, The MWDN significantly performs the baselines even better in the same network.

II. MULTI-WAVELENGTH DIFFRACTIVE NETWORK

Spatial domain implies per-wave in-plane propagation reasoning about diffractive in the particular phase and frequency which can analyze and integrate different direction waves. It operates in the frequency domain. The wave distribution of the observation and

aperture plane can be viewed as the linear combination with a great many monochromatic plane wave of different direction propagation. The amplitude and phase of each plane wave lies on the angular spectrum. The angular spectrum can be acquired by FFT analysis [24] process. The plane wave propagation is a complex task that take into consider many affecting factors, such as direct, phase and amplitude.

As shown in Fig. 1, we adjust optical grating parameters (height and the complex index of refraction), the height is altered by 3D printed, and the complex index of refraction is altered by laser light with different power. Different power can alter different refraction of phase change materials. We input images in MNIST and MNIST-FASHION simultaneously, the input optical wavelength of MNIST and MNIST-FASHION task is λ_1 and λ_2 , respectively. The diffractive network with the different task has the same optical parameters. The bottom of the figure is the optical carrier. Fig. 2 shows the framework of diffractive network and different color denotes different index of refractive.

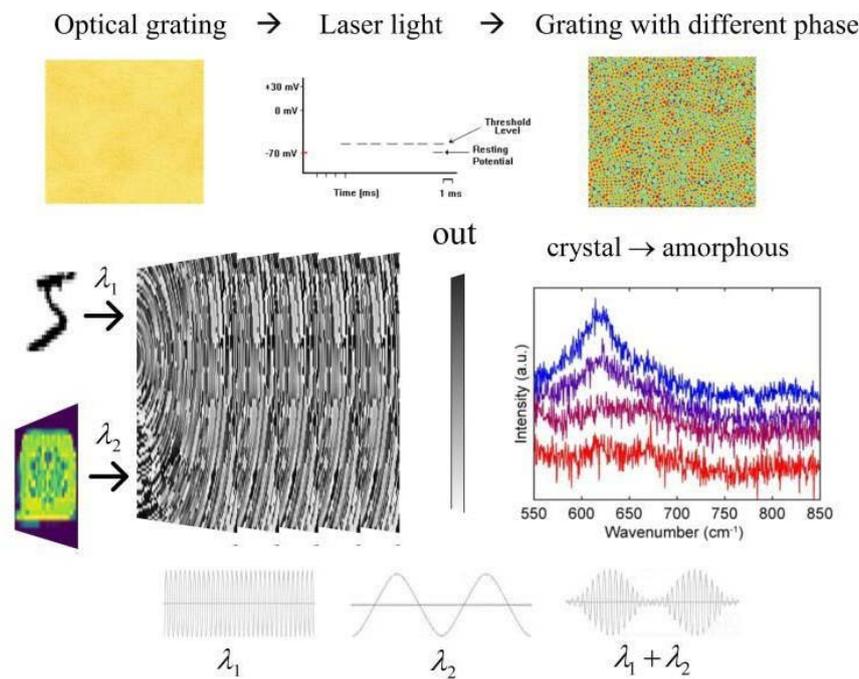


Fig. 1: The architecture of Multi-wavelength diffractive network

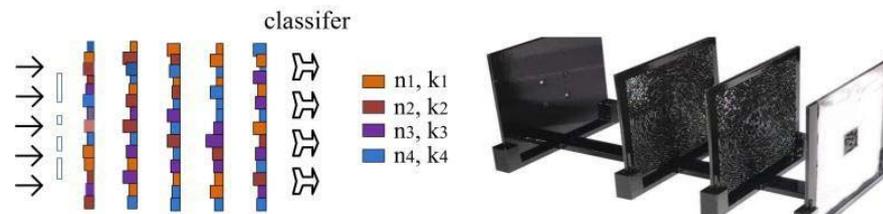


Fig. 2: The framework of Multi-wavelength diffractive network

Firstly, we convert image information to the phase and amplitude of optical information as the input of systems. Then, the optical grating is manufactured by 3D-printing device with different height. In the following sections, we discuss that MWDN tackles the tasks predominately using the angular spectrum. MWDN by the 3D-printing would influence the amplitude and phase of the wave to $0 \sim 1$ and $0 \sim 2\pi$, for two tasks in the same network. For each layer of MWDN, we set the neuron size range $200\mu\text{m}$ to $700\mu\text{m}$, which is an effect tunable.

Following the Fresnel diffraction equation, we can consider the optical signal from the spatial domain to the frequency domain. The angular spectrum method of plane wave explains how wave propagate. It is the primary method of analyzing diffraction in the frequency domain. Based on the angular spectrum, the free space transfer function is to control free propagation. The wave plane can transfer angular spectrum by FFT process, where diffractive data processing is more evident as follows:

$$A_0(f_x, f_y) = FFT(U_0(x, y)) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} U_0(x, y) e^{-j2\pi(\frac{f_x x}{N} + \frac{f_y y}{M})}$$

$$A(f_x, f_y) = A_0(f_x, f_y) \cdot H(f_x, f_y)$$

$$U(x, y) = IFFT(A(f_x, f_y))$$

$$H(f_x, f_y) = \exp(jkz) \exp(j\pi\lambda z (f_x + f_y)^2)$$

Where, f_x, f_y are space frequency correspond to x, y location ($f_x = 1/(x-x_0)$), $x-x_0$ is gaps in the optical map, N and M is the number of grooves on optical grating in height and width direction, $U_0(x, y)$ is the original field distribution, $U(x, y)$ is the field distribution after free space transfer, $H(f_x, f_y)$ is transfer function, $A_0(f_x, f_y)$ is

the original angular spectrum, $A(f_x, f_y)$ is the angular spectrum after free space transfer. The results of the inverse Fourier transform to transfer function, are the impulse response function. The equation can view as the Fourier transform.

$$\begin{aligned} h(x-x_0, y-y_0) &= F^{-1}(H(f_x, f_y)) \\ &= \frac{1}{j\lambda z} \exp(jkz) \exp(j \frac{k}{2z} ((x-x_0)^2 + (y-y_0)^2)) \end{aligned}$$

The output wave plane distribution propagates through 3D material and the field distribution is changed by the refractive index.

$$O(x, y) = U(x, y) \cdot (1 - \alpha \cdot \Delta d) \cdot \exp(j\phi)$$

$$\phi = \frac{2\pi\Delta n_1 \Delta d}{\lambda}, \Delta n_1 = n - n_0, \alpha = \frac{4\pi k}{\lambda}$$

Where α is extinction efficiency, n is the refractive index real part of 3D material, n_0 is the refractive index real part of the vacuum, k is the refractive index imagery part, λ is the wavelength, Δd is the height of material map, ϕ is the phase difference. If we choose transparent material ($k \rightarrow 0$) ignoring the optical losses, the transmission coefficient of a neuron is composed of only phase term; if we select non-transparent material, the

transmission coefficient of a neuron is composed of amplitude and phase in MWDN architecture.

According to the size of input data, an effective and flexible linear interpolation algorithm is to fit the diffractive input layer. The interconnection rate between adjacent layers relates to the distance and diffraction angle, which approach the critical value (1.0).

Furthermore, the number of the network layers and the axial distance is also a tunable. The output layer can part into ten regions corresponding to ten classes, where the summation of light intensity can be detected in the wave plane region. Mean square error (MSE) uses to train MWDN parameters compared to the target. We aim to minimize a loss function, which increases target region wave intensity and decrease other regions. The training batch size set to be 10 for the classifier.

III. THE BACKWARD OF MULTI-WAVELENGTH DIFFRACTIVE NETWORK

To train MWDN, we use the back-propagation algorithm with a dam optimization method.

We focus on the intensity of wave and define loss function with MSE between output and target.

$$E(\Delta d) = \frac{1}{K} \sum_{k=1}^K (o_k - t_k)^2$$

where K is the number of training data, o_k is the output of the MWDN, and t_k is the label of the corresponding input. The optimization problem can be written follows:

$$\min_{\Delta d_i^l} E(\Delta d_i^l)$$

where l is the layer, i is the l_{th} layer location. The gradient of loss to all parameters can be calculated, which is used to update MWDN architecture parameters during the training process. Each batch of the training data is fed into MWDN, where each layer gradient can be calculated to update.

IV. THE BACKWARD OF MULTI-WAVELENGTH DIFFRACTIVE NETWORK

The optical diffractive network and deep neural network are markedly different. The function of the optical diffractive network is determined by wavelength and the parameters of the optical grating (height and complex refractive index). Multi-wavelength diffractive network has a broad range of requirements that differ from the conventional network.

Different wavelength has different effectiveness. We set different wavelength for different tasks. Meanwhile, the network needs to ensure that different wavelengths do not affect each other. By setting one to baseband and the other to the carrier, the diffractive network is used to adjust optical plane wave independently. The algorithm can be considered as an efficient carrier algorithm. The ratio of baseband and carrier wavelengths is 1:30. The short wavelength is little influence to long wavelength and vice versa. If the phase

difference of long wavelength is ϕ_1 and the phase difference of short wavelength is ϕ_2 , the corresponding relationship as follows:

$$\phi_1 \lambda_1 = \phi_2 \lambda_2, \quad \lambda_1 / \lambda_2 = 1:30$$

So, the equation can be as follows:

$$\phi_2 / \phi_1 = 1:30, \quad \phi_2 \ll \phi_1$$

$$\phi_1 = 2\pi n + \phi'_1, n = 0, 1, 2 \dots$$

$$\phi_2 = \pi n / 15 + \frac{\phi'_1}{30}, \phi'_1 = 0 \sim 2\pi$$

The second terms of ϕ_2 is relative to the first term can be ignore, the equation can be shown as follows:

$$\phi_2 \approx \pi n / 15$$

The multi-wavelength diffractive network can be effective, and more powerful than deep neural network. Phase difference ϕ_i ($i=1, 2$) can be obtained easily by adjusting the height on the diffractive network. Due to $\phi_2 \ll \phi_1$, then we adjust Δd for minor learning rate for ϕ_2 , as well as for large learning rate for ϕ_1 , without one impacting the other.

$$\Delta d \approx \phi \lambda / 2\pi \Delta n$$

V. EXPERIMENT

In this work, we apply the proposed MWDN to implement on two different dataset MNIST and MNIST-Fashion.

a) Model setup

By comparing to the state-of-the-art methods with accuracy and speed of, MNIST and MNIST-FASION in this method achieve better performance. The size of the network is set to 200×200 , 500×500 and 1500×1500 , each having a trainable height of the map. The optical network possesses two types, one for phase modulation, and the other for complex-modulation. The MNIST and MNIST-FASION tasks with different optical wavelengths, the input is altered by optical grating mask.

Using the backward propagation, the model is trained with two task datasets alternately, validated its effectiveness. We train the network with different learning rate for different tasks, which overcome the drawbacks of local optimum to solve. As well as, all the parameters of the network are adjusted by the gradient descent algorithm to minimize the error.

b) Dataset

We evaluate the approach on two datasets and input information for neurons in the form of phase fed into the network. The two datasets have different data distribution, which is difficult to classify in the same network. The conventional networks require the input information to be independent and identical distribution. The task is to handle two different distribution data in a same network.

c) Experimental analysis

For better performance, we set a different learning rate and different signal frequency to two datasets. The maximum half-cone diffraction angle is formulated as follow:

$$\phi_{max} = \sin^{-1}(\lambda / 2d_f)$$

The light wavelength is 0.4THz, 14.4THz for MNIST and MNIST fashion. The neuron size is set to be 200μm. The height of the map and axial distance between two successive layers are trainable. As comparing the performance of MWDN and DN methods with single task, the results was shown in Table 1. It is clearly that the performance of MWDN would improve the accuracy of 1.15% and 3.2%, independently. To evaluate the multi-wavelength for multi-task, so we compare the multi-task to a single task in Table 2. The multi-task diffractive network enables consistent performance with a single task. The result can perform well in the same parameters. The experiment set of setting 1 is the same wavelength for comparison. Setting 2 is performed by a different wavelength. The DN-FASION and DN-MNIST are evaluated by independent diffractive network.

Table 1: The performance of MWDN and DN methods

Method	MNIST	λ (THZ)	MNIST-Fasion	λ (THZ)
MWDN(PCM)	92.85%	0.4	84.33%	14.4
DN-MNIST	91.75%	0.4	81.13%	0.4

Table 2: The performance of MWDN and DN methods with multi-task

Method	MNIST	MNIST-Fasion	λ1 (THZ)	λ2(THZ)
Setting 1	23.45%	12.12%	0.4	0.4
Setting 2	90.45%	76.67%	14.4	0.144
DN-Fasion	/	78.70%	14.4	/
DN-MNIST	91.75%	/	/	14.4

d) Convergence analysis

Fig3 demonstrate two classifiers for two datasets, where each dataset has ten target class. One classifier is set to classify in the same region and

another is set to a different region. The different frequency setting to different datasets is effective and different regions in the MWDN has high accuracy.

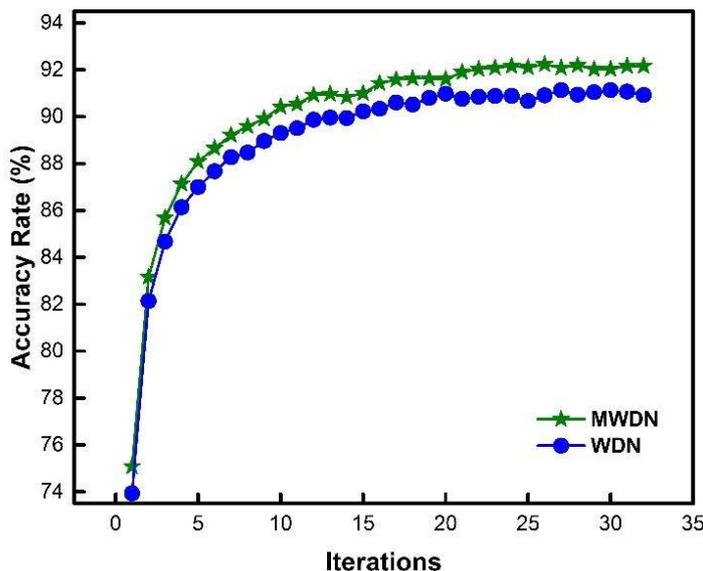


Fig. 3: The training process of different method

Finally, we report the performance on the validation data of MNIST and MNIST_FASION, a challenging task with different datasets. Using only a single network, two datasets classification can be accomplished simultaneously we investigate the effects of various combinations of different datasets for MWDN. The results are shown in Fig 3. We find that we can implement two classes to the same network with MWDN algorithm. Compared to other approach that use only single dataset as input, our approach even yields a boost.

VI. CONCLUSION

In this paper, we propose a novel and multi-tasks optical network named as the multi-wavelength diffractive network (MWDN). Based on plane wave propagation, our method can achieve comparable accuracy against the single-task network. We successfully apply MWDN to multi-tasks with different datasets distribution and provide a multi-wavelength method with different model size. In the future, we aim to develop a more effective network to achieve complex tasks and reach better performance.

ACKNOWLEDGEMENTS

We acknowledge the financial supports by National Key R&D Program of China (Grant No. 2017YFE0112000). The authors wish to thank Professor Du. Furthermore, the authors would like to express sincere thanks for server supported by Light-ca Technology Corporation.

REFERENCES RÉFÉRENCES REFERENCIAS

- Jewell, and J. L., "Regenerative pulsations from an intrinsic bistable optical device," *Applied Physics Letters*, vol. 40, no. 4, pp. 291.
- J. Kraitl, H. Ewald, and H. Gehring, "An optical device to measure blood components by a photoplethysmographic method," *Journal of Optics A Pure & Applied Optics*, vol. 7, no. 6, pp. 0-0.
- Y. Liu, X. Lin, C. Wei, C. Zhang, and Y. S. Zhao, "3D-printed optical-electronic integrated devices," *Science China Chemistry*, 2019.
- S. Kim, S. G. Menabde, V. W. Brar, and M. S. Jang, "Functional Midinfrared Polaritonics in van der Waals Crystals," *Advanced Optical Materials*, pp. 1901194, 2019.
- M. Spencer, J. Eickholt, and J. Cheng, "A Deep Learning Network Approach to ab initio Protein Secondary Structure Prediction," *Computational Biology & Bioinformatics IEEE/ACM Transactions on*, vol. PP, no. 99, pp. 1, 2014.
- X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method," *Medical Physics*, vol. 44, no. 10, pp. 5221, 2017.
- C. Che, R. Lin, X. Zeng, E. Karim, G. John, and M. Xu, "Improved deep learning-based macromolecules structure classification from electron cryo-tomograms," *Machine Vision & Applications*.
- X. S., R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987-5995, 2017.
- D. Chao, C. L. Chen, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution."
- D. Li, "Deep learning: from speech recognition to language and multimodal processing," *Apsipa Transactions on Signal & Information Processing*, vol. 5, pp. -.
- S. P. Singh, A. Kumar, H. Darbari, L. Singh, and S. Jain, "Machine translation using deep learning: An overview."
- T. O. Kwok, and Y. K. Kwok, "On the design, control, and use of a reconfigurable heterogeneous multi-core system-on-a-chip."
- T. F. Cianciulli, J. A. Lax, F. E. Cerruti, G. E. Gigena, H. J. Redruello, M. A. Orsi, J. A. Gagliardi, A. N. Dorelle, M. A. Riccitelli, and H. A. Prezioso., "COMPLEMENTARY VALUE OF TRANSTHORACIC ECHOCARDIOGRAPHY & CINEFLUOROSCOPIC EVALUATION OF MECHANICAL HEART PROSTHETIC VALVES," *Echocardiography*, vol. 21, no. 2, pp. 211-211, 2010.
- A. F. Vincent, J. Larroque, W. Zhao, N. B. Romdhane, and D. Querlioz, "Spin-Transfer Torque Magnetic Memory As a Stochastic Memristive Synapse."
- A. N. Tait, M. A. Nahmias, T. Yue, B. J. Shastri, and P. R. Prucnal, *Photonic Neuromorphic Signal Processing and Computing*: Springer Berlin Heidelberg, 2013.
- C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "Integrated all-photonic non-volatile multi-level memory," *Nature Photonics*.
- R. A. Hut, N. Mrosovsky, and S. Daan, "Nonphotic Entrainment in a Diurnal Mammal, the European Ground Squirrel (*Spermophilus citellus*)," *Journal of Biological Rhythms*, vol. 14, no. 5, pp. 409-420.
- K. Voutsas, G. Langner, J. Adamy, and M. Ochse, "A brain-like neural network for periodicity analysis," *IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics*, vol. 35, no. 1, pp. 0-22.
- M. Dülk, S. Fischer, M. Bitter, M. Caraccia, W. Vogt, E. Gini, H. Melchior, W. Hunziker, A. Buxens, and H. N. Poulsen, "Ultrafast all-optical demultiplexer based on monolithic Mach-Zehnder interferometer

- with integrated semiconductor optical amplifiers,” *Optical & Quantum Electronics*, vol. 33, no. 7-10, pp. 899-906.
20. X. Lin, R. Yair, Y. N. T., V. Muhammed, L. Yi, J. Mona, and O. Aydogan, “All-optical machine learning using diffractive deep neural networks,” *Science*, pp. eaat8084-.
 21. A. d. Campo, and C. Greiner, “SU-8: a photoresist for high-aspect-ratio and 3D submicron lithography,” *Journal of Micromechanics & Microengineering*, vol. 17, no. 6, pp. R81-R95.
 22. Y. Tian, P. Luo, X. Wang, and X. Tang, “Pedestrian Detection aided by Deep Learning Semantic Tasks.”
 23. L. Deng, “The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web],” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. p.141-142.
 24. C. F. V. Loan, *Computational Frameworks for the Fast Fourier Transform*, 2011.

