

Photonic computing enabled end-to-end image compression and reconstruction with adjustable ratio and high-fidelity

Supplementary information

S1. Details regarding the LiPICO-Net

The architecture of LiPICO-Net contains one fully connected layer (FCL), one residual connecting (RC) and three convolutional layer (CLs), where FCL and RC are implemented on FPGA while CLs are executed by the 32×32 FFT-based optical computing core. LiPICO-Net employs 4×4 convolutional kernels at the first and third CL and 1×1 convolution kernel at the second CL. For a 16×16 matrix generated by an optical computing system, it essentially executes 16×16 matrix operations, so we unfold the 4×4 convolution into matrix multiplication. The matrix's 16 rows naturally correspond to 16 distinct kernels, inherently generating 16 channels in a single transmission cycle—this fundamentally determines why we designed LiPICO's CLs with 16 channels. Unlike conventional models whose channel scale exceeding optical computing system capabilities, LiPICO-Net maintains a constant 16-channel width throughout (where channel-wise summation after the second CL ensures the third layer retains 16 channels), thus fully leveraging our system's 16-port parallel processing capacity.

Taking 25% compression ratio as an example, LiPICO-Net processes the compression data of a 32×32 image patch, which is a 16×16 image patch as input. The FCL first generates a preliminary output by expanding dimensions from 16×16 to 32×32 , followed by three CLs for feature extraction. Each CL owns 16 channels and the first and third CL utilize 4×4 kernels with 1 stride while the second CL utilizes 1×1 kernel with 1 stride. The RC adds the output of FCL and CL as the final output. During training, backward propagation optimizes the network by maximizing the evaluation function between the ground truth and model output.

An original image X is firstly split into N 32×32 image patches with stride 14, and these patches are flattened into $x_n \in R^{1024 \times 1} (n = 1, 2, \dots, N)$ corresponding N compression data $x_{in} \in R^{256 \times 1}$. The mapping from x_n to x_{in} is a linear mapping:

$$x_{in} = S_{sample} \cdot x_n (1)$$

where $S_{sample} \in R^{256 \times 1024}$ is a 25% compression ratio matrix. LiPICO-Net takes x_{in} as input, and generate the preliminary output $x_{fc} \in R^{1024 \times 1}$ after FCL:

$$x_{fc} = S_{fc} \cdot x_{in} (2)$$

where $S_{fc} \in R^{1024 \times 256}$ is the mapping matrix of FCL. Then, x_{fc} is unfold as $x'_{fc} \in R^{16}$ to operate convolution by matrix multiply and sent to CLs and get $x_{cov} \in R^{1024 \times 1}$, the relation between x'_{fc} and x_{cov} is:

$$x_{cov} = S_{cov} \cdot x'_{fc} (3)$$

where $S_{cov} \in R^{16 \times 16}$ is the unfolded matrix of CL. By adding a residual connecting between FCL and CL, LiPICO-Net get the final output x_{out}

$$x_{out} = x_{fc} + x_{cov}(4)$$

By combining formulas (1), (2), (3), and (4), we obtain:

$$x_{out} = (1 + S'_{cov}) \cdot S_{fc} \cdot S_{sample} \cdot x_n(5)$$

We use Peak signal-to-noise ratio (PSNR) as evaluation function, which is a common metric to characterize image reconstruction quality and is defined as:

$$PSNR = 20 \cdot \log_{10} \frac{MAX_f}{\sqrt{MSE}}(6)$$

Where MAX_f is the maximal pixel value in the original image and MSE represents the mean square error between the original image and reconstructed image, which can be expressed as:

$$MSE = \frac{1}{N} \sum_{n=1}^N \|x_{out} - x_n\|_2^2(7)$$

By combining formulas (5), (6), and (7), we obtain:

$$PSNR = 20 \cdot \log_{10} \frac{MAX_f}{\frac{1}{N} \sum_{n=1}^N \|x_n - (1 + S'_{cov}) \cdot S_{fc} \cdot S_{sample} \cdot x_n\|_2^2}(8)$$

By backward propagation of PSNR, we achieve the optimization of S_{sample} , S_{fc} and S'_{cov} , which completes the parameters training of FL and CLs. In case of using different compression matrix S_{sample} , LiPICO-Net haven't to retrain total parameters and merely need to train S_{fc} in electronic.

S2. Details of the opto-electronic computing processor

The 32×32 photonic computing chip (PCC) was fabricated on the silicon on insulator (SOI) substrate with the 130nm optical lithography process of silicon photonics at Chongqing United Microelectronics Center (CUMEC). The PCC is first packaged with 32 fiber-array and 192 electrical wire, then co-packaged with the other electronics. Vertical grating couplers are used for fiber/chip coupling and to support fiber-array package. The PCC consists of 80 Mach-Zehnder-Interferometers(MZIs) placed in FFT architecture instead of Clements architecture due to reduced number of PSs and high fabrication tolerance[1]. In comparison, a Clements architecture would require almost 500 MZIs for a 32×32 PCC[2]. 192 PSs based on thermo-optic effect are implemented on-chip to program the matrix of the PCC, with 160PSs for the 80 MZIs in the matrix and 32 for the MZMs. Each PS has a high-power efficiency with $P_{\pi} < 18.4\text{mW}$, thanks to thermal trenches around the PSs. This chip monolithically integrates 32 high-speed modulators to impose image data on optical carrier with a bandwidth over 30GHz, high-speed photodetectors for converting optical signal into electrical domain with a bandwidth over 40GHz. The high-speed modulators on the PCC is based on travelling wave Mach-Zehnder-modulator (MZM). Each MZM consists of a 1×2 and a 2×2 multimode interference (MMI) couplers, and a phase shifter on each arm combined with traveling-wave electrodes and p-n junctions to achieve high modulation speed. The photodetectors are Germanium absorption detectors. Thanks to the optimized structure, it can offer high responsivity ($\sim 0.95\text{A/W}@-1\text{V bias}@1550\text{nm}$) as well as low dark current ($< 10\text{nA}@-1\text{V}$

bias).

The digital-to-analog converters (DACs) can operate at 6 Giga-samples per second (GSPS) with 14-bit resolution, while the analog-to-digital converters (ADCs) can operate at 2 GSPS with 12-bit resolution. The transimpedance amplifiers (TIAs) has a bandwidth of 450 MHz and a gain of 2 k Ω .

The whole computing board integrates the PCC and all necessary control electronics (TIA, ADC/DAC, FPGA etc.). It operates at a rated voltage of 12V and a current of 4A, with a total power consumption of 48W. The power consumption of the laser source is 17.8 dBm.

Component	Key Parameter	Value
Waveguide	Propagation Loss	1.80 dB/cm
Waveguide Crossing	Insertion Loss	0.029 dB
MZI	Insertion Loss	0.1dB
Grating Coupler	Insertion Loss	4dB
	3 dB Bandwidth	60nm
	1 dB Bandwidth	35nm
Photodetector	Responsitivity	0.95A/W
Modulator	Efficiency ($V\pi \cdot L\pi$)	2 V $\cdot$ cm
	doping area	1.5mm

Supplementary Table 1 Key Parameters and their values of optical components.

S3. Estimations of performance metrics

1. Performance metrics of ReconNet, DR²-Net and the model in [3].

ReconNet, DR²-Net are seminal deep-learning models widely adopted in image processing fields. While the work in [3] is the first demonstration using photonic integrated circuit for image compression and reconstruction. Therefore, we choose them for performance comparison and we obtain their performance metrics by running those models at state-of-the-art GPU: Nvidia RTX4090, which has a power of 450W and 1320TOPS.

ReconNet: The image compression part processes a 33 $\times$ 33 image block simultaneously and consumes 26.27ns, corresponding with a total energy of 450W $\times$ 26.27ns=11.82 μ J. Therefore, latency per pixel is 26.27ns/(33 $\times$ 33)=**24.12ps/pixel**, while the energy is **10.85nJ/pixel**. The Similarly, image reconstruction part takes **8.83ns/pixel** and consumes **3.98 μ J/pixel** energy. In total, the latency is **8.85ns/pixel** and the energy is **3.99 μ J/pixel**.

DR²-Net: The image compression part has identical performance with ReconNet, but the reconstruction part is very different: latency is **18.39ns/pixel** and consumes **8.28μJ/pixel** energy. In total, the latency is **18.41ns/pixel** and the energy is **8.29μJ/pixel**.

Model in [3]: The image compression part is performed in photonics with a latency of **3.91ps/pixel** and energy of **72fJ/pixel**. Image reconstruction relies upon a complex deep-learning model, running at RTX 4090 GPU takes 16.17ms for one image block with size of 512×512. Thus, the latency is **61.67ns/pixel** and energy is **27.75μJ/pixel**. In total, the latency is **61.67ns/pixel** and the energy is **27.75μJ/pixel**.

2. Performance metrics of our work

Before performing the calculations, some basic understandings need to be established:

- A 1×32 vector propagates through our 32×32 photonics matrix, and the output is the multiplication between the vector and the matrix, which means one matrix computation performs 32×32=1024 multiply–accumulate operations (MACs) operation simultaneously.
- Our 32×32 PCC integrates modulators with 38GHz bandwidth and detectors with 45GHz bandwidth. Therefore, the image data can be fed into our PCC at the clock frequency of $f_0 = 30GHz$. Given that each computation performs 1024MACs and takes $1/f_0$, each MAC takes 0.032ps.
- Our model LiPICO-Net consists of 1 full-connection-layer (FCL) and 3 convolution layers (CLs). The FCL requires 256×1024 MACs and is executed in electronics, while 3 CLs have kernel size of 4×4, 1×1 and 4×4 and are executed in our PCC. In total 3 CLs performs 786,432 MACs for one image block with size of 32×32 pixels. Thus, the MAC/pixel is 256 by FCL and 768 by CLs.

2.1 Latency

Image compression: Our PCC performs image compression by multiplying a random matrix with the input vector. The latency between two compressions equals clock interval. Since our PCC has 32 inputs, during one clock period it can process $N=32$ pixels simultaneously. Therefore, the latency per pixel is $\tau_0 = \frac{1}{Nf_0} = \mathbf{1.04ps/pixel}$.

Image reconstruction: The FCL takes 24ns when running at RTX4090 and processes image block with size of 32×32, thus, FCL has a latency of **23.44ps/pixel**. The 3 CLs requires 768MAC/pixel, while each MAC takes 0.032ps, thus the latency per pixel is **25ps/pixel**. In total, the latency of image reconstruction is **48.44ps/pixel**.

In summary, the latency of image compression and reconstruction amounts to **49.48ps/pixel**.

2.2 Energy consumption

The power consumption of our opto-electronic computing processor consists of two parts: optical power consumption and electrical power consumption. The optical power consumption for each

matrix multiplication mainly comes from the laser source, which can be expressed as: $E_o = \frac{P_{laser}}{f_0}$,

where P_{laser} is the power consumption of the laser, and f_0 is the operation frequency of the system. We use a laser source operating at 17.8 dBm (i.e., 60.26mW) as the carrier signal, while our 32×32 photonic matrix core operates at a 30 GHz clock frequency thanks to the high-speed modulator and detector. The optical power consumption for each computation can be calculated as 2 pJ. **So the optical power consumption per MAC operation is 2pJ/1024=1.95fJ.**

For the electrical power consumption, main contributions come from the thermal-optic PSs. The other electronics such as ADC/DAC/TIA contributes negligible energy consumption. We can also

get the electrical power consumption for each matrix computation as: $E_e = \frac{P_{resistor}}{f_0}$, where $P_{resistor}$

is the power consumption of the resistors in PSs. Our 32×32 photonic matrix core has 192 PSs, with an average voltage of 3 V and an average resistance of 1500 ohms, resulting in a total power consumption of 1.152W. This will be **38.4pJ per computation and 37.5 fJ per MAC operation.**

So the total energy consumption per **MAC is 1.95fJ+37.5fJ=39.45fJ**. Then the energy consumption of our processor during image compression is **39.45×32=1.26pJ/pixel**. For image reconstruction, the energy consumption comes from two parts: CLs executed in PCC and FCL executed electrically. The CLs consume 768MAC/pixel×39.45pJ/MAC=30.3pJ/pixel, while the FCL running at RTX4090 consumes 10.55nJ/pixel, in total it consumes **10.58nJ/pixel** for image reconstruction.

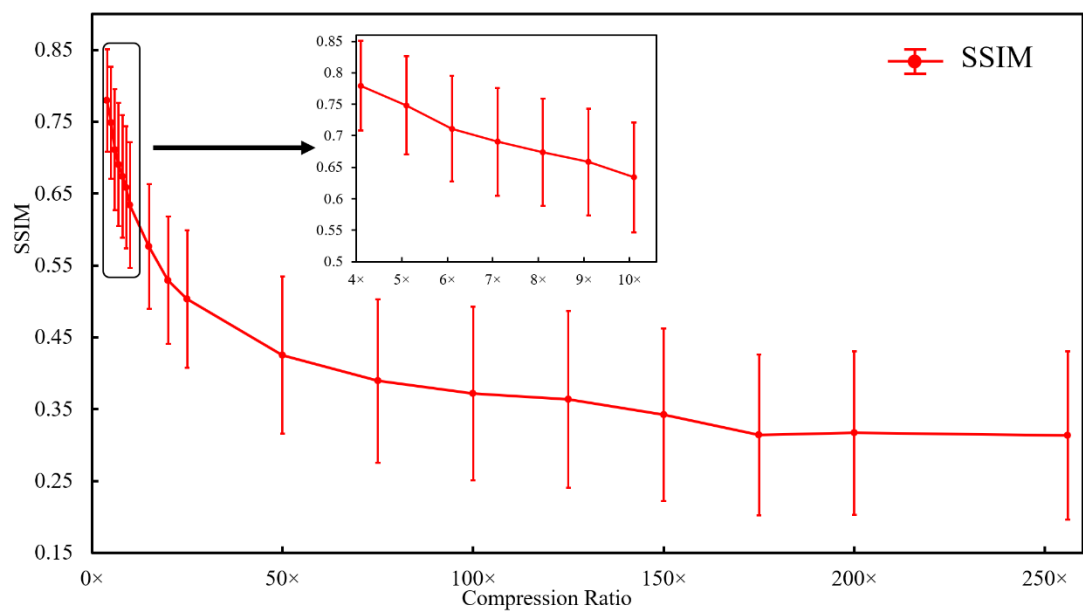
In summary, our processor consumes **1.26pJ/pixel** for image compression and **10.58nJ/pixel** for image reconstruction.

For completeness, we provide here the energy consumption for compression and reconstruction including the control electronics. As mentioned in Section S2, the computing board consumes 48W (including the 1.152W of PSs). Thus, the energy consumption for image compression is calculated as $1.26 \times (48 + 0.06026) / (1.152 + 0.06026) = 49.95$ pJ/pixel. The energy consumption for image reconstruction is $10.55 + 0.0303 \times (48 + 0.06026) / (1.152 + 0.06026) = 11.75$ pJ/pixel.

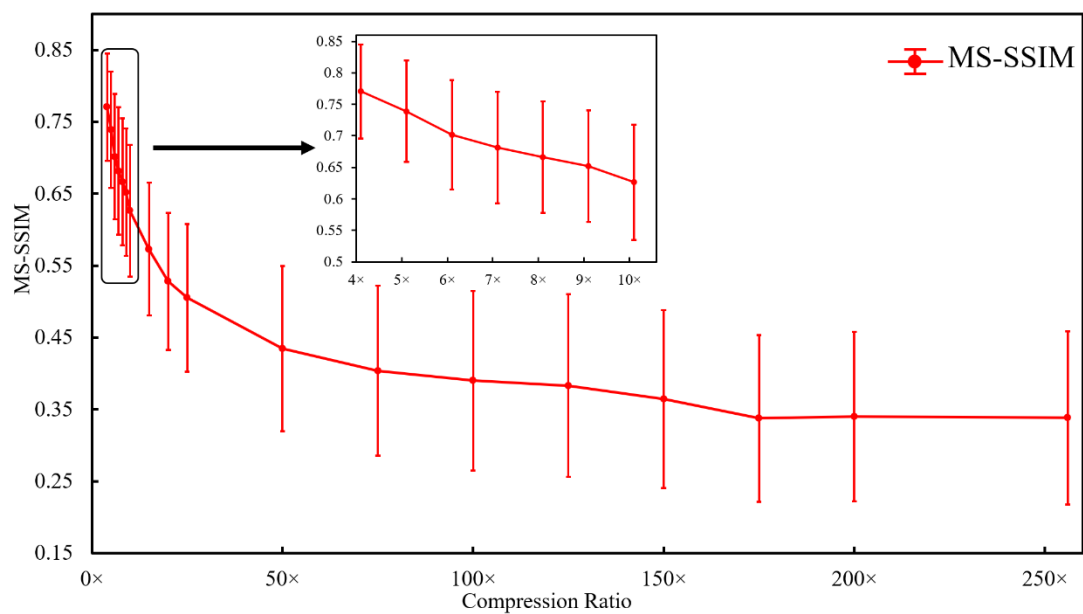
Furthermore, constraints in commercially available electronic components prevent the PCC from operating at its intrinsic 30 GHz frequency, limiting the full system to 1 GHz. Accounting for this limitation, the energy consumption for image compression becomes $49.95 \text{ pJ/pixel} \times 30 = 1.5$ nJ/pixel. The corresponding reconstruction energy is calculated as $10.55 + 1.20 \times 30 = 46.55$ pJ/pixel.

This still represents an order-of-magnitude improvement in compression energy and two orders of magnitude in reconstruction energy compared to conventional electronic approaches.

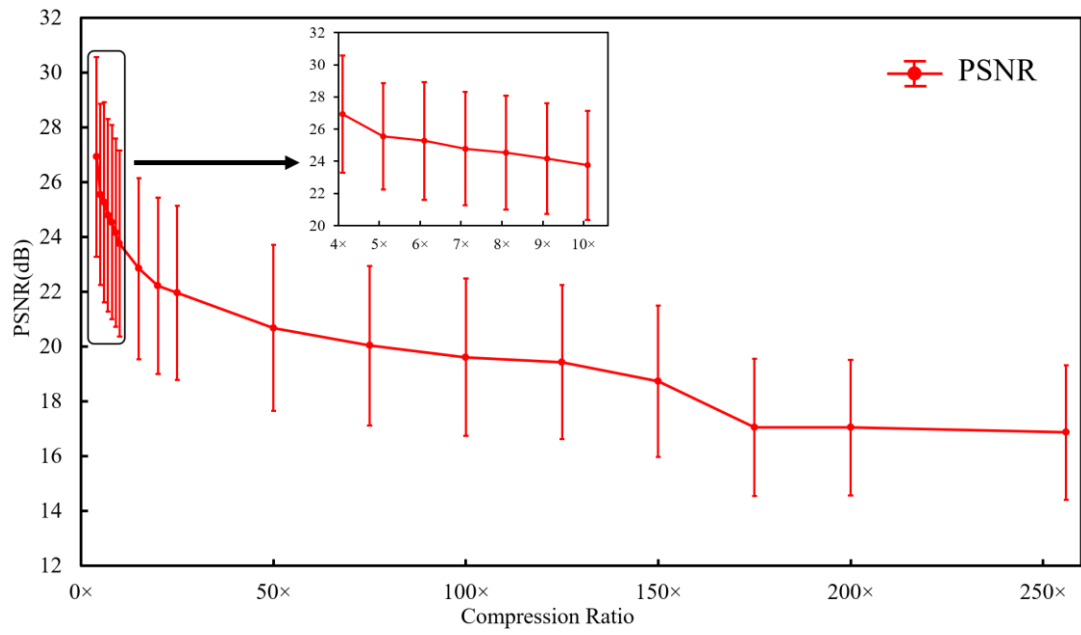
S4. Details of performance metrics



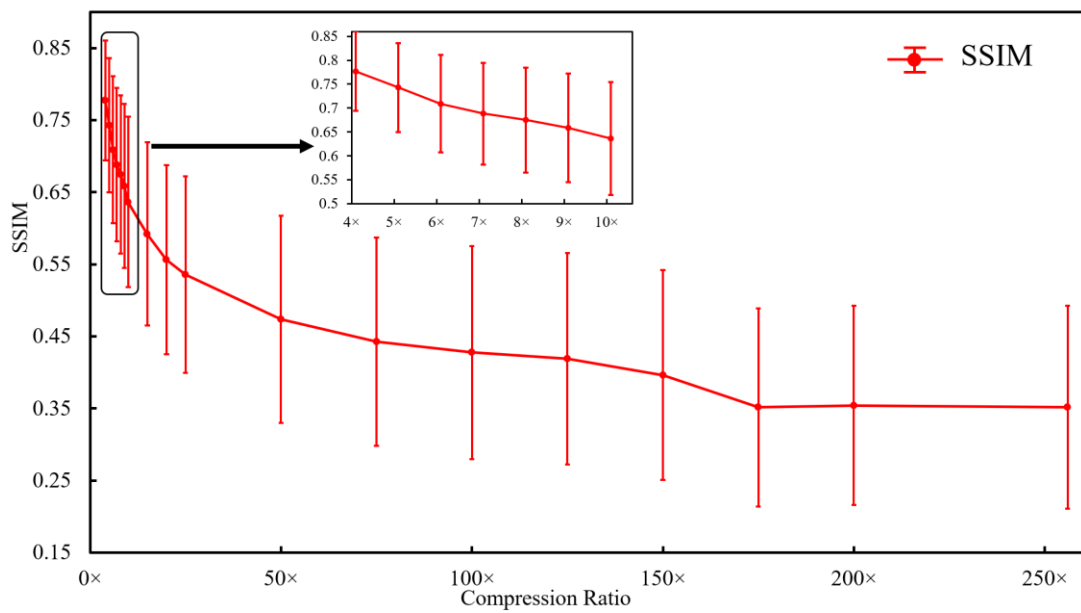
Supplementary Figure. 1 Average SSIM against Compression Ratio of 11 images based on LiPICO-Net(AE).



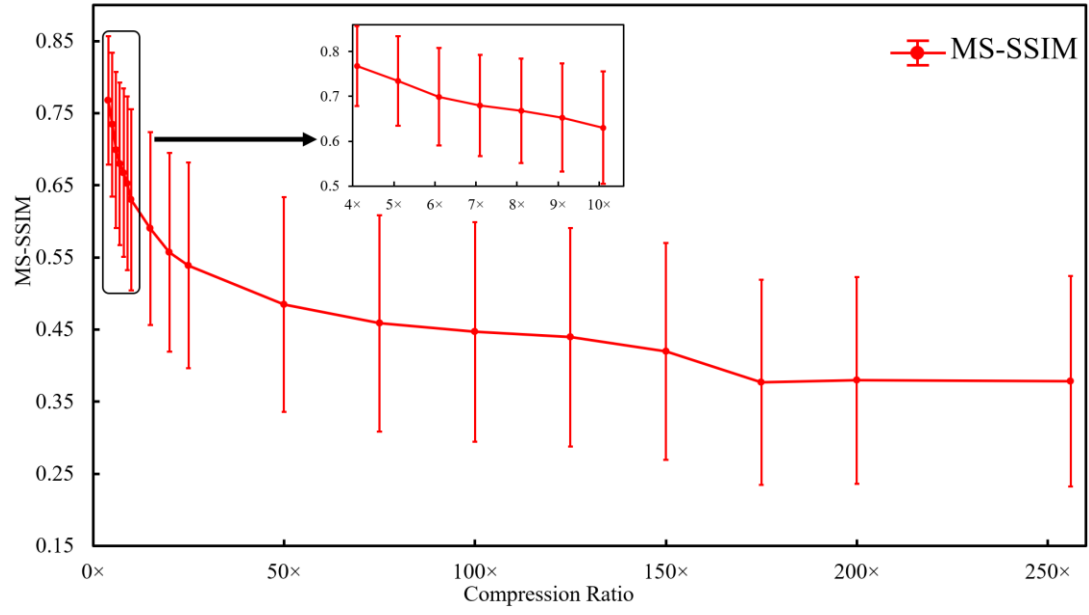
Supplementary Figure. 2 Average MS-SSIM against Compression Ratio of 11 images based on LiPICO-Net(AE).



Supplementary Figure. 3 Average PSNR versus Compression Ratio of BSDS 200 based on LiPICO-Net(AE).



Supplementary Figure. 4 Average SSIM versus Compression Ratio of BSDS 200 based on LiPICO-Net(AE).



Supplementary Figure. 5 Average MS-SSIM versus Compression Ratio of BSDS 200 based on LiPICO-Net(AE).

Compression Ratio		4×	10×	25×	100×
Data Category					
LiPICO-Net(AE)	PSNR(dB)	25.48 ± 2.52	22.31 ± 2.14	20.28 ± 2.14	17.90 ± 1.96
	SSIM	0.7796 ± 0.0714	0.6339 ± 0.0875	0.5034 ± 0.0956	0.3717 ± 0.1206
	MS-SSIM	0.7705 ± 0.0745	0.6263 ± 0.0915	0.5052 ± 0.1025	0.3899 ± 0.1247
LiPICO-Net(RM)	PSNR(dB)	25.17 ± 2.34	22.28 ± 2.09	20.25 ± 2.13	17.79 ± 1.92
	SSIM	0.7637 ± 0.0705	0.6319 ± 0.0822	0.5049 ± 0.0929	0.3690 ± 0.1180
	MS-SSIM	0.7527 ± 0.0723	0.6236 ± 0.0853	0.5066 ± 0.0998	0.3868 ± 0.1222

Supplementary Table 2 Mean and standard deviations of PSNR, SSIM and MS-SSIM of LiPICO-Net(AE) and LiPICO-Net(RM) on 11 images.

Compression Ratio Data Category		4×	8×	16×	256×
Actual Photo	PSNR(dB)	30.57	26.40	25.78	23.13
	SSIM	0.8463	0.7256	0.6699	0.5760
	MS-SSIM	0.8519	0.7480	0.7001	0.6241
Ultra-large Image	PSNR(dB)	34.00	30.35	28.71	26.56
	SSIM	0.9081	0.8191	0.7811	0.7133
	MS-SSIM	0.9124	0.8395	0.8089	0.7558

Supplementary Table 3 Mean and standard deviations of PSNR, SSIM and MS-SSIM of on actual photo and ultra-large image.

References:

1. M. Y.-S. Fang, et al., "Design of optical neural networks with component imprecisions," *Optics express* **27**, 14009-14029 (2019).
2. W. R. Clements, et al., "Optimal design for universal multiport interferometers," *Optica* **3**, 1460-1465 (2016).
3. X. Wang, et al., "Integrated photonic encoder for low power and high-speed image processing," *Nature Communications* **15**, 4510 (2024).