nature communications

Article

https://doi.org/10.1038/s41467-026-71296-0

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

Received: 20 September 2025

Accepted: 18 March 2026

Published online: 27 April 2026

Check for updates

Yuhang Wang[1,3], Ang Li[1,3] ✉, Yihang Shao[1,3], Qiang Li[2], Yang Zhao[2] & Shilong Pan[1] ✉

The rapid development of remote sensing, satellite radar, and medical equipment has created an imperative demand for ultra-efficient image compression and reconstruction. We demonstrate an end-to-end image compression and reconstruction approach using an opto-electronic computing processor, achieving orders-of-magnitude higher speed and lower energy consumption than electronic counterparts. Its core is a 32×32 silicon photonic computing chip, which monolithically integrates 32 high-speed modulators, 32 detectors, and a programmable photonic matrix core, co-packaged with all necessary control electronics. Leveraging the photonic core's programmability, the processor generates trainable compressive matrices, enabling adjustable image compression ratios (up to 256×) to meet diverse application needs. Deploying a customized lightweight photonic integrated circuit-oriented network enables high-quality reconstruction of compressed images. Our approach core parts require end-to-end latency of 49.5 ps/pixel while consuming less than 10.6nJ/pixel. This work not only provides a transformative solution for computational image processing but also opens new avenues for photonic computing application.

The exponential proliferation of massive images—spanning airborne remote sensing data, high-resolution AR/VR renderings, satellite radars, astronomical and medical imaging—has created an unprecedented challenge for transmission and storage[1–8]. With data volumes often exceeding terabytes, these images cannot be directly transmitted or archived, rendering robust compression and high-fidelity reconstruction indispensable[9–15]. Conventional solutions, exemplified by JPEG and its variants[12,13], usually include domain transformation, quantization, and entropy coding modules and are implemented on electronics, which confront insurmountable bottlenecks: prohibitive hardware costs for parallelizing operations, crippling latency in real-time processing, excessive power consumption incompatible with mobile or spaceborne platforms, and limited throughput failing to match data generation rates. These limitations starkly constrain their applicability in scenarios requiring rapid handling of massive images.

Photonics has emerged as a disruptive alternative due to photon's distinct properties compared with electrons: terahertz bandwidth, femtosecond-level latency, and near-zero heat dissipation. Photonics chip has emerged as a compelling platform for implementing a wide range of signal processing primitives, including matrix multiplication, discrete and continuous Fourier transforms, and SAR signal processing[16]. Indeed, over the past decade, photonics chips for image processing have gained momentum, but research has predominantly focused on feature extraction for simple tasks (e.g., vowel recognition, handwritten digit recognition, or animal classification), involving small-scale, low-dimensional image data[17–28]. Monolithic photonic systems for end-to-end image compression and reconstruction remain scarce.

A pioneering step was reported in 2024 by researchers at the University of Arizona, who validated silicon photonics for image

[1]National Key Lab of Microwave Photonics, Nanjing University of Aeronautics and Astronautics, Nanjing, China. [2]United Microelectronics Center, Chongqing, China. [3]These authors contributed equally: Yuhang Wang, Ang Li, Yihang Shao. ✉e-mail: ang.li@nuaa.edu.cn; pans@nuaa.edu.cn

compression using a passive scattering structure, with significantly higher speed and lower power-consumption compared with electronic counterparts[29]. Despite its significance, this passive photonics-based system suffers critical limitations: it relies on random scattering effects, leading to high optical losses and reflections; the compression ratio is fixed at 4×, which is both inflexible and insufficient for high-rate scenarios; the purely passive chip lacks photodetectors, modulators, and supporting electronic control chips, resulting in low integration; and crucially, image reconstruction remains dependent on complex AI model and external computers, limiting practical deployment [29].

Here, we present a fully integrated opto-electronic computing processor (OECP) that addresses these limitations and provide end-to-end image compression and image reconstruction with high speed and low power. At its core is a 32×32 silicon photonic computing chip (PCC) monolithically integrating 32 high-speed modulators, 32 high-speed detectors, and a programmable 32×32 optical matrix, alongside heterogeneously packaged electronic chiplets including ADCs, DACs, TIAs, and an FPGA. For image compression, the PCC's programmability generates trainable compressive matrices 32×N (where N is an integer between 2 and 16) with low transmission loss, enabling adjustable compression ratio from 2× to 256×. For image reconstruction, it deploys our custom lightweight model, Lightweight PIC-Oriented Network (LiPICO-Net), which provides comparable or even better performance compared with mainstream image reconstruction models like ReconNet[30] and DR^2-Net[31], while being optimized for photonic hardware. Validated on diverse images—including a 130 million-pixel airborne remote sensing image (>1 GB size)—this system achieves adjustable-ratio compression and high-fidelity reconstruction (best PSNR > 34 dB) with a latency of only 49.5 ps/pixel, two orders of magnitude lower than classical AI model running on Nvidia RTX 4090. Our system also achieves an energy consumption of 10.58nJ/pixel for image compression and reconstruction, 2–3 orders of magnitude lower than classical AI model running on Nvidia RTX 4090.

Results

In contrast to domain-transformation-based image compression and reconstruction such as JPEG, deep learning-based approach has gained traction in recent years due to its superior performance and ability to adjust to various scenarios[9,32,33]. This approach achieves image compression by multiplying the image data 1×N with a N×M random matrix (where N > M) and image reconstruction using a neural network. The random matrix and the neural network can be co-trained to achieve optimal performance.

Researchers at the University of Arizona demonstrated a passive disordered photonic system that achieves 4× optical image compression via 16×4 random matrices, with image reconstruction handled by external Deep ResUnet50/ResUNet models[29]. Although the passive architecture significantly reduces latency and power consumption compared to digital processors, its inherent scattering losses critically constrain energy-sensitive applications. The fixed 16×4 compression matrix further limits operational flexibility by preventing adjustable compression ratios across diverse scenarios as well as tandem-training with image reconstruction networks for optimal performance. Compounding these issues is the absence of integrated photonic-electronic components – including photodetectors, modulators, and control ICs –which impedes system miniaturization and functional autonomy. Ultimately, the necessity for off-chip AI computation imposes fundamental bottlenecks on processing speed, energy efficiency, and real-world deployability.

In this work, we propose an opto-electronic computing processor (OECP) for end-to-end image compression and reconstruction that overcomes the aforementioned limitations. As shown in Fig. 1, its core is a programmable 32 × 32 photonic computing chip (PCC) on a silicon photonics platform, with 32 high-speed modulators and photodetectors. By controlling on-chip phase shifters (PSs), the PCC's transmission matrix can be dynamically reconfigured. Image data modulates optical carriers across 32 inputs to form a 1 × 32 vector, which is then multiplied at light speed by the 32 × 32 core. This optical-domain operation incurs near-zero static power while fundamentally enabling high throughput, ultralow latency, and minimal power consumption versus electronic systems.

Image compression with an adjustable ratio is realized utilizing the excellent matrix programmability of the PCC. Let's take a 2× compression ratio achieved by a 16 × 16 PCC as an example. The entire PCC matrix can be further divided into 2 submatrices, representing the top 8 outputs and the bottom 8 outputs, respectively, as shown in Fig. 2a. Then we can perform a simple sum operation between two submatrices, and get a new 16×8 matrix, which can be considered as

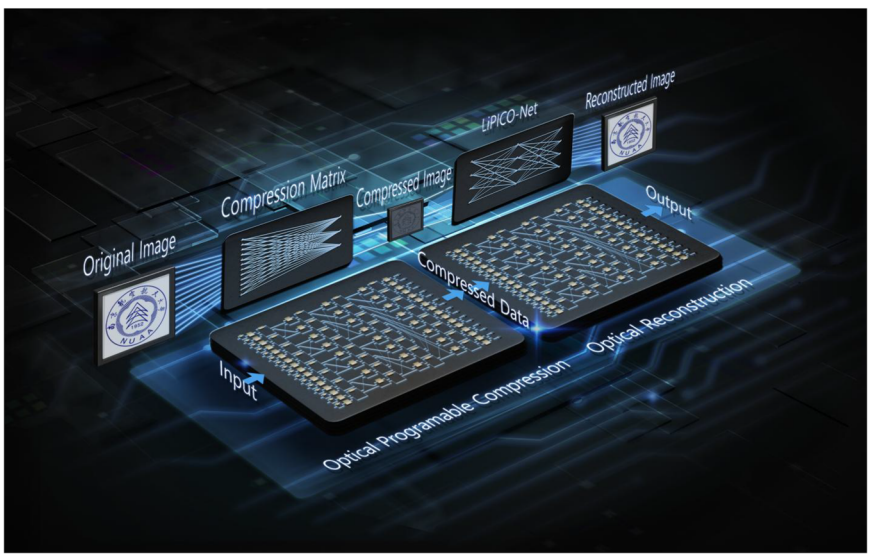

Fig. 1 | Schematic of the optoelectronic computing system for end-to-end image compression and reconstruction. Its core is a 32 × 32 photonic computing chip (PCC). Using the PCC's matrix programmability, image compression with adjustable ratios (2× to 256×) can be performed. By implementing a custom lightweight neural network on the PCC, high-fidelity image reconstruction is achieved.

https://doi.org/10.1038/s41467-026-71296-0

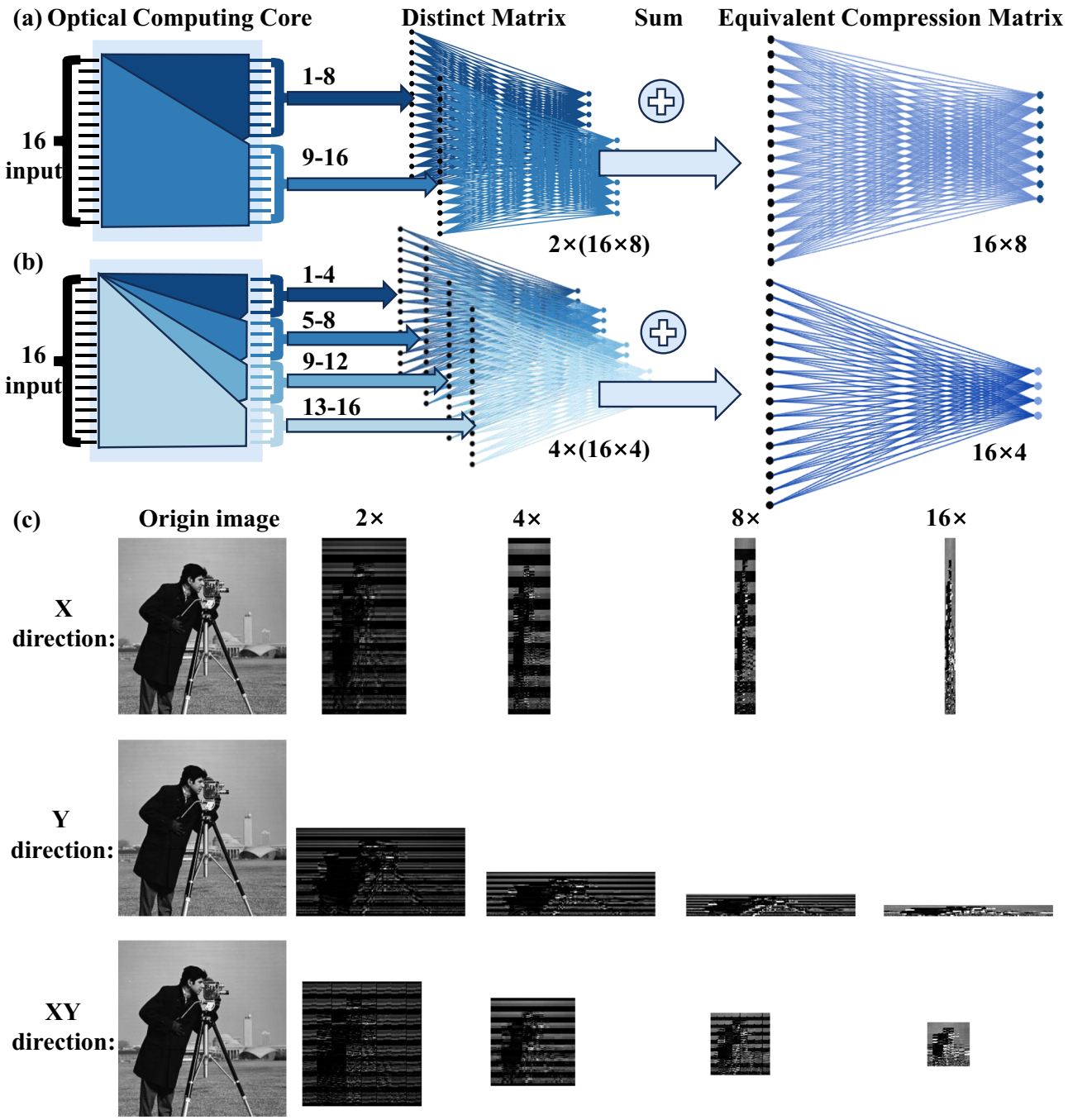

Fig. 2 | Optical matrix compression schemes. a, **b** Illustrations of 2× and 4× compression ratio using 16×16 optical matrices. **c** Illustration of compression in X, Y and XY dimension of an image.

the final compressive matrix for image compression. The image data processed by the PCC will end up being half of its original size. This methodology extends to arbitrary compression ratios through configurable submatrix partitioning and summation as illustrated in Fig. 2b. Compression applies independently to X/Y dimensions or both (Fig. 2c), enabling 2× to 1024× ratios via the 32×32 PCC. This work demonstrates up to 256× compression, e.g., 16× in X/Y dimensions. Note that, by simply truncation of output ports (e.g., retaining data from only selected ports while discarding others) can also lead to image compression, but at the price of information loss and energy loss, leading to poorer performance and low signal-to-noise-ratio.

In terms of image reconstruction, various classical neural networks with impressive performance have been reported. However, their complexity precludes direct PCC implementation. For instance, ReconNet and DR^2-Net employ up to 64-channel convolutions with 194-388 total channels, exceeding optical computing limitations. To address this problem, we propose LiPICO-Net (Lightweight Photonic Integrated Circuits-Oriented Network), an ultralight architecture optimized for PCC. The architecture of LiPICO-Net is shown in Fig. 3a. It comprises a 256×1024 fully-connected layer (FCL, FPGA-executed), one residual connection (RC), and three convolutional layers (CLs, PCC-executed) with 4 × 4, 1 × 1, and 4 × 4 kernels, respectively. More technical details of this network can be found in the Supplementary Information S1.

We use Peak signal-to-noise ratio (PSNR) as evaluation of our model, which is a common metric to characterize image

https://doi.org/10.1038/s41467-026-71296-0

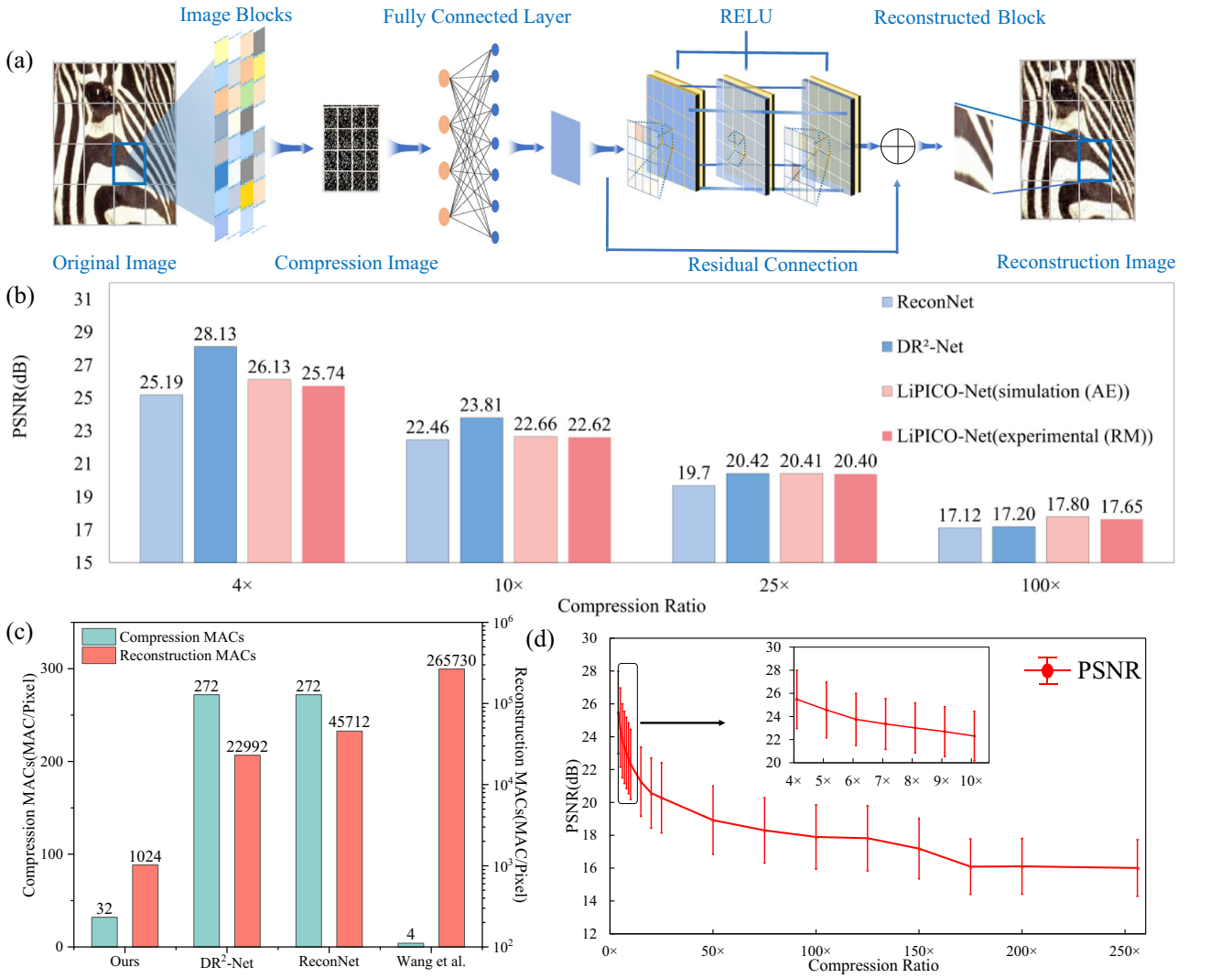

Fig. 3 | LiPICO-Net architecture and performance benchmarking. a Architecture of our custom lightweight neural network for image reconstruction: LiPICO-Net. **b** Performance comparison between our LiPICO-Net and other seminal models, including ReconNet and DR^2-Net, at four compression ratios. AE stands for "as expected" which is obtained through simulations and RM represents "real measurements", which is obtained from experiments. We use eight images existing both in refs. 30 and[31], that is: Barbara, boat, cameraman, fingerprint, flinstones, house, lena, peppers256. The results of ReconNet and DR^2-Net derive from[31]. **c** Computational resources required by our approach and other approaches at 4× compression ratio, characterized by MAC/pixel. Over 95% reductions is achieved compared with DR^2-Net, ReconNet and a pioneer work in ref. 29. **d** Average PSNR versus compression ratio of 11 images based on LiPICO-Net (AE).

reconstruction quality and is defined as:

$$\mathrm{PSNR} = 20 \cdot \log_{10} \frac{\mathrm{MAX_f}}{\sqrt{\mathrm{MSE}}} \tag{1}$$

where $\mathrm{MAX_f}$ 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:

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{n=1}^{N} \| x_{out} - x_n \|_2^2 \tag{2}$$

where x_{out} is the the final output of LiPICO-Net and x_n is flattened vector of the input image patch. The derivation of x_{out} can be found in supplementary information S1. The training of this network is performed by backward propagation of PSNR. During the training procedure, the compression matrix as well as parameters of FCL and CL are optimized simultaneously, leading to optimal performance of the entire system, while previous approach lacks of optimization of the compression matrix[29], which explains our performance improvement. Note that, when using different compression matrix, LiPICO-Net haven't to retrain total parameters and merely need to train full connection layer in electronics domain.

We evaluated LiPICO-Net's theoretical performance through PyTorch simulations using random Gaussian compression matrices. The dataset for training is the same 91 images dataset used in ref. 30 and the dataset for testing is the 11 images dataset used in ref. 30. 4 different sampling rates (e.g. compression ratio) were studied: 25%, 10%, 4%, and 1%. And we provide direct performance comparison with two seminal AI models for image reconstruction: ReconNet and DR^2-Net. The mean PSNR results are shown in Fig. 3b LiPICO-Net(AE), where AE stands for "as expected". Simulation results reveal LiPICO-Net matches ReconNet's PSNR across sampling rates and surpasses DR^2-Net at 1% sampling. Due to discrepancy between the ideal matrix and the generated matrix of PCC, the mean PSNR based on physical implementation of LiPICO-Net on PCC (but image compression is still in electronics) decrease a little compared with the LiPICO-Net(AE), as shown in Fig. 3b LiPICO-Net(RM), where RM is short for "real

https://doi.org/10.1038/s41467-026-71296-0

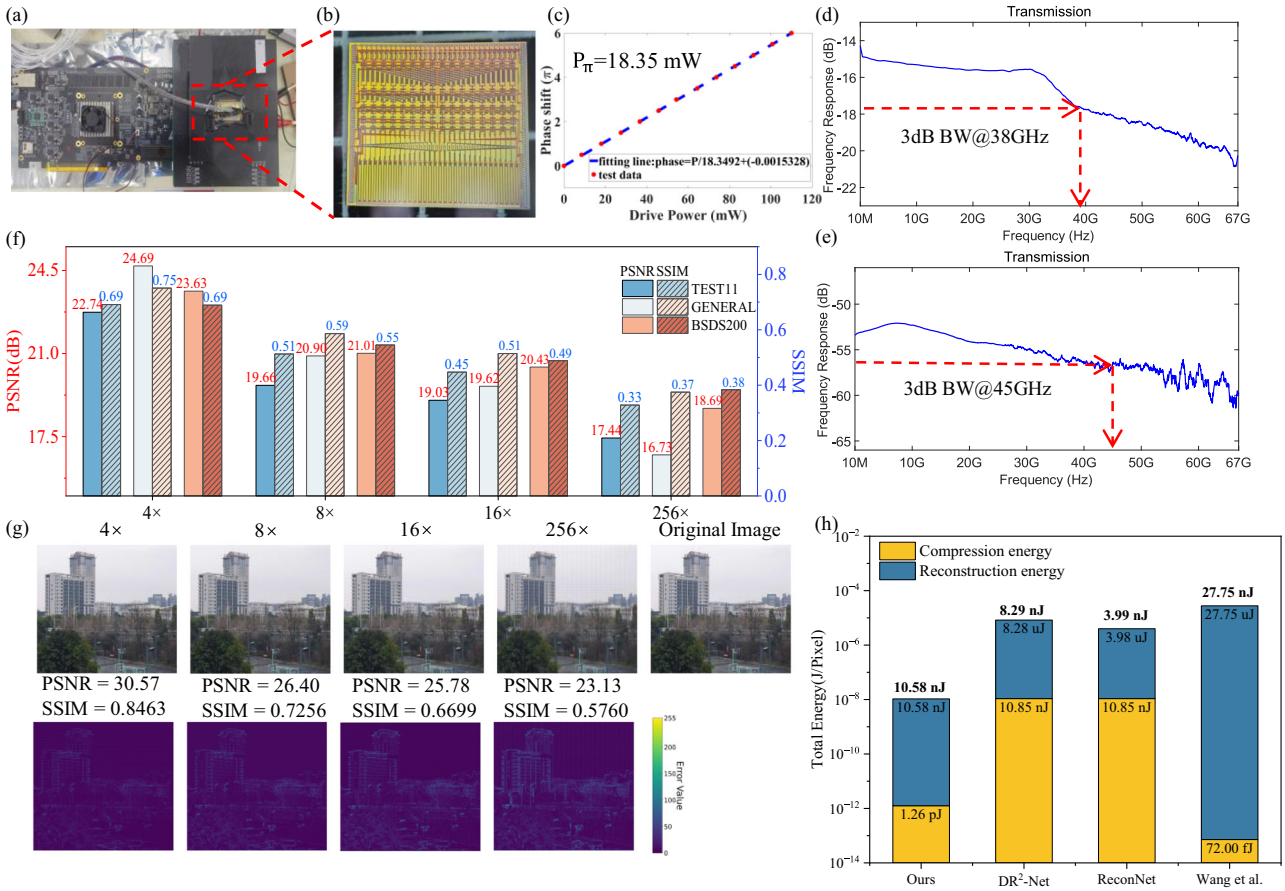

Fig. 4 | Opto-electronic computing processor characterization and system-level validation. Photos of the opto-electronic computing processor (OECP) (**a**) and the 32×32 photonic computing chip (PCC) (**b**), respectively. **c** Measured P_π of on-chip thermo-optic phase shifter. **d** Electrical bandwidth measurement of high-speed modulator. **e** Electrical bandwidth measurement of high-speed detector. **f** Performance comparison of OECP performing end-to-end compression and reconstruction of multiple image datasets at various compression ratios. **g** An actual photo taken in real life, processed by our OECP at various compression ratios, and a corresponding Viridis color map of the actual photo. The error visualization using the Viridis color map demonstrates the effectiveness of our reconstruction. The predominant dark purple hues across the image body indicate minimal reconstruction errors, while the sparse yellow regions are confined to fine details and edges, confirming good overall fidelity. **h** Energy consumption of our approach, the pioneer work in ref. 29, two classical models, ReconNet and DR^2-Net, running on state-of-the-art GPUs.

measurement", but still maintain at a high level. To provide comprehensive evaluation, we also provide SSIM and MS-SSIM results corresponding to Fig. 3b, which can be found in Supplementary Table 2.

Crucially, our approach drastically reduces computational demands, making it ideal for resource-constrained environments like edge devices. We quantify computational resources using multiply-accumulate operations per pixel (MAC/pixel) required by our model and other models. At 4× compression ratio, our image compression procedure requires 32 MAC/pixel via the 32×32 PCC. While for image reconstruction, our custom LiPICO-Net requires 786,432 MACs via the PCC (16×32×32×4×4, 16×32×32×1×1, 16×32×32×4×4 MACs for three CLs, respectively) and 262,144 to via electronics (due to the 256×1024 FCL). Thus, the total number of MACs amount to 1,048,576 for 32×32 pixels, resulting in 1024 MAC/pixel. This represents 95.46% and 97.70% reductions versus ReconNet (23,264 MAC/pixel) and DR^2-Net (45,984 MAC/pixel), respectively (Fig. 3c). When compared with the work in ref. 29, our approach only requires less than 1% computational resources. Moreover, the integration of our lightweight model with photonic computing hardware achieves transformative energy efficiency, detailed in Hardware implementation and characterization. To investigate the variation of performance with compression ratio, we utilize LiPICO-Net(AE) which can easily achieve accurate compression ratios for verification and have accordingly expanded our tests with more compression ratios to generate smoother curves. The average PSNR against compression ratio of 11 images is shown in Fig. 3(d). The average PSNR against compression ratio of BSDS 200 and the average SSIM and MS-SSIM against compression ratio of 11 images and of BSDS 200 are detailed in Supplementary Figs. 1–5.

The photo of the OECP and the PCC are given in Fig. 4a and b, respectively. The OECP co-packaged all necessary electronics to control the PCC, including ADCs, DACs, TIAs, Amplifiers and FPGA enabling standalone operation. The PCC implements an FFT-based mesh comprising only 80 Mach-Zehnder-interferometers (MZIs) versus ~500 required in Clements topology[34]. 192 thermo-optic PSs are implemented on-chip to program the PCC's matrix. Each PS has a high-power efficiency with $P_\pi < 18.4$mW (Fig. 4c), thanks to thermal trenches. This chip monolithically integrates 32 high-speed modulators to impose image data on an optical carrier with a bandwidth over 30 GHz (Fig. 4d), high-speed photodetectors for converting optical signal into the electrical domain with a bandwidth over 40 GHz (Fig. 4e). Image data would be read from an FPGA, converted into analog signals via DACs and applied on modulators to perform amplitude modulations. Modulated optical carriers would then propagate through the photonics matrix, which is pre-defined by the FPGA. Optical signals after matrix multiplication are converted into photocurrents by 32 photodetectors, subsequently amplified and transformed into voltage signals by TIAs, digitized by ADCs and sent into FPGA for further processing. This unified photonic-electronic data path executes

Table 1 | Performance benchmark against prior works

Function / Metrics / Model	Image compression				Image reconstruction			End-to-End	
	CR_{min}	CRmax	Latency	Energy	PSNR	Latency	Energy	Latency	Energy
ReconNet	4×	100×	24.12 ps	10.85nJ	25.54	8.83 ns	3.98uJ	8.85 ns	3.99uJ
DR^2-Net	4×	100×	24.12 ps	10.85nJ	**28.66**	18.39 ns	8.28uJ	18.41 ns	8.29uJ
29	4×	4×	3.91 ps	**72fJ**	/	61.67 ns	27.75uJ	61.67 ns	27.75uJ
Our work	**2×**	**256×**	**1.04 ps**	1.26pJ	24.17	**48.44 ps**	**10.58nJ**	**49.48 ps**	**10.58nJ**

Comparison of compression ratio, latency(per pixel), power consumption(per pixel) and PSNR at 4× between our work and other approaches. CR_{min} and CRmax denotes minimum and maximum compression ratio, respectively. Metrics in red indicate the best. Our work exhibits significant improvement upon compression ratio, total latency and energy consumption.

complete compression-to-reconstruction workflows without external computation. More details regarding the processor can be found in supplementary information S2. The key parameters of other devices, including waveguides, waveguide crossings, MZIs, and grating couplers, are listed in Supplementary Table 1.

We then perform end-to-end image compression and reconstruction of multiple datasets using OECP. The mean PSNR results of different datasets are given in Fig. 4f. Even if the performance degrades sightly compared with simulations, the minimum PSNR at 256× compression ratio is still around 17 dB for all datasets, indicating broad applicability. We also apply the OECP to an actual photo taken in real life. The results shown in Fig. 4g exhibit good reconstruction quality at various compression ratios, further validating the feasibility of our work.

Based on the estimation methods for key performance metrics explained in supplementary information S3, our processor can perform image compression and reconstruction at a latency of only 49.5 ps/pixel, which is 2-3 orders of magnitude faster than ReconNet and DR^2-Net running on state-of-the-art GPUs (Nvidia RTX 4090). In terms of energy consumption, our processor consumes in total 10.58nJ/pixel (with only 1.26pJ/pixel for image compression). The image compression part consumes more energy than the pioneer work in ref. 29 as it is a pure passive photonic chip without any control elements in[29]. But the total energy consumption is 3 orders of magnitude lower than the work in ref. 28 as well as than ReconNet and DR^2-Net running on state-of-the-art GPUs (Nvidia RTX 4090). To provide a clear performance comparison, we list the key performance metrics of our approach, the work in ref. 29, ReconNet and DR^2-Net in Table 1. Note that all the metrics related to the reconstruction model in ref. 29, ReconNet and DR^2-Net are based on real results obtained by running them on a state-of-the-art GPU Nvidia RTX 4090.

Besides testing on open datasets which typically contain small or moderate images, we further demonstrated the feasibility of our OECP in practical application where ultra-large images are generated. Acquired via an airborne large-format remote sensing camera, the image consists of over 130 million-pixel and has a size over 1GB. The PSNR values reach 34.00, 30.35, 28.71, and 26.56, at 4×, 8×, 16×, and 256× compression ratio. as shown in Fig. 5. Thanks to the custom lightweight model, it reduces 2887040 million and 5840640 million MACs compared to ReconNet and DR^2-Net. In combination with photonics computing based hardware, it achieves significant reduction in energy and latency compared with electronics counterparts. This demo proves the value of our system for massive images processing in applications like remote sensing. To provide comprehensive evaluation, we also provide MS-SSIM results corresponding to Fig. 4f, g and Fig. 5, which are listed in Supplementary Table 3.

Discussion

This work overcomes fundamental limitations in existing image processing by demonstrating the first fully integrated optoelectronic computing processor (OECP) for end-to-end compression and reconstruction. At its core, a monolithic 32×32 photonic computing chip (PCC) with an FFT-based matrix architecture enables programmable compression ratios (2×–256×) via dynamic submatrix reconfiguration, resolving the inflexibility of prior fixed-ratio systems. The co-designed LiPICO-Net model reduces computational load by >95% versus state-of-the-art AI models while maintaining competitive PSNR (> 34 dB at 4× and >17 dB at 256× compression ratio), enabling efficient on-chip reconstruction. Heterogeneous integration of all necessary control electronics establishes a unified photonic-electronic pipeline delivering unprecedented latency (49.5 ps/pixel) at 10.58pJ/pixel—2-3 orders of magnitude more efficient than GPU-based implementations of classical models. Validation on a 130 million-pixel aerial image confirms practical viability for latency/power-constrained scenarios where electronics fail. By addressing programmability, integration, and computational efficiency barriers, this work establishes a new paradigm for photonic computing in massive-scale image processing. The FFT architecture currently achieves a maximum scale of 64×64, limited by physical constraints including optical attenuation and thermal noise. Further scaling the matrix dimensions will necessitate more complex and costly approaches, such as multi-chiplet architectures or custom ASIC fabrication.

Methods

LiPICO-Net architecture and training

LiPICO-Net consists of one fully connected layer (FCL), one residual connection (RC), and three convolutional layers (CLs). The FCL and RC are implemented on FPGA, while CLs are executed by the optical computing core. To align with the optical hardware capabilities, the network maintains a constant width of 16 channels throughout the convolutional stages. The first and third CLs utilize 4×4 kernels (unfolded into 16×16 matrix operations) with a stride of 1, while the second CL employs a 1×1 kernel. For training, we utilized the T91 dataset, where original images were cropped into 32×32 pixel grayscale patches with a stride of 14. The network was trained using the Mean Squared Error (MSE) loss function and the Adam optimizer. The training process was conducted on an Nvidia RTX 4090 GPU. To

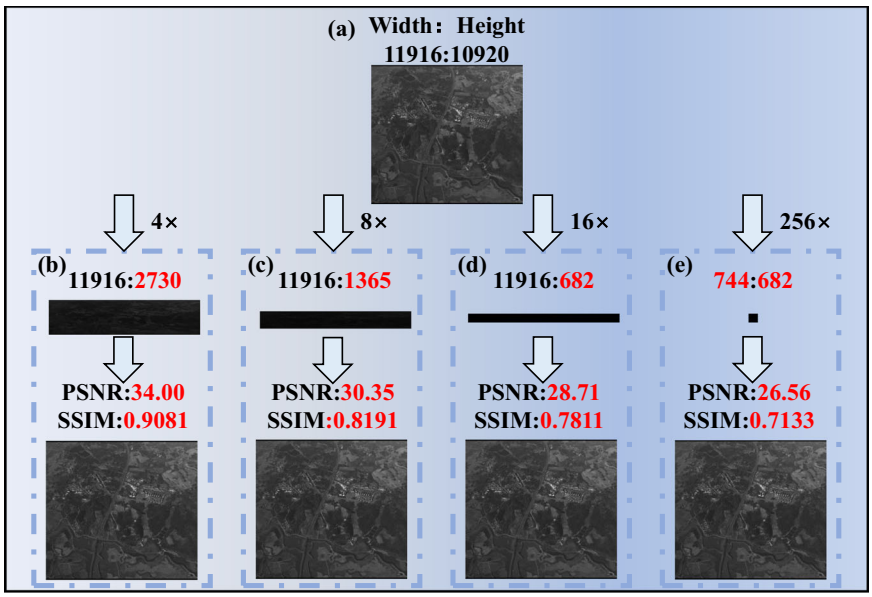

Fig. 5 | Ultra-large aerial image compression and reconstruction. a The original 130 million-pixel ultra-large image acquired by a 500 million-pixel high-precision airbone remote sensing camera. **b**–**d** compression images and reconstruction image using 4×, 8× and 16× compression ratio along Y direction. **e** Compressed images and reconstruction image using 256× compression ratio along XY direction.

evaluate the model performance, we used the Berkeley Segmentation Dataset 200 (BSDS200) and General100 dataset as the testing set.

Fabrication of the photonic computing chip

The 32×32 photonic computing chip (PCC) was fabricated on a silicon-on-insulator (SOI) substrate using Chongqing United Microelectronics Center (CUMEC)'s 130 nm silicon photonics lithography process. The optical core adopts a Fast Fourier Transform (FFT) architecture comprising 80 Mach-Zehnder Interferometers (MZIs), selected over the Clements architecture to minimize the number of phase shifters and enhance fabrication tolerance. The chip integrates 192 thermo-optic phase shifters with thermal trenches to ensure high power efficiency (P_π < 18.4 mW). For high-speed data encoding and readout, the device monolithically integrates 32 traveling-wave Mach-Zehnder modulators (MZMs) with a bandwidth exceeding 30 GHz and Germanium absorption photodetectors with a bandwidth over 40 GHz and high responsivity (~0.95 A/W at −1 V bias).

Opto-electronic system integration

The PCC is co-packaged with control electronics using vertical grating couplers for fiber-to-chip coupling and wire bonding for electrical interconnects. The computing board integrates the necessary driving hardware, including digital-to-analog converters (DACs) operating at 6 GSPS with 14-bit resolution and analog-to-digital converters (ADCs) operating at 2 GSPS with 12-bit resolution. Transimpedance amplifiers (TIAs) with a bandwidth of 450 MHz and a gain of 2kΩ are used to bridge the optical and electrical domains. The complete system operates at a rated voltage of 12 V and 4 A current, utilizing a 17.8 dBm laser source to power the optical computations.

Data availability

The reconstruction results generated in this study have been deposited in the Figshare database under the Creative Commons Attribution 4.0 International Public License and can be accessed via https://doi.org/10.6084/m9.figshare.31338478. This study made use of three publicly available benchmark datasets: BSDS200, GENERAL100, and T91. These datasets are widely used in the image reconstruction community and can be accessed from their official sources. For convenience, the versions used in this work are also available in the GitHub repository associated with ref. 30 at: https://github.com/Chinmayrane16/ReconNet-PyTorch/tree/master/images/all%20images. The 11 testing images used in this work are available in the GitHub repository associated with ref. 30 at: https://github.com/ricedsp/D-AMP_Toolbox/tree/master/TestImages. Readers are advised to refer to the original dataset publications for terms of use.

Code availability

The codes in this study have been deposited on Zenodo database under MIT License at https://doi.org/10.5281/zenodo.18641777.

References

1. Brady, D. J. et al. "Multiscale gigapixel photography. *Nature* **486**, 386–389 (2012).
2. Brady, D. J. et al. Parallel cameras. *Optica* **5**, 127–137 (2018).
3. Misra, A. et al. Mapping global floods with 10 years of satellite radar data. *Nat. Commun.* **16**, 5762 (2025).
4. T. Liu, et al. Astronomical image denoising by self-supervised deep learning and restoration processes. *Nat. Astro.* **9**, 608–615 (2025).
5. M. Kuehl, et al. Pathology-oriented multiplexing enables integrative disease mapping. *Nature*, 1–11, https://doi.org/10.1038/s41586-025-09225-2 (2025).
6. Jin, D. et al. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. *Nat. Mach. Intell.* **3**, 876–884 (2021).
7. Shi, L. et al. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* **591**, 234–239 (2021).
8. Abitbol, J. L. & Karsai, M. Interpretable socioeconomic status inference from aerial imagery through urban patterns. *Nat. Mach. Intell.* **2**, 684–692 (2020).
9. Yuan, X. & Haimi-Cohen, R. Image compression based on compressive sensing: End-to-end comparison with JPEG. *IEEE Trans. Multimed.* **22**, 2889–2904 (2020).
10. Jain, A. K. Image data compression: A review. *Proc. IEEE* **69**, 349–389 (1981).
11. Feng, Y. et al. Memristor-based storage system with convolutional autoencoder-based image compression network. *Nat. Commun.* **15**, 1132 (2024).

12. Skodras, A. et al. "The JPEG 2000 still image compression standard. *IEEE Signal Process. Mag.* **18**, 36–58 (2002).
13. Usevitch, B. E. A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000. *IEEE signal Process. Mag.* **18**, 22–35 (2002).
14. Li, M. et al. Learning content-weighted deep image compression. *IEEE Trans. pattern Anal. Mach. Intell.* **43**, 3446–3461 (2020).
15. J. Ballé, et al. End-to-end optimized image compression, arXiv preprint arXiv:1611.01704 (2016).
16. di Toma, A. Nnarrita et al. LiNbO_3-Based photonic FFT processor: an enabling technology for SAR On-Board processing. *J. Lightwave Technol.* **43**, 912–921 (2025).
17. Ashtiani, F. et al. "An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).
18. Shen, Y. et al. "Deep learning with coherent nanophotonic circuits. *Nat. photonics* **11**, 441–446 (2017).
19. Ahmed, S. R. et al. "Universal photonic artificial intelligence acceleration. *Nature* **640**, 368–374 (2025).
20. Wang, Z. et al. Integrated photonic metasystem for image classifications at telecommunication wavelength. *Nat. Commun.* **13**, 2131 (2022).
21. Zhou, T. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
22. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
23. Xu, Z. et al. Large-scale photonic chiplet Taichi empowers 160-TOPS/W artificial general intelligence. *Science* **384**, 202–209 (2024).
24. Feldmann, J. et al. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
25. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
26. Wang, T. et al. Image sensing with multilayer nonlinear optical neural networks. *Nat. Photonics* **17**, 408–415 (2023).
27. Fu, T. et al. Photonic machine learning with on-chip diffractive optics. *Nat. Commun.* **14**, 70 (2023).
28. Xu, X. et al. "11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
29. Wang, X. et al. Integrated photonic encoder for low power and high-speed image processing. *Nat. Commun.* **15**, 4510 (2024).
30. K. Kulkarni, et al. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 449-458 (2016).
31. Yao, H. et al. Dr2-net: Deep residual reconstruction network for image compressive sensing. *Neurocomputing* **359**, 483–493 (2019).
32. Huszar, F., Theis, L., Shi, W. & Cunningham, A. Lossy Image Compression with Compressive Autoencoders. Apollo - University of Cambridge Repository. https://doi.org/10.17863/CAM.51995 (2020).
33. D. Bank, et al. Autoencoders, Machine learning for data science handbook: data mining and knowledge discovery handbook, 353–374 (2023).
34. Fang, M. Y.-S. et al. Design of optical neural networks with component imprecisions. *Opt. express* **27**, 14009–14029 (2019).

Acknowledgements

This work was supported by Natural Science Foundation of China under grant No. 62422507 and 62375126.

Author contributions

A.L. conceived the idea. A.L. and S.P. supervised the project. Y.W. and Y.S. developed the compression/reconstruction algorithms and performed simulations. Q.L. and Y.W. designed the optical layout. Q.L. and Y.Z. packaged the photonic chip with control electronics. Q.L., Y.Z., and Y.W. performed measurements. A.L., Y.W., and Y.S. analyzed data and calculated performance metrics. Y.W. and Y.S. generated all figures. All authors discussed results and contributed to manuscript writing/revision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-71296-0.

Correspondence and requests for materials should be addressed to Ang Li or Shilong Pan.

Peer review information *Nature Communications* thanks David Brady, Giuseppe Brunetti, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.