Supplementary Material for

Programming Nonlinear Propagation for Efficient Optical Learning Machines

Ilker Oguz,[a,*] Jih-Liang Hsieh,[a] Niyazi Ulas Dinc,[a] Uğur Teğin,[a,b] Mustafa Yildirim,[a] Carlo Gigli,[a] Christophe Moser[a] and Demetri Psaltis[a]

[a] Ecole Polytechnique Fédérale de Lausanne (EPFL), Institute of Electrical and Micro Engineering, Ecublens, Switzerland, 1015
[b] Koc University, Department of Electrical and Electronics Engineering, Istanbul, Turkey, 34450

Contents

Supplementary Note 1: The Stability of the Experiment over Time

Due to the fluctuation in environmental conditions and other noise factors, fluctuation in the laser output power is expected. To decouple this fluctuation from the optical computing setup, light intensity at the output of the multimode fiber is tracked with a power meter and corrections to the angle of polarization before the polarizing beamsplitter are applied according to this reading. This way the light intensity inside the optical fiber can be kept stable. Moreover, when wavefront shaping is applied, the same reading is also used for making sure the intensity of light coupled inside the MMF is directly controlled by the intensity programming parameter. To investigate reproducibility, the inference experiment is redone for the same PPs and RWs on the same task over 15 hours continuously by sending the samples to the optical system. During those experiments, the inference accuracy, the stabilized power level, as well as the average correlation between samples are recorded. This correlation is calculated by the average of correlation coefficients for each input sample, between the initial output beam shape and the output beam shape for the current experiment. As shown in Supplementary Figure 1, the fluctuation in the correlation values and power levels resemble each other closely, and they are both much smaller than 1% over 14 hours. The average beam correlation in 14 hours is 99.9% with a 0.04% standard deviation and compared to its maximum, the power level is 99.7% on average with a 0.14% standard deviation. Consequently, the accuracy fluctuations are minimal, with the first and final test accuracy being exactly the same at 83.7%, the average value is 83.8% and the standard deviation is 0.35%. Moreover, single-step readout weight training allows lightweight recalibration, which can further decrease fluctuations by applying once in many experiments.

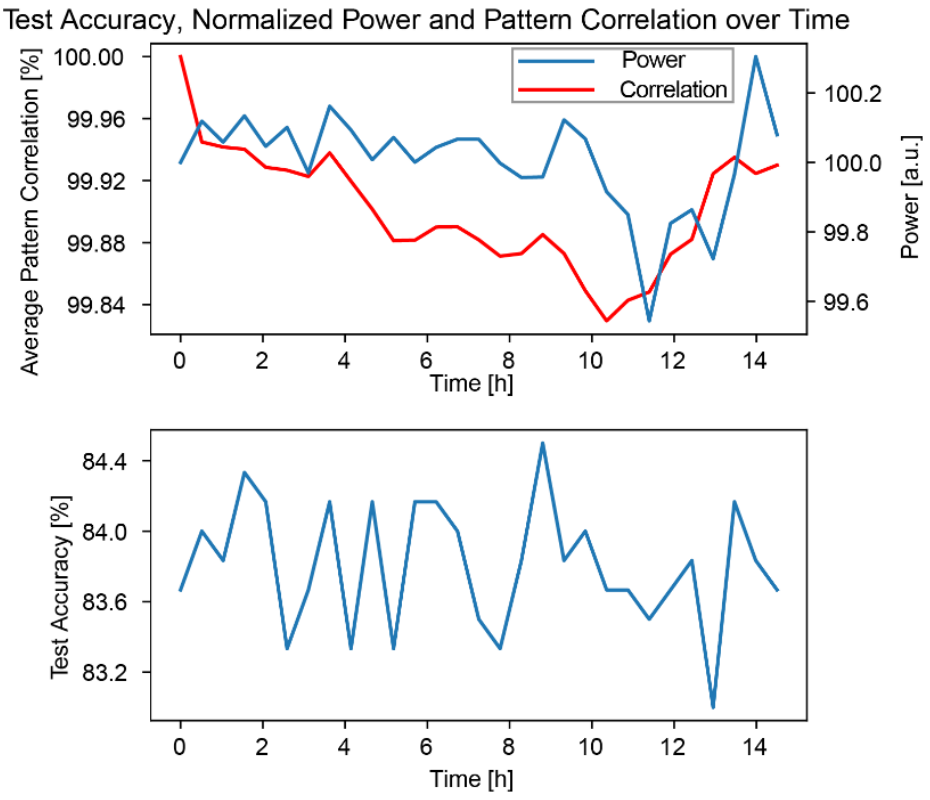

Fig. S1 Coupled power, output pattern correlation, and classification accuracy of the system over time

Supplementary Note 2: Operation Speed and Power Consumption Analysis of the Computation Method

Dimensions and Operation Count

The experimental results on different datasets indicate that programming MMF propagation can achieve similar accuracies with digital NNs requiring ~1 MFLOP/sample. Therefore, for the rest of the document for each image processed with the optical system, this process will be accounted as equivalent to 1 MFLOP of operation on the GPU. And for each operation, the input is assumed to be 22x22 pixels with 8 bit-depth and the output is 45x45 pixels with 8 bit-depth. These resolutions are determined by the supported modes on the fiber. On the input side, the diffraction-limited resolution is sampled by 2 input pixels and on the output side, the sampling is about 4 pixels per diffraction-limited point size because of the spatially displayed temporal information thanks to the diffraction grating.

Optical Power Consumption

For optimal performance, the optical system requires 10 ps long pulses with about 10 kW peak optical power inside the fiber. Therefore with 100 nJ($=10^{-11}$x10^{4}nJ) per pulse, when only the optical energy consumption is considered, the energy efficiency for the equivalent digital computation is 0.1 pJ/FLOP.

Power Consumption for Optoelectrical Conversion

Currently, the data is encoded to optical pulses via a liquid crystal phase modulator controlled by an analog high-speed circuitry, it has Meadowlark HSP1920-600-1300 with 8-bit precision, each pixel sized 9.2 μm x 9.2 μm working at ~ 50 Hz or 20 ms per frame. The SLM consumes in total (24V,1A) 24 W continuously. Therefore, each refresh consumes around 0.5 J (=24Wx20ms), or 217nJ/pixel (0.5/(1920x1152)). For optimal utilization of the SLM, in which all pixel information couples to a fiber, by using multiple fibers or only activating the subgroup of pixels with light incident on them, the consumption could be brought to 22x22 pixels x217 nJ/pixel = 105 μJ/image.

We can replace the phase-only LC SLM with a DMD, for instance, Texas Instruments DLP9500, which can show 23148 patterns per second with a 1920x1080 resolution and 4.5 W electrical consumption[1]. Therefore, the power consumption per frame is 196 μJ and per mirror is 94 pJ. For modulation of 22x22x8 bits per input image, the energy cost of modulation would be 0.4 μJ.

For recording output beam shapes, a commercial CMOS camera is used, the camera has a resolution of 720 x 540 pixels and can reach 522 frames per second, consuming 3W. This amounts to 5.7 mJ for each full frame recording and 15 nJ per each pixel readout, or for 45x45 pixel output images, 30 μJ /image. Again, this camera can be replaced with another commercially available, but more power-efficient version; for instance, LUX 1310, can read out with 1.4nJ/pixel, which can decrease the consumption down to 2.8μJ/image.

The energy consumption of laser is calculated by taking fiber coupling, SLM diffraction, and laser efficiency into consideration. The efficiency of coupling into MMF is around 50%, for both the LC SLM and the DMD light diffraction efficiency greater than 70% and a Yb-doped fiber-based femtosecond laser can convert the electrical energy to light pulses with about 3.3% efficiency[2]. Therefore, the electrical power cost of a 100 nJ pulse inside the fiber is calculated to be 8.7 μJ.

Table S1 The energy budget breakdown of the proposed optical computing method

Optoelectronic Device	Energy Consumption per Image-Consumption of All Pixels on the Device	Energy Consumption per Image-Consumption Scaled to Required Pixels
LC SLM	0.5 J	105 μJ
DMD	196 μJ	0.4 μJ
Camera	5.7 mJ	30 μJ
Laser Consumption	8.7 μJ	8.7 μJ
Total Consumption-LC SLM	506 mJ	144 μJ
Total Consumption-DMD	5.9 mJ	39 μJ

The energy consumption of each component of the experiment is provided on Supplementary Table 1. In the two columns of the table, two different ways of accounting for the energy are presented, the first column accounts for the case where consumption due to all of the possible pixels is included. In the second column, the energy consumption of the device is scaled to the proportion which is actually used. Moreover, in the calculation of total consumption two alternatives are presented, the one with the LC SLM is the current experimental setting, while the DMD option is the case when LC SLM is switched with a DMD while keeping the experiment exactly the same. In the latter, by turning on only the necessary pixels per image energy consumption could be lowered to 39 μJ per image, corresponding to 39 pJ/FLOP. In comparison, benchmark results show that an NVIDIA V100 GPU provides inferences with ResNet-50 neural network at an energy consumption rate of 46 pJ/FLOP[3].

Switching from an LC SLM to a DMD could improve the inference speed by nearly 3 orders of magnitude, reaching up to 23000 frames/second, and allowing 23GFLOP/s. However, as it can be observed in the optical telecommunication systems, the transfer and manipulation of information in the optical domain supports conveniently GHz-level rates. To demonstrate that this potential indeed encompasses the proposed computation method, in the following section we describe an implementation that can reach the processing speed of state-of-the-art GPUs, with a favorable efficiency.

The energy efficiency and high-performance computing potential of the proposed method can be realized with a high-speed system as shown in Supplementary Figure 4.

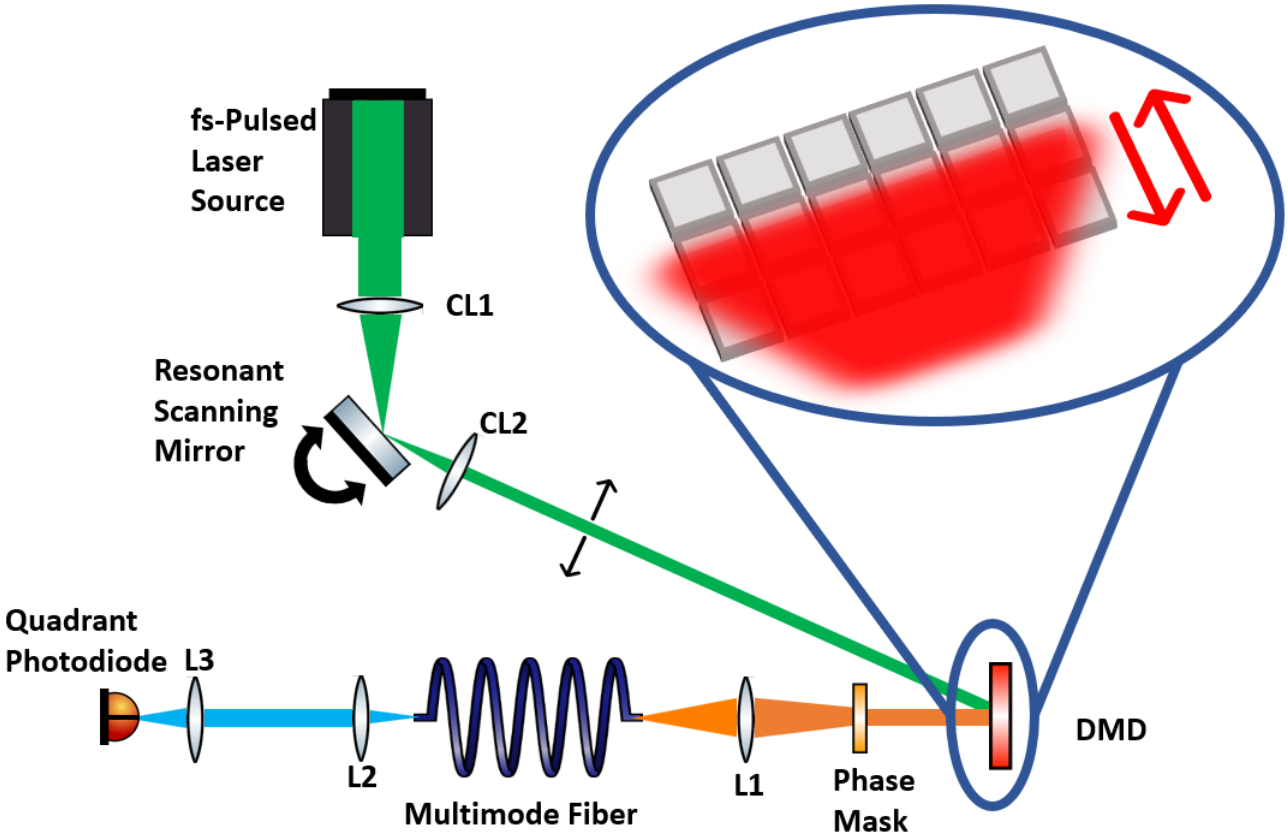

Fig. S2 The experimental schematic of a high-speed implementation of the proposed method

For reaching high data processing speed efficiently, some of the system parameters and components should be changed without changing the main principle of computation. These changes are:

1- Pulse length
2- Spatial modulation device
3- Fiber length
4- Readout device.

As the strength of optical nonlinearities is proportional to the peak power of the optical pulse, the required average optical power for the same strength of the nonlinearity could be decreased by working with a shorter pulse. In the high-speed system, instead of the current 10 ps pulses, 300 fs can be used with the same laser source (Amplitude Laser, Satsuma). This can decrease the light source related consumption by 33 times, down to 264 nJ/image.

In our current proof-of-principle system a liquid-crystal SLM is utilized. Even though their ability to directly control the phase makes them ideal prototyping tools, the slow response of liquid crystals (a few milliseconds) limits the modulation speed with the LC SLMs. On the other hand, digital micromirror devices offer a much faster alternative with their millions of fast switchable mirrors (few microseconds), and they were successfully utilized in optical neural network realizations[4,5]. To benefit fully from DMD's speed and the high number of pixels, while only a few hundreds of channels exist in the multimode fiber, we propose using a narrow beam and sweeping the lines of the DMD with a resonant scanner as shown in Supplementary Figure 4. For instance, the utilization of a ~11.5 kHz resonant mirror (Novanta Photonics, CRS 12) whose upwards and downwards cycles are synchronized to 23148 Hz rate refreshing with a DMD (TI DLP9500), could achieve the processing of individual lines on the DMD, each line consisting of 1920 pixels would encode an input to be processed by the system. Therefore, 1920/240 = 8 bits of information can be coupled to each one of the 240 channels of the fiber and 23148x1080 ≈25 million samples per second can be shown. For the conversion of 1-dimensional data and programming to 2-dimensions, a scattering phase mask can be used[6]. After coupling into the fiber and propagation, the output beam location, which was shown to be programmable to directly provide inference results without any need for a digital readout layer, can be tracked with a sectioned photodiode such as Excelitas C30665GH-4. Therefore, 25 million inferences per second could be performed, and equivalently 25 TFLOP/s processing speed could be reached. In comparison, during the benchmarking on the ResNet50 neural network, as NVIDIA V100 GPU has a peak consumption of 300 W[3], with 46 pJ/FLOP, it can provide 6.5 TFLOP/s.

Table S2 The energy budget breakdown of the high-speed implementation of the optical computing method

Optoelectronic Device	Power Consumption (W)
DMD	4.4
Resonant Mirror	1.5
Quadrant Photodiode	0.1
Femtosecond Laser	6.6
Total Consumption	12.6

In addition to the high-speed computation, the proposed implementation is power efficient. As the breakdown can be seen in Supplementary Table 2, while performing inferences with 25 million samples per second and the power consumption would be 12.6 W, corresponding to 0.5 pJ/FLOP efficiency.

In summary, the overall energy consumption of the system depends on the energy cost of input modulation and reading out of pixels, in addition to the energy required to create a pulsed laser beam. As the pulse repetition rates of femtosecond lasers can reach up to GHz levels, the computation speed is effectively limited by the slowest of the optoelectronic conversion devices. The proof-of-principle system with a

selection of optoelectronic equipment not being optimized for power consumption can achieve computing with power consumption similar to the GPU. Replacing only the SLM of this setup with a DMD is expected to improve further the energy efficiency to an advantageous level and speed, since DMD is faster than LC SLMs and per bit modulation cost is lower. Scanning different locations of the DMD with the laser beam and reading out the inference result all-optically is expected to bring down the energy consumption 2 orders of magnitude while providing significantly faster computation.

The potential of this computing system can be further explored by decreasing the dominant power expenditure on optoelectronic modulation by employing silicon photonics based modulation, which can achieve an efficiency of 0.1 pJ/bit at 10 Gbit/s [7]. In addition to further improving the per bit efficiency 3 orders of magnitude compared to the DMD, with its high speed, it removes the need for the resonant mirror. In this case, the dominant power expenditure would be from the femtosecond laser and this expenditure could be again decreased with an optimal fiber, with which nonlinearities could be enhanced more than 10 times for the same peak power level or could provide the same amount of nonlinearities with 10 times lower peak powers[8].

Furthermore, for the proposed improvements, fiber mode number, hence the input and output resolutions are assumed to be kept the same. However, as we discuss in the next section, the efficiency of the system can be increased by scaling the number of modes.

Discussion on Scalability of the Approach

In the previous sections, different implementations are proposed to improve the speed and power efficiency while the fiber dimensions are kept the same. Another possible approach would be to increase the number of propagation modes, $N = \frac{\alpha}{\alpha+2} n_0^2 k^2 R^2 \Delta$ [9]. This number scales with the square of the core radius, R^2, with the square of the numerical aperture of the fiber ($NA^2 = 2\Delta n_0^2$), and the inverse square of the light wavelength ($\lambda = 2\pi/k$). As N could be increased with simply a different selection of the fiber and light source, the number of effective calculations is expected to scale with at least N^2 and potentially even steeper. This expectation stems from the interaction terms in the Multimode Generalized Nonlinear Schrodinger Equation (Eqn. S1). Even in the absence of nonlinear interactions, linear interactions create some amount of coupling from each mode to every other mode for each infinitesimal propagation step. With nonlinear interactions, these interactions become products of four mode coefficients, which create a number of interactions on the N^4 order. On the other hand, the sample input rate and energy consumption are proportional to N, for instance, if modulation with M features takes t time, with the proposed laser scanning method, modulation with 2M features for 2N modes will take 2t. Similarly, for the same level of nonlinearity the optical peak power per mode (P/N) should be kept the same, so for a 2N mode fiber, the optical power consumption should be 2P, while the number of operations should scale to at least 4C from C. Then, increasing the dimensionality as much as possible should clearly improve the power efficiency and

operation speed. Even with the same wavelength, a 100-fold increase in the dimensionality and an improvement in the operation speed and power efficiency at the same ratio can easily be achieved by switching the current fiber with the core diameter of 50 μm to another commercially available fiber with a 500 μm core diameter.

Supplementary Method 1: Comparison with Digital Neural Networks

To create a one-to-one comparison between the presented optical computation method and electronics-based digital implementations, we used the same portions of the same dataset for training and testing. The digital neural networks were implemented with the Keras software library, with stochastic gradient descent, 64 sample batches. In addition to LeNet-1 and LeNet-5 [10] implementations, a larger version of LeNet-5 with 9 layers and higher numbers of filters per layer was benchmarked. This neural network has the same input size and nonlinear activations as the LeNet-5 while it consists of 2 consecutive 2-dimensional convolution layers with 48, 3x3 kernels, a 2-dimensional Max Pooling, 2 consecutive 2-dimensional convolution layers with 96, 3x3 kernels, a 2-dimensional Max Pooling, a 2-dimensional convolution layers with 192, 3x3 kernels, a 2-dimensional Max Pooling, a dense layer with 512 neurons, and finally a dense layer with neurons as many as output classes.

EfficientNetB6[11] benchmark is run by importing the model from Keras library with the weights trained on the ImageNet dataset. Input dimensions are set as 45x45, the same with the optical system's output sampling. The output of the model is formed with a fully connected layer with softmax activation. Initially this layer is trained for 20 epochs with Adam optimizer at 10^{-3} learning rate, while EfficientNetB6 weights are fixed. Then, in the fine-tuning step, whole model is set as trainable and trained for 20 epochs with Adam optimizer, 10^{-5} learning rate. Batches are again composed of 64 samples.

For each one of the models and datasets, the training is done with 5 different random starting of the neural network, the test set accuracies are recorded, and the mean and the standard deviation of these trials are reported.

Supplementary Method 2: Detailed Experimental Setup

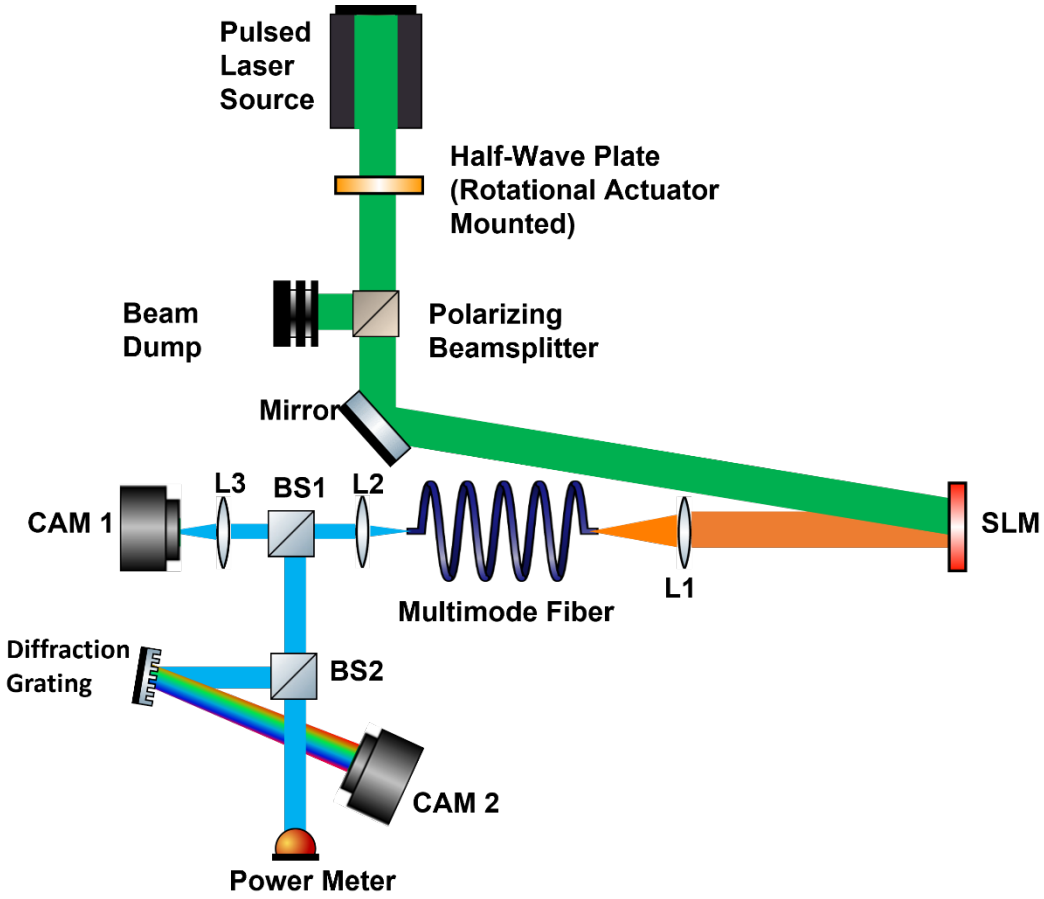

Fig. S3 The detailed schematic of the experimental setup

The experimental system shapes the wavefront of a pulsed IR laser beam, couples it into a graded-index fiber, then collimates and sends it to two different imaging arms. The laser pulses with 10 ps length, 125 kHz repetition rate, and 1030 nm center wavelength are provided by a mode-locked ytterbium fiber laser (Amplitude Laser, Satsuma). The intensity of the laser beam is controlled from the computer via a half-wave plate mounted on a motorized rotation stage (Thorlabs PRM1) followed by a polarizing beam splitter. Then, the beam is incident on the reflective phase-only two-dimensional spatial light modulator with 1920-by-1152 pixels and the pitch size (Meadowlark HSP1920-600-1300). The 8-bit SLM is calibrated to provide phase modulation between 0 to 2π at each pixel location, for transmitted pixel values between 0 to 255. The laser beam has a circular shape with approximately 6.7 mm diameter on the SLM surface, therefore all images shown on the SLM are converted to grayscale, upsampled to 520-by-520 to reach the size of the beam, and has the format of 8-bit unsigned integer. To isolate diffracted beam from the undiffracted portion, all images are shown on the SLM after elementwise addition with a blazed grating phase pattern and its phase depth is varied over the beam area to control the diffraction efficiency for each location, hence

achieving amplitude modulation on the diffracted beam. The first-order diffracted beam is clearly separated from the undiffracted beam after propagating 150 mm and focused by a plano-convex lens with a focal length of 15 mm. The input facet of commercially available graded-index MMF (OFS, bend-insensitive OM2) of 50 μm core diameter, 0.20 NA, 5 m long, is placed in the focal plane of the lens and its position is fine-tuned with a 3-dimensional alignment stage. After propagating through the fiber, the beam is collimated with a 20 mm focal length plano-convex lens. Then, it is separated into two beams by a non-polarizing 50-50 beam splitter. The transmitted beam is imaged onto a monochrome CMOS camera (FLIR BFS-U3-31S4M-C) with an achromatic doublet lens of 100 mm focal length. This camera is used for directly inspecting the beam shape and for calculating the center location of the beam. The reflected beam again goes through a beamsplitter, one of the branches arrive at an InGaAs power sensor (Thorlabs S145C), and the other branch is reflected from the diffraction grating (600 lines/mm, Thorlabs GR25-0610) and is incident onto the same model of CMOS camera. Neutral density filters are placed before the cameras in the optical setup to avoid saturation. The detailed schematic of the experiment is shown on Supp. Fig 5. All the electrical equipment is connected to the same general-purpose computer via USB and PCI-E ports.

Supplementary Method 3: Implementation of Programming Parameters

During this study, the programming of propagation is combined with the data on a digital computer, and the combinations were converted to optical signals, even though the combination of PP-controlled shaping and data is also realizable in a fully optical manner by means of a fixed 2-dimensional plate with a pre-determined spatially-varying distribution of phase and magnitude transfer properties. Three ways to program the propagation are presented. In addition to those programming methods based on wavefront shaping, for each experiment, 6 additional parameters controlling light intensity (I), displacement of the data on the SLM (Δx, Δy), diffraction angle on the SLM with respect to the optical axis of the system ($\Delta\theta$, $\Delta\phi$), and focal length of defocusing (f) are optimized.

The first programming method is illustrated in Figure 4.a. and performs an elementwise multiplication of fields due to the data and programming. The programming pattern is formed by the linear combination of analytically calculated propagation modes of the GRIN MMF. For determining the programming pattern as a combination of N modes, N parameter pairs are selected each for real and imaginary coefficients. Therefore the programming complex pattern is formed as $P(x,y) = \sum_{i=1}^{23} F_i\,(x,y) * (A_i + jB_i)$, where A_i and B_i are the coefficients for the i -th mode and $F_i(x,y)$ is the scalar field of the i -th mode. These mode fields could be expressed in terms of Laguerre polynomials[12], $L_p(x)$: $F_i(\rho,\phi) =$

$\sqrt{\frac{p!}{\pi(p+|m|)!}}\frac{\rho^{|m|}}{\rho_0^{|m|+1}}\mathrm{e}^{-\rho/2\rho_0^2}L_p^{|m|}\left(\frac{\rho^2}{\rho_0^2}\right)\mathrm{e}^{im\phi}$. p and m are radial and angular numbers of the modes, $\rho_0^2 = \frac{R}{k\sqrt{2\Delta}}$, R is the radius of the fiber, k is the wavenumber for the center of the fiber, and $\Delta = \frac{\mathrm{n}_1^2 - \mathrm{n}_0^2}{2n_1^2}$, relative refractive number difference. The input data, $D(x,y)$, is scaled to be between 0 and 2π, then the

combination of the programming pattern and other PPs constitutes the diffracted electric field: $E(x,y) = \sqrt{I}\ |P(x-\Delta x, y-\Delta y)| \exp\left[j(D(x-\Delta x, y-\Delta y) + \arg\big(P(x-\Delta x, y-\Delta y)\big) + \frac{(x-\Delta x)^2+(y-\Delta y)^2}{f} + k\cos(\Delta\phi)\sin(\Delta\theta)\,x + k\sin(\Delta\phi)\sin(\Delta\theta)\,x\right]$.

In the second case, shown in Figure 4.f, the phase of the programming pattern is pointwise multiplied with the data, in this case, the diffracted electric field is $(x,y) = \sqrt{I}\ \exp\left[j(D(x-\Delta x, y-\Delta y)\arg\big(P(x-\Delta x, y-\Delta y)\big) + \frac{(x-\Delta x)^2+(y-\Delta y)^2}{f} + k\cos(\Delta\phi)\sin(\Delta\theta)\,x + k\sin(\Delta\phi)\sin(\Delta\theta)\,x\right]$.

For the convolutional programming (the third case) in Figure 4.k, the $c\ x\ c$ convolution kernel, $C(x,y)$ is formed with c^2 PPs. Then, the total field is

$E(x,y) = \sqrt{I}\ \exp\left[j(\sum_{x_m}^{c}\sum_{y_m}^{c}(D(x-\Delta x-x_m, y-\Delta y-y_m)C(x_m,y_m)) + \frac{(x-\Delta x)^2+(y-\Delta y)^2}{f} + k\cos(\Delta\phi)\sin(\Delta\theta)\,x + k\sin(\Delta\phi)\sin(\Delta\theta)\,x\right]$.

Supplementary References

1. DLP9500 data sheet, product information and support | TI.com. https://www.ti.com/product/DLP9500.
2. PHAROS Lasers. *LIGHT CONVERSION* https://lightcon.com/product/pharos-femtosecond-lasers/.
3. Yao, C. *et al.* Evaluating and analyzing the energy efficiency of CNN inference on high-performance GPU. *Concurr. Comput. Pract. Exp.* **33**, e6064 (2021).
4. Zhou, T. *et al.* Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
5. Spall, J. *et al.* Hybrid training of optical neural networks. *Optica* **9**, 803–811 (2022).
6. Whitehead, J. E. M. *et al.* 2D beam shaping via 1D spatial light modulator using static phase masks. *Opt. Lett.* **46**, 2280–2283 (2021).
7. Asghari, M. & Krishnamoorthy, A. V. Energy-efficient communication. *Nat. Photonics* **5**, 268–270 (2011).
8. Hirano, M., Nakanishi, T., Okuno, T. & Onishi, M. Silica-Based Highly Nonlinear Fibers and Their Application. *IEEE J. Sel. Top. Quantum Electron.* **15**, 103–113 (2009).

9. Mafi, A. Pulse Propagation in a Short Nonlinear Graded-Index Multimode Optical Fiber. *J. Light. Technol. Vol 30 Issue 17 Pp 2803-2811* **30**, 2803–2811 (2012).

10. LeCun, Y. *et al.* Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1**, 541–551 (1989).

11. [1905.11946] EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. https://arxiv.org/abs/1905.11946.

12. Mafi, A. Bandwidth Improvement in Multimode Optical Fibers Via Scattering From Core Inclusions. *J. Light. Technol.* **28**, 1547–1555 (2010).