Science Advances
AAAS

Supplementary Materials for

Photonic unsupervised learning variational autoencoder for high-throughput and low-latency image transmission

Yitong Chen *et al.*

Corresponding author: Qionghai Dai, daiqh@tsinghua.edu.cn; Xing Lin, lin-x@tsinghua.edu.cn; Lu Fang, fanglu@tsinghua.edu.cn

Sci. Adv. **9**, eadf8437 (2023)
DOI: 10.1126/sciadv.adf8437

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Supplementary Note 1: The mathematical modeling of the mapping from the image transmission link to the PED

We considered the physical OFC system as an end-to-end network, encoding the input information to a hyper-space and decoding the transmitted information from the varied hyper-space. Different coding schemes are corresponding to miscellaneous hyper-spaces. The complicated signal impairment during the transmission is the transformation in the hyper-spaces. The designs of different coding schemes are essentially pursuing better mapping hyper-spaces. As a result, we formulate the PED to end-to-end learn the optimal hyper-space with the ideal encryption, compression and noise-resistance capabilities.

We train the PED to extract features Z from the input data that obey certain distribution or manifold in an optical latent space. By variation in this distribution $P(z)$, we obtain reconstructed data with features from training datasets. Because any distribution can be mapped into another arbitrary distribution in the latent domain Ψ, we are able to assume a reasonable $P(z)$ in Ψ. The encoder calculates $Q(Z|X)$, which is the feature distribution indicated by each sample. We employ Kullback-Leibler divergence (KL divergence) to constrain $Q(Z|X)$ close enough to $P(z)$. According to the fact that the value that the PED outputs are all positive when we only measure the intensity, we employ a biased Gaussian distribution as $Q(Z|X)$ to fit the optical latent space. Therefore, the KL divergence becomes:

$$\begin{aligned} &\mathrm{D}(\mathrm{Q}(z|x)||\mathrm{P}(z)) \\ &= \mathrm{D}(\mathrm{N}(M, \Sigma^2)||\mathrm{N}(M_0, I) \\ &= \frac{1}{2}[tr(\Sigma^2) + (M - M_0)^T(M - M_0) - k - \log(\det(M^2))] \end{aligned} \quad \text{(S1)}$$

where k is the dimensionality of the optical latent space.

The encoder extracts features and embeds the input data $\mathrm{X} = \{x_1, x_2, \ldots, x_n\}$ to an optical latent space Ψ with the dimensionality k. We sample in $\mathrm{N}(M_\mathrm{m}, \Sigma_\mathrm{m}^2)$ in the optical latent space to simulated the transmission noise and use the decoder to reconstruct the latent values to the image domain. As the fact that sampling Z from $\mathrm{N}(M_\mathrm{m}, \Sigma_\mathrm{m}^2)$ is equivalent to sampling ε from $\mathrm{N}(0, \mathrm{I})$ and let $\mathrm{Z} = M + \varepsilon \times \Sigma$, we have a derivable training process. After joint unsupervised training, the appropriate optical latent space, i.e. the coding scheme, is learned automatically.

Supplementary Note 2: Modeling of fiber transmission noise and impairment

Here we take the image transmission in optical fiber communication systems as an example. The causes of the noise in OFC are related to the whole process of signal transmission and detection, such as the thermal noise, reflection noise and quantum noise (*6*). In the case of the PED, some of the above sources are faint and the impairment of the signal during transmission is dominant to the SNR when the characteristics of the fibers, emitters, detectors and transmission distances are certain. We model the impairment of fiber transmission based on the following three main sources in this case: 1) fiber attenuation, 2) amplified spontaneous emission (ASE) noise, 3) dispersion and nonlinearity. Compared with communications with multimode fibers reported in cutting-edge papers (*47*), single-mode fibers are prominently robust to changes in the environment.

Fiber attenuation

Fibers attenuate the amplitude of light during transmission. The attenuation is usually modeled as follows:

$$\alpha = -\frac{1}{P}\left(\frac{dP}{dz}\right) (14), \quad \text{(S2)}$$

where α is the attenuation ratio; P is the power; and z is the transmission distance. α is normally related to the intrinsic absorption of the fiber material, the absorption of impurity ions, scattering loss and bending loss. Attenuation decreases the SNR when the noise level in the fiber link is fixed.

Amplified spontaneous emission (ASE) noise

To deal with the attenuation in long-distance transmission, amplifiers are widely used to cascade the OFC signal in fiber links. The erbium-doped optical fiber amplifier (EDFA) is currently the most commonly used optical amplifier. It introduces ASE noise and adds up as cascaded and observably affects the signal transmission quality. ASE noise usually appears as additive white Gaussian noise (AWGN) whose mean value is zero, and variance is a constant depending on the ASE noise level (*48*).

Dispersion and Nonlinearity

Dispersion is the phenomenon that light of different frequencies has different phase velocities. It stretches the pulse and therefore increases the error ratio. Dispersion depicts the pulse broadening of the pulses transmitted in the fiber. The dispersion is subtle if the transmission distance is short but accumulates as the transmission distance becomes longer. We can normally classify the dispersion in fibers as: 1) modal dispersion, caused by the different group velocities in different modes in multi-mode fibers. 2) polarization mode dispersion (PMD), caused by the two different polarization modes transmitted in single-mode fibers. 3) chromatic dispersion, caused by the different group velocities at different frequencies, including the waveguide dispersion and material dispersion (*14*). As a result, the pulse broadening Δt is composed of Δt_{modal}, $\Delta t_{chromatic}$ and $\Delta t_{polarization-mode}$ (*14*).

Because the PED utilizes single-mode fibers, we do not consider modal dispersion here. Additionally, the value of the polarization mode dispersion is usually smaller than chromatic dispersion and often ignorant when the modulation speed is under 2.5G/s (*49*). The PED uses a modulation speed of up to 1.6G/s and controls the polarization mode transmitted in the fiber via fiber polarization controllers, so we mainly consider the chromatic dispersion here. The chromatic dispersion is the slope of the group delay (*14*). The group velocity V_g is the velocity of the pulse along the fiber and can be calculated as

$$V_g = \frac{d\omega}{d\beta},$$

where ω is the frequency and β is the propagation constant of the fiber. Therefore, the group delay Δt is

$$\Delta t = \frac{1}{V_g} = \frac{d\beta}{d\omega}.$$

The propagation constant β can be calculated with:

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_o) + \frac{1}{2}\beta_2(\omega - \omega_o)^2,$$

where $\beta_n = \frac{d^n\beta(\omega)}{d\omega^n}\Big|_{\omega=\omega_o}$. ω_o is the center frequency of the taylor expansion. In practice, people usually use the dispersion coefficient D(ps/nm/km) instead of β_2(ps^2/km). They two satisfy:

$$D = \frac{d\beta_1}{d\lambda} = -\frac{2\pi c}{\lambda^2}\beta_2.$$

For standard single-mode fiber at a central wavelength of 1550nm ($\frac{\omega}{2\pi} = 193\text{THz}$), β_2 =-22 ps^2/km (D=17 ps/nm/km) (*50*). The modulation array achieves a modulation rate up to 1.6GHz when the bit rate of the PED is 24.6Tbit/s. Fig. S9 shows the received signal with dispersion after transmitting 2000km at 1550nm, modulated both by a square wave. The dispersion is quite subtle because the modulation frequency is relatively low compared with the frequency of the carrier wave. We also add the interference of the dispersion into the simulation of the PED and the precision remains satisfying (Table S1, the first and second rows). Additionally, we note that dispersion is a linear distortion so it can be compensated with linear PED even if the distortion is more severe (*51*).

The nonlinear response of the material becomes observable when the input optical intensity is high. As a result, nonlinearity is not negligible in long-distance transmission. It will distort the signal when the signal power is strong enough. Nonlinearity normally includes Brillouin scattering, self-phase modulation, cross-phase modulation, four-wave mixing and Raman scattering (*14*). According to published analysis and experiments (*52*, *53*), the nonlinearity can be modeled as Gaussian noise (GN) with high accuracy in uncompensated OFC. The perturbative model of nonlinear propagation in uncompensated OFC we leverage here can be depicted as:

$$G_{NLI}(f) = \sum_{n=1}^{N_s} G_{NLI}^{m}(f),$$

where $G_{NLI}^{m}(f)$, indicating the power spectral density of the *m*th span propagated through the link, which can further be written as (*52*):

$$G_{NLI}^{m}(f)=\frac{16}{27}\gamma_n^2 L_{eff,n}^2 \prod_{k=1}^{n-1} e^{6\int_0^{L_{s,k}} g_k(\zeta)d\zeta} e^{-6\alpha_k L_{s,k}} \Gamma_k^3 \prod_{k=n}^{N_s} e^{2\int_0^{L_{s,k}} g_k(\zeta)d\zeta} e^{-2\alpha_k L_{s,k}} \Gamma_k$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(f_1)\, G(f_2) G(f_1+f_2-f)\rho_{n_s}(f_1,f_2,f)df_1 df_2,$$

where *n* and *k* are the integer subscripts indicating the span; γ is the fiber nonlinearity coefficient; α is the fiber field loss coefficient; L_{eff} is the span effective length; L_s is the span length; N_s is the total number of the spans in the link; $g(\zeta)$ is the fiber field gain coefficient; Γ is the lumped power gain, for example via an EDFA; ζ is the longitudinal spatial coordinate; G is the power spectral density of the wave-division multiplexing of the transmitted signal; ρ is the non-degenerate four-wave mixing efficiency of the beating after normalization. It is based on the assumption that the nonlinearity interference is much smaller than the signal. Therefore, we can find the approximate analytical solution to the nonlinear Schrodinger equation or the Manakov equation with perturbative models. Different models converge to the same result and formula under the same assumptions, which additionally validates the GN modelling (*52*).
The prediction of the GN model matches well with the nonlinear simulation over both QPSK and QAM-16 as shown in ref. 52. The normalized NLI noise variance η over various numbers of spans in both simulations with and without pre-dispersion (the red line and blue line) are well predicted by both the coherent and incoherent GN-models. As a result, we use the GN model as an accurate enough prediction of the nonlinearity in the PED.

Therefore, in this case we employ an ASE noise dominated AWGN channel model to establish the transmission noise. Even though many corrections have been published (*54*, *55*), it remains one of the most commonly used models in optical transmission. The mean value is zero because of the statistical distributions of the noise mentioned above. The variance depends on the intrinsic characteristics and distance of the transmission link. The noise figure (NF) of the EDFA is defined as follows:

$$NF = 10\log(\frac{SNR_I}{SNR_O}), \quad \text{(S3)}$$

where SNR_I and SNR_O are the signal-to-noise ratios of the input and output signals of the EDFA. According to the specifications of commercial equipment, we set the attenuation of the fiber as 0.18 dB/km@1550 nm; the distance between each EDFA as 100 km; the NF of the EDFA as 3.5 dB@-18 dB; and the SNR of the laser signal as 65 dB. We can therefore derive some corresponding transmission distances based on the channel SNR. For example, 24 dB, 17dB, 14 dB and 9 dB corresponds to about 1171km, 1371 km, 1457 km and 1600 km respectively. Adjusting the intervals between EDFAs makes it possible to apply the PED to a larger range of transmission distances.

Because SNR is defined based on the average power and does not indicate the phase noise, we add additive Gaussian noise to the phase domain as phase noise, which increases as the transmission distance rises. For example, the standard deviations we use are $\frac{\pi}{20}, \frac{\pi}{15}$ and $\frac{\pi}{12}$, corresponding to SNRs of 13.5 dB, 11 dB and 9 dB respectively, according to typical phase noise levels in OFC links. The phase noise does not interfere with the w/o PED instances because the OOK and PAM detect only the intensity, while the PED detects both the intensity and phase.

Supplementary Note 3: Calculation of transmission throughput and computation latency

On the adrenal dataset, the PED transmits a 2^8-level grayscale image of 784 pixels with 3×3 channels. Compared with 8-level pulse amplitude modulation (PAM-8), which transmits 2^3 bits each pulse and has comparable fidelity (~0.85) at the same noise level (24 dB), the PED increases the throughput by $\frac{28\times28\times8/3}{3\times3} = 232.3$ times at the same pulse frequency.

We also compare the PED with coherent modulation QAM-512, which is one of the most high-throughput coherent modulation methods in long-distance OFC. When applying the same level of noise, QAM-512 achieves comparable fidelity (~0.85) transmission fidelity but using $\frac{28\times28\times8/9}{3\times3} = 77.4$ times more throughput than the PED.

Compared with the mainstream compression method DCT, which achieves comparable reconstruction fidelity (~0.85) with 7×7 float16 values, the PED improves the throughput by $\frac{7\times7\times16}{3\times3} = 87.1$ times. The PED requires no quantification because the PED decoder directly processes the optical signals output by the PED encoder after transmission while DCT demands quantification to deal with the noise in transmission and introduced by digital signal processing systems.

We test the procedures of compression and reconstruction by DCT with a state-of-the-art CPU (Intel, CA, Intel(R) Core(TM) i7 6500U CPU @2.5GHz), the average computation latency for each image over 10k test dataset is 3.0703×10^{-4} s. Because the coding procedure of transferring the digits into bit codes and transferring back is quite quick by mainstream OFC digital processors, we take the latency of compression and reconstruction as a lower bound estimation of the whole processing latency. The computation latency of the all-optical PED is dependent on the diffractive distance (45 cm both in the encoder and decoder) and the velocity of

light, which is $\frac{0.45\times2}{3\times10^8} = 3 \times 10^{-9}$ s. Therefore, the PED improves the computation latency by five orders of magnitude.

Additionally, we calculate the end-to-end latency of the complete system of PED, and show that the PED optimizes the time-consuming bottleneck that restricts the improvement of throughput, instead of some parts that are fast enough already. The signal flow in the end-to-end system in both traditional OFC system and the PED are shown in Fig. S11.

In traditional OFC system, digital signals are compressed and encrypted by digital pre-process and modulated to analog optical signals by a digital-analog converter (DAC) and an electrical-optical modulator (EOM). After the transmission in the fiber link, the analog optical signals are demodulated to digital electronic signals by a photodetector (i.e. optical to electronic converter, OEC) and an analog-digital converter (ADC) and decompressed and decrypted by digital post-process.

In the PED system, the original digital signals are first converted by a DAC and an EOM to analog optical signals and processed by the PED all-optically. After transmission in the fiber link, the analog signals are post-processed by the PED all-optically to decompress and decrypt. Finally, the analog optical signals are converted to digital electronic signals with an OEC and an ADC. We calculate the latency of both systems to transmit a same 28×28 8-bit grayscale image and achieve the same fidelity, respectively, in Table S2. To validate the advance of PED, we choose state-of-the-art CPU (Intel, CA, Intel(R) Core(TM) i7 6500U CPU @2.5GHz) to compare with. We note that the commercial OFC system can only achieves a computing speed much lower than it considering the cost.

When using discrete cosine transformation (DCT) as the pre-process to compress and encrypt the input. The 2D-DCT can be depicted as

$$F(u,v) = c(u)c(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f(i,j)\cos\left[\frac{(i+0.5)\pi}{N}u\right]\cos\left[\frac{(i+0.5)\pi}{N}v\right],$$

where N is the dimension of the image; u, v are the coordinates after DCT. The cosine is calculated with Taylor expansion and the quantitative accuracy after DCT is binary16 and the number of channels is 7×7. The quantization accuracy and number of channels are designed like this to achieve the same reconstruction fidelity of the PED. As a result, the operation number is ~2×10^6 and the latency of digital pre-process is ~153.5 μs as measured above. If implementing the pre-process with a state-of-the-art GPU with a computing speed of 47 TOPS/s (*43*), the latency of digital pre-process can be reduced to ~42.6 ns. The PED does not require extra digital pre-processes.

Traditional OFC uses a Mach–Zehnder modulator (MZM) as the EOM in traditional OFC. The latency of DAC is ~2ns including interface delay (Analog Devices, Inc, MA, USA, AD9789) and the modulation speed of MZM is ~14ps/bit (state-of-the-art) (*56*). The PED uses a spatial light modulator with a modulation speed of 1.6GHz/frame (state of the art) (*57*).

The latency of optical pre-process of the PED depends on the light speed and diffractive distance because the PED is all-optical. The diffractive distance is 0.45m for two layers so the latency is ~1.5ns. Traditional OFC does not require extra optical pre-process.

The optical post-process and digital post-process have the same computational complexity and latency as the pre-processes.

Traditional OFC uses an SOA-PIN detector with a latency of ~25ps/bit (state of the art) (*58*) and ADC with a latency of ~2.4ns (Analog Devices, Inc, MA, USA, AD9680). The PED uses the same OEC and ADC array in parallel, but the OEC and ADC in the traditional system are in

serial, which would be slower. We here use parallel latency for both systems for a fair comparison.
As a result, the end-to-end latency of the traditional system and the PED are ~307μs and ~7.66ns, respectively. The main reason is that the PED leverages ultra-high speed optical computation to replace digital signal process, which occupies the main time cost in traditional OFC systems. Both the end-to-end latency (~307μs/~7.66ns) and computation-only latency (~307μs/~3ns) are improved by more than four orders of magnitudes.

Additionally, for the computation latency of optoelectronic PED, because optoelectronic PED implements all computation with light and the electronic parts are only for image capture to reuse the devices, the computation latency of optoelectronic PED only depends on the diffraction distance (300 mm for each layer). Therefore, the computation latency of optoelectronic PED is about 4 ns.

For system latency, because we use on-shelf commercial devices to implement the optoelectronic PED, the system latency of optoelectronic PED is mainly restricted by the frame rate of the amplitude-modulation spatial light modulator (SLM) (HOLOEYE Photonics AG, HES6001). It has a frame rate of 60Hz. While all the other processes for one frame, i.e. one layer, including light propagation (about 1 ns), phase change in the other SLM (about 1.4 ms) and camera exposure and storage (about 6.9 ms) can all be completed within this interval (about 16.7 ms). As a result, the system latency of each layer and each frame in optoelectronic PED is 16.7 ms and 66.7 ms, respectively. If we use state-of-the-art devices, i.e. 1.6GHz SLM*(57)* and 27.1Gpixels/s high-speed camera (i-speed 7, iX Cameras, United Kingdom), the system latency can further be improved to 23.6 μs. We also note that all-optical PED can be mass-produced with nanofabrication and we just use optoelectronic PED to evaluate the performance.

To better demonstrate the improvement in transmission throughput due to latency reduction by the PED, we also supplement a numerical demonstration of video transmission. As shown in Movie S2, we use a 6×6 fiber bundle to transmit 100 24-fps video with a resolution of 28×280 8-bit pixels per frame simultaneously (one example shown in the first row in Movie S2). Discrete Cosine Transform (DCT) uses 16-QAM while PED does not requires digital modulation. The transmission latency of each frame with the PED is about 7.7 μs with the dimensionality of the optical latent space as 6×6. While the transmission latency of each frame with DCT, which is also compressed to 6×6, is about 307 ms. Computation details in Supplementary material Note 3 and Table S2. The transmitted videos with PED and DCT with their latency respectively are shown in the second and third row of Movie S2. The video transmitted with the PED is as fluent as the original one while the video transmitted with DCT is delayed heavily due to the much longer latency of electronic processing. We use an example of only 100 videos for better demonstration effect, or the display would become too laggard. And we note that the PED is able to improve the computation latency by four orders of magnitude as shown in Table S2.

Supplementary Note 4: Transmission bit rate of the PED

Considering the set-up of the PED, the bit rate R of the PED can be calculated by $R = ndf$, where n is the number of coding modules that a frame can include; d is the depth each coding module transmits; f is the modulation frequency. n depends on the resolution of the modulator and the size of each coding module. We use an amplitude-modulation spatial light modulator (SLM, HOLOEYE Photonics AG, HES6001) or a digital micro mirror device (DMD, DLP9000X, Texas Instruments) to generate the grayscale input and a phase-only spatial light modulator (Meadowlark Optics, Inc. P1920-400-800-PCIE) as the diffractive layer in the optical

neural network. Each PED coding module that transmits 9 bits requires as small as 20×20 pixels and a 5-pixel interval on each side to eliminate crosstalk. Fig. S14 shows that the PED achieves satisfying performance at such a small size. Therefore, each frame is able to contain 1705 PEDs. According to reported advances in the modulation techniques, the modulation frequency can reach up to 1.6GHz (*57*). As a result, the bit rate would be able to reach 1.705×9×1.6 = **24.6TBit/s** with a single wavelength. In Fig. 3, limited to the frame rate of the DMD (20kHz) we use, the bit rate we utilize in Fig.3 is 1.705×9×0.02 = 0.31GBit/s. The PED has the potential to achieve higher bit rates by further adding division methods such as wavelength division multiplexing (WDM).

Supplementary Note 5: The compatibility of the PED with existing OFC systems.
We demonstrate the compatibility of the PED with real-world fiber communication networks in two aspects: the hardware (e.g. fiber links) and the software (e.g. coding algorithms).
In terms of hardware, fiber bundles are common carriers in existing fiber links (*6*, *59*), which already perfectly satisfy the demand of the PED. The wave front modulators in the PED can also be integrated into chip-scale metamaterials easily in coupling systems in base stations (*60*). Additionally, we can use delay lines to switch between the parallel transmission in fiber bundles and serial transmission in independent fibers during different segments in a link. Therefore, the PED is compatible with the real-world hardware of OFC.
In terms of software, it is correct that we use the mutual information in several fibers instead of letting each fiber carry independent information. By such innovation, the PED achieves high gain in throughput and reduction of transmission errors. The PED provides an all-optical decoder to post-process the information coupled from different fibers to replace the digital signal processors (DSPs) in conventional communication systems. Additionally, coupling improves the encryption security because eavesdropping requires information in all fibers and their correct order.
As a result, coupled information in fiber bundles brings many advantages compared with independent single fibers and the PED provides a corresponding processor for the change. The main difference in dependent fiber bundle transmission compared to independent fibers is synchronization. We analyze it in three aspects:

1) Polarization synchronization;
The polarization synchronism is pre-calibrated, via the same method in coherent transmission, e.g. QAM, which is commonly used especially in long-distance transmission. For short links such as data centers, we can also use polarization-maintaining fibers to maintain polarization during transmission.

2) Phase and amplitude synchronization;
The phase and amplitude synchronism can also be pre-calibrated via the similar method of coherent transmission. However, unlike the polarization may change relatively slowly in fixed fiber bundles, the phase and amplitude variation are composed of systematic and random fluctuation. The systematic part usually varies slowly enough for pre-calibration. For the random fluctuation that is difficult to pre-calibrate, we take it into consideration during the training of the PED and the PED therefore establishes robustness to certain scale phase noise.

3) Time synchronization.

The time synchronism can be pre-calibrated with delay lines (*55*). When the bit rate of the PED achieves ~24.6Tbit/s, the modulation frame rate is ~1.6GHz (the detailed calculation of frame rate and bit rate is listed in SI Note 4), which corresponds to an optical path of ~187.5mm for each interval and satisfies the compensation scale of delay lines.

In conclusion, the PED is compatible with the existing hardware in real-world OFC while introducing advanced architecture for coding and software.

Supplementary Note 6: Compatibility of the PED with current multiplexing methods

The PED is a novel space-division multiplexing method that has great potential to be combined with other current multiplexing methods to further improve the transmission throughput. The combination may generate further functions such as the following:

Space division multiplexing

Current space division multiplexing usually increases the number of fiber bundles or cores. The PED is able to not only be combined with it but also encrypt and compress information using the correlation between cores, which leads to higher throughput and security as well as lower transmission error.

Wave division multiplexing (WDM) and Dense Wavelength Division Multiplexing (DWDM)

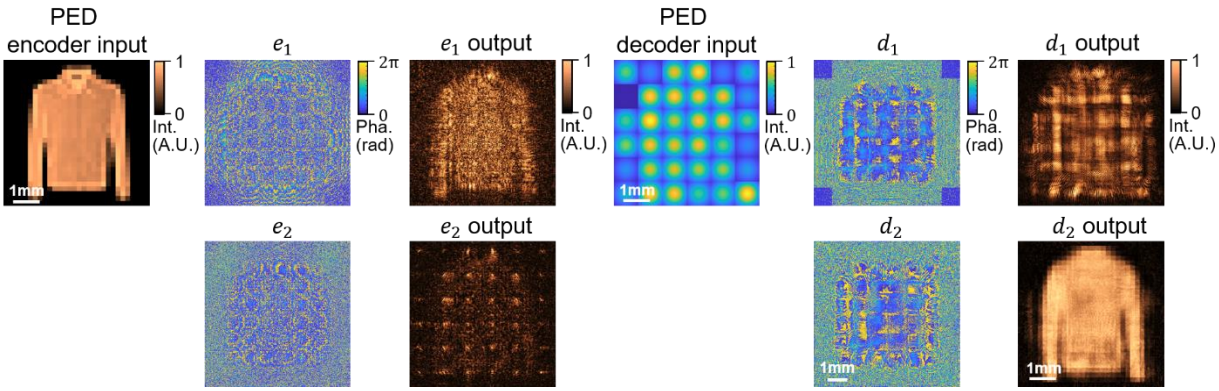
WDM/DWDM employs multiple wavelengths and is compatible with the modulation of the PED. The appropriate employment of dispersion helps establish intelligent photonic networks (*28*) . Therefore, the PED combined with WDM/DWDM may result in more complicated and powerful optical intelligence besides increasing throughput with WDM/DWDM. The effective coding with the PED also facilitates the existing point-to-point links evolve into many-to-many networks.

Time division multiplexing (TDM)

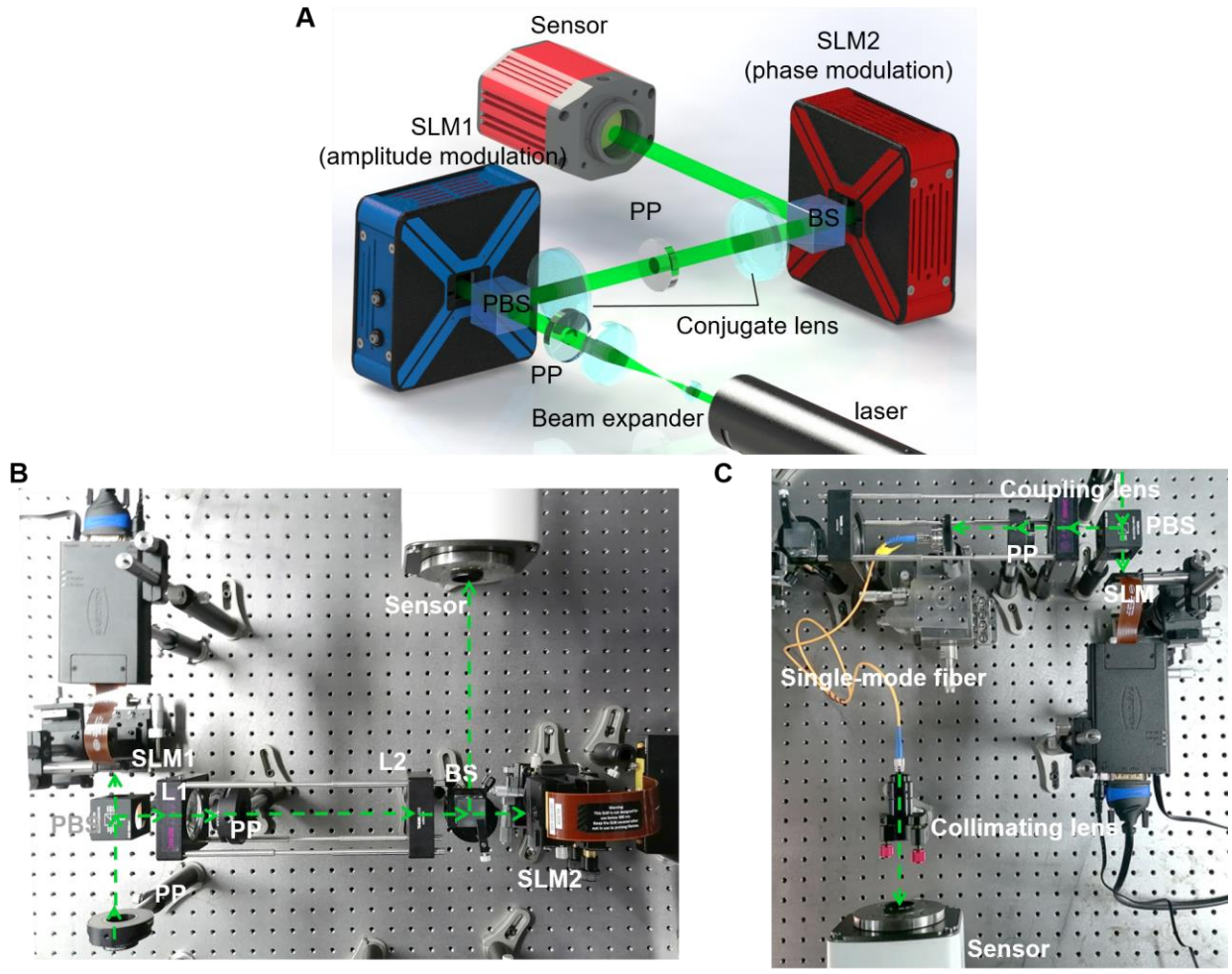
The PED combined with TDM can markedly improve the throughput and precision of transmission. In addition, reconfigurable modulators such as the SLMs that we introduce in this work enable convenient and frequent shifts for data-specific coding.

Phase coding

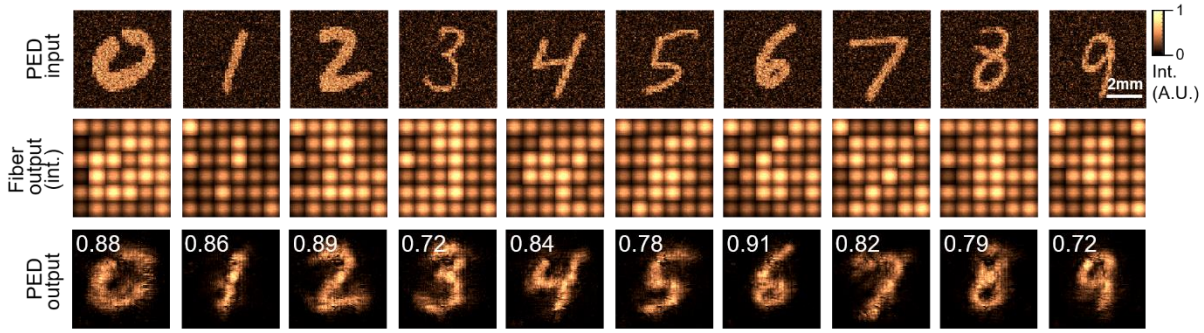
The outstanding performance of the all-optical PED demonstrates its powerful ability to encode and decode phase information at almost no cost, which indicates great potential in combining the PED with phase coding methods such as QAM. It may not only increase the transmission throughput by leveraging the phase information but also provide a way to decode and process the phase information with ultra-low latency and high convenience.



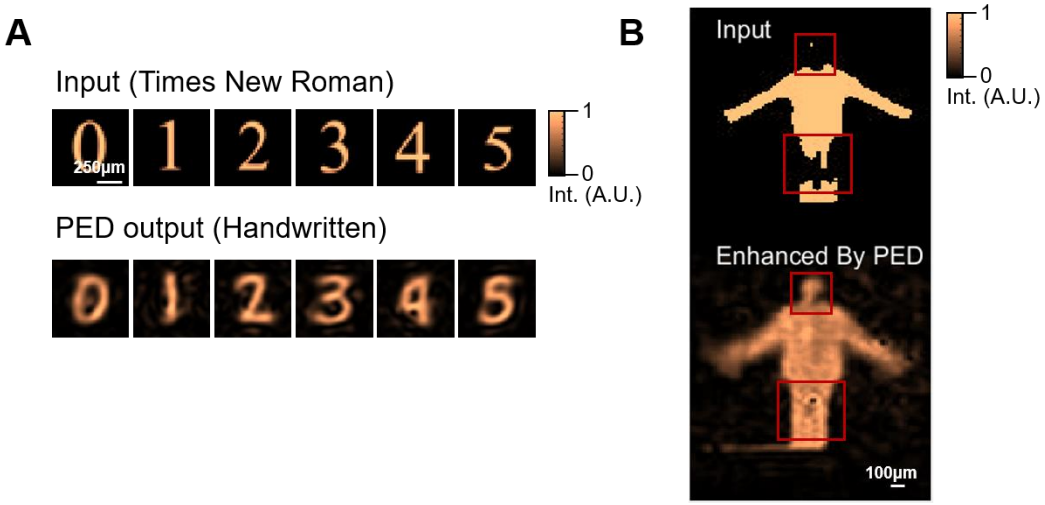
Supplementary Fig. S1. A transmission example over fashion-MNIST with the PED. The input image, i.e. the PED encoder input, is encoded all-optically by the PED encoder with two diffractive layers, i.e. $\boldsymbol{e_1}, \boldsymbol{e_2}$ into the optical latent space after being coupled into the fiber bundle. The collimated output of the fiber bundle is shown as the PED decoder input, which is decoded by a two-layer all-optical decoder to reconstruct the input information. $\boldsymbol{e_1}, \boldsymbol{e_2}$ is the first and the second masks of the encoder. $\boldsymbol{d_1}, \boldsymbol{d_2}$ is the first and the second masks of the decoder. The $\boldsymbol{e_1}$ output, $\boldsymbol{e_2}$ output, $\boldsymbol{d_1}$ output and $\boldsymbol{d_2}$ output are the intensity of the output images relatively. The input image is an example from the 'shirt' class in the test dataset of Fashion-MNIST.



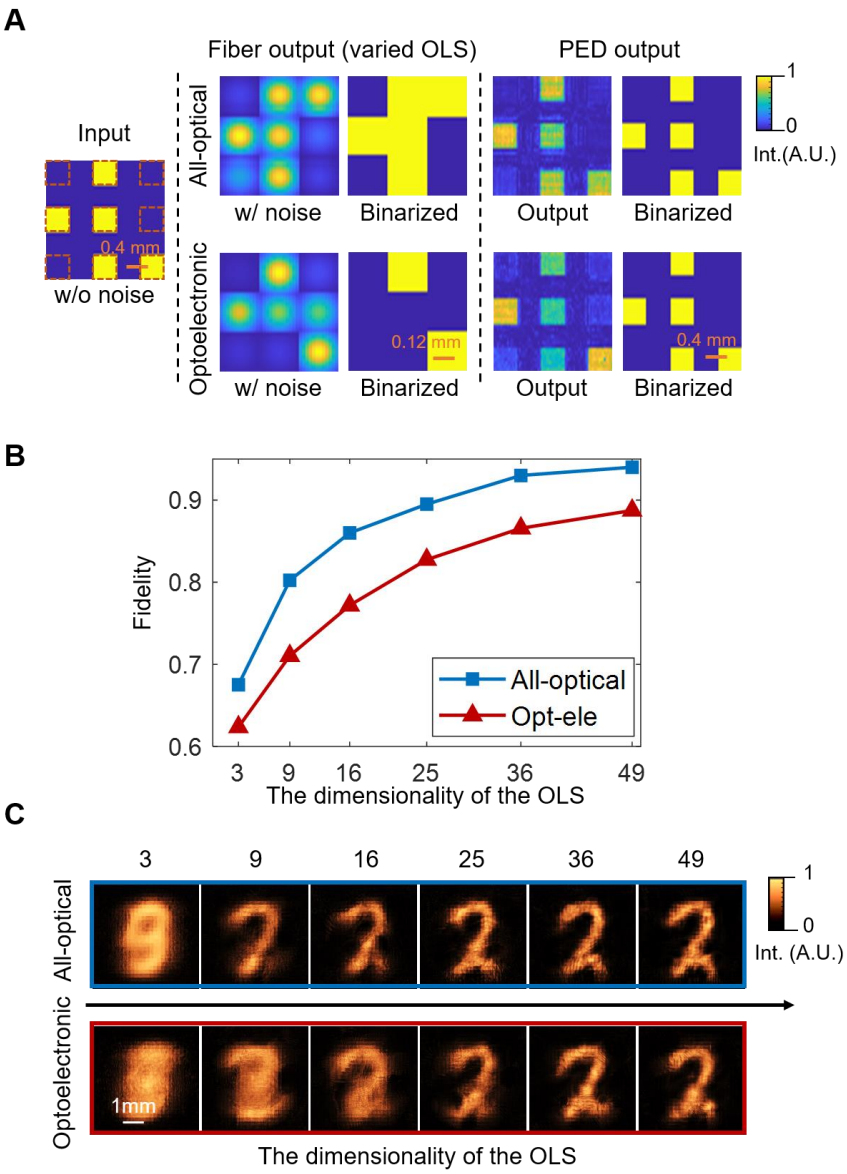
Supplementary Fig. S2. Set-up of the experimental system. a. The sketch of the optical set-up. We use a spatial light modulator as the input and another as a phase modulator. **b-c.** The photos of the experimental optical set-up and fiber-coupling set-up. PP: polarizing plate. PBS: polarizing beam splitter. BS: beam splitter. SLM: spatial light modulator. L: lens.



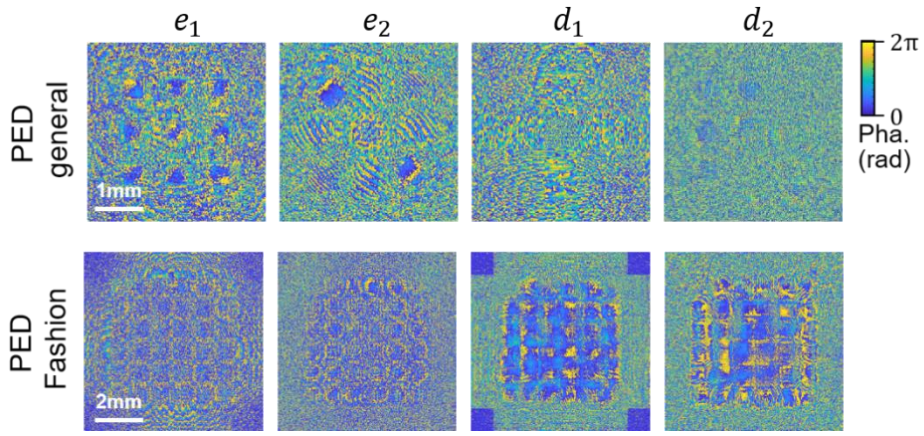
Supplementary Fig. S3. Experimental results of data-specific mode PED with contaminated input. We add Gaussian noise to the input image to mimic the noisy input in multi-level optical communications. The optical signals transmitted in OLS endure both amplitude noise (24 dB) and phase noise (additive Gaussian noise). The reconstruction of PED remains the correct digits (the third row). The image fidelity of the reconstructed results is labeled on the corner.



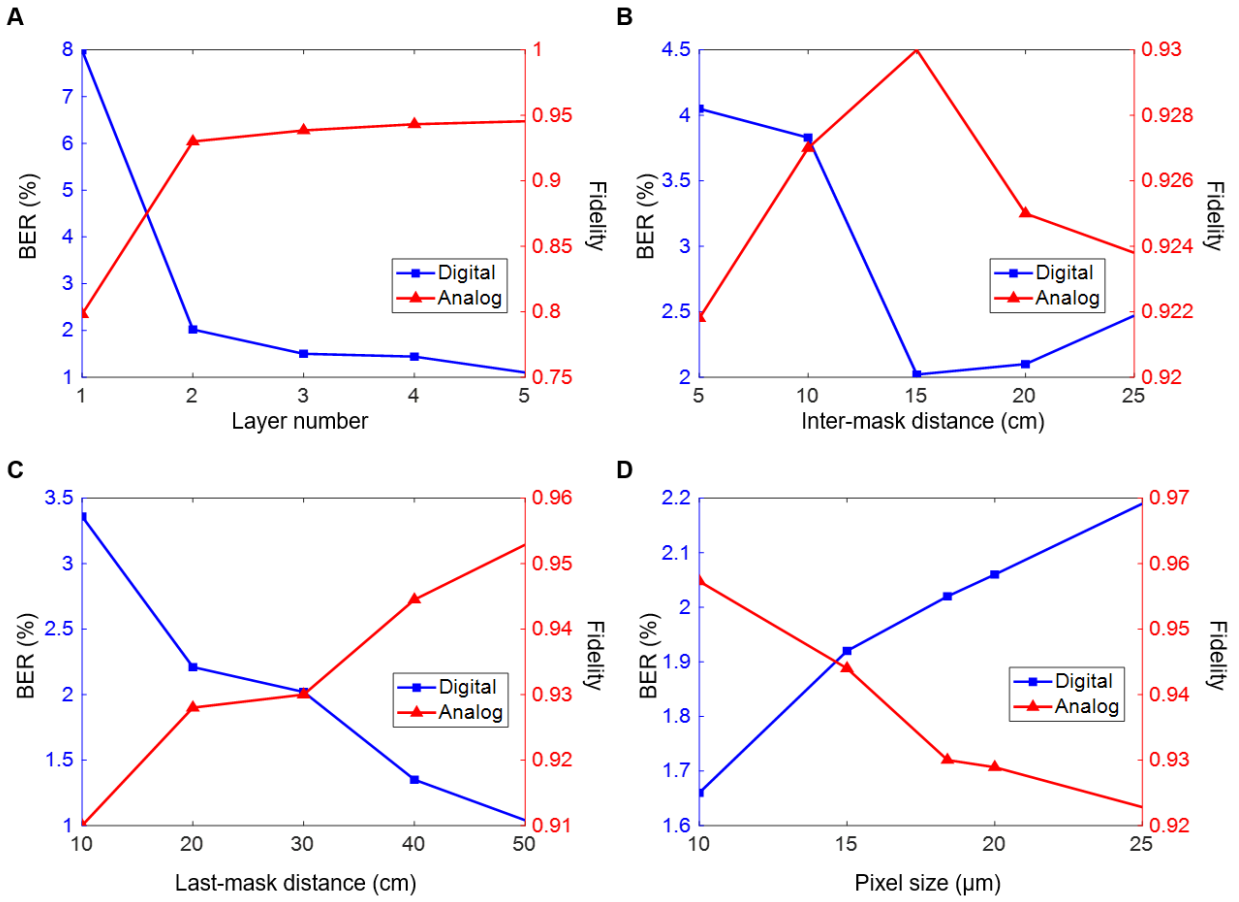
Supplementary Fig. S4. Other applications of the PED as a generative unsupervised photonic neural network. (A) An example of unsupervised style transformation with the PED on digits, from Times New Roman to hand-written styles. (B) Video enhancement with the PED (See Movie S1 for the complete video). The lost parts such as the head and trousers are intelligently repaired by the PED after unsupervised training. The broken input is binarized human action videos of handwaving action from KTH (*61*).



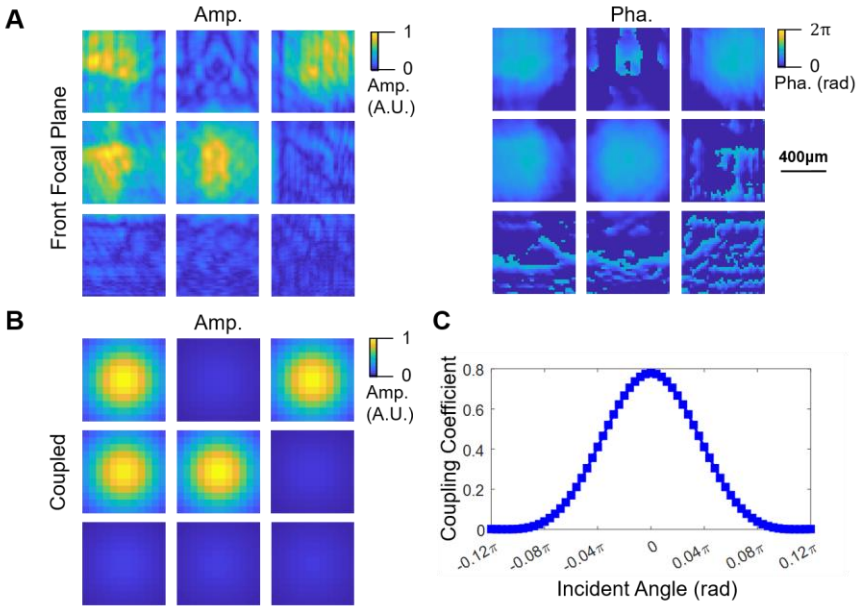
Supplementary Fig. S5. Performance comparison of the all-optical and optoelectronic PED. **a.** In the general mode, the encoder encrypts the inputs (the first column) into their OLS space representations, where the decoder is able to correct the error induced by the transmission noise (the second and third columns) and reduce the bit error ratio (the fourth and fifth columns) in both all-optical (the first row) and optoelectronic (the second row) PED. **b-c.** Both the all-optical and optoelectronic PED achieves comparably satisfactory performance on data-specific transmission, and improve with the increasing of OLS dimensionality. This example is demonstrated over the test dataset of MNIST. All-optical: all-optical PED. Opt-ele: optoelectronic PED.



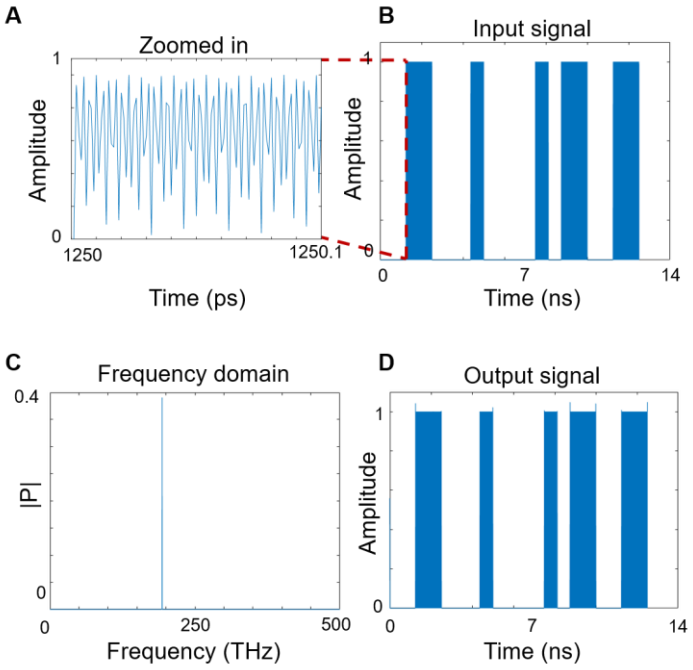
Supplementary Fig. S6. Mask patterns of the all-optical PED for general and data-specific (for Fashion-MNIST as an example) modes. The PED we utilize in both modes are composed of a two-layer encoder and a two-layer decoder. e_n $(n = 1,2)$ is the nth layer of the encoder and d_n $(n = 1,2)$ is the nth layer of the decoder. The neural number of each diffractive layer is 200×200 for the general mode and 400×400 for the data-specific mode.



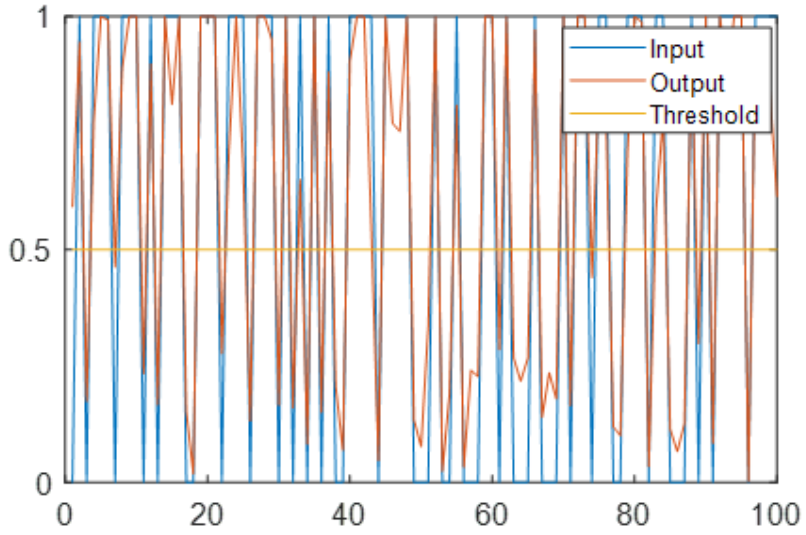
Supplementary Fig. S7. Parameter analysis of the all-optical PED. (**A**) The performance of the PED improves significantly when the layer number of the PED decoder increases from 1 to 2 and grows gradually when the layer number continues to increase for both general and data-specific modes. (**B**) We analyze the influence of the distance between the diffractive masks to the PED. 15cm achieves the best performance for both general and data-specific modes in all-optical PED. We use MNIST as an example dataset for data-specific mode here. (**C**) The performance of the PED increases as the diffractive distance after the last layer of the encoder/decoder increases in the range of 10cm to 50 cm. (**D**) The PED shows better performance with smaller pixel size in the range of 10 to 25 μm. We use 18.4 μm in the manuscript due to the restrictions of the devices.



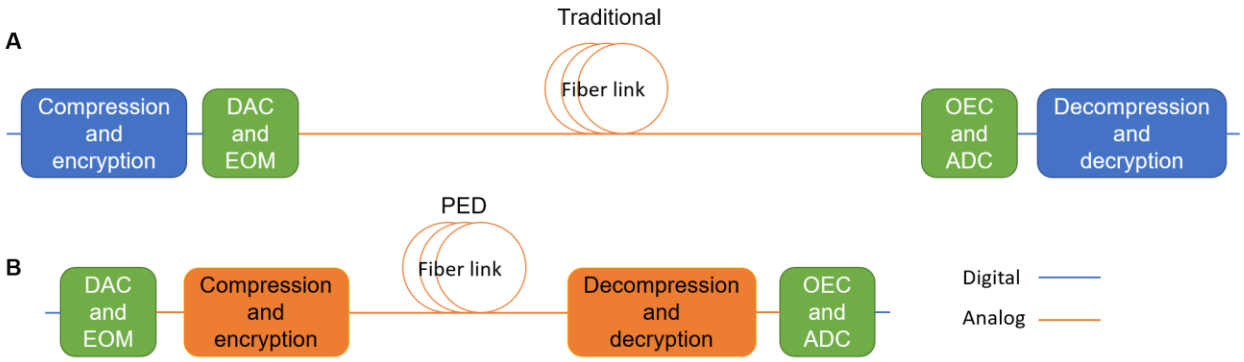
Supplementary Fig. S8. Fiber Coupling in the all-optical PED. a. An example of the light field on the front focal plane of the lens array in the general mode. The Encoder embeds the input light field to this and the lens array couples the light field into fiber bundles with corresponding complex coupling coefficients. b. The amplitude of the signal in the single-mode fiber bundle after coupling the light field into a. c. The amplitude of the complex coupling coefficient in our fiber simulated by FDTD. The incident angle refers to the angle between the incident direction and the perpendicular direction. Because it decays sharply, we only display a narrow range of incident angles.



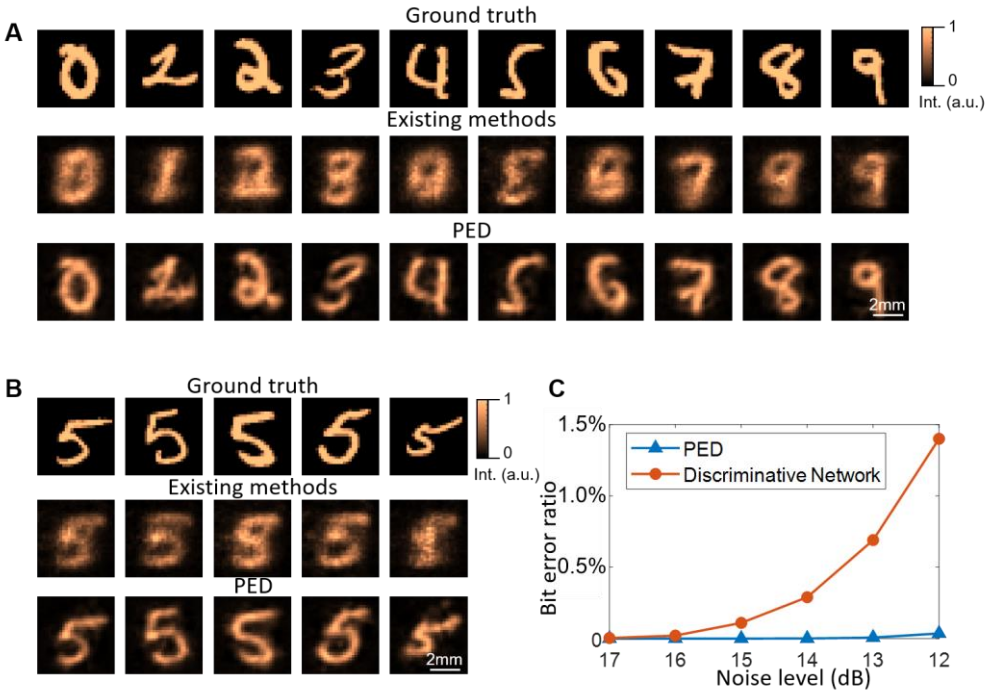
Supplementary Fig. S9. Dispersion analysis in the PED transmission. a-b. An example of input signal with square modulation wave. The modulation frequency is 1.6GHz. The frequency of the carrier wave is 193THz. The sampling frequency is 1 PHz. According to Nyquist's sampling law, the zoomed in waveform appears to be the waveform in a. c. The signal in frequency domain after Fourier transformation. The dispersion constant of the fiber is 17 ps/nm/km. d. The output signal with dispersion.



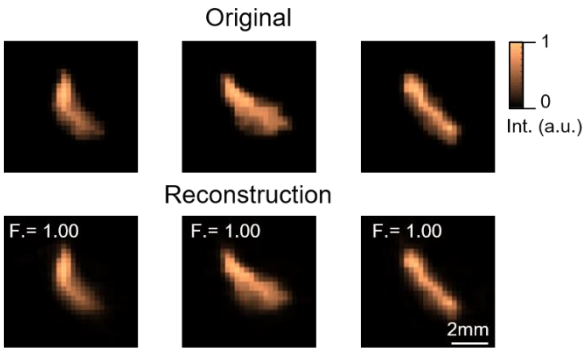
Supplementary Fig. S10. Examples of signals at 11dB. The input is random binary signal and output is the signal distorted by Gaussian noise. We limit the output in the range of [0, 1] for display. The judgement threshold is 0.5. 100 example points are shown here.



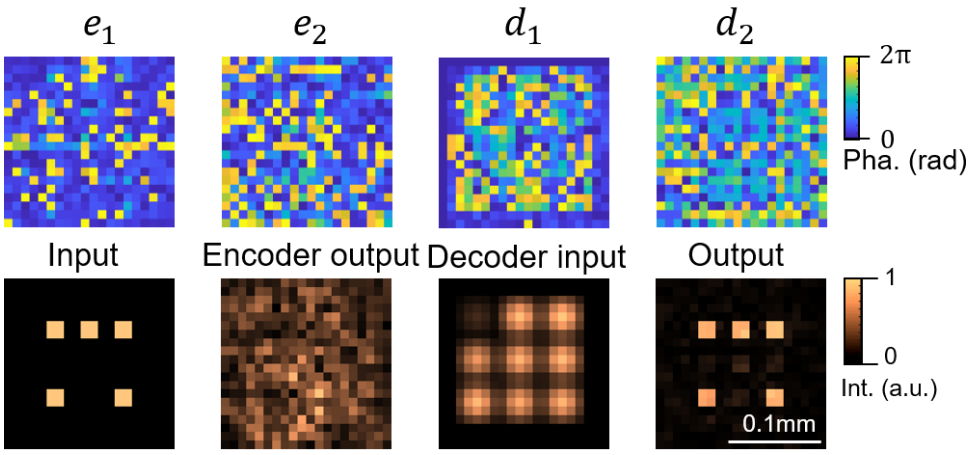
Supplementary Fig. S11. The signal flow in the end-to-end system in both traditional OFC system and the PED. a. In traditional flow, the signals are processed in digital electronic domain and transmitted in optical domain after digital-analog conversions and electronic-optical modulators. DAC: digital to analog converter. ADC: analog to digital converter. EOM: electronic to optical modulator. OEC: optical to electronic converter.



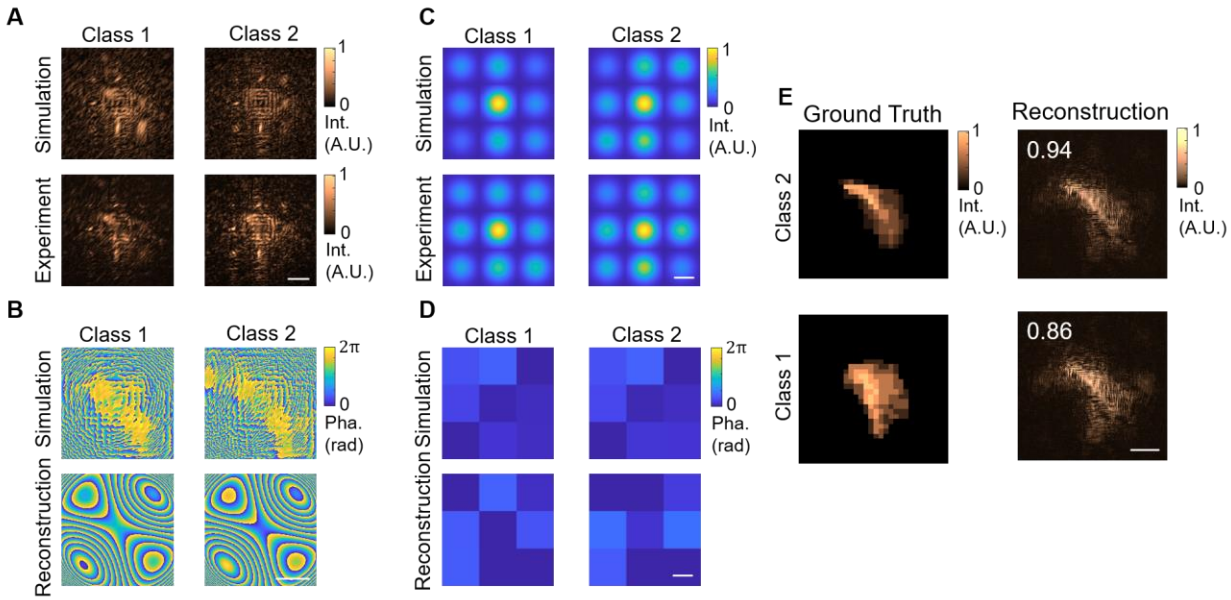
Supplementary Fig. S12. Comparison of the PED and existing methods. a. The PED reconstructs all 10 classes of hand-written digits successfully while the existing methods fails to reconstruct correct digits. b. For single class of hand-written digits, the PED maintains the style and details while existing methods is almost unrecognizable. c. Bit error ratio of the PED and existing discriminative optical network. PED: all-optical PED with a two-layer encoder and a two-layer decoder. Existing methods: discriminative optical neural network with the same number of trainable parameters.



Supplementary Fig. S13. Reconstruction with the PED when the fidelity is 1.00. We use all-optical PED with a fifteen-layer encoder and fifteen-layer decoder while the dimensionality of the optical latent space is 64. F: reconstruction fidelity. The images are from the testing dataset of MedMNIST.



Supplementary Fig. S14. The PED achieves satisfying error ratio with mask of 20×20 pixels. We use all-optical PED with a two-layer encoder and two-layer decoder. $\boldsymbol{e_i}$ is the *i*th mask of the encoder; $\boldsymbol{d_i}$ is the *i*th mask of the decoder. The PED achieves satisfying performance with modulation pixels as few as 20×20.



Supplementary Fig. S15. Experimental results of equivalent all-optical PED on medical data. a-b, The encoder output of intensity and phase. **c-d**, the decoder input of intensity and phase. **e**, the output of all-optical PED. The reconstruction fidelity is labeled on the corner. The experiment is implemented with fiber (Daheng Optics, DH-FSM450-FC) transmission and the phase is measured with coherent reference light. Both classed of images (healthy and pathological) are transmitted successfully with high fidelity. Scale bar: 1mm.

SNR(dB)	17	16.5	16	15.5	15	14.5	14
PED w/ dispersion	0	0	0	0	4.8×10^{-7}	4.3×10^{-6}	6.7×10^{-6}
PED w/o dispersion	0	0	0	0	4.8×10^{-7}	4.3×10^{-6}	6.7×10^{-6}
OOK	4.8×10^{-7}	1.4×10^{-6}	6.2×10^{-6}	2.0×10^{-5}	5.4×10^{-5}	1.3×10^{-4}	3.0×10^{-4}
FEC	0	0	0	4.8×10^{-7}	2.4×10^{-6}	1.5×10^{-5}	5.2×10^{-5}

Supplementary Table S1. Bit error ratio of on-off keying (OOK), forward error correction (FEC), the PED w/ and w/o dispersion modelling. Because the modulation frequency is relatively low compared to the frequency of the carrier wave, dispersion show little influence in error ratio. The PED w/ dispersion modelling remains to show exceeding performance compared with OOK. We use all-optical PED with a ten-layer encoder and ten-layer decoder. We use Hamming code with 1% overhead as the forward error correction code based on OOK in FEC. The test signal is a sequence of 2^{21} random bits.

	Digital pre-process	**DAC and EOM**	**Optical pre-process**	**Optical post-process**	**OEC and ADC**	**Digital post-process**	**Total**
Traditional	153.5μs	2.03ns	0	0	2.43ns	153.5μs	~307μs
PED	0	2.23ns	1.5ns	1.5ns	2.43ns	0	~7.66ns

Supplementary Table S2. The end-to-end latency of the PED and traditional transmission methods. DAC: digital to analog conversion. ADC: analog to digital conversion. EOM: electronic to optical modulation. OEC: optical to electronic converter.

SNR(dB)	**17**	**16.5**	**16**	**15.5**	**15**	**14.5**	**14**
PED	0	0	0	0	4.8×10^{-7}	4.3×10^{-6}	6.7×10^{-6}
OOK	4.8×10^{-7}	1.4×10^{-6}	6.2×10^{-6}	2.0×10^{-5}	5.4×10^{-5}	1.3×10^{-4}	3.0×10^{-4}
QAM-16	5.7×10^{-3}	7.4×10^{-3}	9.5×10^{-3}	1.3×10^{-2}	1.6×10^{-2}	1.9×10^{-2}	2.4×10^{-2}

Supplementary Table S3. Bit error ratio of the PED, OOK and QAM-16. See detailed modelling of transmission noise and impairment in SI Note2. Because QAM-16 encodes more bits than OOK with both amplitude and phase information, QAM-16 has higher bit error ratio than OOK. The test signal is a sequence of 2^{21} random bits.

Supplementary Movie S1. Video enhancement with PED. The lost parts such as head and trousers are intelligently repaired by the PED after unsupervised training. The broken input is binarized human action videos of handwaving action from KTH.

Supplementary Movie S2. Numerical video transmission with all-optical PED. We use a 6×6 fiber bundle to transmit 100 24-fps video with a resolution of 28×280 8-bit pixels per frame simultaneously (one example shown in the first row here). The video transmitted with the PED is as fluent as the original one while the video transmitted with DCT is delayed heavily due to the much longer latency of electronic processing.

REFERENCES AND NOTES

1. A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, 7068349 (**2018**).

2. S. Li, L. D. Xu, S. Zhao, The internet of things: A survey. *Inf. Syst. Front.* **17**, 243–259 (2015).

3. Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, X. Shen, A survey on metaverse: Fundamentals, security, and privacy. *arXiv:2203.02662 [cs.CR]* (5 Mar 2022).

4. P. Winzer, Scaling optical fiber networks: Challenges and solutions. *Opt. Photonics News* **26**, 28–53 (2015).

5. L. Wanhammar, DSP integrated circuits, in *Academic Press Series in Engineering*, (Academic Press, 1999), pp. 1–29.

6. K. Kikuchi, Fundamentals of coherent optical fiber communications. *J. Light. Technol.* **34**, 157–179 (2016).

7. D. J. Richardson, Beating the electronics bottleneck. *Nat. Photonics* **3**, 562–564 (2009).

8. S. Dhawan, A review of image compression and comparison of its algorithms. *Int. J. Electron. Commun. Technol.* **2**, 22–26 (2011).

9. M. Rabbani, P. W. Jones, in *Digital Image Compression Techniques* (SPIE Press, 1991), p. 240.

10. M. M. P. Petrou, C. Petrou, in *Image Processing: The Fundamentals* (John Wiley & Sons, ed. 2, 2010).

11. P. Bayvel, Kuen Charles Kao (1933–2018): Engineer who proposed optical fibre communications that underpin the Internet. *Nature* **563**, 326 (2018).

12. P. Sharma, M. Singh, A review of the development in the field of fiber optic communication systems. *Int. J. Emerg. Technol. Adv. Eng.* **3**, 113–119 (2013).

13. D. A. Morero, M. A. Castrillon, A. Aguirre, M. R. Hueda, O. E. Agazzi, Design tradeoffs and challenges in practical coherent optical transceiver implementations. *J. Light. Technol.* **34**, 121–136 (2016).

14. G. P. Agrawal, in *Fiber-Optic Communication Systems* (John Wiley & Sons, ed. 4, 2011).

15. L. R. Rabiner, B. Gold, in *Theory and Application of Digital Signal Processing* (Prentice-Hall, 1975).

16. O. Ishida, K. Takei, E. Yamazaki, Power efficient DSP implementation for 100G-and-beyond multi-haul coherent fiber-optic communications, in *Proceedings of the 2016 Optical Fiber Communications Conference and Exhibition (OFC)* (IEEE, 2016).

17. K. Roberts, I. Roberts, DSP: A disruptive technology for optical transceivers, in *Proceedings of the 2009 35th European Conference on Optical Communication (ECOC)* (IEEE, 2009).

18. E. M. Ip, J. M. Kahn, Fiber impairment compensation using coherent detection and digital signal processing. *J. Light. Technol.* **28**, 502–519 (2010).

19. A. S. Raja, S. Lange, M. Karpov, K. Shi, X. Fu, R. Behrendt, D. Cletheroe, A. Lukashchuk, I. Haller, F. Karinou, B. Thomsen, K. Jozwik, J. Liu, P. Costa, T. J. Kippenberg, H. Ballani, Ultrafast optical circuit switching for data centers using integrated soliton microcombs. *Nat. Commun.* **12**, 5867 (2021).

20. A. E. Willner, S. Khaleghi, M. R. Chitgarha, O. F. Yilmaz, All-optical signal processing. *J. Light. Technol.* **32**, 660–680 (2014).

21. L. Yan, A. E. Willner, X. Wu, A. Yi, A. Bogoni, Z.-Y. Chen, H.-Y. Jiang, All-optical signal processing for ultrahigh speed optical systems and networks. *J. Light. Technol.* **30**, 3760–3770 (2012).

22. M. P. Fok, Z. Wang, Y. Deng, P. R. Prucnal, Optical layer security in fiber-optic networks. *IEEE Trans. Inf. Forensics Secur.* **6**, 725–736 (2011).

23. A. Sludds, S. Bandyopadhyay, Z. Chen, Z. Zhong, J. Cochrane, L. Bernstein, D. Bunandar, P. B. Dixon, S. A. Hamilton, M. Streshinsky, A. Novack, T. Baehr-Jones, M. Hochberg, M. Ghobadi, R. Hamerly, D. Englund, Delocalized photonic deep learning on the internet's edge. *Science* **378**, 270–276 (2022).

24. C. Liu, Q. Ma, Z. J. Luo, Q. R. Hong, Q. Xiao, H. C. Zhang, L. Miao, W. M. Yu, Q. Cheng, L. Li, T. J. Cui, A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat. Electronics* **5**, 113–122 (2022).

25. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, A. Ozcan, All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).

26. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, M. Soljačić, Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).

27. J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, H. Bhaskaran, Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).

28. X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, D. J. Moss, 11 TeraFLOPs per second photonic convolutional accelerator for deep learning optical neural networks. *Nature* **589**, 44–51 (2021).

29. J. Liu, Q. Wu, X. Sui, Q. Chen, G. Gu, L. Wang, S. Li, Research progress in optical neural networks: Theory, applications and developments. *PhotoniX* **2**, 5 (2021).

30. M. Li, H.-S. Jeong, J. Azaña, T.-J. Ahn, 25-terahertz-bandwidth all-optical temporal differentiator. *Opt. Express* **20**, 28273–28280 (2012).

31. E. Cohen, D. Malka, A. Shemer, A. Shahmoon, Z. Zalevsky, M. London, Neural networks within multi-core optic fibers. *Sci. Rep.* **6**, 1–14 (2016).

32. F. Ashtiani, A. J. Geers, F. Aflatouni, An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).

33. D. Brunner, M. C. Soriano, C. R. Mirasso, I. Fischer, Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat. Commun.* **4**, 29080 (2013).

34. B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, P. R. Prucnal, Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**, 102–114 (2021).

35. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, W. H. P. Pernice, All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).

36. C. Huang, S. Fujisawa, T. F. de Lima, A. N. Tait, E. C. Blow, Y. Tian, S. Bilodeau, A. Jha, F. Yaman, H.-T. Peng, H. G. Batshon, B. J. Shastri, Y. Inada, T. Wang, P. R. Prucnal, Silicon photonic-electronic neural network for fibre nonlinearity compensation. *Nat. Electron.* **4**, 837–844 (2021).

37. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks. *Commun. ACM* **63**, 139–144 (2020).

38. M. Welling, Auto-encoding variational bayes. arXiv:1312.6114 [stat.ML] (20 December 2013).

39. C. Wu, X. Yang, H. Yu, R. Peng, I. Takeuchi, Y. Chen, M. Li, Harnessing optoelectronic noises in a photonic generative network. *Sci. Adv.* **8**, eabm2956 (2022).

40. H. Xiao, K. Rasul, R. Vollgraf, Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. arXiv:1708.07747 [cs.LG] (25 August 2017).

41. J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, M. Ritter, Audio set: An ontology and human-labeled dataset for audio events, in *Proceeedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2017).

42. J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, B. Ni, MedMNIST v2: A large-scale lightweight benchmark for 2D and 3D biomedical image classification, arXiv:2110.14795 [cs.CV] (27 October 2021).

43. Nvidia. Inference Platforms for HPC Data Centers | NVIDIA Deep Learning AI; www.nvidia.cn/deep-learning-ai/inference-platform/hpc/.

44. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in *Proceeedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015).

45. T. Zhou, X. Lin, J. Wu, Y. Chen, H. Xie, Y. Li, J. Fan, H. Wu, L. Fang, Q. Dai, Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).

46. C. Ruilier, A study of degraded light coupling into single-mode fibers. *Proc. SPIE* **3350**, 319–329 (1998).

47. B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, C. Moser, Multimode optical fiber transmission with a deep learning network. *Light Sci. Appl.* **7**, 69 (2018).

48. A. Splett, C. Kurtske, K. Petermann, Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities, in *Proceedings of the 1993 19th European Conference on Optical Communication (ECOC)* (IEEE, 1993).

49. J. Hecht, in *Understanding Fiber Optics* (Laser Light Press, 2015).

50. V. J. Urick, J. D. Mckinney, K. J. Williams, in *Fundamentals of Microwave Photonics* (Wiley, 2015).

51. S. Xing, G. Li, Z. Li, N. Chi, J. Zhang, Study of efficient photonic chromatic dispersion equalization using MZI-based coherent optical matrix multiplication. arXiv:2206.10317 [physics.optics] (20 May 2022).

52. P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, F. Forghieri, The GN-model of fiber non-linear propagation and its applications. *J. Light. Technol.* **32**, 694–721 (2014).

53. P. Poggiolini, The GN model of non-linear propagation in uncompensated coherent optical systems. *J. Light. Technol.* **30**, 3857–3879 (2012).

54. B. Goebel, R. J. Essiambre, G. Kramer, P. J. Winzer, N. Hanik, Calculation of mutual information for partially coherent Gaussian channels with applications to fiber optics. *IEEE Trans. Inf. Theory* **57**, 5720–5736 (2011).

55. T. Fehenberger, A. Alvarado, G. Böcherer, N. Hanik, On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel. *J. Light. Technol.* **34**, 5063–5073 (2016).

56. C. Haffner, W. Heni, Y. Fedoryshyn, J. Niegemann, A. Melikyan, D. L. Elder, B. Baeuerle, Y. Salamin, A. Josten, U. Koch, C. Hoessbacher, F. Ducry, L. Juchli, A. Emboras, D. Hillerkuss, M. Kohl, L. R. Dalton, C. Hafner, J. Leuthold, All-plasmonic Mach-Zehnder modulator enabling optical high-speed communication at the microscale. *Nat. Photonics* **9**, 525–528 (2015).

57. A. Forouzmand, M. M. Salary, G. Kafaie Shirmanesh, R. Sokhoyan, H. A. Atwater, H. Mosallaei, Tunable all-dielectric metasurface for phase modulation of the reflected and transmitted light via permittivity tuning of indium tin oxide. *Nanophotonics* **8**, 415–427 (2019).

58. C. Caillaud, P. Chanclou, F. Blache, P. Angelini, B. Duval, P. Charbonnier, D. Lanteri, G. Glastre, M. Achouche, Integrated SOA-PIN detector for high-speed short reach applications. *J. Light. Technol.* **33**, 1596–1601 (2015).

59. B. J. Puttnam, G. Rademacher, R. S. Luís, Space-division multiplexing for optical fiber communications. *Optica* **8**, 1186–1203 (2021).

60. Z. Wang, L. Chang, F. Wang, T. Li, T. Gu, Integrated photonic metasystem for image classifications at telecommunication wavelength. *Nat. Commun.* **13**, 2131 (2022).

61. C. Schüldt, I. Laptev, B. Caputo, Recognizing human actions: A local SVM approach, in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004 (ICPR)* (IEEE, 2004).