All-optical machine learning using diffractive deep neural networks

Xing Lin[1,2,3]*, Yair Rivenson[1,2,3]*, Nezih T. Yardimci[1,3], Muhammed Veli[1,2,3], Yi Luo[1,2,3], Mona Jarrahi[1,3], Aydogan Ozcan[1,2,3,4]†

Deep learning has been transforming our ability to execute advanced inference tasks using computers. Here we introduce a physical mechanism to perform machine learning by demonstrating an all-optical diffractive deep neural network (D[2]NN) architecture that can implement various functions following the deep learning–based design of passive diffractive layers that work collectively. We created 3D-printed D[2]NNs that implement classification of images of handwritten digits and fashion products, as well as the function of an imaging lens at a terahertz spectrum. Our all-optical deep learning framework can perform, at the speed of light, various complex functions that computer-based neural networks can execute; will find applications in all-optical image analysis, feature detection, and object classification; and will also enable new camera designs and optical components that perform distinctive tasks using D[2]NNs.

Deep learning is one of the fastest-growing machine learning methods (*1*). This approach uses multilayered artificial neural networks implemented in a computer to digitally learn data representation and abstraction and to perform advanced tasks in a manner comparable or even superior to the performance of human experts. Recent examples in which deep learning has made major advances in machine learning include medical image analysis (*2*), speech recognition (*3*), language translation (*4*), and image classification (*5*), among others (*1*, *6*). Beyond some of these mainstream applications, deep learning methods are also being used to solve inverse imaging problems (*7–13*).

Here we introduce an all-optical deep learning framework in which the neural network is physically formed by multiple layers of diffractive surfaces that work in collaboration to optically perform an arbitrary function that the network can statistically learn. Whereas the inference and prediction mechanism of the physical network is all optical, the learning part that leads to its design is done through a computer. We term this framework a diffractive deep neural network (D[2]NN) and demonstrate its inference capabilities through both simulations and experiments. Our D[2]NN can be physically created by using several transmissive and/or reflective layers (*14*), where each point on a given layer either transmits or reflects the incoming wave, representing an artificial neuron that is connected to other neurons of the following layers through optical diffraction (Fig. 1A). In accordance with the Huygens-Fresnel principle, our terminology is

based on each point on a given layer acting as a secondary source of a wave, the amplitude and phase of which are determined by the product of the input wave and the complex-valued transmission or reflection coefficient at that point [see (*14*) for an analysis of the waves within a D[2]NN]. Therefore, an artificial neuron in a D[2]NN is connected to other neurons of the following layer through a secondary wave modulated in amplitude and phase by both the input interference pattern created by the earlier layers and the local transmission or reflection coefficient at that point. As an analogy to standard deep neural networks (Fig. 1D), one can consider the transmission or reflection coefficient of each point or neuron as a multiplicative "bias" term, which is a learnable network parameter that is iteratively adjusted during the training process of the diffractive network, using an error back-propagation method. After this numerical training phase, the D[2]NN design is fixed and the transmission or reflection coefficients of the neurons of all layers are determined. This D[2]NN design—once physically fabricated using techniques such as 3D-printing or lithography—can then perform, at the speed of light, the specific task for which it is trained, using only optical diffraction and passive optical components or layers that do not need power, thereby creating an efficient and fast way of implementing machine learning tasks.

In general, the phase and amplitude of each neuron can be learnable parameters, providing a complex-valued modulation at each layer, which improves the inference performance of the diffractive network (fig. S1) (*14*). For coherent transmissive networks with phase-only modulation, each layer can be approximated as a thin optical element (Fig. 1). Through deep learning, the phase values of the neurons of each layer of the diffractive network are iteratively adjusted (trained) to perform a specific function by feeding training data at the input layer and then computing the network's output through optical

diffraction. On the basis of the calculated error with respect to the target output, determined by the desired function, the network structure and its neuron phase values are optimized via an error back-propagation algorithm, which is based on the stochastic gradient descent approach used in conventional deep learning (*14*).

To demonstrate the performance of the D[2]NN framework, we first trained it as a digit classifier to perform automated classification of handwritten digits, from 0 to 9 (Figs. 1B and 2A). For this task, phase-only transmission masks were designed by training a five-layer D[2]NN with 55,000 images (5000 validation images) from the MNIST (Modified National Institute of Standards and Technology) handwritten digit database (*15*). Input digits were encoded into the amplitude of the input field to the D[2]NN, and the diffractive network was trained to map input digits into 10 detector regions, one for each digit. The classification criterion was to find the detector with the maximum optical signal, and this was also used as a loss function during the network training (*14*).

After training, the design of the D[2]NN digit classifier was numerically tested using 10,000 images from the MNIST test dataset (which were not used as part of the training or validation image sets) and achieved a classification accuracy of 91.75% (Fig. 3C and fig. S1). In addition to the classification performance of the diffractive network, we also analyzed the energy distribution observed at the network output plane for the same 10,000 test digits (Fig. 3C), the results of which clearly demonstrate that the diffractive network learned to focus the input energy of each handwritten digit into the correct (i.e., the target) detector region, in accord with its training. With the use of complex-valued modulation and increasing numbers of layers, neurons, and connections in the diffractive network, our classification accuracy can be further improved (figs. S1 and S2). For example, fig. S2 demonstrates a Lego-like physical transfer learning behavior for D[2]NN framework, where the inference performance of an already existing D[2]NN can be further improved by adding new diffractive layers—or, in some cases, by peeling off (i.e., discarding) some of the existing layers—where the new layers to be added are trained for improved inference (coming from the entire diffractive network: old and new layers). By using a patch of two layers added to an existing and fixed D[2]NN design (*N* = 5 layers), we improved our MNIST classification accuracy to 93.39% (fig. S2) (*14*); the state-of-the-art convolutional neural network performance has been reported as 99.60 to 99.77% (*16–18*). More discussion on reconfiguring D[2]NN designs is provided in the supplementary materials (*14*).

Following these numerical results, we 3D-printed our five-layer D[2]NN design (Fig. 2A), with each layer having an area of 8 cm by 8 cm, followed by 10 detector regions defined at the output plane of the diffractive network (Figs. 1B and 3A). We then used continuous-wave illumination at 0.4 THz to test the network's inference performance (Figs. 2, C and D). Phase values of

[1]Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095, USA. [2]Department of Bioengineering, University of California, Los Angeles, CA 90095, USA. [3]California NanoSystems Institute (CNSI), University of California, Los Angeles, CA 90095, USA. [4]Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: ozcan@ucla.edu

each layer's neurons were physically encoded using the relative thickness of each 3D-printed neuron. Numerical testing of this five-layer D²NN design achieved a classification accuracy of 91.75% over ~10,000 test images (Fig. 3C). To quantify the match between these numerical testing results and our experiments, we 3D-printed 50 handwritten digits (five different inputs per digit), selected among the same 91.75% of the test images for which numerical testing was successful. For each input object that is uniformly illuminated with the terahertz source, we imaged the output plane of the D²NN to map the intensity distribution for each detector region that is assigned to a digit. The results (Fig. 3B) demonstrate the success of the 3D-printed diffractive neural network and its inference capability: The average intensity distribution at the output plane of the network for each input digit clearly reveals that the 3D-printed D²NN was able to focus the input energy of the beam and achieve a maximum signal at the corresponding detector region assigned for that digit. Despite 3D-printing errors, possible alignment issues, and other experimental error sources in our setup (14), the match between the experimental and numerical testing of our five-layer D²NN design was found to be 88% (Fig. 3B). This relatively small reduction in the perform-

ance of the experimental network compared to our numerical testing is especially pronounced for the digit 0 because it is challenging to 3D-print the large void region at the center of the digit. Similar printing challenges were also observed for other digits that have void regions; e.g., 6, 8, and 9 (Fig. 3B).

Next, we tested the classification performance of D²NN framework with a more complicated image dataset—i.e., the Fashion-MNIST dataset (19), which includes 10 classes, each representing a fashion product (t-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots; see fig. S3 for sample images). In general, for a coherently illuminated D²NN, we can use the amplitude and/or phase channels of the input plane to represent data to be classified or processed. In our digit classification results reported earlier, input objects were encoded by using the amplitude channel, and to demonstrate the utility of the phase channel of the network input, we encoded each input image corresponding to a fashion product as a phase-only object modulation (14). Our D²NN inference results (as a function of the number of layers, neurons, and connections) for classification of fashion products are summarized in figs. S4 and S5. To provide an example of our performance, a phase-only and a complex-

valued modulation D²NN with $N = 5$ diffractive layers (sharing the same physical network dimensions as the digit classification D²NN shown in Fig. 2A) reached an accuracy of 81.13 and 86.33%, respectively (fig. S4). By increasing the number of diffractive layers to $N = 10$ and the total number of neurons to 0.4 million, our classification accuracy increased to 86.60% (fig. S5). For convolutional neural net–based standard deep learning, the state-of-the-art performance for Fashion-MNIST classification accuracy has been reported as 96.7%, using ~8.9 million learnable parameters and ~2.5 million neurons (20).

To experimentally demonstrate the performance of fashion product classification using a physical D²NN, we 3D-printed our phase-only five-layer design and 50 fashion products used as test objects (five per class) on the basis of the same procedures employed for the digit classification diffractive network (Figs. 2A and 3), except that each input object information was encoded in the phase channel. Our results are summarized in Fig. 4, revealing a 90% match between the experimental and numerical testing of our five-layer D²NN design, with five errors out of 50 fashion products. Compared with digit classification (six errors out of 50 digits; Fig. 3), this experiment yielded a slightly better match
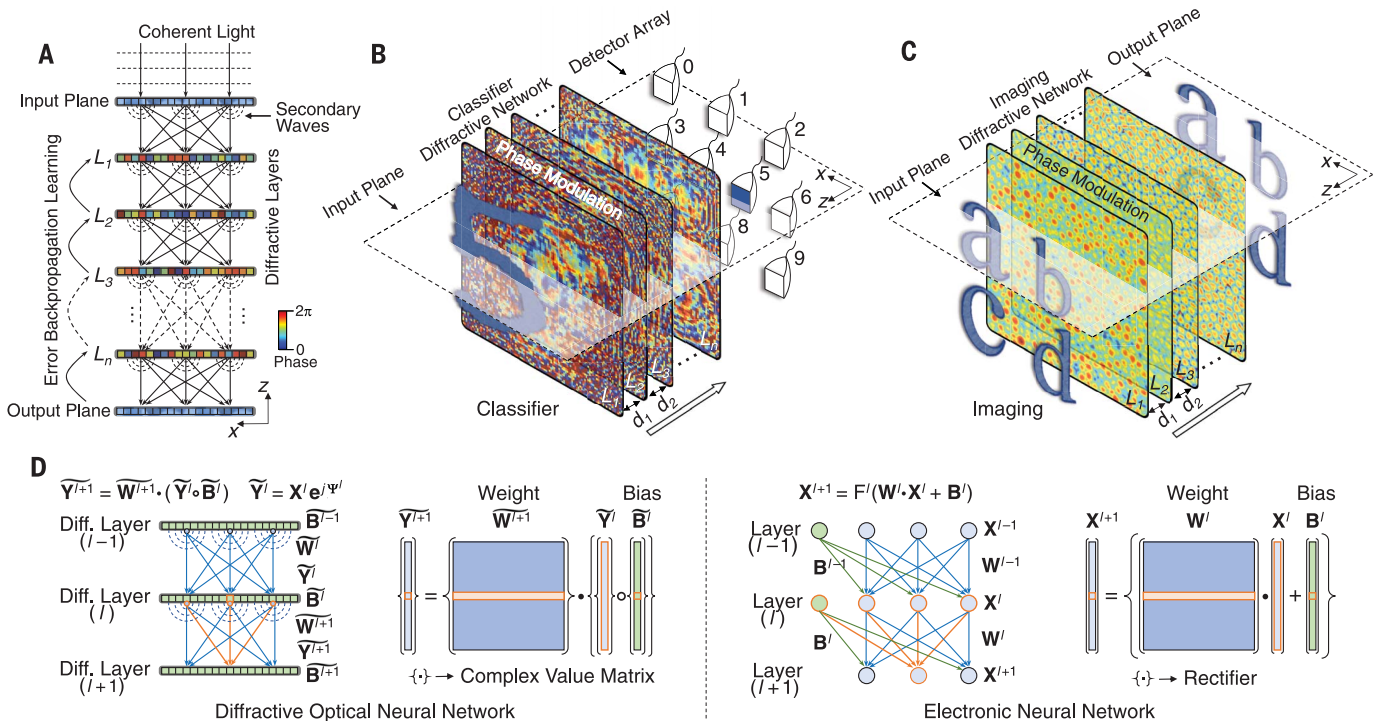
Fig. 1. Diffractive deep neural networks (D²NNs). (A) A D²NN comprises multiple transmissive (or reflective) layers, where each point on a given layer acts as a neuron, with a complex-valued transmission (or reflection) coefficient. The transmission or reflection coefficients of each layer can be trained by using deep learning to perform a function between the input and output planes of the network. After this learning phase, the D²NN design is fixed; once fabricated or 3D-printed, it performs the learned function at the speed of light. L, layer. (**B** and **C**) We trained and experimentally implemented different types of D²NNs: (B) classifier (for handwritten digits and fashion products) and (C) imager.

d, distance. (**D**) Comparison between a D²NN and a conventional neural network (14). Based on coherent waves, the D²NN operates on complex-valued inputs, with multiplicative bias terms. Weights in a D²NN are based on free-space diffraction and determine the interference of the secondary waves that are phase- and/or amplitude-modulated by the previous layers. "o" denotes a Hadamard product operation. "Electronic neural network" refers to the conventional neural network virtually implemented in a computer. **Y**, optical field at a given layer; **Ψ**, phase of the optical field; **X**, amplitude of the optical field; F, nonlinear rectifier function [see (14) for a discussion of optical nonlinearity in D²NN].
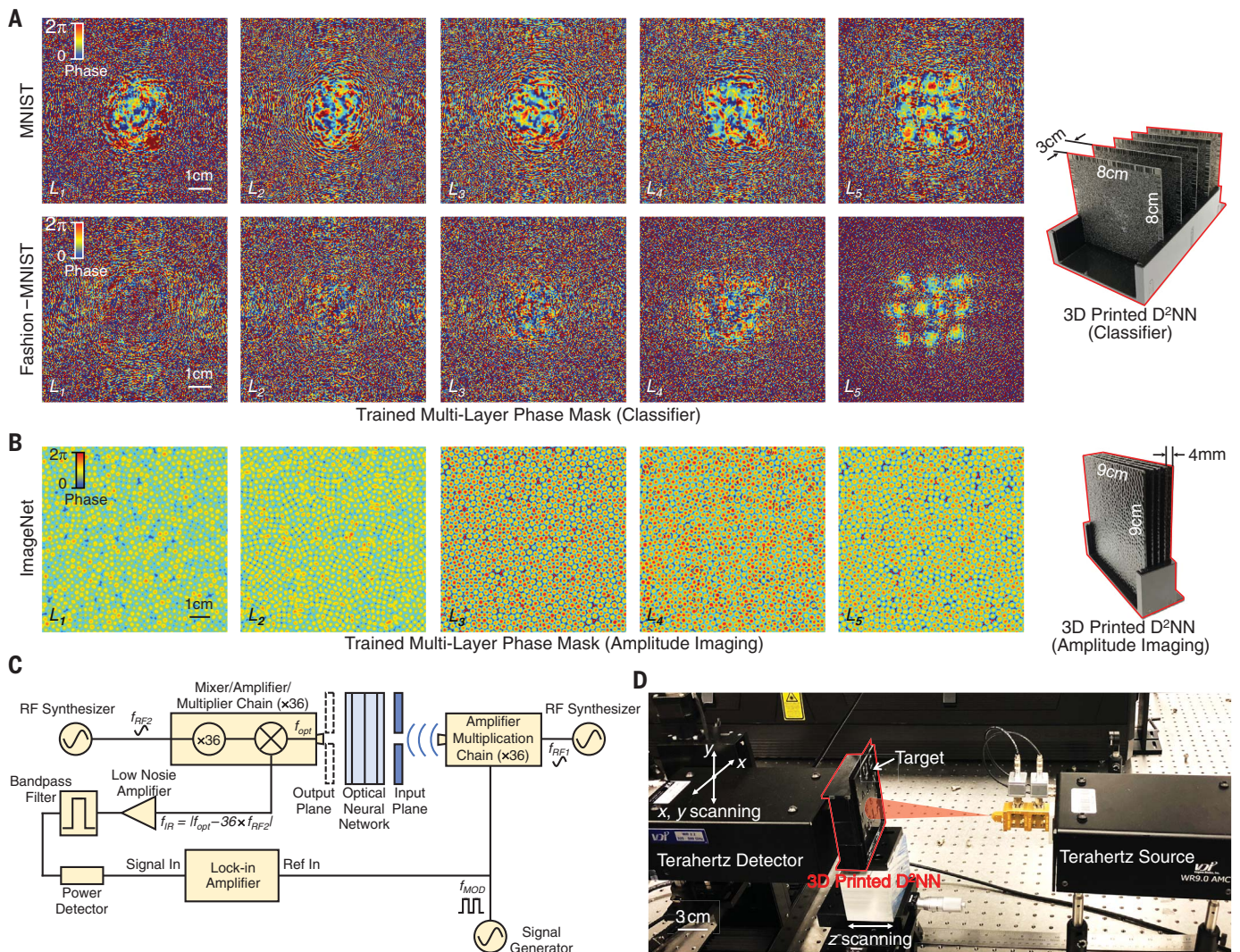
Fig. 2. Experimental testing of 3D-printed D²NNs. (**A** and **B**) After the training phase, the final designs of five different layers (L_1, L_2, ..., L_5) of the handwritten digit classifier, fashion product classifier, and the imager D²NNs are shown. To the right of the network layers, an illustration of the corresponding 3D-printed D²NN is shown. (**C** and **D**) Schematic (C) and photo (D) of the experimental terahertz setup. An amplifier-multiplier chain was used to generate continuous-wave radiation at 0.4 THz, and a mixer-amplifier-multiplier chain was used for the detection at the output plane of the network. RF, radio frequency; f, frequency.

between the experimental and numerical testing results (despite the more challenging nature of Fashion-MNIST dataset), perhaps because we used the phase channel, which does not suffer from the challenges associated with 3D-printing of void regions [such as in digits 0, 6, 8, and 9 (Fig. 3)], to encode input image information for fashion products.

Next, we tested the performance of a phase-only D²NN, composed of five 3D-printed transmission layers to implement amplitude imaging (Fig. 2B). This network was trained using the ImageNet database (*21*) to create a unit-magnification image of the input optical field amplitude at its output plane (~9 cm by 9 cm)—that is, the output image has the same physical size as the input object (*14*). As illustrated in fig. S6, A and C, the trained network initially connects every amplitude point at the input plane to various neurons and features of the fol-

lowing layers, which then focus the light back to a point at the output (i.e., image) plane, which is, as expected, quite different than the case of free-space diffraction (i.e., without the presence of the diffractive network), illustrated in fig. S6, B and D.

After training and blind testing, which served to numerically prove the imaging capability of the network (figs. S6 and S7), we then 3D-printed this designed D²NN. Using the same experimental setup shown in Fig. 2, C and D, we imaged the output plane of the 3D-printed D²NN for various input objects that were uniformly illuminated by continuous-wave radiation at 0.4 THz. Figure S8 summarizes our experimental results achieved with this 3D-printed D²NN, which successfully projected unit-magnification images of the input patterns at the output plane of the network, learning the function of an imager, or a physical auto-encoder. To evaluate the point spread

function of this D²NN, we imaged pinholes with different diameters (1, 2, and 3 mm), which resulted in output images, each with a full width at half maximum of 1.5, 1.4, and 2.5 mm, respectively (fig. S8B). Our results also revealed that the printed network can resolve a linewidth of 1.8 mm at 0.4 THz (corresponding to a wavelength of 0.75 mm in air), which is slightly worse in resolution compared with the numerical testing of our D²NN design, where the network could resolve a linewidth of ~1.2 mm (fig. S7C). This experimental degradation in the performance of the diffractive network can be due to factors such as 3D-printing errors, potential misalignments, and absorption-related losses in the 3D-printed network (*14*).

Optical implementation of machine learning in artificial neural networks is promising because of the parallel computing capability and power efficiency of optical systems (*22–24*). Compared

Fig. 3. Handwritten digit classifier D²NN. (**A**) A 3D-printed D²NN successfully classifies handwritten input digits (0, 1, ..., 9) on the basis of 10 different detector regions at the output plane of the network, each corresponding to one digit. As an example, the output image of the 3D-printed D²NN for a handwritten input of "5" is demonstrated, where the red dashed squares represent the trained detector regions for each digit. Other examples of our experimental results are shown in fig. S9. (**B**) Confusion matrix and energy distribution percentage for our experimental results, using 50 different handwritten digits (five for each digit) that were 3D-printed, selected among the images for which numerical testing was successful. (**C**) Same as (B), except summarizing our numerical testing results for 10,000 different handwritten digits (~1000 for each digit), achieving a classification

accuracy of 91.75% using a five-layer design. Our classification accuracy increased to 93.39% by increasing the number of diffractive layers to seven, using a patch of two additional diffractive layers added to an existing and fixed D²NN (fig. S2).
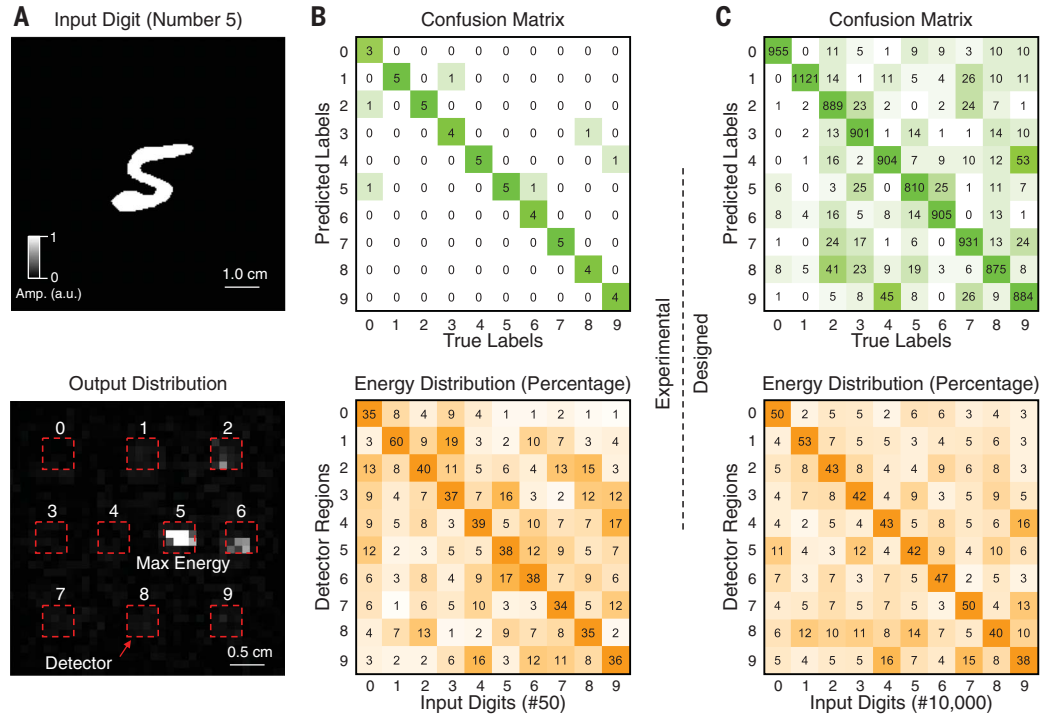
Fig. 4. Fashion product classifier D²NN. (**A**) As an example, the output image of the 3D-printed D²NN for a sandal input (Fashion-MNIST class 5) is demonstrated. The red dashed squares represent the trained detector regions for each fashion product. Other examples of our experimental results are shown in fig. S10. (**B**) Confusion matrix and energy distribution percentage for our experimental results, using 50 different fashion products (five per class) that were 3D-printed, selected among the images for which numerical testing was successful. (**C**) Same as (B), except summarizing our numerical testing results for 10,000 different fashion products (~1000 per class), achieving a classification accuracy of 81.13% using a five-layer design. By increasing the number of diffractive layers to 10, our classification accuracy increased to 86.60% (fig. S5).
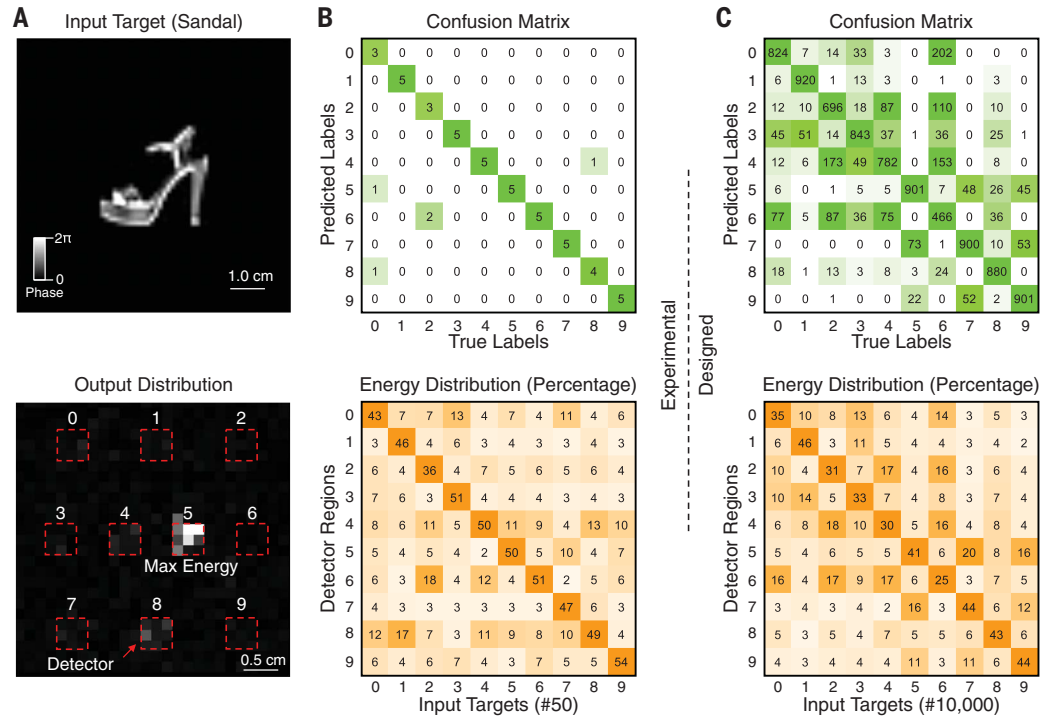
with previous optoelectronics-based learning approaches (*22, 25–27*), the D²NN framework provides a distinctive all-optical machine learning engine that efficiently operates at the speed of light using passive components and optical diffraction. An important advantage of D²NNs is that they can be easily scaled up using various high-throughput and large-area 3D-fabrication methods (such as soft lithography and additive manufacturing), as well as wide-field optical components and detection systems, to cost-effectively reach tens to hundreds of millions of neurons and hundreds of billions of connections in a scalable and power-efficient manner. For example, integration of D²NNs with lensfree on-chip imaging systems (*28, 29*) could provide extreme parallelism within a cost-effective and portable platform. Such large-scale D²NNs may

be transformative for various applications, including image analysis, feature detection, and object classification, and may also enable new microscope or camera designs that can perform specific imaging tasks using D^2NNs. To achieve these new technologies, nonlinear optical materials (14) and a monolithic D^2NN design that combines all layers of the network as part of a 3D-fabrication method would be desirable. Among other techniques, laser lithography based on two-photon polymerization (30) can provide solutions for creating such D^2NNs.

REFERENCES AND NOTES

1. Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**, 436–444 (2015).
2. G. Litjens *et al.*, *Med. Image Anal.* **42**, 60–88 (2017).
3. A. Graves, A. Mohamed, G. Hinton, in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2013), pp. 6645–6649.
4. K. Cho *et al.*, arXiv:1406.1078 [cs.CL] (3 June 2014).
5. A. Krizhevsky, I. Sutskever, G. E. Hinton, in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger, Eds. (Curran Associates, 2012), pp. 1097–1105.
6. D. Silver *et al.*, *Nature* **529**, 484–489 (2016).
7. U. S. Kamilov *et al.*, *Optica* **2**, 517 (2015).
8. Y. Rivenson *et al.*, *Optica* **4**, 1437 (2017).
9. K. H. Jin, M. T. McCann, E. Froustey, M. Unser, *IEEE Trans. Image Process.* **26**, 4509–4522 (2017).
10. Y. Rivenson, Y. Zhang, H. Gunaydin, D. Teng, A. Ozcan, *Light Sci. Appl.* **7**, 17141 (2018).
11. A. Sinha, J. Lee, S. Li, G. Barbastathis, *Optica* **4**, 1117 (2017).
12. K. Hammernik *et al.*, *Magn. Reson. Med.* **79**, 3055–3071 (2018).
13. Y. Rivenson *et al.*, *ACS Photonics* **5**, 2354–2364 (2018).
14. Materials and methods are available as supplementary materials.
15. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, *Proc. IEEE* **86**, 2278–2324 (1998).
16. D. Ciregan, U. Meier, J. Schmidhuber, in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2012), pp. 3642–3649.
17. C.-Y. Lee, P. W. Gallagher, Z. Tu, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* (AISTATS, 2016), pp. 464–472.
18. M. A. Ranzato, C. Poultney, S. Chopra, Y. LeCun, in *Advances in Neural Information Processing Systems 20 (NIPS 2007)*, J. C. Platt, D. Koller, Y. Singer, S. T. Roweis, Eds. (Curran Associates, 2007), pp. 1137–1144.
19. H. Xiao, K. Rasul, R. Vollgraf, arXiv:1708.07747 [cs.LG] (25 August 2017).
20. S. Zeng, B. Zhang, J. Gou, arXiv:1802.07589 [cs.CV] (21 February 2018).
21. ImageNet database, www.image-net.org.
22. Y. Shen *et al.*, *Nat. Photonics* **11**, 441–446 (2017).
23. D. Psaltis, D. Brady, X.-G. Gu, S. Lin, *Nature* **343**, 325–330 (1990).
24. K. H. Wagner, in *Proceedings of Frontiers in Optics 2017* (OSA, 2017), FW2C.1.
25. B. J. Shastri *et al.*, arXiv:1801.00016 [cs.ET] (29 December 2017).
26. M. Hermans, M. Burm, T. Van Vaerenbergh, J. Dambre, P. Bienstman, *Nat. Commun.* **6**, 6729 (2015).
27. D. Brunner, M. C. Soriano, C. R. Mirasso, I. Fischer, *Nat. Commun.* **4**, 1364 (2013).
28. A. Greenbaum *et al.*, *Nat. Methods* **9**, 889–895 (2012).
29. A. Ozcan, E. McLeod, *Annu. Rev. Biomed. Eng.* **18**, 77–102 (2016).
30. M. Emons *et al.*, *Opt. Mater. Express* **2**, 942–947 (2012).