Universal approximation capabilities of coherent diffractive systems

Anonymous Author(s)
Affiliation
Address
email

Abstract

Coherent optical computing systems are a promising avenue for increasing computation speed and solving energy requirements for machine learning applications. These systems utilize the diffraction process of coherent waves to perform calculations in the optical domain. While the diffraction process is linear in complex space $\mathbb{C}$, it empirically has been shown, that these systems are able to outperform standard linear matrix multiplications in $\mathbb{R}$, because photo-sensors project from complex space to real space. Here we give theoretical insights, of why this is the case and show, that a system consisting of multiple phase-plates, two output photo-detectors, and the appropriate input encoding is theoretically able to learn any one-dimensional function. We furthermore use these theoretical insights to show that encoding the input information solely in the intensity of the diffractive system is never enough to make the system a universal aproximator. These results are useful to understand the capabilities of diffractive optical systems and improve their training.

1 Introduction

In recent years large artificial neural networks have set new standards in research and industrial applications. The rise of these large models has been largely made possible by an increase in memory and computing power. Training these large neural networks requires general purpose computing devices that are used to calculate huge matrix multiplications and other operations in parallel. Usually these devices are GPUs. However, since many artificial neural network architectures show increased performance with increasing number of parameters [1], the best models are in general those with the maximum possible number of trainable parameters, only constrained by available data and training time. This leads to huge energy and computational requirements during training and deployment. To tackle the ever increasing computing power requirements for large neural networks new computation architectures are needed to change the way a computer handles these large amount of floating point operations.

One promising approach for such devices is optical computing [2–4]. Approaches can be divided into those using incoherent light [5, 6], and those using coherent waves [7–12]. Some of these systems rely on optical waveguides, while other systems utilize free space propagation of light [13–18]. Systems that utilize coherent free space propagation and coherent holographic plates are also called diffractive deep neural networks (DDNN) [13]. An input signal is encoded in the amplitude and/or phase of a coherent wave that propagates through multiple plates, which are masks that change the phase or amplitude of a wave across its wavefrong. These plates are designed in such a way, that the complete network performs a desired operation directly in the analog space of the wave. These networks have been realized with optical light [19], terahertz waves [13] and with ultrasound [20].

Submitted to the Second Workshop on Machine Learning with New Compute Paradigms at NeurIPS (MLNCP 2024). Do not distribute.

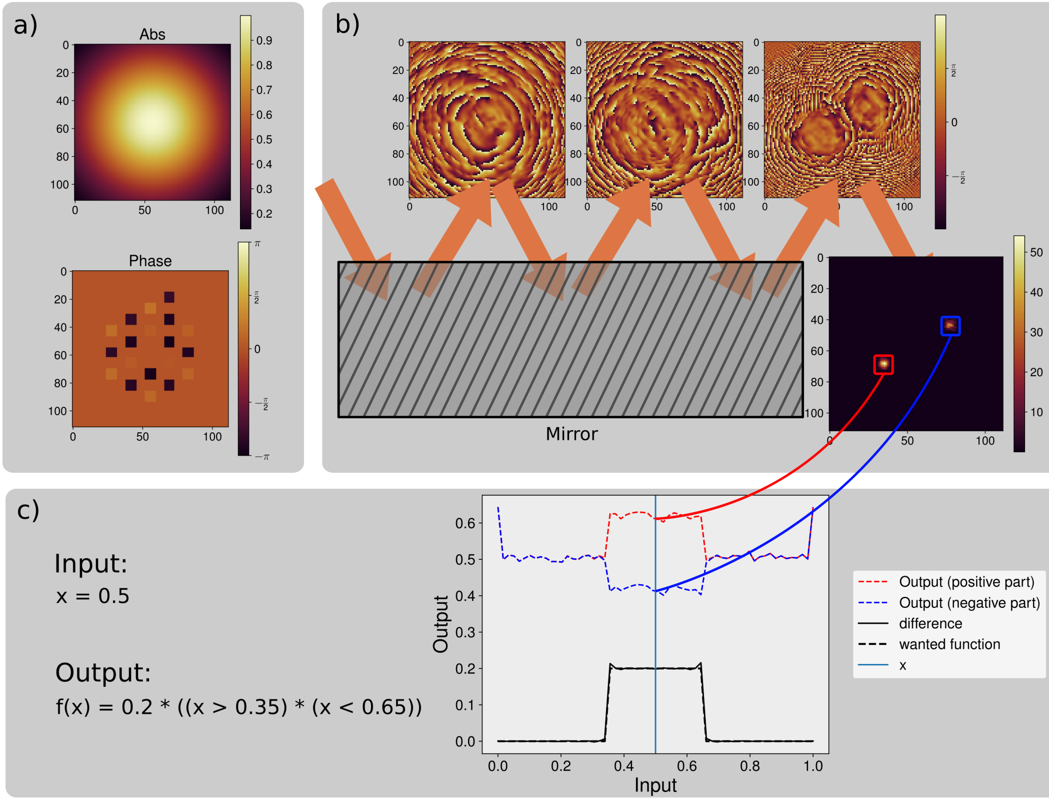

Figure 1: Overview of how a diffractive deep neural network is able to create arbitrary onedimensional functions. a) Encoding of the input ($x = 0.5$) with 20 different fixed coefficients in the phase domain of an incoming gaussian beam. b) This encoding propagates through the network, hits 3 trained phaseplates while being reflected by a mirror and is finally captured by two regions on a CCD. c) The two regions on the CCD are subtracted and thus the result is obtained. Using this encoding the system is able to be trained to perform arbitrary functions.

They are able to perform a whole array of different computations, ranging from image classification and mode conversion [18] to learning of logical functions [14, 15]. However such networks are inherently limited by the lack of a nonlinearity, since free space wave diffraction and hologram plates are linear in complex space, in which the wave resides in. It has been shown, that diffractive deep neural networks without a nonlinearity can perform a complex matrix multiplication [18]. Optical nonlinearities in optical neural networks are an active research field [21, 22]. However one of the easiest implementable nonlinearities is a photosensor. Such a photosensor maps the complex field to a voltage, that is proportional to the intensity reaching the sensor. It has numerically been shown, that such a system is able to learn certain nonlinear mappings, like an XOR logical function [15]. This raises questions about the theoretical capabilities of such systems. Here we show, that a network with a Fourier encoding is able to learn arbitrary onedimensional functions, and that information should be encoded in the phase of the coherent wave, instead of the intensity, because intensity encodings can never be universal function approximators. This is promising for the further development of coherent optical computational systems, and gives suggestions, on how to best set up such a system.

2 Results

2.1 Universal function approximators

A function that fulfills the universal function approximation property can approximate any continuous function $f(x) \in C(\mathbb{R}^n)$. This means that for every $f(x) \in C(\mathbb{R}^n)$ there exists an $g(x;\theta)$ and θ

such that $\underset{x \in K}{max} |f(x) - g(x;\theta)| < \epsilon$ for any $\epsilon > 0$. θ are learnable parameters. Crucially this property does not determine how to calculate these learnable parameters, just that a combination exists, for which f and g are arbitrarily close everywhere in $K \subseteq \mathbb{R}^m$. Most of these theorems are used to describe artificial neural networks, and estimate bounds and limitations on the depth and width of the layers of deep neural networks[23–25]. These theorems give us a framework to analyze the theoretical capability of diffractive deep neural network and similar systems.

2.2 Deep Diffractive Networks

The input of a diffractive system can be described as a vector of complex numbers, that represent the amplitude and phase of a wavefield at certain input positions.

$$I = \begin{pmatrix} I_1 e^{i\psi_1} \\ \vdots \\ I_n e^{i\psi_n} \end{pmatrix} \tag{1}$$

This input propagates through the system of diffractive plates, until it hits a photodiode at which point only the intensity, which is the square of the absolute value of the complex field, is measured.

To further derive the capability of DDNNs, we need to take a look at the output of the system. The first result we need, is that with enough plates diffractive deep neural networks can perform a complex matrix multiplication (see [18]). This allows us to write the propagation of waves through the diffractive system as a matrix multiplication

$$\begin{aligned} Y &= \mathbf{W} I \qquad (2) \\ &= \begin{pmatrix} A_{1,1}e^{i\phi_{1,1}} & \dots & A_{1,n}e^{i\phi_{1,n}} \\ \vdots & \ddots & \vdots \\ A_{n,1}e^{i\phi_{n,1}} & \dots & A_{n,n}e^{i\phi_{n,n}} \end{pmatrix} \begin{pmatrix} I_1 e^{i\psi_1} \\ \vdots \\ I_n e^{i\psi_n} \end{pmatrix} \qquad (3) \end{aligned}$$

.

Secondly, the measurement of the output wave only measure the intensity of the light. We can thus calculate the output for a single pixel on the photodiode as follows:

$$\begin{aligned} O_j &= \left| \sum_i^n A_{i,j} I_i e^{i(\phi_{i,j} + \psi i)} \right|^2 \qquad (4) \\ &= \sum_i^n A_{i,j}^2 I_i^2 + 2 \sum_k^n \sum_i^{k-1} A_{k,j} I_k A_{i,j} I_i \cos(\delta) \qquad (5) \end{aligned}$$

, with $\delta = \phi_{k,j} + \psi_k - \phi_{i,j} - \psi_i$. This results enables us to make further statements about the universal approximation capability of diffractive plates.

2.2.1 Intensity Encoding

Equation 5 lets us make statements about the nature of the input encoding that should be used. The first input encoding is the most natural one, encoding the information in the input intensity. The following result states that these networks are never universal function approximators. To prove this we will need the following result. Neural networks with one hidden layer, that is of the form $y(x) = \sum_i^n c_i \sigma(\mathbf{A}\mathbf{x} - \mathbf{b})$ with $\mathbf{A}, \mathbf{x} \in \mathbf{R}^n$, $c_i \in \mathbf{R}$ and σ a single valued function applied elementwise are universal function approximators, if and only if σ is a not a polynomial [26]. Equation 5 shows us, how the output of a single pixel depends on the input encoding $\mathbf{I}$. If the encoding is only in the intensity of the incoming wave field

$$\mathbf{I} = \begin{pmatrix} I_1 e^{i\psi_1} \\ \vdots \\ I_n e^{i\psi_n} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \tag{6}$$

, equation 5 reduces to

$$O_j = \sum_i^n A_{i,j}^2 x_i^2 + 2 \sum_k^n \sum_i^{k-1} A_{k,j} A_{i,j} x_k x_i \cos(\phi_{k,j} - \phi_{i,j}) \tag{7}$$

which is a second degree polynomial in $A_i x_i$, since $\cos(\phi_{k,j} - \phi_{i,j})$ is not dependent on the input encoding. Thus no diffractive deep neural network with intensity encoding and no other nonlinearity other than the output diode can be an universal function approximator. This result holds for the onedimensional case, as well as the multidimensional case.

2.3 Phase encoding

Since encoding the information only in the intensity is insufficient to make the network an universal function approximator, the phase can be included in the encoding. Using an input encoding of

$$\mathbf{I} = \begin{pmatrix} I_1 e^{i\psi_1} \\ \vdots \\ I_n e^{i\psi_n} \end{pmatrix} = \begin{pmatrix} e^{ix_1} \\ \vdots \\ e^{x_n} \end{pmatrix} \tag{8}$$

, equation 5 reduces to

$$O_j = \sum_i^n A_{i,j}^2 + 2 \sum_k^n \sum_i^{k-1} A_{k,j} A_{i,j} cos(\phi_{k,j} + x_k - \phi_{i,j} - x_i) \tag{9}$$

This is equivalent to a single layer neural network with fixed weights of 1 in the first layer and offsets that can be trained.

$$O = \sum_i^n c_i \sigma(x - x_i) \tag{10}$$

Here c_i are the weights in the last layer, x the input variable and x_i biases at different positions. According to some results, networks of this type are arbitrary function approximators [27] for single valued functions, if σ is a even periodic continuous function. Since cos fullfills these requirements, this result holds for diffractive deep neural networks with phase encoding. However in numerical experiments we have been unable to sufficiently approximate functions that contained higher order frequencies, or where different from the family of cosine functions. This is possibly due to the finite number of output pixels.

2.4 Fourier encoding

Instead of using only one input for each variable the signal can be encoded with different input weights w_i. This method was used by Yildrim. M et. al. [28] to improve the performance of a diffractive system in image classification tasks on multiple datasets. Instead of learning the input weights however, we opted to simply use full integer increments to introduce higher frequencies into the system. The encoding for a onedimensional variable in the optical system is displayed in equation 11

$$\mathbf{I} = \begin{pmatrix} I_1 e^{i\psi_1} \\ \vdots \\ I_n e^{i\psi_n} \end{pmatrix} = \begin{pmatrix} e^{i2\pi x_1} \\ e^{i4\pi x_1} \\ e^{i6\pi x_1} \\ \vdots \\ e^{in2\pi x_1} \end{pmatrix} \tag{11}$$

Note that all intensities are assumed to be 1. Using this encoding in equation 5 gives

$$O_j = \sum_i^n A_{i,j}^2 + 2 \sum_k^n \sum_i^{k-1} A_{k,j} A_{i,j} cos(2\pi(k-i)x + \phi_{k,j} - \phi_{i,j}) \tag{12}$$

With this encoding higher order frequencies exist and can be used to approximate desired output functions. Due to the notable similarities to the Fourier series

$$F = C_0 + \sum_{i=1}^n C_i \cos(2\pi \frac{i}{P} x - \phi_n) \tag{13}$$

we call this encoding Fourier encoding. Since the Fourier series is able to approximate a function arbitrarily well on an interval P, a diffractive neural network with this encoding should be able to do the same.

However, in equation 12 the coefficients A_i are all ≥ 0, meaning that we can only create positive coefficients. The Fourier series coefficients on the other hand are $\in \mathbf{R}$. To create coefficients that can be negative, two photodiodes can be used, the first acting as a measurement device for the positive coefficients, and the second one capturing the negative coefficients. The output of these photodiodes can easily be subtracted in an analog electrical system.

3 Experimental Results

The following numerical and experimental results have been performed with a diffractive deep neural network consisting of 3 layers of 112×112 pixel on an area with a sidelength of 0.1792 mm corresponding to a pixel size of 16 μm. A laser with a wavelength of 781 nm focused in a Gaussian beam with a radius of 0.09 mm was used as the input source. The input was encoded on a plate with the same dimensions as the trainable plates. The experimental setup used a single reflective spatial light modulator and a mirror slightly angled to reflect the beam back to the SLM. The setup is similar to the one used in [28]. This system was implemented and trained with Tensorflow [29]. Fourier encodings with 1,2,5, 8,10,20,50 and 100 components have been trained to compare the influence of more complicated encodings on the capability of the system. Results for four different nonlinear functions are displayed in 2 It can be seen that the system is able to learn all four of the

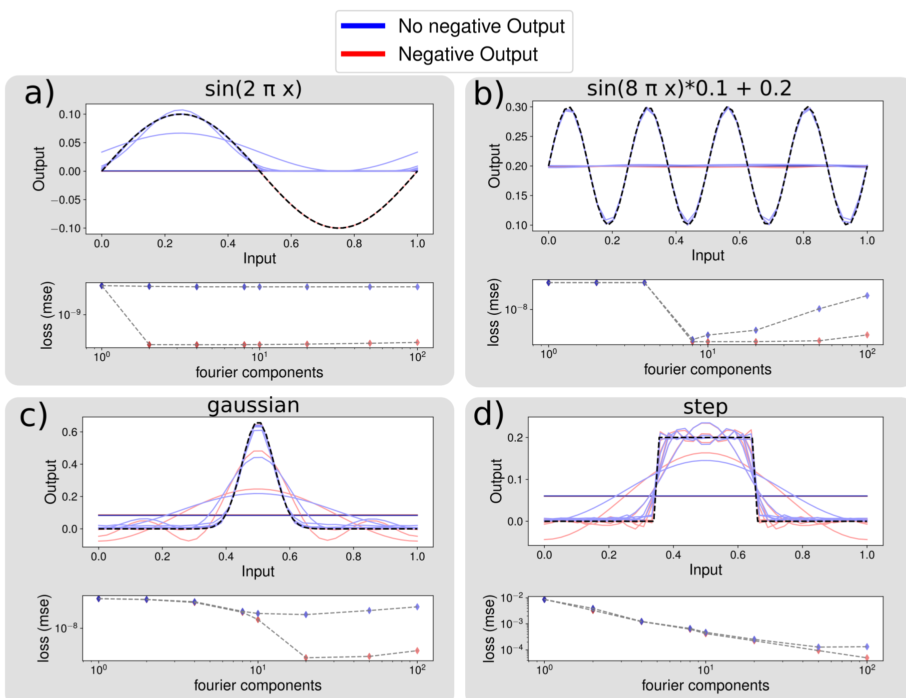

Figure 2: Numerical Results for four different nonlinear functions. The loss improves with an increasing number of Fourier components added to the optical system. a) $sin(2\pi x) * 0.1$. This function contains negative values. It cannot be learned by the system, that only considers positive output values. Only when adding negative output components can the function be fully approximated. b) $sin(8\pi x) * 0.1 + 0.2$. This function shows that multiple Fourier components are needed to approximate functions with higher frequencies c) $\frac{1}{2\sqrt{2\pi}} e^{-0.5\frac{(x-0.5)^2}{0.2^2}}$. This function needs even more Fourier coefficients to be well approximated. It can be seen, that only the system that takes negative output coefficients into account is able to approximate the function well. d) 0.2 if $x \in [0.35, 0.65]$. This function is the hardest one to approximate, probably because it is not smooth. In all cases, the network that takes negative coefficients into account gives a better approximation.

functions. Higher frequencies are needed to approximate all functions. This is especially clear when comparing the sinusoidal functions with increasing frequencies. While 2 frequencies are enough to train a simple sinus function, the requirements get harder with increasing frequencies. It can also be seen, that the negative output values improve the results in all cases, and are necessary to learn functions with negative values (see plot a) of figure 2). Overall the system with negative output values manages to learn all functions to a high degree of accuracy.

These trained phase plates where tested in a physical system, and the output values where measured. For this a mirror and a single SLM was used that influenced the phase of the incoming wave. The output was measured with a CCD sensor and the two output regions where estimated and summed up. Since the output values form the camera sensor are not normed, a multiplicative offset, that corrects for the overall energy in the system and an additive offset that corrects for the base noise level in the camera sensor where manually chosen for all measured output values. The results are displayed in figure 3 The results confirm the numerical experiments. They show that it is possible

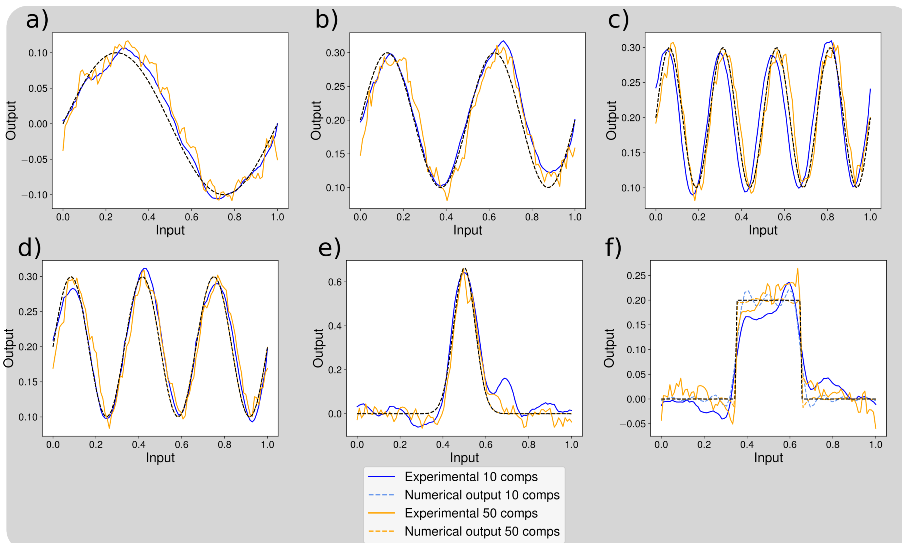

Figure 3: Real world experimental results. The functions for the plots a) to f) are the same was in figure 2. It can be seen, that the measured functions match the numerical ones. Some noise is introduced during measurement.

to train a diffractive deep neural network to perform arbitrary nonlinear functions.

4 Discussion

We have shown, that a coherent optical system with diffractive plates, also called deep diffractive neural network, is capable of performing arbitrary nonlinear functions in one dimension. We utilized an encoding, that embeds the input variable into multiple phase inputs, that have different factors, that match those of the Fourier series. Furthermore we have shown a mathematical proof that a DDNN with only an intensity encoding can never be considered an universal function approximator. A pure phase encoding that does not introduce higher frequencies into the system has been shown to be theoretically sufficient for the system to be an universal function approximator, as long as ref [27] holds. However we have not been able to numerically train a real world system that purely relies on phase encoding to perform arbitrary nonlinear functions. This is potentially due to the finite size of the system. Using the Fourier encoding however made it possible to train any onedimensional function. A negative output region was useful for all trained functions, but made the biggest difference for functions that had negative values for obvious reasons.

Using the results mentioned above should also be helpful when training DDNNs for different tasks, that are potentially more demanding than a onedimensional function. In general the information should not be encoded solely in the intensity of the incoming wave, to ensure cosine nonlinearities in the output. Furthermore a region with negative values has been shown to be important for all functions, thus should be considered when training a DDNN for other tasks. Such a negative output region can be realized with an analog electrical system, consisting of two photodiodes.

References

1. Brown, T. *et al.* Language models are few-shot learners. *Advances in neural information processing systems* **33,** 1877–1901 (2020).
2. Wetzstein, G. *et al.* Inference in artificial intelligence with deep optics and photonics. *Nature* **588,** 39–47 (2020).
3. Caulfield, H. J. & Dolev, S. Why future supercomputing requires optics. *Nature Photonics* **4,** 261–263 (2010).
4. Shastri, B. J. *et al.* Photonics for artificial intelligence and neuromorphic computing. *Nature Photonics* **15,** 102–114 (2021).
5. Gu, H. & Warde, C. Evaluation of organic light-emitting diodes as light sources for a compact optoelectronic integrated neural coprocessor. *Applied optics* **55,** 8576–8581 (2016).
6. Lu, T., Wu, S., Xu, X. & Francis, T. Two-dimensional programmable optical neural network. *Applied optics* **28,** 4908–4913 (1989).
7. Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nature Photonics* **11,** 441–446 (2017).
8. Xu, X. *et al.* 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589,** 44–51 (2021).
9. Feldmann, J. *et al.* Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589,** 52–58 (2021).
10. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606,** 501–506 (2022).
11. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569,** 208–214 (2019).
12. Zhang, H. *et al.* An optical neural chip for implementing complex-valued neural network. *Nature Communications* **12,** 1–11 (2021).
13. Lin, X. *et al.* All-optical machine learning using diffractive deep neural networks. *Science* **361,** 1004–1008. eprint: `https://www.science.org/doi/pdf/10.1126/science.aat8084`. `https://www.science.org/doi/abs/10.1126/science.aat8084` (2018).
14. Zarei, S. & Khavasi, A. Realization of optical logic gates using on-chip diffractive optical neural networks. *Scientific Reports* **12,** 15747 (2022).
15. Qian, C. *et al.* Performing optical logic operations by a diffractive neural network. *Light: Science & Applications* **9,** 59 (2020).
16. Bernstein, L. *et al.* Freely scalable and reconfigurable optical hardware for deep learning. *Scientific reports* **11,** 1–12 (2021).
17. Bernstein, L. *et al.* Single-Shot Optical Neural Network. *arXiv preprint arXiv:2205.09103* (2022).
18. Kulce, O., Mengu, D., Rivenson, Y. & Ozcan, A. All-optical information-processing capacity of diffractive surfaces. *Light: Science & Applications* **10,** 25 (2021).
19. Chen, H. *et al.* Diffractive deep neural networks at visible wavelengths. *Engineering* **7,** 1483–1491 (2021).
20. Athanassiadis, A. G., Schlieder, L., Melde, K., Volchkov, V. & Fischer, P. Animating sound using neurally multiplexed holograms. *The Journal of the Acoustical Society of America* **148,** 2807–2807 (2020).
21. Zuo, Y. *et al.* All-optical neural network with nonlinear activation functions. *Optica* **6,** 1132–1137 (2019).
22. Li, G. H. *et al.* All-optical ultrafast ReLU function for energy-efficient nanophotonic deep learning. *arXiv preprint arXiv:2201.03787* (2022).
23. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural networks* **4,** 251–257 (1991).
24. Funahashi, K.-I. On the approximate realization of continuous mappings by neural networks. *Neural networks* **2,** 183–192 (1989).

25. Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **2,** 359–366. ISSN: 0893-6080. https://www.sciencedirect.com/science/article/pii/0893608089900208 (1989).
26. Leshno, M., Lin, V. Y., Pinkus, A. & Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks* **6,** 861–867 (1993).
27. Lin, S., Guo, X., Cao, F. & Xu, Z. Approximation by neural networks with scattered data. *Applied Mathematics and Computation* **224,** 29–35 (2013).
28. Yildirim, M., Dinc, N. U., Oguz, I., Psaltis, D. & Moser, C. Nonlinear processing with linear optics. *Nature Photonics,* 1–7 (2024).
29. Abadi, M. *et al.* {*TensorFlow*}*: a system for* {*Large-Scale*} *machine learning* in *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (2016), 265–283.