nature photonics

Article

https://doi.org/10.1038/s41566-024-01493-0

Nonlinear optical encoding enabled by recurrent linear scattering

Received: 16 August 2023

Accepted: 1 July 2024

Published online: 31 July 2024

Check for updates

Fei Xia[1] ✉, Kyungduk Kim[2], Yaniv Eliezer[2], SeungYun Han[2], Liam Shaughnessy[2], Sylvain Gigan[1] ✉ & Hui Cao[2] ✉

Optical information processing and computing can potentially offer enhanced performance, scalability and energy efficiency. However, achieving nonlinearity—a critical component of computation—remains challenging in the optical domain. Here we introduce a design that leverages a multiple-scattering cavity to passively induce optical nonlinear random mapping with a continuous-wave laser at a low power. Each scattering event effectively mixes information from different areas of a spatial light modulator, resulting in a highly nonlinear mapping between the input data and output pattern. We demonstrate that our design retains vital information even when the readout dimensionality is reduced, thereby enabling optical data compression. This capability allows our optical platforms to offer efficient optical information processing solutions across applications. We demonstrate our design's efficacy across tasks, including classification, image reconstruction, keypoint detection and object detection, all of which are achieved through optical data compression combined with a digital decoder. In particular, high performance at extreme compression ratios is observed in real-time pedestrian detection. Our findings open pathways for novel algorithms and unconventional architectural designs for optical computing.

Optical information processing leverages the unique properties of light, such as its parallelism, which allows the simultaneous processing of multiple data streams, as well as low energy consumption[1–5]. Moreover, light possesses a vast frequency spectrum, enabling ultrahigh bandwidth and data throughput[4–7]. By exploiting these characteristics, optical information processors have the potential to unlock new levels of performance, scalability and energy efficiency, which could transform the landscape of information processing in the optical domain[4,5]. It has enabled new applications when coupled with existing optical instruments, such as imaging system[8] to enhance the performance.

However, the full potential of optical processors can only be realized by overcoming certain challenges, one key requirement being optical nonlinear mapping[6,9]. Nonlinear mapping is essential to approximate arbitrary function and has been a powerful element in neural networks as it allows models to recognize complex patterns and approximate any given function[10]. It plays a vital role in representation and feature learning, as it facilitates the discovery of higher-level, more informative and discriminative features for a task[11,12]. The application of nonlinear mappings allows the extraction of abstract and nonlinear features, thereby enhancing the input data representation[13,14]. In existing optical computing platforms, optical nonlinear mapping has been primarily achieved using nonlinear optical materials, which provide a nonlinear relationship between the input and output fields[15–22]. Optical nonlinearity often requires intense pumping and high peak power, which can be energy demanding, necessitates design and engineering of nonlinear or active materials and is generally restricted to lower-order nonlinear mapping with limited tunability[6,9]. Alternatively, the conversion of signals from optical to electrical and back to optical is used for the nonlinear processing of optical data, but with limited speed.

[1]Laboratoire Kastler Brossel, ENS-Universite PSL, CNRS, Sorbonne Université, Collège de France, Paris, France. [2]Department of Applied Physics, Yale University, New Haven, CT, USA. ✉e-mail: fei.xia.oe@gmail.com; sylvain.gigan@lkb.ens.fr; hui.cao@yale.edu

Here we propose to exploit the passive nonlinear optical mapping inside a multiple-scattering cavity[23], akin to the steady state of a reservoir computer, for rapid optical information processing. High-order nonlinearity fosters the generation of low-dimensional latent feature space and facilitates strong data compression. Previously, propagation through a multiple-scattering material has been exploited to perform linear optical random projections[24], followed by an intensity detection. It can be regarded as a single-random-layer neural network, and has been used for multiple machine learning tasks[25–28], but remains limited in performance by its intrinsic linear mapping behaviour. By introducing multiple scatterings in a cavity design, we enable multiple bounces on the same input pattern, effectively creating an optical nonlinear transformation of the input data, without the need of nonlinear optical materials or optical–electrical–optical conversion typically used for nonlinearity in optical information processing. We demonstrate high computing performances across tasks from classification, image reconstruction, keypoint detection and object detection, with the optically compressed output fed into a digital decoder. In particular, we show that our system exhibits high performance even at a mode compression ratio (defined by the input macropixel numbers on a digital micromirror device (DMD) to output the number of speckle grains on the camera) of ~3,000:1 for high-level computing tasks, as evidenced in real-time pedestrian detection with bounding box generation. Our work illuminates the role of varying nonlinear orders in optical data compression based on mutual information analysis, and paves the way for tunable optical nonlinear mapping and energy-efficient computing.

Results

Nonlinear random mapping with tunable nonlinearity

Introducing nonlinearity has long been a challenge and simultaneously a necessity in optical computing platforms. Nonlinearity is a key element for enabling complex operations and boosting computational power[4,5]. It is particularly important for approximating arbitrary functions—a task critical in machine learning. In this study, we present a novel approach to address this challenge by utilizing nonlinear mapping provided by multiple linear scatterings of light within an optical cavity[23]. We constructed the multiple-scattering cavity using an integrating sphere (Fig. 1a), which features a rough inner surface that scatters light. A continuous-wave laser operating at low power is injected into the cavity via the first port, resulting in an output speckle pattern from the second port. The third port integrates a DMD to display the input patterns. In general, a Born series can be used to describe the scattering process in the cavity:

$$E_{\text{out}} = \mathbf{T}E_{\text{in}} = \left[\mathbf{V} + \mathbf{V}(\mathbf{G}_{\text{o}}\mathbf{V}) + \mathbf{V}(\mathbf{G}_{\text{o}}\mathbf{V})^2 + \ldots\right]E_{\text{in}}. \quad (1)$$

Here matrix $\mathbf{T}$ represents a linear mapping from the input optical field E_{in} to the cavity to the output field E_{out}. $\mathbf{V}$ is the matrix that denotes the scattering potential inside the cavity, and $\mathbf{G}_{\text{o}}$ is Green's matrix representing light propagation within the cavity in between bounces off the boundary. The notation $(\mathbf{G}_{\text{oV}})^n$ represents the matrix $\mathbf{G}_{\text{oV}}$ multiplied by itself n times. The final intensity image formed on the camera is given by $I_{\text{cam}} = |E_{\text{out}}|^2$, where $||^2$ represents an element-wise operation. The $\mathbf{T}$ expansion begins with a term indicative of single scattering and subsequent terms indicate multiple scatterings in the cavity. In the cases where a single scattering is the dominant event, the mapping from $\mathbf{V}$ to E_{out} is predominantly linear. In our case with multiple scatterings, the relation between the scattering potential configuration $\mathbf{V}$ and output field E_{out} becomes nonlinear. The Born series can also be reformulated as $\mathbf{T} = \mathbf{V}\sum_{m=1}^{\infty}(\mathbf{G_0V})^{m-1} = \mathbf{V}\sum_{m=1}^{\infty}\mathbf{U}\mathbf{\Lambda}^{m-1}\mathbf{U}^{-1}$, where $\mathbf{G_0V} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, $\mathbf{\Lambda}$ is a diagonal matrix of elements equal to eigenvalues of $\mathbf{G_0V}$ and the corresponding eigenvectors are columns of $\mathbf{U}$. For high-order m, the largest eigenvalue $\lambda_{\max}$ dominates over all the other eigenvalues, and the polynomial orders in $\mathbf{T}$ can be approximated as $\sum_{m=1}^{\infty}\lambda_{\max}^{m-1}$. Thus, the nonlinear coefficient decays exponentially with order m. Experimentally, due to chaotic ray dynamics in our cavity, it is difficult to extract the largest eigenvalues for different active areas on the DMD. Furthermore, since part of the surface area of the cavity can be modified by the DMD using the input (modulation) patterns, this also provides a reconfigurable scattering potential inside the cavity. Multiple bounces of light off the modulated area of the DMD results in a nonlinear mapping from the input pattern displayed on the DMD to the output speckle pattern. As the number of bounces on the DMD increases, the order of this nonlinear mapping increases (Fig. 1b,c). It is this nonlinear relationship that forms the foundation for the passive nonlinear encoding technique that we explore in this work.

Light scattering within the cavity can be solely adjusted by altering the pattern displayed on the DMD as an input pattern (Fig. 1a). Each micromirror on the DMD can be toggled between two angles. This action effectively modifies the scattering potential $\mathbf{V}$ for light, determining mapping from the input pattern on the DMD to the output optical field E_{out}. A larger modulation area boosts the probability of light scattering by the modulated part of the scattering potential, thereby enhancing the nonlinear mapping. The more times light is scattered by the DMD pattern, the more chance it samples the input pattern (Fig. 1b). Each scattering event effectively mixes the information from different parts of the DMD, resulting in a complex optical encoding of the entire pattern. The longer the light remains in the cavity, the more encoding and mixing occur, effectively ensuring that light in each output mode (speckle grain) carries information about a multitude of input data (Fig. 1b). The interaction due to multiple scatterings results in a nonlinear mapping where the intensity of each output mode (speckle grain) becomes a highly nonlinear function of the input pattern (Fig. 1c). The number of bounces determines the order of nonlinearity. To further enhance the nonlinear order, the number of bounces is increased by covering the output port by a partial reflector to increase the dwell time of light inside the cavity. Such nonlinear mapping induced by multiple scatterings is purely passive (no need for high power) and is fundamentally distinct from the traditional nonlinear optics that rely on an intrinsic material response.

This scheme offers an efficient means of achieving tunable high-order nonlinear random mapping at a constant low power with a continuous-wave laser in a passive manner (Methods), compared with conventional optical nonlinearities that rely on the material response at high optical intensity[29,30]. In our case, the nonlinear order is independent of the input power, and can be rapidly tuned (~20 kHz) by altering the DMD-modulated area. This rapid tuning capability outperforms many known nonlinear effects, such as thermo-optical nonlinearity[31,32]. Additionally, our scheme avoids dynamic chaos and instabilities commonly associated with conventional nonlinear optical systems and lasers[29,30,33].

To comprehend and characterize the tunable nonlinear mapping introduced by our system, we explore how deep neural networks can function as a proxy to understand the nonlinear random mapping in our system. As detailed in Supplementary Section 3, we find that the higher-order nonlinear mapping, provided by a larger area of modulation on the DMD, can be approximated by a deeper neural network (the 'Further explanation of Born series' section explains the reformation of the Born series in terms of its proxy as deep neural networks with fixed random weights).

Enhanced image classification

To evaluate whether this nonlinear mapping can indeed provide any computational benefits, we begin by testing on a simple but widely recognized machine learning benchmark task, namely, the Fashion MNIST dataset[34]. Fashion MNIST is a popular fashion image classification challenge that includes 60,000 training samples and 10,000 test samples, each image measuring 28 × 28 pixels.

We input the Fashion MNIST data on the DMD and directly read the output speckle pattern to obtain both higher- and lower-dimensional

https://doi.org/10.1038/s41566-024-01493-0

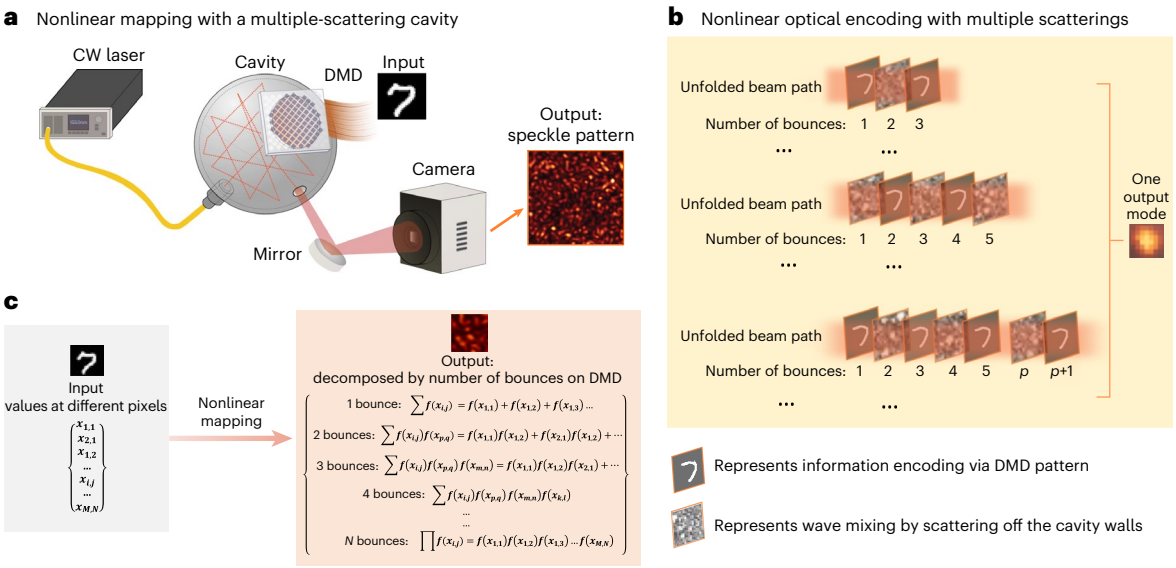

Fig. 1 | Concept of using a multiple-scattering cavity as a passive, tunable nonlinear optical information processor. a, Experimental setup in which the key component for creating the passive nonlinear random mapping is a DMD mounted on an integrating sphere. The output of the cavity produces a fully developed speckle pattern, with its response being nonlinear in the geometric configuration of the DMD. **b**, Representative figure showing that the cavity essentially encodes the input pattern on the DMD by optically mixing different areas of input through multiple bounces to create a highly nonlinear feature—a speckle recorded by a camera (input pattern is adapted from the MNIST dataset[62]). **c**, Mathematical representation of a nonlinear mapping process that transforms a set of input elements on the DMD into a collection of nonlinear features in the output speckle pattern. Multiple scatterings in the cavity generate mixed terms of input values at different pixels with various high nonlinear orders, which provide rich nonlinear features that can be optimally trained to enhance performance in complex computational tasks. $f(x)$ denotes the operation of scaling the configuration of a DMD macropixel $x_{i,j}$.

representations. These representations, which we refer to as nonlinear features of the input information, can be utilized to execute computing tasks. To achieve nonlinear random mapping with tunable nonlinearity, given a dataset with fixed input size, we either adjust the modulated area of the DMD (Fig. 2a) or partially close the output port to change the number of times light is scattered by the DMD. During the training phase, we train only the linear digital layer using the nonlinear features generated from the training dataset at each given configuration. In the inference/test stage, we forward the output images from the cavity to the trained linear digital layer to generate predictions (Methods).

In Fig. 2b, we present the classification performance in the Fashion MNIST dataset using a linear classifier. To quantitatively compare the performance of different nonlinear strengths in the optical encoder, we fixed the linear decoder and used test accuracy as a metric for comparison. We observe that stronger nonlinearity leads to improved classification performance, particularly when the number of optical modes/speckles is smaller. This indicates that each speckle from higher-order nonlinear mapping embeds more information. These findings further suggest that our device may possess a unique advantage in optical data compression.

To more comprehensively quantify the information within each spatial mode (speckle grain) in our output images, we employ the concept of mutual information. Compared with regression, mutual information includes both linear and nonlinear dependencies and does not make assumptions about the underlying data distribution[35]. It is widely used in compressive sensing[36], a technique focused on efficiently acquiring and reconstructing sparse or compressible signals, and has found important applications in machine learning for tasks such as feature selection[37], model interpretation[38] and understanding variable dependencies[39]. In our case, we calculate the mutual information between the output features and target classes for the dataset (Methods and Supplementary Section 4). This quantifies how well the nonlinear optical features contain the abstract information that is useful for high-level computing tasks (Supplementary Section 4). In Fig. 2c,d. the violin plots effectively illustrate the distribution of mutual information between the speckle grains and classification targets. A notable observation from the results is the onset of saturation of mutual information required for the Fashion MNIST classification with 4–25 modes/speckles (Fig. 2c). This saturation occurs under the highest-order nonlinear mapping in our experiments. Further, Fig. 2d underscores the benefit of escalating to higher-order nonlinearity. We observe that, indeed, higher-order nonlinear mapping creates stronger mutual information between the features and targeted classes, given the same number of output modes/speckles. This observation implies that our system can more effectively capture the underlying relationships between the features and target classes when higher-order nonlinear mapping is introduced.

Demonstration with complex tasks

Image reconstruction. Building on the enhanced information provided by the nonlinear features from our system, we pose the question: can this enhanced information (within a few output modes) yield superior image reconstruction? To address this, we conduct a comparative analysis of the nonlinear features generated in two distinct scenarios: one featuring a higher-order nonlinear optical random mapping induced in the multiple-scattering cavity (Fig. 3c), and another that presents a linear optical random projection (Fig. 3a)[24] with nonlinearity only at the detection stage (intensity measurement).

To most efficiently extract the embedded information from a few speckles, we deviate from the traditional approach of employing a digital linear layer for classification and, instead, introduce a customized and optimized multilayer perceptron (MLP) as a decoder for image reconstruction. The architecture of this decoder is finely tuned using neural architecture search to optimize the image reconstruction. Subsequently, we train two digital decoders, each featuring optimal architectures, on the Fashion MNIST dataset under high compression ratios of ~31:1 (only ~25 modes) (Fig. 3).

It is noteworthy that despite the optimally trained decoder in each case, the quality of the reconstructed images varies (Fig. 3b,d and Supplementary Figs. 4 and 5). We observe that augmented nonlinear

https://doi.org/10.1038/s41566-024-01493-0

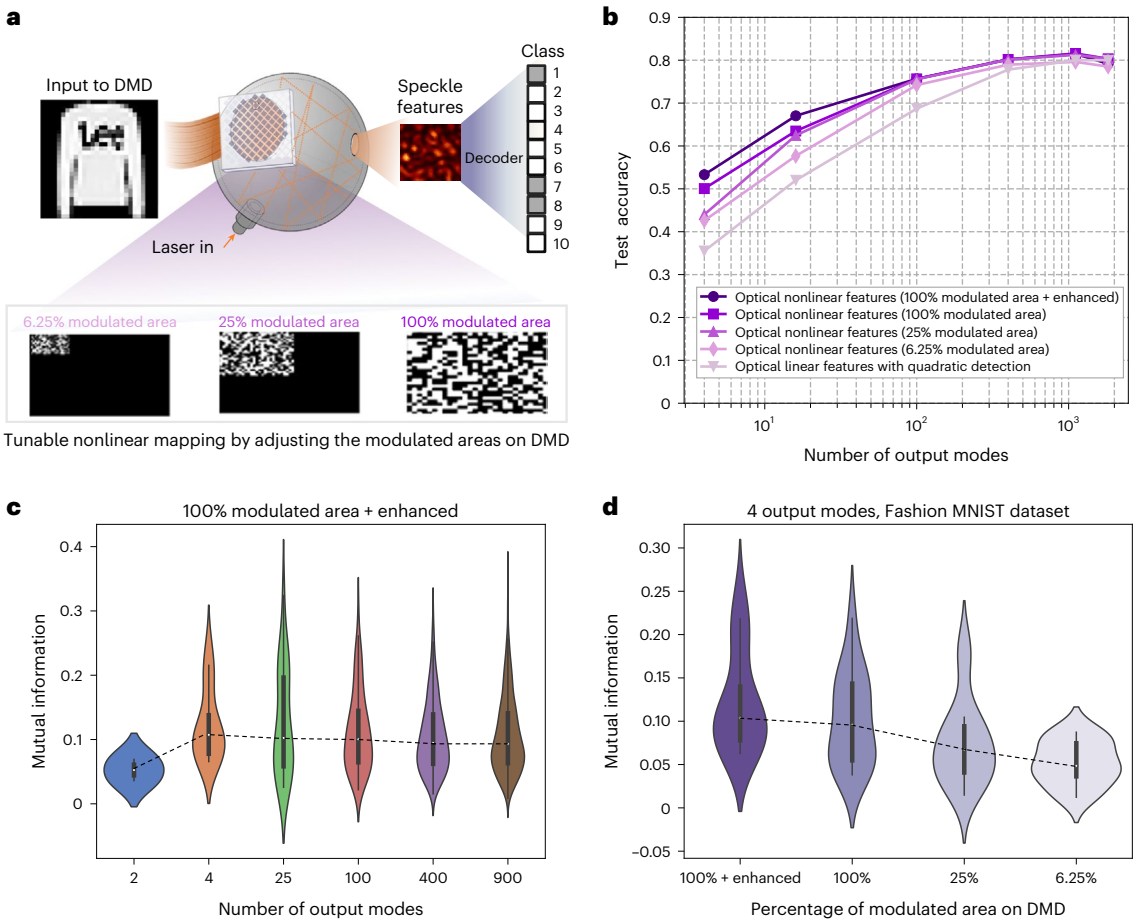

Fig. 2 | Classification with nonlinear mapping. a, Training data from the Fashion MNIST datasets are used to train a one-layer neural network as a digital decoder for classification tasks. Additionally, the percentage of the modulated area on the DMD is changed among 6.25%, 25% and 100% to adjust the order of nonlinear mapping. With full (100%) modulation of DMD, the nonlinear order is further enhanced by covering the output port with a partial reflector (silicon wafer). **b**, Fashion MNIST classification results with a linear classifier are presented under different numbers of output modes (speckle grains) and varying nonlinear strengths. The optical linear features with quadratic detection are simulated by scattering from a single layer with intensity detection to create a quadratic nonlinear response. Note that a linear regression for binarized Fashion MNIST data cannot exceed 77.6% with the same number of modes. **c**,**d**, Violin plots representing the distributions of mutual information between the speckle grains and classification targets under varying numbers of output modes (**c**) and differing orders of nonlinear mapping by changing the modulated area on the DMD or partially closing the cavity (enhanced) (**d**). For n speckle mode (n on the x axis), $4n$ replicated measurements from the same input were performed in **c** and **d**. The dashed line plots depict the median values of the mutual information. Each violin's width reflects the distribution of the mutual information values of the speckle grains and its probability density. Within each violin, the slim black vertical line represents the range of minimum and maximum values; the black box represents the first to third percentile; the white dot represents the median. **c**, Mutual information analysis when the number of output modes (speckle grains) varies under the highest-order nonlinear mapping. **d**, Mutual information analysis with low-dimensional speckle features (four output modes) for Fashion MNIST as a function of the nonlinear orders varied by modulated area on the DMD, showing the advantage of going to higher-order nonlinear mapping.

random mapping indeed facilitates improved image reconstruction, with a mean squared error of ~1.4 in the test set (Fig. 3d and Supplementary Fig. 5) compared with that of ~2.0 in linear optical features (Fig. 3b and Supplementary Fig. 4), each under a separately optimized decoder architecture. When using the same decoder architecture, the nonlinear features still outperform with the mean squared error of ~1.5 (Supplementary Fig. 6).

Our findings show that the nonlinear optical mapping in our system can efficiently compress and retain vital information as well as decrease data dimensionality. Motivated by these results, we are prompted to explore the potential of nonlinear features in executing other high-level computing tasks.

Keypoint detection. A key advantage emerging from our work is that optical data compression, facilitated by multiple scatterings in the cavity, generates mixtures of highly nonlinear features. These are particularly useful for applications that require high-speed analysis and responses of high-dimensional data. Our DMD contains 4 million pixels and can accommodate large images. However, in our image reconstruction demonstration, the input dimensions of the Fashion MNIST dataset are limited to 28 × 28 pixels, creating an inherent upper limit for the maximum compression ratio that can be demonstrated. A major strength of our system is its ability to easily scale up the size of the input data as well as the effective neural network's depth of the optical encoder without increasing the input power, thereby allowing for an efficient representation of the input information in an energy-efficient way. This adaptability and scalability facilitates tackling more complex tasks and processing larger datasets without losing crucial information.

Pushing the compression further and exploring other high-level computing tasks, we delve into two specific applications where we scale up the input images. A notable example (Fig. 3e) demonstrates that we can extract 15 keypoints from human face images[40] with an order of magnitude improvement in the mean squared error, which decreases from 0.208 (using 25% modulated area in the DMD) to 0.014 (using

https://doi.org/10.1038/s41566-024-01493-0

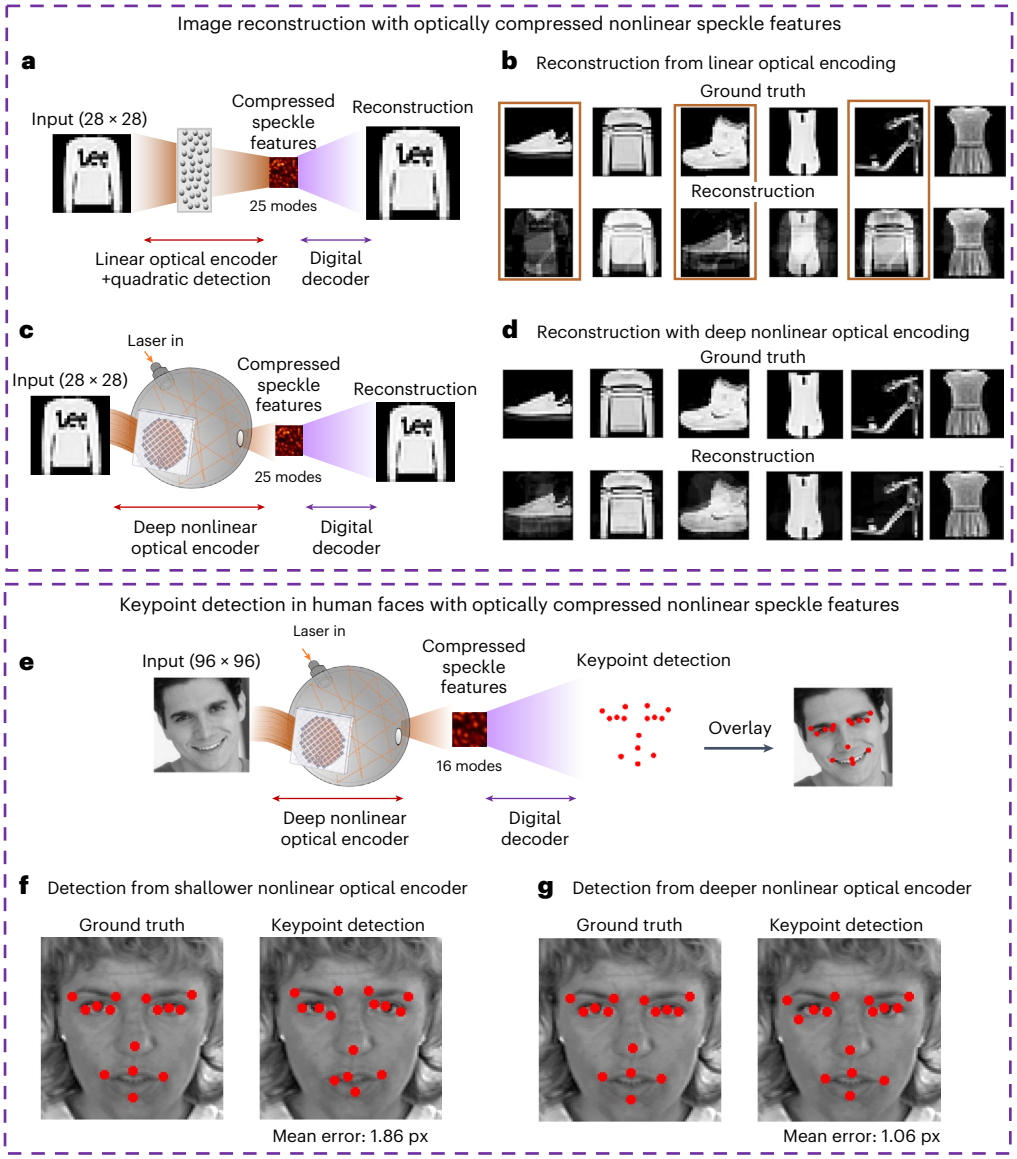

Fig. 3 | Computing performance enhanced by nonlinear optical data compression. a, Concept of image reconstruction using linear optical complex media for linear encoding and camera detection with quadratic response. **b**, Reconstruction using the speckle features from **a**. The orange boxes represent the wrongly reconstructed pairs. **c**, Multiple-scattering cavity as a nonlinear optical encoder along with camera detection and employing compressed speckle features for digital reconstruction of the original image data. **d**, Reconstruction from speckle features generated by the multiple-scattering cavity. In **b** and **d**, approximately 25 speckle grains are used with a compression ratio of 31:1 and are used to train two digital decoders (Methods). It is demonstrated that given the same number of compressed output modes (speckle grains), nonlinear features generated from the cavity can provide a reduced mean squared error by 0.6, resulting in a better reconstruction of the images in **d** compared with **b**. More results are provided in Supplementary Figs. 4–6. **e**, Concept of keypoint detection in human faces (images with 96 × 96 pixels) with compressed speckle features. **f**, Keypoint detection with a mode compression ratio of 576:1, using 16 output modes with relatively weaker nonlinearity (25% modulated areas in the DMD) and a five-layer MLP decoder. **g**, Improved keypoint detection with a reduced mean error in pixels across 15 keypoints (1.06 pixels compared with 1.86 pixels errors in **f**), using 16 output modes (speckle grains) with relatively stronger nonlinearity (full modulated areas in the DMD) and a nine-layer MLP decoder.

100% modulated area in the DMD enhanced with the partial reflector), due to the incorporation of stronger nonlinearity with a larger modulation area, even when the number of output modes (speckle grains) is reduced to 16 (Fig. 3f–g). In both cases, the architectures of the decoders are separately optimized and trained for optimal performance. Even when we use the same architecture (a five-layer MLP) that was optimized for features from a 25% modulated area in the DMD for the decoder to train features from the latter case, the mean squared error associated with these higher nonlinear features remains low (~0.3). This task—crucial for various applications such as facial recognition, emotion detection and other human–computer interaction systems—illustrates the robustness of our approach in dealing with high-level tasks and maintaining a high compression ratio. An additional advantage of our methodology lies in its implications for privacy protection and adversarial robustness, as our method can securely encode facial information in random speckle grains.

Real-time video analytics. The last application we demonstrate is real-time video analytics, using the benchmark dataset known as Caltech Pedestrian[41], including real-time video recordings (Fig. 4a). The images from the videos displayed on the DMD have dimensions of 240 × 320 pixels (Methods). Using our multiple-scattering cavity,

https://doi.org/10.1038/s41566-024-01493-0

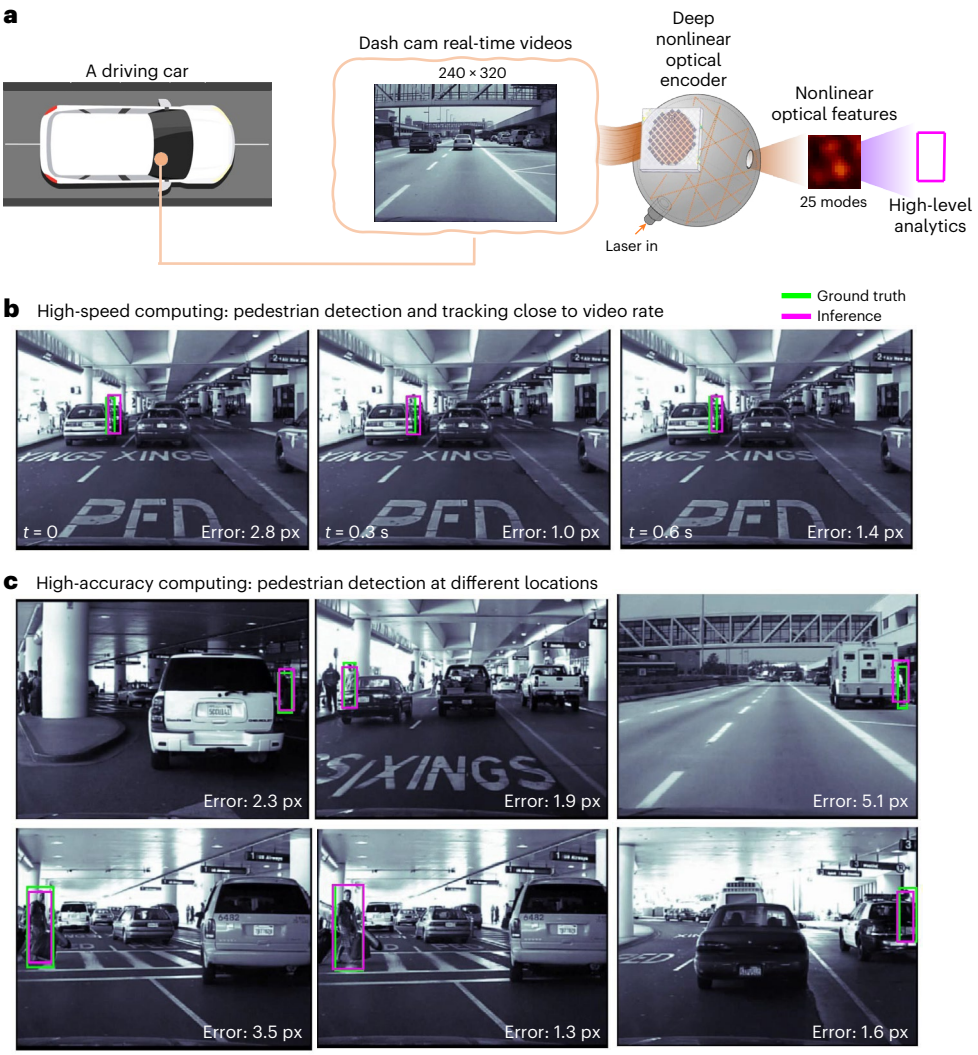

Fig. 4 | Real-time video pedestrian detection in driving with high mode compression ratio using only 25 output modes. a, Schematic of real-time pedestrian detection using video data from a dash camera during driving. The multiple-scattering cavity functions as an optical data compressor, and compressed nonlinear optical features are utilized for pedestrian detection with a digital decoder. **b**, Demonstration of pedestrian detection at a rate close to a real-time video. The magenta boxes represent the inference results from the speckle. The green boxes represent the ground truth. The speed of optical processing, that is, nonlinear feature generation, is as fast as light, and its readout speed is limited by only the camera. With only 25 modes, our camera can currently reach at least 800 Hz. The inference time with the 25 modes in pedestrian detection is 0.0035 s, leading to a total response time (inference + generation of optical features) of less than 0.1000 s, which is faster than the typical human response time of ~0.2000–22.0000 s. The error unit is in pixels (px). **c**, Demonstration of pedestrian detection at various locations during continuous video streaming; the mean detection error with only 25 modes remains within 1.92 pixels (px).

we can compress the data to achieve a compression ratio of up to 3,072:1 (that is, using only 25 output modes), as well as maintain high positional accuracy (Fig. 4b) within mean squared errors of 1.92 pixels in identifying pedestrian positions at a high speed and 0.0035 s in total response time (including compressed optical nonlinear feature generation and inference time) with an optimized digital backend (a ten-layer MLP) per frame (Fig. 4c and Supplementary Videos 1 and 2).

This application is particularly critical in the field of autonomous vehicles and advanced driver-assistance systems, where high-speed pedestrian detection is essential to ensure safety and allow fast reaction time. The high compression ratio of our system, combined with its rapid processing speed, shows great promise for such applications where fast and accurate detections are imperative.

To further estimate the gain we have in terms of optical data compression, we calculated the number of parameters and operations in the digital domain with and without an optical encoder. In human face keypoint detection, our method with an optical encoder demonstrated a mean pixel error of 1.06, slightly surpassing the performance of a widely utilized conventional convolutional neural network (CNN) architecture (which is still widely used as a benchmark for vision tasks). With a CNN model (one convolutional layer + one pooling + one convolutional layer + three fully connected layers), we achieved a mean pixel error of approximately 1.23. The digital CNN model comprises over 74 million parameters and requires around 83 million operations, and are two orders of magnitude higher than the digital operations/parameters used in our system (approximately 310,000 digital trainable parameters/operations). This comparison underlines the enhanced accuracy of our approach, as well as pointing to a substantial reduction in computational complexity and resource utilization inherent to our method. For pedestrian tracking, our method exhibited mean pixel errors ranging from 1.3 to 3.6, closely matching the performance of a conventional CNN model used for comparison, with which we obtained mean pixel errors between 1.37 and 3.33. The comparison model in this instance incorporates three convolutional layers and two

fully connected layers, involving more than 39 million parameters and necessitating approximately 172 million operations. In contrast, our decoder, after nonlinear optical projection, used only about 1 million parameters and a similar number of operations (1 million) to achieve comparable performance. In general, the higher the input dimension and the more we compress, the larger the number of digital operations we can allocate into the optical domain and therefore better leverage the advantage of information processing of light.

Discussion

Exploiting optics for computing, which brings benefits such as high speed, large bandwidth and parallelization, has traditionally been impeded by the challenge of addressing optical nonlinearity. Conventional all-optical methods typically involve complex experimental conditions, using nonlinear materials, like nonlinear crystals or polymers, pumped by high-power short-pulsed lasers, or semiconductor lasers operating in continuous or pulsed modes[42,43]. Although these have shown optical computing benefits in a variety of platforms including multimode fibre[18], integrated photonics[20] and free-space optics, limitations regarding their robustness, energy efficiency and stability persist.

In this work, we completely avoid the limitations of conventional optical nonlinearity by proposing a unique approach to achieve optical nonlinear random mapping by utilizing multiple scatterings within an optical cavity. This strategy enables us to institute nonlinear random mapping, where the adjustment of nonlinearity is entirely dependent on the geometrical configuration and quality factor of the cavity, thereby influencing the scattering potential. The intrinsic mixing of input information within the dataset at varying nonlinear orders permits us to generate highly nonlinear features compared with traditional optical nonlinear mappings, especially those with solely lower-order (2–3) nonlinearity that most nonlinear materials practically permit. From a machine learning perspective, by expanding to higher-order nonlinear mapping, we essentially generate an augmented optical feature space, incorporating more mixtures of higher-level input information. This expansion increases mutual information between the subspace of the feature space and the entire input pattern (evident by the image reconstruction task) and output targets (Fig. 2 and other high-level tasks), facilitating a higher compression ratio for complex tasks. In essence, our system demonstrates the capability to execute optical data compression by harnessing multiple scatterings of light in a reconfigurable cavity. This approach allows for the efficient preservation of critical information as well as stringently reducing data dimensionality.

We have demonstrated that our multiple-scattering cavity, equipped with passive and tunable nonlinear optical random mapping capabilities, can act as an optical nonlinear encoder with adjustable nonlinearity. Our system can deliver enhanced computing performance in a low-dimensional latent feature space for a range of computing tasks, from image classification to higher-level tasks such as image reconstruction, keypoint detection and object detection, when trained with a lightweight digital backend. This approach might offer considerable benefits for high-speed analytics in both scientific and real-world applications. Our system permits easy scaling of both input data and effective depth of the neural networks approximating the optical encoders, providing an efficient optical representation of large-scale input patterns using a limited number of output modes. This versatility helps to manage more intricate tasks and process larger datasets without substantial loss of vital information.

Our nonlinear mapping system functions as a reservoir computer in a steady state. This is also the case for other systems that have been realized before[18,44,45] but comparatively, our design allows for easy scaling up and tuning of nonlinearity without varying the input power. Furthermore, our system may serve as a trainable physical neural network[46], if one part of the DMD is utilized for an input pattern and another is tuned or trained for direct readout without the need for digital processing. The performance of our computing tasks can be further improved by, for example, replacing the binary DMD with an analogue spatial light modulator for information encoding. The detection part of our system can be further improved by replacing the camera with a fast photodetector array, given the small number of output modes that need to recorded for decoding.

Our current optical computing architecture is beyond one-to-one architectural mapping of the digital neural network. It may inspire next-generation optical computing to exploit nonlinear mappings beyond conventional schemes and promote the development of more energy-efficient neuromorphic computing platforms including[47] and beyond[48–51] optics, where nonlinearity can be effectively harnessed and utilized. Our findings could also spark new research directions in fields such as optical data compression for imaging[52–54] and sensing[12,55], optical communication[56] and quantum computing[57], where innovative nonlinear mechanisms can substantially enhance performance, efficiency and potential opportunity in enhancing data privacy and adversarial robustness[28,58,59].

During the final stage of this work, we became aware of 2 independent works of very different optical machine learning implementations that are based on the same principle of realizing nonlinear processing with linear optics[60,61].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41566-024-01493-0.

References

1. Prucnal, P. R. & Shastri, B. J. *Neuromorphic Photonics* (CRC Press, 2017).
2. Kues, M. On-chip generation of high-dimensional entangled quantum states and their coherent control. *Nature* **546**, 622–626 (2017).
3. Xu, X. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
4. Wetzstein, G. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
5. Shastri, B. J. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* **15**, 102–114 (2021).
6. Shen, Y. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
7. Rotter, S. & Gigan, S. Light fields in complex media: mesoscopic scattering meets wave control. *Rev. Mod. Phys.* **89**, 015005 (2017).
8. Chang, J., Sitzmann, V., Dun, X., Heidrich, W. & Wetzstein, G. Hybrid optical–electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* **8**, 12324 (2018).
9. Hughes, T. W., Minkov, M., Shi, Y. & Fan, S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* **5**, 864–871 (2018).
10. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
11. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
12. Wang, T. Image sensing with multilayer nonlinear optical neural networks. *Nat. Photon.* **17**, 408–415 (2023).
13. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Proc. Advances in Neural Information Processing Systems* Vol. 25 (eds Pereira, F. et al.) (Curran Associates, 2012).
14. Lin, X. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).

15. Tait, A. N. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
16. Miller, D. A. B. Are optical transistors the logical next step? *Nat. Photon.* **9**, 10–13 (2015).
17. Wang, M. M., Pagani, M. & Eggleton, B. J. A chip-integrated coherent photonic-phononic memory. *Nat. Commun.* **9**, 574 (2018).
18. Teğin, U., Yıldırım, M., Oğuz, İ., Moser, C. & Psaltis, D. Scalable optical learning operator. *Nat. Comput. Sci.* **1**, 542–549 (2021).
19. Williamson, I. A. D. Reprogrammable electro-optic nonlinear activation functions for optical neural networks. *IEEE J. Sel. Topics Quantum Electron.* **26**, 7700412 (2019).
20. Li, G. H. Y. All-optical ultrafast ReLU function for energy-efficient nanophotonic deep learning. *Nanophotonics* **12**, 847–855 (2022).
21. Zhou, T., Scalzo, F. & Jalali, B. Nonlinear Schrödinger kernel for hardware acceleration of machine learning. *J. Lightwave Technol.* **40**, 1308–1319 (2022).
22. Shirdel, M. & Mansouri-Birjandi, M. A. Photonic crystal all-optical switch based on a nonlinear cavity. *Optik* **127**, 3955–3958 (2016).
23. Eliezer, Y., Ruhrmair, U., Wisiol, N., Bittner, S. & Cao, H. Tunable nonlinear optical mapping in a multiple-scattering cavity. *Proc. Natl Acad. Sci. USA* **120**, e2305027120 (2023).
24. Saade, A. et al. Random projections through multiple optical scattering: approximating kernels at the speed of light. In *Proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 6215–6219 (IEEE, 2016).
25. Rafayelyan, M., Dong, J., Tan, Y., Krzakala, F. & Gigan, S. Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction. *Phys. Rev. X* **10**, 041037 (2020).
26. Dong, J., Rafayelyan, M., Krzakala, F. & Gigan, S. Optical reservoir computing using multiple light scattering for chaotic systems prediction. *IEEE J. Sel. Topics Quantum Electron.* **26**, 7701012 (2019).
27. Brossollet, C. et al. LightOn optical processing unit: scaling-up AI and HPC with a non von Neumann co-processor. In *Proc. 2021 IEEE Hot Chips 33 Symposium (HCS)* 1–11 (IEEE, 2021).
28. Ohana, R. Photonic differential privacy with direct feedback alignment. *Adv. Neural Inf. Process. Syst.* **34**, 22010–22020 (2021).
29. Agrawal, G. P. in *Nonlinear Science at the Dawn of the 21st Century* 195–211 (Springer, 2000).
30. Boyd, R. W., Gaeta, A. L. & Giese, E. in *Springer Handbook of Atomic, Molecular, and Optical Physics* 1097–1110 (Springer, 2008).
31. Wang, J. Thermo-optic effects in on-chip lithium niobate microdisk resonators. *Opt. Express* **24**, 21869–21879 (2016).
32. Ryou, A. Free-space optical neural network based on thermal atomic nonlinearity. *Photon. Res.* **9**, B128–B134 (2021).
33. Ohtsubo, J. *Semiconductor Lasers: Stability, Instability and Chaos* 2nd edn, Vol. 111 (SSOS, 2013).
34. Xiao, H., Rasul, K. & Vollgraf, R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. Preprint at https://arxiv.org/abs/1708.07747 (2017).
35. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (Wiley, 2012).
36. Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
37. Peng, H., Long, F. & Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1226–1238 (2005).
38. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
39. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E* **69**, 066138 (2004).
40. Petterson, J. & Cukierski, W. Facial keypoints detection. *Kaggle* https://kaggle.com/competitions/facial-keypoints-detection (2013).
41. Dollár, P., Wojek, C., Schiele, B. & Perona, P. Pedestrian detection: a benchmark. In *Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition* 304–311 (IEEE, 2009).
42. Skalli, A. Photonic neuromorphic computing using vertical cavity semiconductor lasers. *Opt. Mater. Express* **12**, 2395–2414 (2022).
43. Van der Sande, G., Brunner, D. & Soriano, M. C. Advances in photonic reservoir computing. *Nanophotonics* **6**, 561–576 (2017).
44. Boikov, I. K., Brunner, D. & De Rossi, A. Evanescent coupling of nonlinear integrated cavities for all-optical reservoir computing. *New J. Phys.* **25**, 093056 (2023).
45. Porte, X. A complete, parallel and autonomous photonic neural network in a semiconductor multimode laser. *J. Phys. Photonics* **3**, 024017 (2021).
46. Wright, L. G. Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022).
47. Chen, Z. et al. Deep learning with coherent VCSEL neural networks. *Nat. Photon.* **17**, 723–730 (2023).
48. Momeni, A., Guo, X., Lissek, H. & Fleury, R. Physics-inspired neuroacoustic computing based on tunable nonlinear multiple-scattering. Preprint at https://arxiv.org/abs/2304.08380 (2023).
49. del Hougne, P. & Lerosey, G. Leveraging chaos for wave-based analog computation: demonstration with indoor wireless communication signals. *Phys. Rev. X* **8**, 041037 (2018).
50. Momeni, A., Rahmani, B., Malléjac, M., Del Hougne, P. & Fleury, R. Backpropagation-free training of deep physical neural networks. *Science* **382**, 1297–1303 (2023).
51. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Volume 1 (Long and Short Papers)* 4171–4186 (Association for Computational Linguistics, 2019).
52. Chen, C. L., Mahjoubfar, A. & Jalali, B. Optical data compression in time stretch imaging. *PLoS ONE* **10**, e0125106 (2015).
53. Weng, X., Feng, J., Perry, A. & Vuong, L. T. Non-line-of-sight full-Stokes polarimetric imaging with solution-processed metagratings and shallow neural networks. *ACS Photonics* **10**, 2570–2579 (2023).
54. Li, J. Spectrally encoded single-pixel machine vision using diffractive networks. *Sci. Adv.* **7**, eabd7690 (2021).
55. Muminov, B. & Vuong, L. T. Fourier optical preprocessing in lieu of deep learning. *Optica* **7**, 1079–1088 (2020).
56. Wu, B., Shastri, B. J. & Prucnal, P. R. in *Emerging Trends in ICT Security* 173–183 (Elsevier, 2014).
57. Venkataraman, V., Saha, K. & Gaeta, A. L. Phase modulation at the few-photon level for weak-nonlinearity-based quantum computing. *Nat. Photon.* **7**, 138–141 (2013).
58. Bezzam, E., Vetterli, M. & Simeoni, M. Privacy-enhancing optical embeddings for lensless classification. Preprint at https://arxiv.org/abs/2211.12864 (2022).
59. Cappelli, A. et al. Adversarial robustness by design through analog computing and synthetic gradients. In *Proc. ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 3493–3497 (IEEE, 2022).
60. Yildirim, M., Dinc, N. U., Oguz, I., Psaltis, D. & Moser, C. Nonlinear processing with linear optics. *Nat. Photon.* **18** (2024).
61. Wanjura, C.C. & Marquardt, F. Fully non-linear neuromorphic computing with linear wave scattering. Preprint at https://arxiv.org/abs/2308.16181 (2023).
62. LeCun, Y. The MNIST database of handwritten digits. *Yann Lecun* http://yann.lecun.com/exdb/mnist/ (1998).

Methods

Further explanation of Born series

To better connect the Born series with a neural network, such as an MLP, the Born series can also be rewritten as

$$E_{\text{out}} = \mathbf{S}^n E_{\text{in}} = \mathbf{S}_u(\mathbf{S}_{u-1}(\ldots(\mathbf{S}_1(\mathbf{V}))\ldots))E_{\text{in}}, \quad (2)$$

where $\mathbf{S}_n(\cdot)$ represents a scattering operator and $\mathbf{S}_1(\mathbf{V}) = \mathbf{V}$, for $n > 1$, $\mathbf{S}_{n+1}(\mathbf{V}) = \mathbf{V} + \mathbf{S}_n(\mathbf{V})\mathbf{G}_0\mathbf{V}$. The iterative expression of the Born series involving the scattering operator $\mathbf{S}$ can be seen to structurally resemble the iterative operation in an MLP, where data are transformed across multiple layers in an iterative way. However, $\mathbf{V}$ and $\mathbf{G}_0$ are identical in all the layers.

Setup information

The multiple-scattering cavity is an integrating sphere with a rough inner surface and a diameter of 3.75 cm. The cavity has three ports on its boundary: one port is attached to a DMD (Texas Instruments DLP9000X), which provides a reconfigurable scattering potential, whereas the other two ports serve as the input and output ports of the cavity. A single-frequency continuous-wave laser (Agilent 81940A; wavelength, 1,550 nm) at 21.3 mW is coupled through a single-mode fibre into the cavity through the input port. On multiple scatterings at the rough inner surface, the output light escapes the cavity via the output port. From the spectral correlation width of the output speckle pattern, the average path length of light inside the cavity is estimated to be approximately 100 m. The average number of bounces off the cavity boundary is on the order of 5,000. To capture the output intensity pattern, a mirror is positioned adjacent to the output port, directing the output light towards an InGaAs camera (Xenics Xeva FPA-640). A linear polarizer is placed in front of the camera to record the speckle intensity patterns, which represent a complex nonlinear relationship between the configuration of the DMD and the resulting output speckle.

Experimental procedure of computing with the multiple-scattering cavity

Experimentally, we couple a continuous-wave single-frequency laser through a single-mode fibre into the cavity. An input image is loaded onto the DMD, which modifies the scattering potential in a reconfigurable manner. The entire modulation area of the DMD consists of 2,560 × 1,600 micromirrors with a pixel pitch of 7.6 μm. Each micromirror can be tilted by +15° or −15°, representing the binary states +1 and −1, respectively. The input port of the cavity covers a portion of the DMD area (1,260 × 784 micromirrors), which is exposed to light in the cavity; thus, we only modulate the micromirrors within this region. Instead of controlling individual micromirrors, we group micromirrors into macropixels, where all the micromirrors in a single macropixel have the same tilt angle (binary state). To make images compatible for loading onto the DMD, we employed a binary thresholding method using the Floyd–Steinberg dithering algorithm[63].

We control the order of nonlinear mapping in our cavity in two ways, both involving changing the number of scattering events on the modulated area of the DMD. First, we reduce the DMD area where micromirros are toggled. Outside the modulated area, the micromirror configuration remains fixed. The number of bounces of light from a smaller modulated area is lower. By shrinking the dimension of macropixels by a factor of 4 or 16, the total modulated area is reduced by the same factor. Alternatively, we can enhance the number of bounces with the DMD by increasing the dwell time of light inside the cavity. This is realized by covering the output port of the cavity with a partial reflector—a silicon wafer (thickness, 0.63 mm), which partially reflects light at 1,550 nm. As a result, the order of nonlinear random mapping increases.

The temporal coherent length of light exceeds the typical optical path length inside the cavity. The output light maintains high spatial coherence, resulting in a relatively high intensity contrast (~0.8) of the output speckle pattern (after passing through a linear polarizer). Compared with the nonlinearity introduced by optical effects such as harmonic generation and self-phase modulation, a broadband pulsed laser is necessary to achieve the high pulse energy required for these nonlinear processes, producing much lower contrast. In addition, our system's nonlinear response is insensitive to optical power, more stable and more energy efficient. The output images recorded by our camera consistently display stable speckle patterns, with each speckle grain representing a distinct spatial mode. The number of output modes (speckle grains) in the camera image is determined by dividing the total area of the speckle grains used for computation by the average size of one speckle grain, which is derived from the full-width at half-maximum of the intensity correlation function.

Fashion MNIST classification task

In the Fashion MNIST[34] classification task, we study the impact of the number of modes and the size of the modulated area on the DMD in terms of classification accuracy and mutual information between the output modes and ground-truth target classes. To vary the number of modes, we crop the output camera image, controlling the number of output modes. This is achieved in PyTorch using the nn.transform.CenterCrop function. We manipulate the modulated area on the DMD by adjusting the macropixel size for the input data. For example, for the Fashion MNIST dataset, we use a 45 × 28 micropixels for each macropixel on the DMD, corresponding to one of the 28 × 28 Fashion MNIST image when we utilize the full modulated area. For a 25% modulated area, we use 22 × 14 micropixels for one macropixel on the DMD. The entire set of 60,000 training data and 10,000 testing data are sequentially input on the DMD, and the corresponding camera speckle images are collected. We further applied a filter based on system stability (Supplementary Section 2) to select images with a speckle stability over the threshold of 0.96. The data are then split in a 9:1 ratio to form training and testing datasets for classification. For classification, we employ ridge regressor from the keras package to train and infer with the output modes. Regarding the calculation of mutual information, detailed information on the algorithm is provided in Supplementary Section 4. We use the mutual_info_regression function, which takes vectors of pixel values in output modes and class labels, from the feature selection module in scikit-learn.

Programmable optical and digital parameters

The maximum number of programmable optical parameters is given by the number of mirrors of the DMD, which is approximately 4 million. The count of digitally programmable parameters, however, depends on the decoder utilized. Specifically, for the Fashion MNIST classification task, the linear classifier requires only 1,000 parameters. In the task of Fashion MNIST reconstruction, the parameter count increases to approximately 90,000. For human face detection, the requirement is around 310,000 parameters, and for pedestrian tracking, the model uses around 1 million programmable parameters.

Training of digital decoder

For the tasks beyond classification, we start with low-dimensional vectors derived from the deep optical encoder—the multiple-scattering cavity. Using these vectors, we train a digital decoder based on a neural network, with the objective of minimizing the mean squared loss relative to the ground-truth target values in our training dataset. The dimensions for each target differed according to the tasks: 28 × 28 for Fashion MNIST image reconstruction, 15 sets of keypoints for human face keypoint detection and four bounding box coordinates for pedestrian detection. The architecture selected for the decoding neural network is an MLP, which incorporates batch normalization before each activation function. The ideal depths and widths of the hidden layers are determined by conducting a neural architecture search,

randomly initialized at least 100 times to select the best architecture for the digital decoder. The same activation function, chosen among relu, tanh and sigmoid functions, is used during each search. All training instances are conducted on the NVIDIA A100 Tensor Core GPU via Google Colab.

Fashion MNIST image reconstruction task

In the Fashion MNIST[34] image reconstruction task, we train an MLP as a digital decoder using pairs of 16 output modes (inputs) and ground-truth Fashion MNIST images (targets) to reconstruct the original images from the speckle patterns. To optimize the decoder's architecture, we employ neural architecture search, varying both depth and width to identify the best architecture for the decoder. We primarily study and compare two cases. Case 1, speckle features generated from a linear random projection through complex media, followed by quadratic detection on the optical field (to generate linear optical features, we follow the methods described elsewhere[24]); case 2, speckle features generated from nonlinear random mapping via a multiple-scattering cavity, again followed by quadratic detection on the optical field. In both scenarios, we ensure that the number of modes remains consistent, making the reconstruction quality comparable between the two cases. For the first case, the optimized decoder comprises a two-layer MLP. For the second case, the optimized decoder utilizes a four-layer MLP, both with the same activation sigmoid function. We further evaluate the reconstruction using a test dataset.

Human face keypoint detection task

The human face keypoint detection dataset at Kaggle[40] consists of facial keypoints, each characterized by a real-valued pair (x, y) indicating its position in the domain of pixel indices. This dataset identifies 15 specific keypoints corresponding to facial features, including centres of the left and right eyes, inner and outer corners of both eyes, inner and outer ends of both eyebrows, the tip of the nose, corners of the mouth on both sides, and the top and bottom centres of the lips. It is noteworthy to mention that the terms 'left' and 'right' are based on the subject's point of view. Some data points might not have all the keypoints, which are represented as missing entries in the dataset. Each image in the dataset contains a list of pixels, with values ranging from 0 to 255, formatted for a resolution of 96 × 96 pixels. The training set includes 7,049 images. Each row in this file provides the (x, y) coordinates of the 15 keypoints and image data in a row-ordered list of pixels. Conversely, the test set comprises 1,783 images: each row lists an ImageId and the corresponding row-ordered list of pixels for the image.

For data preprocessing, entries without keypoint information are excluded. In cases where an image had fewer than 15 keypoints, we duplicated some keypoints to ensure that all the labels consisted of 15 target points. This procedure ensures a consistent size of the MLP output layer.

Following this, the data are sent into a multiple-scattering cavity, with different modulated areas, corresponding to variable nonlinearity strengths, reminiscent of deep neural networks encoding (Supplementary Section 3). Only 16 output modes (using nn.transform.CenterCrop) are extracted from this system. Using these modes, a digital decoder is developed based on neural architecture search, aiming to train on and infer the 15 facial keypoints.

Our analysis mainly compared two scenarios: one with a modulated area of ~6.25% and another termed '100% + enhanced', which is the full modulated area bolstered by an extra partial reflector for enhanced scattering (Supplementary Section 1). Our findings indicated that even with a decoder trained to its optimal capacity, the 100% + enhanced setup yielded better results.

Pedestrian detection task

In this task, we use the Caltech Pedestrian dataset[41], one of the pioneering collections in the domain of computer vision, specifically designed for pedestrian detection tasks. This dataset has played an instrumental role in shaping the research trajectories in pedestrian detection, serving as a benchmark for numerous detection algorithms over the years. The dataset offers a wide variety of real-world scenarios captured from urban settings, including pedestrians in various poses, occlusions and varying light conditions. It provides an invaluable resource for the development and evaluation of algorithms, with its meticulous annotations and diverse challenges it poses.

Within this dataset, bounding boxes are utilized to accurately locate individual pedestrians in frames. These boxes are characterized by a set of four real-valued positions: (x_1, y_1) for the top-left corner and (x_2, y_2) for the bottom-right corner. Given the dynamic nature of urban environments, a single frame can contain multiple pedestrians, which results in multiple bounding boxes within that image.

To preprocess the dataset, we adopt a simplification strategy. Regardless of the number of bounding boxes present in the original image, we ensure that only one bounding box is retained per image. For images that contain multiple bounding boxes, only the first bounding box is selected and used as a label. In the cases where an image lacks a bounding box, it is removed from the dataset. All the images from the Caltech dataset inherently possess a resolution of 640 × 480 pixels. In our case, we downsampled the images to 320 × 240 pixels. Our curated version of the dataset, divided into training and test segments, encapsulates a total of 10,000 images.

Following the preprocessing steps, images are then introduced into a multiple-scattering cavity, with the full modulated area being enhanced by the partial reflector—the silicon wafer. From this system, a total of 25 output modes (using nn.transform.CenterCrop) are derived. Harnessing these modes, we engineered a digital decoder rooted in the principles of neural architecture search. The overarching objective of this decoder is to train and subsequently infer the solitary bounding box in the images.

We also generated Supplementary Videos 1 and 2 using the test dataset from various video locations. In these movies, the frame rate was reduced from the actual 30 fps to 9 fps for visualization purposes. The green boxes indicate the ground-truth bounding boxes, whereas the magenta boxes represent the inferences. The actual inference time is well under 0.1 s.

Data availability

Partial example data pertaining to this study are available via GitHub at https://github.com/comediaLKB/learning_with_passive_optical_nonlinear_mapping. Additional data are available from the corresponding authors upon reasonable request. The training datasets used are publicly available via Fashion MNIST[34], the Human Face Keypoint Kaggle dataset[40] and the Caltech Pedestrian dataset[41].

Code availability

Code is available via GitHub at https://github.com/comediaLKB/learning_with_passive_optical_nonlinear_mapping.

References

63. Floyd, R. W. & Steinberg, L. An adaptive algorithm for spatial greyscale. In *Proc. Society for Information Display* 36–37 (1976).

Acknowledgements

The research conducted at Yale University is partly funded by the US National Science Foundation (NSF) under grant no. ECCS-1953959 and the US Air Force Office of Scientific Research (AFOSR) under grant no. FA9550-21-1-0039. H.C. acknowledges discussions with U. Rührmair and M. Lachner. S.G. and F.X. acknowledge funding from the Chan Zuckerberg Initiative (2020-225346). F.X. acknowledges funding from the Optica Foundation Challenge Award 2023. We acknowledge P. McMahon, L. Wright, T. Wang, J. Hu, J. Dong and Z. Nie for feedback on the initial paper. We also acknowledge

J. Hu for carefully proofreading the paper and B. Loureiro for insightful discussions.

Author contributions

F.X., S.G. and H.C. designed the research. F.X. further developed the concept and performed the research and prepared the paper with input from all authors. F.X. and S.H. performed the analysis. Y.E. built the experimental setup. K.K., Y.E., S.H. and L.S. collected the experimental data. All authors discussed the results. K.K., Y.E., S.G. and H.C. edited the paper. S.G. and H.C. initiated and supervised the project.

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/s41566-024-01493-0.

Correspondence and requests for materials should be addressed to Fei Xia, Sylvain Gigan or Hui Cao.

Peer review information *Nature Photonics* thanks Gordon Wetzstein and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.