Perspective

Inference in artificial intelligence with deep optics and photonics

https://doi.org/10.1038/s41586-020-2973-6

Received: 28 November 2019

Accepted: 20 August 2020

Published online: 2 December 2020

Check for updates

Gordon Wetzstein[1✉], Aydogan Ozcan[2], Sylvain Gigan[3], Shanhui Fan[1], Dirk Englund[4], Marin Soljačić[4], Cornelia Denz[5], David A. B. Miller[1] & Demetri Psaltis[6]

Artificial intelligence tasks across numerous applications require accelerators for fast and low-power execution. Optical computing systems may be able to meet these domain-specific needs but, despite half a century of research, general-purpose optical computing systems have yet to mature into a practical technology. Artificial intelligence inference, however, especially for visual computing applications, may offer opportunities for inference based on optical and photonic systems. In this Perspective, we review recent work on optical computing for artificial intelligence applications and discuss its promise and challenges.

The capacity of computing systems is in an arms race with the massively growing amount of visual data they seek to understand. In a range of applications—including autonomous driving, robotic vision, smart homes, remote sensing, microscopy, surveillance, defence and the Internet of Things—computational imaging systems record and process unprecedented amounts of data that are not seen by a human but instead are interpreted by algorithms built on artificial intelligence (AI).

Across these applications, deep neural networks (DNNs) are rapidly becoming the standard algorithmic approach for visual data processing[1–3]. This is primarily because DNNs achieve state-of-the-art results across the board—often by a large margin. Recent breakthroughs in deep learning have been fuelled by the immense processing power and parallelism of modern graphics processing units (GPUs) and the availability of massive visual datasets that enable DNNs to be efficiently trained using supervised machine learning strategies.

However, high-end GPUs and other accelerators running increasingly complex neural networks are hungry for power and bandwidth; they require substantial processing times and bulky form factors. These constraints make it challenging to adopt DNNs in edge devices, such as cameras, autonomous vehicles, robots or Internet of Things peripherals. Consider vision systems in autonomous cars, which have to make robust decisions instantaneously using limited computational resources. When driving at high speed, split-second decisions can decide between life or death. Indeed, virtually all edge devices would benefit from leaner computational imaging systems, offering lower latency and improvements in size, weight and power.

The computing requirements of the two stages of a DNN—training and inference—are very different. During the training stage, the DNN is fed massive amounts of labelled examples and, using iterative methods, its parameters are optimized for a specific task. Once trained, the DNN is used for inference where some input data, such as an image, is sent through the network once, in a feedforward pass, to compute the desired result. GPUs are used for inference in some applications, but for many edge devices this is impractical, owing to the aforementioned reasons.

Despite the flexibility of electronic AI accelerators, optical neural networks (ONNs) and photonic circuits could represent a paradigm shift in this and other machine learning applications. Optical computing systems promise massive parallelism[4] in conjunction with small device form factors and, in some implementations, little to no power consumption[5]. Indeed, optical interconnects that use light to achieve communications in computing systems are already widely used in data centres today, and the increasing use of optical interconnects deeper inside computing systems is probably essential for continued scaling. Unlike electrical interconnect technologies, optical interconnects offer the potential for orders of magnitude improvements in bandwidth density and in energy per bit in communications as we move to deeper integration of optics, optoelectronics and electronics[4,6–8]. Such improved interconnects could allow hybrid electronic–optical DNNs, and the same low-energy, highly parallel[4] integrated technologies could be used as part of analogue optical processors.

General-purpose optical computing has yet to mature into a practical technology despite the enormous potential of optical computers and about half a century of focused research efforts[9,10]. However, inference tasks—especially for visual computing applications—are well suited for implementation with all-optical or hybrid optical–electronic systems. For example, linear optical elements can calculate convolutions, Fourier transforms, random projections and many other operations ‘for free’—that is, as a byproduct of light–matter interaction or light propagation[11–15]. These operations are the fundamental building blocks of the DNN architectures that drive most modern visual computing algorithms. The possibility of executing these operations at the speed of light, potentially with little to no power requirements, holds transformative potential that we survey in this Perspective.

Historical overview of optical computing

Research into neuromorphic computing was intense in the 1980s (Fig. 1). Following early pioneering work[16–21], Rumelhart, Hinton and Williams published a deeply influential paper in 1986 describing the error-backpropagation method for training multi-layer networks[22].

[1]Stanford University, Stanford, CA, USA. [2]University of California, Los Angeles, Los Angeles, CA, USA. [3]Laboratoire Kastler Brossel, Sorbonne Université, École Normale Supérieure, Collège de France, CNRS UMR 8552, Paris, France. [4]Massachusetts Institute of Technology, Cambridge, MA, USA. [5]University of Münster, Münster, Germany. [6]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. ✉e-mail: gordon.wetzstein@stanford.edu

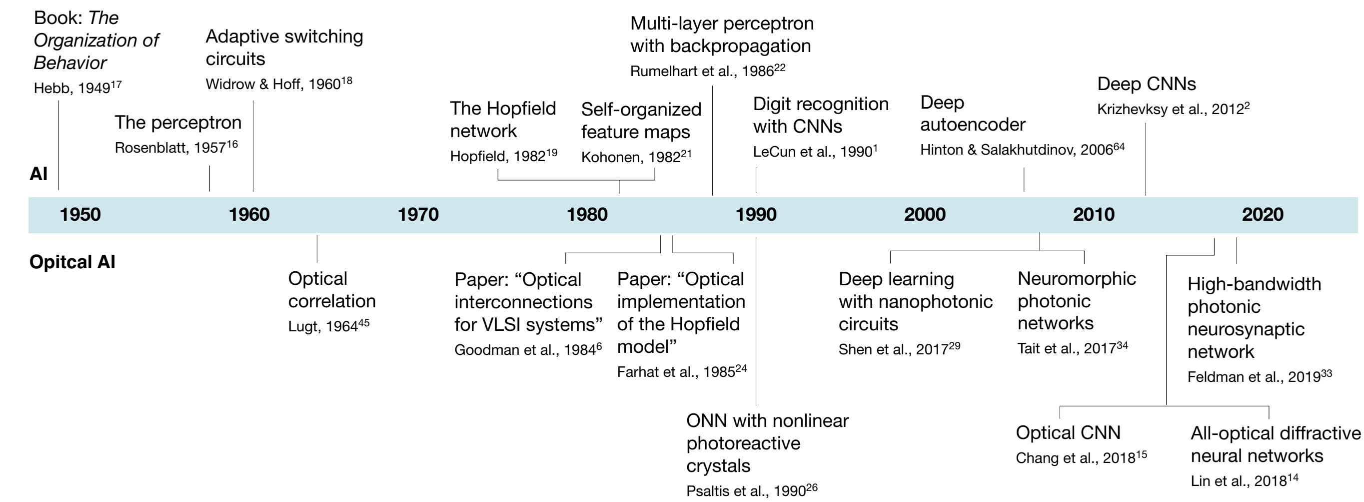

Fig. 1 | Timeline of artificial intelligence and related optical and photonic implementations. Selected key milestones and publications are shown, with an emphasis on recent developments[1,2,6,14–19,21,22,24,26,29,33,34,45,64].

Analogue implementations of neural networks emerged as a promising approach for dealing with the high computational load in training and reading large neural networks[23]. Several analogue very-large-scale integration circuit implementations were demonstrated and, in parallel, analogue optical realizations were pursued. The first optical neural network was a modest demonstration of a fully connected network of 32 neurons with feedback[24]. This demonstration triggered interesting new research in optical neural networks, reviewed by Denz[25]. The next major step for optical neural networks was the introduction of dynamic nonlinear crystals for the implementation of adaptive connections between optoelectronic neurons arranged in planes[26]. In addition to their dynamic nature, nonlinear crystals are inherently three-dimensional (3D) devices, and they allow the storage of a much larger number of weights. In an ambitious demonstration published in 1993, for example, an optical two-layer network was trained to recognize faces with very good accuracy by storing approximately 1 billion weights in a single photorefractive crystal[27].

Despite promising demonstrations of analogue hardware implementations, interest in custom optical hardware waned in the 1990s. There were three main reasons for this: (1) the advantages (power and speed) of the analogue accelerators are useful only for very large networks; (2) the technology for the optoelectronic implementation of the nonlinear activation function was immature; and (3) the difficulty in controlling analogue weights made it difficult to reliably control large optical networks.

The situation has changed in the intervening years. DNNs have emerged as one of the dominant algorithmic approaches for many applications. Moreover, major improvements in optoelectronics and silicon photonics, in particular coupled with the emergence of extremely large networks, have led many researchers to revisit the idea of implementing neural networks optically.

Photonic circuits for artificial intelligence

Modern DNN architectures are cascades of linear layers followed by nonlinear activation functions that repeat many times over. The most general type of linear layer is fully connected, which means that each output neuron is a weighted sum of all input neurons—a multiply–accumulate (MAC) operation. This is mathematically represented as a matrix–vector multiplication, which can be efficiently implemented in the optical domain. One specific change that has occurred since earlier optical computing work is the understanding that meshes of Mach–Zehnder interferometers (MZIs) in specific architectures (for example, those based on singular value matrix decomposition) can implement arbitrary matrix multiplication without fundamental loss; these architectures are also easily configured and controlled[28].

Specifically, recent silicon photonic neuromorphic circuits have demonstrated such singular value matrix decomposition implementations of matrix–vector products using coherent light[29]. In this case, MZIs fabricated on a silicon chip implement the element-wise multiplications. This design represents a truly parallel implementation of one of the most crucial building blocks of neural networks using light, and modern foundries could easily mass-fabricate this type of photonic system.

One of the challenges of such a design is that the number of MZIs grows as N^2 with the number of elements N in the vector, a necessary consequence of implementing an arbitrary matrix. As the size of the photonic circuits grows, losses, noise and imperfections also become larger issues. As a result, it becomes increasingly difficult to construct a sufficiently accurate model to train it on a computer. Approaches to overcoming this difficulty include designing the circuit with robustness to imperfections[30], automatically 'perfecting' the circuit[31], or training the photonic neuromorphic circuit in situ[32].

As an alternative to MZI-based MACs, Feldmann et al. recently introduced an all-optical neurosynaptic network based on phase-change materials (PCM)[33]. In this design, PCM cells implement the weighting of the linear layer and a PCM cell coupled with a ring resonator implements a nonlinear activation function akin to a rectified linear unit (ReLU). Micro-ring weight banks were also used by Tait et al.[34] to implement a recurrent silicon photonic neural network.

Incorporating all-optical nonlinearities into photonic circuits is one of the key requirements for truly deep photonic networks. Yet, the challenge of efficiently implementing photonic nonlinear activation functions at low optical signal intensities was one of the primary reasons that interest in ONNs waned in the 1990s. Creative approaches from the last decade, such as nonlinear thresholders based on all-optical micro-ring resonators[35], saturable absorbers[29,36], electro-absorption modulators[37], or hybrid electro-optical approaches[38], represent possible solutions for overcoming this challenge in the near future. Earlier 'self-electrooptic-effect' device concepts[39] may also offer hybrid solutions, especially with recent advances towards foundry-enabled mass fabrication[40] of silicon-compatible versions of the energy-efficient quantum-confined Stark effect electro-absorption modulators on which they are based.

Comprehensive reviews of neuromorphic photonics[41] and photonic MACs for neural networks[42] were recently published. In one review[42], the authors provide a detailed comparison of photonic linear computing systems and their electronic counterparts, taking metrics such as energy, speed and computational density into account. The primary insight of this study was that photonic circuits exhibit advantages over electronic implementations in all of these metrics when considering large processor sizes, large vector sizes and low-precision operations. However, the authors also point to the long-standing challenge of the high energy cost of electro–optical conversion, which is now rapidly approaching that of electronic links[42].

Photonic circuits could become fundamental building blocks of future AI systems. Although much progress has been made in the last 20 years, major challenges still lie ahead. Electronic computing platforms today offer programmability, mature and high-yield mass-fabrication technology, opportunities for 3D implementation, built-in signal restoration and gain, and robust memory solutions. Moreover, modern digital electronic systems offer high precision, which cannot be easily matched by analogue photonic systems. Yet, AI systems often do not require high precision[43], especially when used for inference tasks. Although programmability has traditionally been more difficult with photonic systems, first steps towards simplifying the process have recently been demonstrated[44].

Overall, the capabilities of photonic circuits have increased considerably over the last decade, and we have seen progress on some of the most crucial challenges that have hindered their utility in the past. Yet, to compete with their electronic counterparts, photonic computing systems still face fundamental engineering challenges. One direction that seems particularly well suited for optical and photonic processing is optical inference with incoherent light to rapidly process scene information under ambient lighting conditions. Such an approach presents many exciting opportunities for autonomous vehicles, robotics and computer vision, which we discuss next.

Computing with free space, lenses and complex media

An alternative to photonic circuits is to build computing capabilities directly on top of an optical field propagating through free space or some medium (see Fig. 2). Mathematically, wave propagation in free space is described by Kirchhoff's diffraction integral, which amounts to a convolution of the field with a fixed kernel[4,11]. This operation represents one of the basic building blocks of convolutional neural networks (CNNs)—the neural network architecture of choice for most visual computing applications. However, to make wave propagation a useful tool for optical computing, we require programmability—for example we should be able to implement a designed convolution kernel. This can be achieved with Fourier optics, wherein a specific lens configuration of an optical setup applies a physical forward or inverse Fourier transform to the wave field. An optical element inserted into the Fourier plane of the setup implements an element-wise multiplication of the input field with the amplitude and phase of the optical element. Via the convolution theorem, this corresponds to a convolution of the input field with the inverse Fourier transform of the element. Thus, we can convolve an image's wave field with an arbitrary convolution kernel at the speed of light using lenses and other optical elements.

This insight has been used to design optical correlators in the past. These devices implement a single convolution that performs template matching directly on incoherent optical images, for example for target detection and tracking[45–49]. Although this idea represents a valuable step towards optical implementations of the convolutional blocks of modern CNNs, convolutions with a single kernel are very restrictive. A CNN typically performs convolutions with many kernels simultaneously in each of its layers. To address this discrepancy, the classical Fourier optics setup can be adjusted to implement parallel convolutions in optics and mimic the functionality of a single CNN block[15] (see Fig. 2). Thus, recent years have seen rapid progress in enabling optical computing capabilities that closely match modern CNNs. However, some of the remaining challenges of the Fourier optics approach include the difficulty of implementing optical nonlinear activation functions and the large device form factor relative to photonic circuits.

The former challenge can be addressed using a hybrid optical–electronic computing approach, again making the efficiency of the electro–optical conversion process the primary bottleneck, although the potential for highly integrated energy-efficient optoelectronics could address such efficient conversion[8].

To achieve a more compact device form factor than classical Fourier optics setups, other wave–matter interactions can be leveraged for optical computing. For example, scattering layers can be used in place of lenses. With such layers, several optimized phase or amplitude patterns can be mounted with some distance between them to implement an all-optical classification algorithm[14]. Interestingly, more complicated shapes of optimized inhomogeneous media can be used to implement recurrent neural networks, for example for vowel classification[50]. However, this is not the only configuration in which we can take advantage of scattering media.

The propagation of light in a dense, complex medium is in many cases akin to mixing the input field with a random matrix. This represents an interesting computational operation, and has been shown to be nearly ideal for compressive sensing. In this application each output pixel is a random projection of the input[12], much like the single-pixel camera paradigm[51]. This approach also retains a lot of information, allowing retrieval of some functional signals at depth, without imaging, which could be particularly interesting for neuroscience[52]. The approach is also amenable to training a neural network, as demonstrated for imaging through a multimode fibre[53,54] or imaging through thin[55] or thick[56] scattering media. Moreover, it was found that the complex medium itself could be seen as an optical realization of a neural network: the connection weights are the coefficients of the random matrix and the nonlinearity is the conversion to intensity during detection by the camera, allowing direct classification without imaging[13,57]. This mathematical reformulation of light propagation could open very interesting avenues of research in optical computing, particularly in any computing problem where large-scale, random-matrix multiplication is used[58], including in reservoir computing[59], phase retrieval[60] and computational imaging.

Inference with deep computational optics and imaging

Computational imaging is a field focused on the co-design of optics and image processing, for example to enhance the capabilities of computational cameras. Despite performing many different tasks, cameras today are designed to mimic the human eye. They capture a two-dimensional (2D) projection of a 3D environment, typically with three colour channels. However, the eyes of other animals have evolved in very different ways, each perfectly adapted to their environment. For example, certain species of mantis shrimp are reported to have photoreceptors that not only are sensitive to the polarization state of light, but also contain up to 12 different spectral bands, features that are fitting for their spectrally vibrant coral reef environment[61,62]. Cameras could thus be adapted to unique environments or optimized for specific tasks, as are animal eyes.

One of the challenges of using conventional sensors to capture the world as the mantis shrimp sees it is that they integrate visual data over various dimensions. A conventional 2D sensor integrates information over the incident plenoptic function[63]—that is, over the wavelength spectrum, over the incident angle and scene depth, over a certain time window, and it is also limited in its dynamic range. Therefore, we can interpret existing sensors as a bottleneck, preventing some visual information from being captured. Optical engineers have the freedom to design camera lenses with specific point spread functions (PSFs), to design the spectral sensitivities of sensor pixels using spectrally selective optical filters, or to choose other properties. Yet, the challenge for developing application-specific imaging systems is how to best design such devices and make use of these engineering capabilities.

In this context, it is helpful to interpret a camera as an encoder–decoder system[64]. The lens or lenses optically encode a scene on the

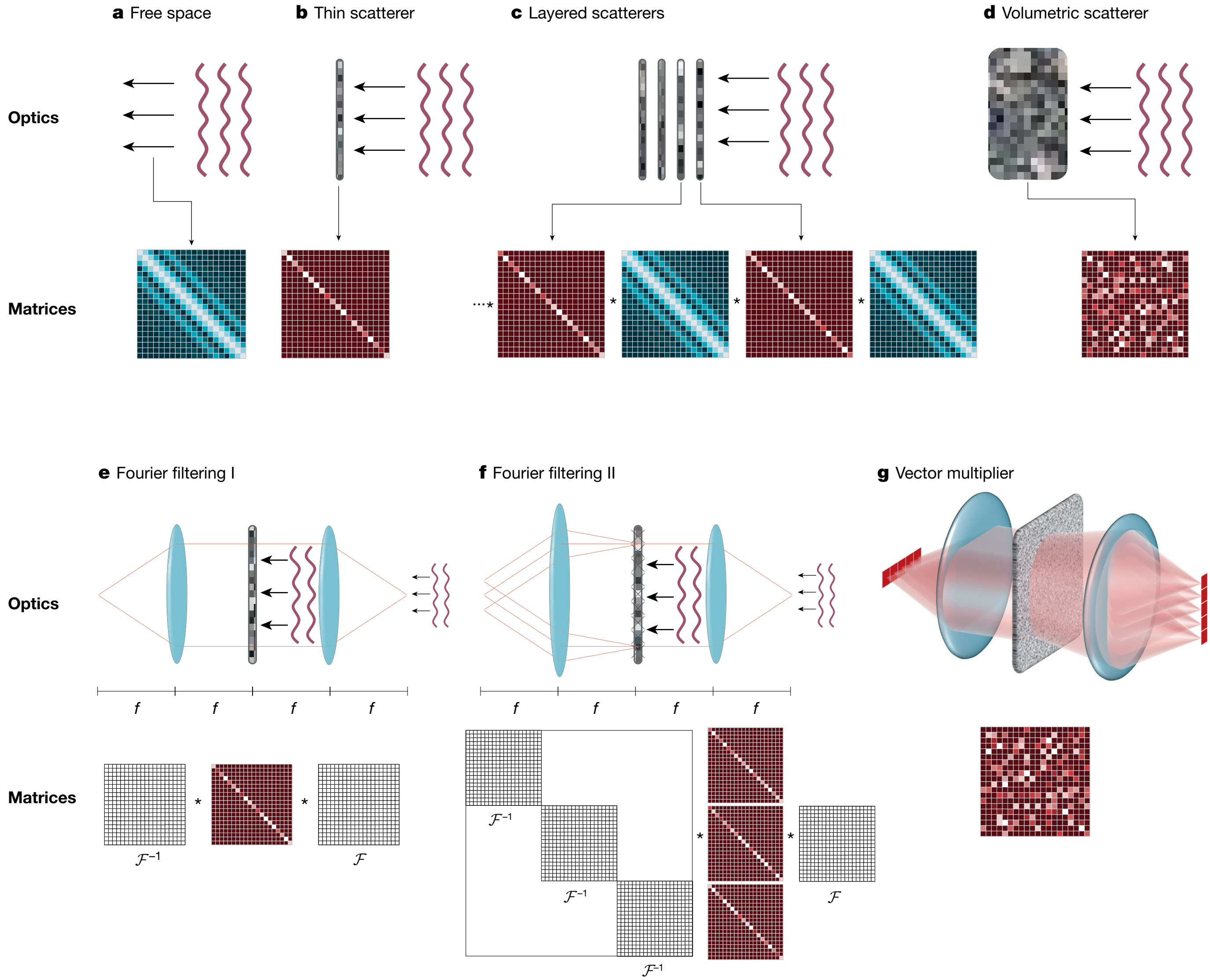

Fig. 2 | Overview of optical wave propagation. Wave propagation in free space and through different media are shown in the top rows, and the corresponding linear matrix operation(s) are given in the bottom rows. **a**, Propagation in free space is mathematically described by a convolution of the wave field with a complex-valued kernel. **b**, The interaction of a field with a thin scattering layer corresponds to multiplication with a diagonal matrix[11]. **c**, A field propagating through multiple thin scatterers with spacing in between them is a concatenation of diagonal matrices and convolution matrices[14]. **d**, A thick (volumetric) scatterer can implement a dense, pseudo-random matrix with a structure that corresponds to the physical properties of the scattering medium[13]. **e**, A traditional optical 4*f* system with a scattering layer implements an element-wise product in the Fourier domain, which corresponds to a convolution in the primal domain via the convolution theorem[11]. **f**, Modified 4*f* systems can also be used to copy the input field multiple times with a grating and convolve each copy with a different kernel[15]. Techniques in **a**–**f** can map a 2D input field to a 2D output field. **g**, A dense matrix–vector multiplication, mapping a one-dimensional input field to a one-dimensional output field, can be implemented with a 4*f*-type system[100]. The complex-valued matrices are colour-coded in red whenever the amplitude terms are most relevant and blue whenever the phase terms dominate.

sensor by projecting them via their depth-varying point spread functions onto a 2D sensor, and spectral filters then determine how the colour spectrum is integrated. Typically, an electronic decoder then estimates certain properties from the raw sensor measurements. Using a differentiable image formation model, we can simulate the optical projection of a 3D multispectral scene on a sensor; algorithms then process that data. Therefore, we can treat the problem of camera design in a holistic manner as the end-to-end optimization of optics and imaging processing[65] (see Fig. 2). Such a 'deep' computational camera can be trained in an offline stage to optimize the performance of a high-level loss function, such as image classification or object detection. Similar to conventional computer vision approaches, such a training procedure optimizes the weights of a neural network or the parameters of another differentiable algorithm. However, our encoder–decoder interpretation goes one step further in allowing the error of a high-level loss function to be backpropagated all the way into the physical parameters of the camera. Thus, a physical lens and a deep neural network can be jointly optimized for a specific task, as defined by a loss function and a training dataset (see Fig. 3). Once optimized, the physical layer (in this example, the lens) can be fabricated and used to perform inference tasks, such as classifying captured images more robustly, faster or using less power than conventional digital layers. We dub this end-to-end optimization of optics and image processing 'deep optics'.

Over the last year, several deep optics approaches have been proposed for various applications. For example, this strategy applies to optimizing the spatial layout of the sensor's colour filter array[66], the pixel exposures of emerging neural sensors[67], structured illumination patterns for microscopy and depth sensing[68–71], and the shape of

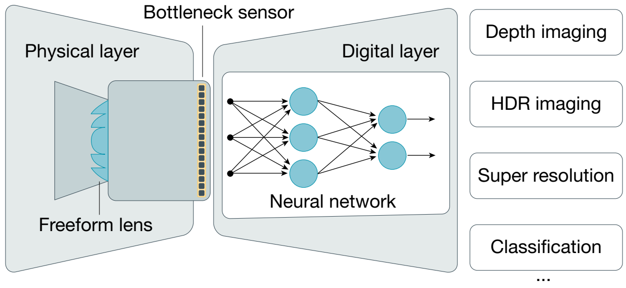

Fig. 3 | Illustration of an optical encoder–electronic decoder system. The sensor acts as a bottleneck, integrating over angle, wavelength spectrum, exposure time, phase and other quantities of the incident light. Freeform lenses or custom sensor electronics can be optimized offline for a specific task, then fabricated and used to optically and electronically code a recorded image. A neural network or another differentiable image processing algorithm then extracts the desired information from the measurements. Together, the encoder and decoder form a hybrid optical–electronic neural network. Adapted with permission from ref. [67]. HDR, high dynamic range.

freeform lenses for extended depth of field[65], image classification[13–15], flat cameras[72], high dynamic range imaging[73], wavelength demultiplexing[74], or depth sensing with conventional 2D cameras[75–77]. Depth awareness, in particular, is crucial for many tasks, including autonomous driving, robotic vision, medical imaging and remote sensing.

Although the optical encoder–electronic decoder interpretation provides an intuitive motivation for end-to-end camera design, it is not the only interpretation of cameras used in deep optical imaging approaches. We can also interpret the principle of operation of the optics as a type of computation, that is, as a pre- or co-processor that works alongside an electronic platform that processes recorded data. With this interpretation, we could try to optimize the latency and power requirements of a computational imaging system by letting the optics do as much of the work as possible. Recent research has demonstrated that this interpretation allows for a single convolutional layer[15], a fully connected layer or an otherwise parameterized[14] layer of a deep network to be implemented in optics. Implementing parts of a neural network or another AI algorithm in optics has transformative potential for improving system latency, memory usage, power efficiency, robustness to noise or other measurement degradations, and accuracy for the task at hand. One of the challenges of developing truly deep optical imaging approaches for computer vision and imaging applications, however, is again the difficulty of implementing nonlinear activation layers in optics that efficiently operate at the low light intensity and broad bandwidth of incoherent light typically captured by a camera.

Applications to microscopy

Another field that deep learning methods have made remarkable impact on is optical microscopy, spanning various modalities, including coherent imaging as well as brightfield and fluorescence microscopy. Solving inverse problems for reconstruction and enhancement of microscopy images has been a hot topic in research for decades[78]; a key component of the previous approaches is the establishment of a forward model of the imaging system[79]. Data-driven approaches based on deep learning have been providing an alternative route for solving inverse problems in optical microscopy[72,79–86]. After its training, which is a one-time effort, a DNN can provide an extremely fast framework within which to perform image reconstruction and enhancement tasks, without the need for any iterations, parameter tuning or a physical forward model. Applications of deep learning in optical microscopy cover brightfield microscopy[80], lensless microscopy[72,81,87], fluorescence microscopy[82,83,85,86,88,89], super-resolution microscopy[80,86,90–92], confocal microscopy[85,86,93], structured illumination microscopy[86] and many others[84,94–96].

There are also emerging applications of deep learning in microscopy, where the establishment of an accurate forward model is not possible with our current understanding of light–matter interaction. One example of this is cross-modality image transformations[84–86,96–98], where a DNN is trained with input and ground-truth image data that come from two different imaging modalities, without the possibility of establishing an accurate physical relationship between the two. For example, recent works have used DNNs to transform auto-fluorescence[96] or quantitative phase images[98] of label-free tissue samples into brightfield equivalent images, matching the labelled images of the same samples after they were stained with chromophores. Here, not only does the imaging modality change from fluorescence (or phase imaging) to brightfield, but also the sample goes through some transformation through the staining process, which makes the establishment of an accurate physical forward model extremely difficult. Another such cross-modality image transformation network was used to reconstruct monochrome holograms of samples with the spatial and spectral contrast of brightfield microscopy[97]; this work transformed a hologram acquired at a single wavelength into a brightfield equivalent image that is spatially and temporally incoherent, free from the coherence artefacts of holographic imaging.

From the perspective of deep learning-based computational imaging, what really sets microscopy apart from macro-scale imaging is the precision and repeatability of microscopes in terms of, for example, their hardware, illumination properties, light–matter interaction, sample properties and dimensions as well as imaging distances, which are all at the heart of this emerging success in data-driven computational microscopy techniques. Furthermore, automated scanning microscopes can generate, even in a single day, sufficiently large image data, containing for example greater than 100,000 patches of training images to robustly train a model.

An important concern regarding the use of deep learning-based approaches in microscopy is the possibility of hallucinations and creation of artefacts. In general, artefacts are known to microscopists as they contain features that do not look real. Hallucinations, by contrast, refer to features that cannot be easily distinguished from the 'real' features of a sample. DNNs can be regularized by various physics-driven constraints, by engineering their training loss functions to include physical terms; therefore, the merging of physical insights and related constraints with learning-based image transformations could form a powerful hybrid method for future computational microscopy approaches. We also believe that deep learning-based solutions to inverse problems in computational microscopy will lead to new understandings in the design of better forward models as well as better image formation and reconstruction theories.

Furthermore, there are potential strategies to mitigate hallucinations or artefacts for a DNN model, to at least warn the users when to amend or fine-tune their models. For example, this can be achieved by monitoring the statistical distance of new input data and the corresponding network output from the training or validation input and output pairs, which can be used to quantify the deviations of the imaging system from the assumptions and the state of the training phase. Transfer learning could be used to efficiently fine-tune an existing model when needed. In fact, this type of 'periodic servicing and calibration' of the network model through additional data and transfer learning is not conceptually new for advanced measurement instrumentation.

We should also consider the portability of an established model from one instrument to another. A network model that has been trained on a certain microscopy hardware should in principle be useful in other instruments that share the same design and components. However, this has not yet been widely explored in the literature and the trade off in imaging performance of a model from one microscopy instrument to others remains to be quantified at large scale to better understand the level of transfer learning and the calibration methodology that are required to faithfully run a trained model on new instruments with the same optical design and components.

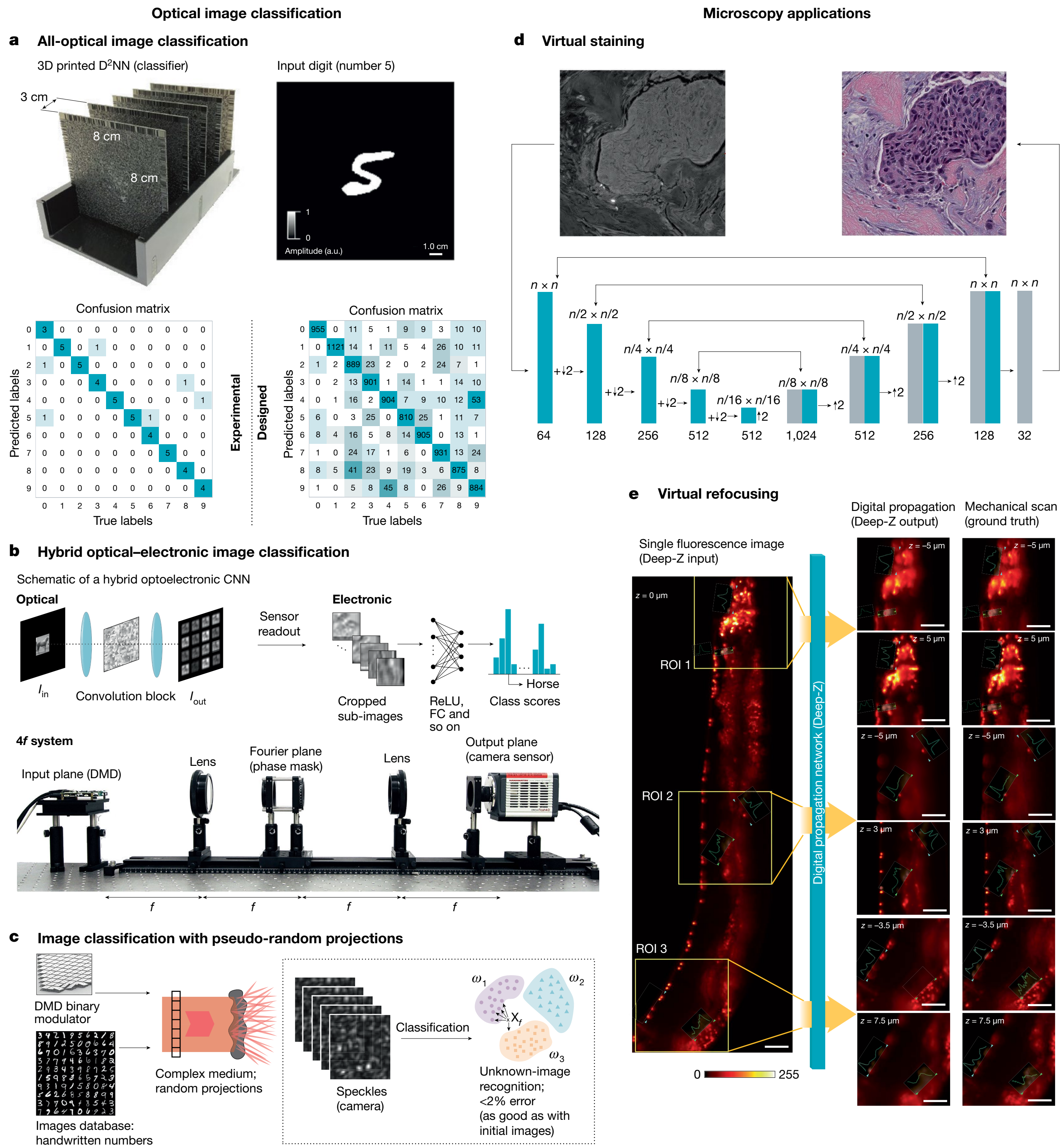

Fig. 4 | Overview of deep optics and photonics applications I. a–**c**, All-optical or hybrid optical–electronic image classification can be achieved by propagating an optical wave field through optimized layers of scatterers (**a**), through a modified optical 4*f* system that implements a single layer of a convolutional neural network (in this case, an image, I_{in}, of a horse; **b**), or through a complex medium that creates speckle (**c**). Amp., amplitude; a.u., arbitrary units; ReLU, rectified linear unit; FC, fully connected; DMD, digital micromirror device; $\omega_{1,2,3}$, pre-trained classes of numbers; X_f, measurement. **d**, **e**, In the microscopy domain, AI has been demonstrated to enable applications such as virtual histological staining of unlabelled tissue (**d**) or 3D virtual refocusing of fluorescence microscopy images using the Deep-Z algorithm compared to a ground-truth scan captured with an appropriately focused widefield microscope (**e**). *n*, number of pixels of input image; ↓2, downsampling by a factor of 2; →, concatenation; ROI, region of interest; *z*, depth of scan. Figures adapted or reproduced with permission from ref. [14] (**a**; copyright IEEE), ref. [13] (**c**), ref. [96] (**d**) and ref. [85] (**e**); adapted under a CC BY 4.0 licence from ref. [15] (**b**).

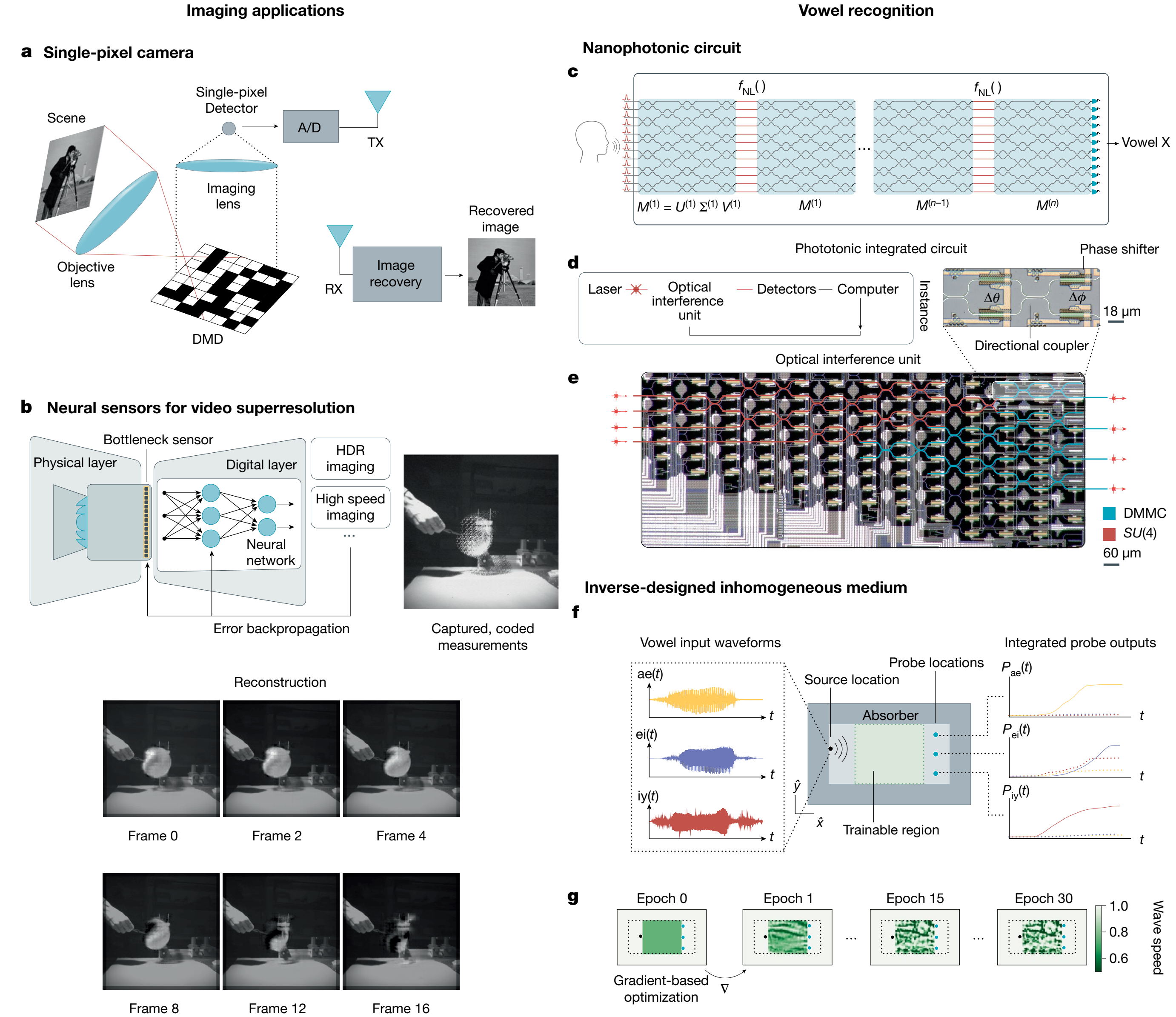

Fig. 5 | Overview of deep optics and photonics applications II. a, **b**, In imaging applications, deep optics enables single-pixel cameras that capture coded projections of a scene with a single photodetector and computationally recover them (**a**) or neural sensors that use optimized pixel exposures to capture temporally superresolved videos (**b**). DMD, digital micromirror device; A/D, analogue-to-digital converter; RX, receiver; TX, transmitter. **c**–**g**, Vowel recognition can be achieved using nanophotonic circuits that use meshes of MZIs to implement dense matrix–vector multiplications (**c**–**e**) or using optimized structures of inhomogeneous media (**f**, **g**). In **c**, each blue block represents a (linear) matrix–vector multiplication and $f_{NL}()$ are nonlinear activation functions. The matrices in each layer, $M^{(1)}, \ldots, M^{(n)}$, can be decomposed by a singular-value decomposition as $M^{(i)} = U^{(i)}\Sigma^{(i)}V^{(i)}$. This photonic-circuit implementation of each linear block follows the mathematical formulation of the singular-value decomposition. In **f**, X(*t*) represent vowel input waveforms and $P_X(t)$ their corresponding outputs. In **g**, each epoch is a cycle of the training phase. $\Delta\theta$, $\Delta\phi$, phase shifts; DMMC, diagonal matrix multiplication core; $SU(4)$, special unitary group (4); ∇, gradient. Figures adapted or reproduced with permission from ref. [51] (**a**; copyright IEEE), ref. [67] (**b**) and ref. [29] (**c**–**e**); reprinted with permission from AAAS from ref. [50] (**f**, **g**; copyright the authors, some rights reserved, exclusive licensee AAAS, distributed under a CC BY NC 4.0 licence).

Deep learning also creates new opportunities to make optical microscopy task-specific, in which the function of the microscope will expand beyond the observation of object features to also include inference—for example recognition of spatial or temporal features of interest through an optimized integration of optical and electronic computation. We believe that deep learning-enabled microscope designs of the future will use task-specific optical processors at their front end. Depending on the nature of the specific microscopic imaging task, the front-end computational optical interface that connects the illumination to the sample or the sample to the optoelectronic detector array will be optimized similar to the recent demonstrations of diffractive systems[14,15,99] that perform computation through diffraction of light. This paradigm will also change the design of the optoelectronic detector array itself (for example, the configuration of the pixels and their positions, shapes and count), making the detector interface between optics and electronics another trainable parameter space. Therefore, the optical front end, the optoelectronic detectors and the back-end electronic computation form an entirely trainable microscope.

We argue that these new types of 'thinking microscopes' can mitigate some of the challenges associated with current microscope designs, which often acquire unnecessarily large amounts of data, creating a massive burden for data sampling, storage, processing and related power requirements. By holistically optimizing the design of a microscope through deep learning methods, a task-specific microscope can potentially perform the desired inference or imaging operations with fewer pixels (or in three dimensions, voxels), at much higher framerates and using much less power, and also considerably reducing the data-storage requirements. Rather than following the traditional sequence of image formation, digitization and then processing, deep learning-enabled microscopes will merge and diffuse all these functions including inference to all the aspects of its design, working as a single task-optimized system.

Discussion

Optical and photonic computing systems have made great strides in the last two decades (see Figs. 4, 5). Creative solutions to some of the most challenging problems have recently been explored, including all-optical nonlinearities, reliable control of large photonic networks, the efficiency of electro-optical conversion and programmability. Yet, developing general-purpose optical computing systems may remain challenging to achieve in the foreseeable future. However, the emergence of AI and its requirements, especially for inference, have created new opportunities for optical components to augment their electronic counterparts. Optical computing remains an exciting research area at the convergence of physics, engineering and computer science.

The question of what breakthroughs will be necessary to propel optical computers to their full potential will certainly spark diverse and controversial answers. It is clear, however, that certain issues have to be carefully considered. Similar to a few decades ago, these issues include more efficient implementations of all-optical nonlinearities, especially those operating at the broad optical bandwidths and low intensity levels of naturally occurring optical signals. It should also be noted that digital electronic systems are the dominant platform for computing systems today. Optical computers are analogue and so far most implementations operate in a purely passive mode. Their analogue nature makes optical computers fundamentally limited by shot noise, which is different—and perhaps more challenging to manage—than the thermal noise limits of electronic accelerators. Their passive nature makes optical computers akin to purely resistive electronic circuits. All-optical approaches for signal restoration through optical gain or more efficient electro–optical conversion mechanisms are promising research directions that could address some of the remaining challenges of optical computing systems. Finally, optimistic assumptions on energy consumption of photonic systems often assume linear transforms within lossless media and without considering electro–optical conversion. Although possible solutions for these challenges exist, these issues should be carefully taken into consideration when discussing possible energy benefits and other advantages of optical computers. However, the potential for transformative use of dense energy-efficient integrated optoelectronics could solve such electro–optical conversion energy problems[8].

We see hybrid optical computing systems as one of the most promising directions in this area. Hybrid systems combine the bandwidth and speed of optical computing with the flexibility of electronic computing, and could exploit a common energy-efficient technology base across analogue and digital optical/optoelectronic/electronic systems. Applied to AI inference in computer vision, robotics, microscopy and other visual computing tasks, hybrid optical–electronic inference machines could realize some of the transformative capabilities long hailed for optical computers.

1. LeCun, Y. et al. Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems 2 (NIPS 1989)* (ed. Touretzky, D. S.) 396–404 (1990).
2. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)* (eds Pereira, F. et al.) 1097–1105 (2012).
3. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
4. Miller, D. A. B. Waves, modes, communications, and optics: a tutorial. *Adv. Opt. Photonics* **11**, 679–825 (2019).
5. Brunner, D., Soriano, M. C., Mirasso, C. R. & Fischer, I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat. Commun.* **4**, 1364 (2013).
6. Goodman, J. W., Leonberger, F. J., Kung, S.-Y. & Athale, R. A. Optical interconnections for VLSI systems. *Proc. IEEE* **72**, 850–866 (1984).
The first paper to provide a substantial analysis and reasons for the use of optics in interconnection (rather than for logic) in digital systems.
7. Miller, D. A. B. Rationale and challenges for optical interconnects to electronic chips. *Proc. IEEE* **88**, 728–749 (2000).
8. Miller, D. A. B. Attojoule optoelectronics for low-energy information processing and communications. *J. Lightwave Technol*. **35**, 346–396 (2017).
9. Miller, D. A. B. Are optical transistors the logical next step? *Nat. Photon*. **4**, 3–5 (2010).
10. Athale, R. & Psaltis, D. Optical computing: past and future. *Opt. Photon. News* **27**, 32–39 (2016).
11. Goodman, J. W. *Introduction to Fourier Optics* (Roberts and Co, 2005).
12. Liutkus, A. et al. Imaging with nature: compressive imaging using a multiply scattering medium. *Sci. Rep*. **4**, 5552 (2014).
13. Saade, A. et al. Random projections through multiple optical scattering: approximating kernels at the speed of light. In *2016 IEEE Intl Conf. Acoustics, Speech and Signal Processing (ICASSP)* 6215–6219 (IEEE, 2016).
14. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
An optical implementation using multiple optimized layers for all-optical image classification.
15. 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).
An optical implementation of a single CNN layer demonstrated for hybrid optical–electronic image classification.
16. Rosenblatt, F. *The Perceptron, A Perceiving and Recognizing Automaton* Report no. 85-460-1 (Project Para, Cornell Aeronautical Laboratory, 1957).
17. Hebb, D. O. *The Organization of Behavior* (Wiley, 1949).
18. Widrow, B. & Hoff, M. E. Adaptive switching circuits. In *1960 IRE WESCON Convention Record* 96–104 (Institute of Radio Engineers, 1960).
19. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
20. Carpenter, G. A. & Grossberg, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Comput. Vis. Graph. Image Process*. **37**, 54–115 (1987).
21. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern*. **43**, 59–69 (1982).
22. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
23. Mead, C. Neuromorphic electronic systems. *Proc. IEEE* **78**, 1629–1636 (1990).
24. Farhat, N. H., Psaltis, D., Prata, A. & Paek, E. Optical implementation of the Hopfield model. *Appl. Opt*. **24**, 1469–1475 (1985).
Optical implementation of content-addressable associative memory based on the Hopfield model for neural networks and on the addition of nonlinear iterative feedback to a vector–matrix multiplier.
25. Denz, C. *Optical Neural Networks* (Springer Science & Business Media, 2013).
26. Psaltis, D., Brady, D., Gu, X.-G. & Lin, S. Holography in artificial neural networks. *Nature* **343**, 325–330 (1990).
Introduction of nonlinear photorefractive crystals for optical computing.
27. Li, H.-Y. S., Qiao, Y. & Psaltis, D. Optical network for real-time face recognition. *Appl. Opt*. **32**, 5026–5035 (1993).
28. Miller, D. A. B. Self-configuring universal linear optical component. *Photon. Res*. **1**, 1–15 (2013).
Proof that arbitrary linear operations such as singular value decompositions can be performed in optics—not just Fourier transforms and convolutions as in early optical computing.
29. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon*. **11**, 441 (2017).
A silicon photonic neural network using meshes of MZIs for vowel recognition.
30. Fang, M. Y.-S., Manipatruni, S., Wierzynski, C., Khosrowshahi, A. & DeWeese, M. R. Design of optical neural networks with component imprecisions. *Opt. Express* **27**, 14009–14029 (2019).
31. Wilkes, C. M. et al. 60 dB high-extinction auto-configured Mach–Zehnder interferometer. *Opt. Lett*. **41**, 5318–5321 (2016).
32. 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).
33. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
A photonic circuit that exploits wavelength division multiplexing techniques for pattern recognition directly in the optical domain.
34. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep*. **7**, 7430 (2017).
35. Huang, C. et al. Giant enhancement in signal contrast using integrated all-optical nonlinear thresholder. In *2019 Optical Fiber Communications Conference and Exhibition (OFC)* 415–417 (IEEE, 2019).
36. Nahmias, M. A., Shastri, B. J., Tait, A. N. & Prucnal, P. R. A leaky integrate-and-fire laser neuron for ultrafast cognitive computing. *IEEE J. Sel. Top. Quantum Electron*. **19**, 1800212 (2013).
37. Amin, R. et al. ITO-based electro-absorption modulator for photonic neural activation function. *APL Mater*. **7**, 081112 (2019).
38. Williamson, I. A. D. et al. Reprogrammable electro-optic nonlinear activation functions for optical neural networks. *IEEE J. Sel. Top. Quantum Electron*. **26**, 7700412 (2020).
39. Miller, D. A. B. Novel analog self-electrooptic-effect devices. *IEEE J. Quantum Electron*. **29**, 678–698 (1993).

40. Srinivasan, S. A. et al. High absorption contrast quantum confined stark effect in ultra-thin Ge/SiGe quantum well stacks grown on Si. *IEEE J. Quantum Electron*. **56**, 5200207 (2020).
41. Ferreira de Lima, T., Shastri, B. J., Tait, A. N., Nahmias, M. A. & Prucnal, P. R. Progress in neuromorphic photonics. *Nanophotonics* **6**, 577–599 (2017).
42. Nahmias, M. A. et al. Photonic multiply–accumulate operations for neural networks. *IEEE J. Sel. Top. Quantum Electron*. **26**, 7701518 (2020).
A review article on the state-of-the-art of photonic MACs along with detailed characterizations and comparisons of the performance of photonic and comparable electronic hardware.
43. Gupta, S., Agrawal, A., Gopalakrishnan, K. & Narayanan, P. Deep learning with limited numerical precision. In *Proc. 32nd Intl Conf. Machine Learning* (eds Bach, F. & Blei, D.) 1737–1746 (PMLR, 2015).
44. Hamerly, R., Bernstein, L., Sludds, A., Soljačić, M. & Englund, D. Large-scale optical neural networks based on photoelectric multiplication. *Phys. Rev. X* **9**, 021032 (2019).
45. Lugt, A. V. Signal detection by complex spatial filtering. *IEEE Trans. Inf. Theory* **10**, 139–145 (1964).
The introduction of optical correlators.
46. Gregory, D. A. Real-time pattern recognition using a modified liquid crystal television in a coherent optical correlator. *Appl. Opt*. **25**, 467–469 (1986).
47. Manzur, T., Zeller, J. & Serati, S. Optical-correlator-based target detection, recognition, classification, and tracking. *Appl. Opt*. **51**, 4976–4983 (2012).
48. Javidi, B., Li, J. & Tang, Q. Optical implementation of neural networks for face recognition by the use of nonlinear joint transform correlators. *Appl. Opt*. **34**, 3950–3962 (1995).
49. Koppal, S. J., Gkioulekas, I., Zickler, T. & Barrows, G. L. Wide-angle micro sensors for vision on a tight budget. In *2011 IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2011)* 361–368 (IEEE, 2011).
50. Hughes, T. W., Williamson, I. A. D., Minkov, M. & Fan, S. Wave physics as an analog recurrent neural network. *Sci. Adv*. **5**, eaay6946 (2019).
51. Duarte, M. F. et al. Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag*. **25**, 83–91 (2008).
52. Moretti, C. & Gigan, S. Readout of fluorescence functional signals through highly scattering tissue. *Nat. Photonics* **14**, 361–364 (2020).
53. Rahmani, B., Loterie, D., Konstantinou, G., Psaltis, D. & Moser, C. Multimode optical fiber transmission with a deep learning network. *Light Sci. Appl*. **7**, 69 (2018).
54. Caramazza, P., Moran, O., Murray-Smith, R. & Faccio, D. Transmission of natural scene images through a multimode fibre. *Nat. Commun*. **10**, 2029 (2019).
55. Li, Y., Xue, Y. & Tian, L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* **5**, 1181–1190 (2018).
56. Horisaki, R., Takagi, R. & Tanida, J. Learning-based imaging through scattering media. *Opt. Express* **24**, 13738–13743 (2016).
57. Ando, T., Horisaki, R. & Tanida, J. Speckle-learning-based object recognition through scattering media. *Opt. Express* **23**, 33902–33910 (2015).
58. Mahoney, M. W. *Randomized Algorithms for Matrices and Data* (Now Publishers, 2011).
59. Dong, J., Rafayelyan, M., Krzakala, F. & Gigan, S. Optical reservoir computing using multiple light scattering for chaotic systems prediction. *IEEE J. Sel. Top. Quantum Electron*. **26**, 7701012 (2019).
60. Gupta, S., Gribonval, R., Daudet, L. & Dokmanić, I. Don't take it lightly: phasing optical random projections with unknown operators. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)* (eds Wallach, H. et al.) 14855–14865 (2019).
61. Marshall, J. & Oberwinkler, J. The colourful world of the mantis shrimp. *Nature* **401**, 873–874 (1999).
62. Thoen, H. T., How, M. J., Chiou, T.-H. & Marshall, J. A different form of color vision in mantis shrimp. *Science* **343**, 411–413 (2014).
63. Wetzstein, G., Ihrke, I., Lanman, D. & Heidrich, W. Computational plenoptic imaging. *Comput. Graph. Forum* **30**, 2397–2426 (2011).
64. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
65. Sitzmann, V. et al. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph*. **37**, 114 (2018).
The first demonstration of end-to-end optimization of optics and image processing for a computational camera design with computer vision applications.
66. Chakrabarti, A. Learning sensor multiplexing design through back-propagation. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)* (eds Lee, D. D. et al.) 3081–3089 (2016).
67. Martel, J. N. P., Muller, L. K., Carey, S., Dudek, P. & Wetzstein, G. Neural sensors: learning pixel exposures for HDR imaging and video compressive sensing with programmable sensors. *IEEE Trans. Pattern Anal. Mach. Intell*. **42**, 1642–1653 (2020).
68. Horstmeyer, R., Chen, R. Y., Kappes, B. & Judkewitz, B. Convolutional neural networks that teach microscopes how to image. Preprint at https://arxiv.org/abs/1709.07223 (2017).
69. Marco, J. et al. DeepToF: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. *ACM Trans. Graph*. **36**, 219 (2017).
70. Su, S., Heide, F., Wetzstein, G. & Heidrich, W. Deep end-to-end time-of-flight imaging. In *2018 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)* 6383–6392 (IEEE, 2018).
71. Kellman, M., Bostan, E., Repina, N. & Waller, L. Physics-based learned design: optimized coded-illumination for quantitative phase imaging. *IEEE Trans. Comput. Imaging* **5**, 344–353 (2019).
72. Sinha, A., Lee, J., Li, S. & Barbastathis, G. Lensless computational imaging through deep learning. *Optica* **4**, 1117–1125 (2017).
73. Metzler, C. A., Ikoma, H., Peng, Y. & Wetzstein, G. Deep optics for single-shot high-dynamic-range imaging. In *2020 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)* 1372–1382 (IEEE, 2020).
74. Luo, Y. et al. Design of task-specific optical systems using broadband diffractive neural networks. *Light Sci. Appl*. **8**, 112 (2019).
75. Haim, H., Elmalem, S., Giryes, R., Bronstein, A. M. & Marom, E. Depth estimation from a single image using deep learned phase coded mask. *IEEE Trans. Comput. Imaging* **4**, 298–310 (2018).
76. Chang, J. & Wetzstein, G. Deep optics for monocular depth estimation and 3D object detection. In *2019 IEEE/CVF Intl Conf. Computer Vision (ICCV)* 10192–10211 (IEEE, 2019).
77. Wu, Y., Boominathan, V., Chen, H., Sankaranarayanan, A. & Veeraraghavan, A. Phasecam3D—learning phase masks for passive single view depth estimation. In *2019 IEEE Intl Conf. Computational Photography (ICCP)* 19–30 (IEEE, 2019).
78. Bertero, M. & Boccacci, P. *Introduction to Inverse Problems in Imaging* (CRC Press, 1998).
79. Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
80. Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
81. Wu, Y. et al. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* **5**, 704–710 (2018).
82. Nehme, E. & Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458–464 (2018).
83. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol*. **36**, 460–468 (2018).
84. Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803 (2018).
85. Wu, Y. et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nat. Methods* **16**, 1323–1331 (2019).
86. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
87. Rivenson, Y., Zhang, Y., Günaydın, H., Teng, D. & Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl*. **7**, 17141 (2018).
88. Boyd, N., Jonas, E., Babcock, H. & Recht, B. DeepLoco: Fast 3D localization microscopy using neural networks. Preprint at https://doi.org/10.1101/267096 (2018).
89. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090 (2018).
90. Nehme, E. et al. DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nat. Methods* **17**, 734–740 (2020).
An end-to-end optimization approach for point spread function engineering and neural-network-based locations for 3D fluorescence superresolution microscopy.
91. Liu, T. et al. Deep learning-based super-resolution in coherent imaging systems. *Sci. Rep*. **9**, 3926 (2019).
92. Zhang, H. et al. High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network. *Biomed. Opt. Express* **10**, 1044–1063 (2019).
93. Escudero, M. C. et al. Digitally stained confocal microscopy through deep learning. In *Proc. 2nd Intl Conf. Medical Imaging with Deep Learning* (eds Cardoso, M. J. et al.) 121–129 (PMLR, 2019).
94. Rivenson, Y. et al. Deep learning enhanced mobile-phone microscopy. *ACS Photonics* **5**, 2354–2364 (2018).
95. Goy, A., Arthur, K., Li, S. & Barbastathis, G. Low photon count phase retrieval using deep learning. *Phys. Rev. Lett*. **121**, 243902 (2018).
96. Rivenson, Y. et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng*. **3**, 466–477 (2019).
97. Wu, Y. et al. Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci. Appl*. **8**, 25 (2019).
98. Rivenson, Y. et al. PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci. Appl*. **8**, 23 (2019).
99. Mengu, D., Luo, Y., Rivenson, Y. & Ozcan, A. Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J. Sel. Top. Quantum Electron*. **26**, 3700114 (2019).
100. Dagenais, M., Sharfin, W. F. & Seymour, R. J. Optical digital matrix multiplication apparatus. EU patent EP0330710A1 (1988).

Acknowledgements We thank E. Otte for help designing figures. G.W. was supported by an NSF CAREER Award (IIS 1553333), a Sloan Fellowship, by the KAUST Office of Sponsored Research through the Visual Computing Center CCF grant, and a PECASE by the US Army Research Office. A.O. was supported by an NSF ERC (PATHS-UP) grant. S.G. acknowledges funding from the European Research Council (ERC; H2020, SMARTIES-724473) and support from the Institut Universitaire de France. S.F. was supported by the US Air Force Office of Scientific Research (AFOSR) through the MURI project (grant no. FA9550-17-1-0002). D.E. and M.S. were in part supported by the US Army Research Office through the Institute for Soldier Nanotechnologies (grant no. W911NF-18-2-0048). D.E. also acknowledges support from an NSF EAGER programme. D.A.B.M. was supported by the Air Force Office of Scientific Research (award no. FA9550-17-1-0002). P.D. acknowledges discussions and a long-term collaboration with N. Farhat.

Author contributions G.W. conceived the idea, coordinated the writing process, wrote parts of the paper, and edited all sections. A.O., S.G., S.F., D.E., M.S., C.D., D.A.B.M. and D.P. wrote parts of the paper and provided feedback on all other parts.

Competing interests M.S. owns stocks of Lightelligence, Inc. S.G. owns stocks of LightOn. D.E. and D.A.B.M. own stocks in Lightmatter Inc. The other authors declare no competing financial interests.

Additional information

Correspondence and requests for materials should be addressed to G.W.

Peer review information *Nature* thanks Geoffrey W. Burr and Nathan Youngblood for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.