Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Orbital-Angular-Momentum-Preserving Vortex-Comb Multiplexing for Massively Parallel Optical Computing

Wolfram Pernice

wolfram.pernice@kip.uni-heidelberg.de

Heidelberg University https://orcid.org/0000-0003-4569-4213

Zhe Zhao

Huazhong University of Science and Technology https://orcid.org/0000-0002-1819-5518

Frank Brückerhoff-Plückelmann

University of Heidelberg https://orcid.org/0009-0009-7713-6735

Xiansong Meng

Heidelberg University

Zhongyu Tang

Heidelberg University

Rongyang Xu

Heidelberg University https://orcid.org/0000-0001-8914-4454

Caghan Ünlüer

Heidelberg University https://orcid.org/0009-0003-9033-2024

Xinyu Ma

Kirchhoff-Institute for Physics, Heidelberg University https://orcid.org/0000-0002-7406-7289

Zichao Han

Heidelberg University

Seongmin Jo

Heidelberg University

Akhil Varri

University of Münster https://orcid.org/0009-0008-5948-4181

Ivonne Bente

University of Münster https://orcid.org/0000-0003-3198-4244

Article

Keywords:

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11591553/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Orbital-Angular-Momentum-Preserving Vortex-Comb Multiplexing for Massively Parallel Optical Computing

Zhe Zhao[1,2], Frank Brückerhoff-Plückelmann[1], Xiansong Meng[1], Zhongyu Tang[1], Rongyang Xu[1], Caghan Ünlüer[1,2], Xinyu Ma[1], Zichao Han[1,2], Seongmin Jo[2,3], Akhil Varri[2,3], Ivonne Bente[2], and Wolfram Pernice[1,2,*]

1. *Kirchhoff-Institut für Physik, University of Heidelberg, 69120 Heidelberg, Germany*

2. *Institute of Physics, University of Münster, Heisenbergstr. 11, 48149 Münster, Germany*

3. *Center of Soft Nanoscience, University of Münster, Busso-Peus-Street. 10, 48149 Münster, Germany*

[*]Corresponding author: wolfram.pernice@kip.uni-heidelberg.de

As generative artificial intelligence (AI) advances, next-generation computing faces growing demands on capacity and energy efficiency. Free-space diffractive optical neural networks (ONNs) offering large multidimensional parallelism are potential candidates. Unlike other degrees of freedom (DoFs) such as wavelength and polarization, no technology exists at present in free-space diffractive ONNs that can preserve the modal information and manipulate it in a programmable manner for parallel optical computing. In those ONNs, optical neurons normally encode spatial phase and amplitude information, introducing a distortion which degrades or even destroys the modal information. Although mode-division multiplexing (MDM) based on orthogonal modal basis sets such as orbital angular momentum (OAM) could offer theoretically unbounded parallelism, the modal information degradation inherently limits its applications in parallel optical computing. Yet MDM in diffractive ONNs faces two bottlenecks: modal information is not preserved through optical computation, and the number of simultaneously multiplexed modes remains limited. Inspired by optical frequency combs, we demonstrate an ONN employing a momentum-space matching approach combined with vortex-comb multiplexing—arrays of OAM-carrying Bessel beams—to preserve modal information and enable programmable large-scale parallel processing. Experimentally, our system supports 45 parallel modes and achieves a total parallelism of approximately 35.3k across spatial and modal dimensions (784 spatial positions and 45 spatial modes). We validate this architecture through parallel 25-mode image convolution and 45-mode classification tasks, achieving modal crosstalk values of -14.11 dB and -15.37 dB, and accuracies of 94.0% and 84.2% for MNIST and Fashion-MNIST, respectively. These results establish a framework for enhancing the parallelism of diffractive ONNs, advancing the potential for large-scale multidimensional optical computing.

Introduction

The rapid development of high-performance artificial intelligence (AI) models is driving a new wave of technological transformation, with profound implications for how we live, learn, and work[1,2]. However, the escalating computational demands of modern AI come with substantial energy and environmental costs, motivating the development of next-generation computing hardware with far greater capacity and energy efficiency[3]. At the computational level, modern neural networks are dominated by tensor operations, most notably matrix multiplications, making efficient tensor processing central to improve both computing

performance and energy efficiency. This computational structure has renewed interest in analogue computing, which exploits physical processes to carry out the linear transformations that dominate tensor workloads directly, rather than reproducing them through sequences of digital operations. Recent analogue in-memory processors, including IBM 64-core phase-in-memory system, show that this approach can scale to complex neural-network inference[4].

Optical wave propagation offers an intriguing physical mechanism for analog computing, supporting high bandwidth, low propagation latency and multiple degrees of freedom (DoFs) for data encoding and processing[5,6,7]. Unlike electronic hardware[4,8], optical computing cores can expand parallelism across multiple DoFs, including mode[9,10,11], wavelength[12,13,14,15], space[16,17,18,19,20,21,22], time[23] and polarization[24]. Optical neural networks (ONNs) have therefore attracted growing interest, with free-space diffractive ONNs being a particularly promising route because of their large multidimensional parallelism[25,26,27]. In principle, this multiplies the computing capability of an ONN by the total parallelism N with limited additional power consumption[28,29]. Therefore, maintaining the advantage of optical computing depends on increasing total parallelism well beyond current levels.

However, unlike other available DoFs, it remains unclear at present how modal information can be preserved through a free-space diffractive ONN and manipulated programmably for parallel computing. Optical neurons in these ONNs encode spatial phase and amplitude to perform specific computing tasks[25,26,27], thus introducing an inherent modal distortion. The modal information thus gets degraded or even lost through these ONNs, which highly limits the application of mode-division multiplexing[30] (MDM) for parallel optical computing.

Recently, MDM built on modal basis sets such as Laguerre-Gaussian[31,32,33] and Bessel beams[34] has accordingly gained interest in ONNs[35,36,37,38,39,40,41]. Notably, each beam in these modal basis sets features a helical phasefront of $e^{j\ell\theta}$, carrying an orbital angular momentum (OAM) of $\ell\hbar$ per photon, where ℓ is the theoretically unbounded topological charge and θ is the azimuthal angle. OAM-based diffractive ONNs were demonstrated for the recognition[35], switching[36] and encoding[37,38,39] of OAM topological charges ℓ. However, the original ℓ values were not preserved and the number of parallel tasks on a single optical computing core remained one. More recently, a single trained diffractive ONN was experimentally demonstrated to perform one of 10 image classification tasks depending on the input image's mode, selected from a group of 10 OAM modes[41]. However, its output beams no longer showed ring-like profiles, indicating its modal information was not preserved (Fig. 1a). Additional retraining was therefore needed to achieve the parallel demultiplexing of OAM modes, and the reported number of parallel modes fell to three[41]. Two bottlenecks for MDM-based ONNs consequently remain: modal information is generally not preserved through optical computation, and the number of simultaneously multiplexed modes remains limited.

Here, we demonstrate a diffractive ONN using a momentum-space matching approach to preserve the modal information and manipulate it in a programmable manner for parallel image processing and recognition. The modal information preservation throughout the ONN is achieved via matching the momentum space among the optical inputs, neurons in diffractive layers, and outputs (Fig. 1d). Inspired by optical frequency combs for massive wavelength-division multiplexing, we use vortex combs[47]—arrays of OAM-carrying beams—rather than single OAM beams, enabling momentum-space matching and enhancing both spatial and modal parallelism. Furthermore, instead of relying only on one-dimensional modal index (the azimuthal OAM value

ℓ), we extend MDM for diffractive ONNs using Bessel beams with variable two-dimensional (2D) indices, defined by both the azimuthal OAM value ℓ and the radial wave vector k_r (Fig. 1h).

Based on these schemes, our ONN supports up to 45-vortex-comb multiplexing in experiment, with each comb comprising 784 Bessel beams. This configuration yields a total parallelism of about 35.3k across 784 inputs, 784 outputs, and 45 modes. Exploiting spatial and modal parallelism alone, we experimentally validate two representative tasks: (i) 25-parallel-mode image processing using convolution operations, with a peak signal-to-noise ratio (PSNR) above 20.27 dB and a multiscale structural similarity (MS-SSIM) index above 0.728; and (ii) 45-parallel-mode image classification on MNIST and Fashion-MNIST, achieving accuracies of 94.0% and 84.2%, respectively. We further show in simulation that the modal parallelism could be extended to 1,000 modes for image classification across seven parallel datasets. When combined with additional parallelism with other DoFs and high-speed modulation, this architecture could provide massive parallelism for multidimensional optical computing.

Results

Principle of vortex-comb multiplexing for ONNs

The experimental workflow of the OAM-preserving vortex-comb multiplexing diffractive ONN proceeds as follows. Multiple vortex combs, each consisting of an array of 28 × 28 OAM-carrying Bessel beams, are multiplexed at the ONN input interface. A vortex comb is generated by projecting a single Gaussian beam onto a spatial light modulator (SLM1) with a phase-only mask plus a Bessel-beam related phase term of $e^{j\ell\theta}e^{jk_r r}$ (Supplementary Note 2 and Fig. S1). Here ℓ is the OAM topological charge and k_r is the radial wave vector, representing the phase changes over the azimuthal and radial directions, respectively. Each Bessel beam in one vortex comb carries one data stream, and each vortex comb represents one independent data matrix or image. Figure 1e shows a 2 × 2 zoom-in view of the multiplexed vortex combs. Each unit in the zoom-in represents one discrete input, which is a superposition of multiple Bessel beams carrying different data streams (12 for the illustration). We then multiplex all vortex combs into a light field and propagate it in parallel through one discrete diffractive layer (Fig. 1f). A lens Fourier-transforms (FFT) the multiplexed light field at the input interface and projects the field onto the diffractive layer. Encoded with a phase-only mask (Supplementary Note 3 and Fig. S2), the diffractive layer acts as a convolution kernel that processes all multiplexed fields in parallel, and introduces an inverse Fourier transform to convert the fields back to vortex combs at the output interface. An additional conjugated phase pattern written as $e^{-j\ell\theta}e^{-jk_r r}$ finally transforms the vortex comb with indices (ℓ, k_r) back into a Gaussian comb for detection of the processed data matrix or image. An additional lightweight electronic classification head further employs these optical computing results for image recognition tasks (Fig. 1g). Further details on the generation, (de)multiplexing, and detection of the vortex combs are given in Supplementary Notes 2,3.

We demonstrate a new concept of momentum-space matching among the optical inputs, neurons in diffractive layers, and outputs to preserve the modal information throughout diffractive ONNs. Figure 1a-d shows the comparison of modal information in diffractive ONNs without and with momentum-space matching for mode preserving. Mathematically, a complex amplitude and phase field can describe each layer in diffractive ONNs. When transforming all the fields into the momentum space via FFT (Fig. 1c,d), each OAM mode occupies a

small continuous momentum window (CMW), while continuous images and diffractive layers generally occupy a much wider CMW. The addition of all these fields leads to a wide CMW with neighbouring OAM modes overlapping (Fig. 1c), resulting in high modal crosstalk[42,43,44,45]. The modal information is thus degraded, and the output beams in real space no longer exhibit ring-like amplitude profiles (Fig. 1a). In contrast, the modal information can be preserved throughout diffractive ONNs in discrete cases (Fig. 1b,d). Sampling the continuous fields with delta functions converts them into discrete ones. Notably, the modal information becomes separable with low crosstalk at the output (Fig. 1d), when the momentum spacing between neighbouring momentum components is: (i) identical for all layers in diffractive ONNs, and (ii) wider than the CMW of a single OAM mode. See Supplementary Note 1 for the mathematical details.

Our OAM-preserving vortex-comb multiplexing diffractive ONN yields a total parallelism of about 35.3k across 784 inputs, 784 outputs, and 45 modes in experiment. First, the momentum-space matching approach permits MDM based on orthogonal modal basis sets with 2D modal indices. Previous OAM-mediated ONNs exploited only the azimuthal index ℓ and ignored the radial modal information (Fig. 1h). In contrast, our approach uses both the azimuthal and radial modal indices of Bessel modal basis set. Here the number of supported modes should not be conflated with the number of modes processed in parallel. We therefore distinguish modal diversity, defined as the number of distinct spatial modes supported by an ONN, from modal parallelism, defined as the number of modes that can be simultaneously processed throughout the ONN (See Discussion section for more details). Figure 1i compares the modal diversity and parallelism of our and previous work. Our approach experimentally achieves both a modal diversity and a modal parallelism of 45 (See Table S2 for detailed comparisons).

Second, instead of multiplexing multiple coaxial OAM modes, we employ vortex-comb multiplexing to further enhance the spatial parallelism, thereby pushing the total parallelism to the tens-of-thousands scale. In particular, our ONN features discrete-port input and output (I/O) interfaces, in which each I/O port is spatially separated from the others and supports high-order modes. Potential advantages of such discrete-port I/O interfaces are discussed in the Outlook section. Figure 1j compares the total parallelism of ONNs with discrete-port I/O interfaces. We characterize an ONN by two quantities: spatial parallelism, the maximum number of spatially separated I/O ports, and optical parallelism, defined as the number of multiplexed channels across multiple DoFs except space. The total parallelism is thus the product of the spatial and optical parallelism. Our ONN demonstrates an approximately tenfold increase in total parallelism for ONNs with discrete-port I/O interfaces (See Table S3 for detailed comparisons). Overall, these comparisons highlight the potential of our ONN for high-throughput optical computing.

Parallel image convolutional processing

We first perform a parallel image-processing task to validate both the preservation of modal information and the high parallelism of our ONN. As a proof of concept, we use the OAM-preserving vortex-comb multiplexing ONN to demonstrate a 25-mode parallel image convolution (Fig. 2a). An input video of walking zebras, containing 25 frames with 162 × 162 pixels each, is divided into smaller patches with 28 × 28 pixels. We then multiplex 25 patches from 25 different frames onto vortex combs using SLM1. Each vortex comb consists of an array of 28 × 28 Bessel beams, with each comb labelled by a 2D modal index pair (ℓ, k_r). The indices are formed by all pairwise combinations of $\ell \in (0, \pm4, \pm8)$ and $k_r \in (0, \pm2k_{r0}, \pm4k_{r0})$, yielding 25 distinct Bessel modes. Here, $k_{r0} = 2\pi/p_{r0}$ is the unit radial wave vector with a radial period p_{r0} = 2.8 mm.

The multiplexed vortex combs then propagate through the ONN, where a single diffractive layer loaded on SLM2 processes all vortex combs in parallel. A single-layer phase-only mask applies the same convolution kernel to all vortex combs. To demultiplex a selected output vortex comb, an additional conjugate Bessel-related phase mask of $\Phi(-\ell, -k_r) = e^{-j\ell\theta}e^{-jk_r r}$ is loaded onto SLM2, converting the vortex comb with indices (ℓ, k_r) back into an array of 28 × 28 Gaussian beams. The filtered Gaussian comb is then recorded by a camera, representing a processed patch with 28 × 28 pixels. This convolution process is repeated over all patch positions until the full frame is covered. Finally, the processed patches are stitched together to reconstruct the processed video.

We validate the preservation of modal information by comparing the modal spectra before and after kernel convolutional processing in our ONN. Due to imperfect system alignment and phase modulation[42], the power of an input vortex comb in a given Bessel mode can partially leak into other modes at the output. A modal spectrum, defined as the normalized power distribution over different output modes for each input mode, therefore provides a quantitative evaluation of modal preservation. To measure the modal spectrum, a vortex comb of 28 × 28 Bessel beams on a single selected Bessel mode first propagates through the ONN. Conjugate phase masks of $\Phi(-\ell, -k_r)$, with indices (ℓ, k_r) sequentially scanned over the 25 modes, then convert the corresponding modal components into Gaussian combs for power measurements. Figure 2c,d shows the measured modal spectra before and after convolution using a 3 × 3 kernel, respectively, with each element averaged over 49 measurements. The detected power on the input mode, namely, the diagonal elements in Fig. 2c,d, is normalized to unity. Before convolution, the average modal crosstalk, *i.e.*, the average normalized power on modes other than the input mode, is approximately -17.97 dB (Fig. 2c). This result indicates that our system is capable of generating and detecting vortex combs with low modal crosstalk. After convolution, the modal spectrum exhibits a similar modal distribution (Fig. 2d), with only a slight change in color contrast corresponding to an increase in average modal crosstalk from approximately -17.97 dB to -16.03 dB. Because the two spectra are measured in the same modal basis, the low crosstalk before and after convolution confirms that modal information is largely preserved throughout the ONN. Further details on modal information preservation are provided in Supplementary Note 5.

We next show that our ONN can perform convolutional processing with a varying number of multiplexed modes. We first implement a 3 × 3 bottom-edge-detection convolution kernel on the input video, with the number of multiplexed modes N varied from 1 to 25. Figure 2e shows the convolution results from a representative processed frame using the vortex-comb multiplexing ONN with N = 1, 12, and 25, respectively. All processed video frames are provided in Supplementary Movie 1. The processed images are quantitatively evaluated using PSNR and MS-SSIM. As N increases from 1 to 25, the average PSNR decreases from 22.0 dB to 21.27 dB, and the average MS-SSIM decreases from 0.835 to 0.728 (Fig. 2h,i). This degradation is attributed to the accumulated modal crosstalk from other multiplexed modes into the target mode as N increases. The measured versus expected convolution results with N = 1 and 25 (Fig. 2f,g) further confirm this trend. The standard deviation of the error, defined as the difference between the measured and expected values, increases from 0.241 to 0.327 as N increases from 1 to 25. Nevertheless, the processed image quality (PSNR and MS-SSIM) remains high even for N = 25, indicating that our ONN could potentially support highly parallel convolutional processing for subsequent neural-computing tasks.

We further demonstrate the programmability and scalability of our ONN by changing the convolution kernel, increasing the kernel size and multiplexing multiple kernels within a single-diffractive-layer ONN. First, we show convolution results from a representative processed frame using four distinct 3 × 3 convolution kernels (Fig. S14). We then scale up the kernel size to 4 × 4 and 16 × 16 and experimentally measure the corresponding kernel matrices (Figs. S4,S5). Finally, we design a pure-phase mask loaded onto a single diffractive layer to implement five 3 × 3 convolution kernels in parallel through OAM multiplexing of five OAM modes (Fig. S6). The design algorithm and experimental details are provided in Supplementary Notes 2 and 3.

Parallel image classification for 10^{45} numbers and 450 objects

We experimentally implement an optoelectronic neural network comprising four parallel 45-vortex-comb multiplexing ONNs, with the optically extracted features fed into an electronic classification head for image classification (Fig. 3a). The input data consists of either: (i) a 45-digit number, where each digit is an image selected from the ten MNIST classes, yielding possible 10^{45} digit combinations; or (ii) a 45-object sequence, where each object is an image selected from the ten Fashion-MNIST classes, yielding a total of 45 × 10 = 450 object class scores. The 28 × 28 × 45 input tensor is first multiplexed using 45 vortex combs, each comprising a 28 × 28 array of Bessel beams and assigned a distinct modal index pair (ℓ, k_r). The 45 modal index pairs are formed by all pairwise combinations of $\ell \in (0, \pm4, \pm8, \pm12, \pm16)$ and $k_r \in (0, \pm2k_{r0}, \pm4k_{r0})$. The multiplexed vortex combs are then distributed to four parallel ONNs, which perform convolutional feature extraction using upper-, lower-, left- and right-edge detection kernels, respectively. Each ONN uses a single diffractive layer for convolutional processing, and the output of four ONNs is a processed 28 × 28 × 4 × 45 tensor. The processed tensor is then fed into a lightweight electronic classification head consisting of one batch-normalization layer, one ReLU activation layer, one 2 × 2 max-pooling layer, one dropout layer and one 784-to-10 fully connected (FC) layer. The last FC layer sequentially processes one 14 × 14 × 4 feature tensor at a time for ten-class image classification. The electronic classification process is repeated 45 times until all 45 digits or objects are classified.

We first measure and quantitatively evaluate the optically extracted convolutional features of our ONNs. Compared with the ground truth, the experimentally measured convolutional features (Fig. 3b) contain noticeable noise even without vortex-comb multiplexing. The measured versus expected values (Fig. 3d,e) further confirm this observation. For ONNs using a single vortex comb, the errors between the measured and expected values have standard deviations of 0.292 and 0.289 for MNIST and Fashion-MNIST, respectively. We believe this noise is mainly attributed to the Gerchberg-Saxton (GS) algorithm[48] used to calculate the phase masks for encoding the input images and convolution kernels. The GS algorithm can introduce spatial inhomogeneous errors between the designed and target images and kernels[48]. Such errors could potentially be reduced by using more advanced optimization algorithms[49,50], such as gradient-descent-based methods[50], or by encoding the data through complex spatial modulation with both amplitude and phase control[51,52].

As the number of multiplexed vortex combs N increases, the quality of the convolutional feature tensors degrades. When N increases from 1 to 45, a slight increase in noise can be observed in edge-detection feature images (Fig. 3b). Quantitatively, the standard deviation of the error between the measured and expected

values increases by 0.09 and 0.208 for MNIST and Fashion-MNIST, respectively. Correspondingly, for MNIST, the PSNR and SSIM structure decrease from 22.14 dB to 15.49 dB and from 0.856 to 0.596, respectively; for Fashion-MNIST, the PSNR and SSIM structure decrease from 24.51 dB to 15.32 dB and from 0.892 to 0.572, respectively. These degradations are mainly attributed to accumulated modal crosstalk as N increases. Figure 3c shows the modal spectra of the processed vortex combs, indicating average modal crosstalk values of approximately -14.11 dB and -15.37 dB for MNIST and Fashion-MNIST, respectively. Despite the degradation in feature quality at larger N, the relatively low modal crosstalk enables highly parallel optical feature extraction for subsequent image classification.

We next demonstrate parallel image classification using our ONN combined with an electronic classification head. The training details are provided in Methods section and Supplementary Note 8. Figure 3h shows the loss and classification accuracy as a function of training epoch, with both values calculated using testing sets containing 1,000 images evenly divided into N groups (*i.e.*, N independent vortex combs with distinct modal index pairs). A higher loss and lower classification accuracy are observed: (i) when the classification task becomes more complex from MNIST to Fashion-MNIST; and (ii) when N increases. This trend is expected because more complex tasks and higher noise generally require networks with greater representational capacity to achieve comparable performance[53]. Figure 3i shows the classification accuracy as a function of N, with the shaded regions indicating the accuracy ranges of all-electronic neural-network baselines. These baselines have the same network structures as the optoelectronic network in Fig. 3a, except that the convolutional features are computed electronically on a CPU, either using the same edge-detection kernels or using unit kernels. The latter case corresponds to the case without convolutional feature extraction. As N increases, accumulated modal crosstalk degrades the optically extracted features, and the classification accuracy decreases from 96.7% to 94.0% for MNIST and from 85.6% to 84.2% for Fashion-MNIST. Nevertheless, as N decreases from 45 to 1, the classification accuracies of our networks are 2.1-4.8% and 4.4-5.8% higher than those of the all-electronic baseline without a convolutional layer for MNIST and Fashion-MNIST, respectively. More details about the ONN performance as a function of modal index pair are provided in Supplementary Note 6. Figure 3g shows the confusion matrices with $N = 45$. These results indicate that the vortex-comb multiplexing ONN could accelerate image classification tasks in parallel.

Scale the modal number up to 1,000 in simulation

Having experimentally demonstrated 45-mode vortex-comb multiplexing, we next explore the modal parallelism of our architecture in simulation. In the experiment, the number of available modes is mainly constrained by: (i) the imperfect encoding of images and kernels[48,51]; and (ii) the finite aperture, pixel resolution and alignment accuracy of the optical system[42], rather than the momentum-space matching principle itself. To evaluate this capability beyond the current experimental limit, we numerically scale the number of multiplexed vortex combs N up to 1,000 and assess the performance of the ONN for parallel image classification of seven datasets.

Figure 4a shows the workflow of the simulated optoelectronic network based on 1,000-mode vortex-comb multiplexing. We consider seven image classification datasets, including MNIST, Fashion-MNIST, and five Med-MNIST[54]. Images from these datasets are first divided into N groups, and then multiplexed into one optical field using vortex combs labelled with distinct index pairs (ℓ, k_r). The up to 1,000 modal index pairs are formed by all pairwise combinations of $\ell \in (0, \pm1, \pm2, \ldots, \pm23, \pm24, +25)$ and $k_r \in (0, \pm2, \pm4, \ldots, \pm16,$

±18, +20) × k_{r0}. All the other network parameters are kept the same as those in the 45-mode vortex-comb multiplexing optoelectronic network (Fig. 3a), except that the final FC layer is replaced by seven task-specific FC layers for different image classification tasks. The simulation model and ONN implementation details are provided in Supplementary Note 7.

We show in simulation that the 1,000-mode vortex-comb multiplexing optoelectronic network can process up to 1,000 classification tasks in parallel across multiple datasets. The training details for different datasets and N values are provided in Fig. S15. Figure 4b shows the PSNR and SSIM structure as functions of N. The average PSNR and SSIM structure remain above 28.4 dB and 0.97, respectively, when N is less than 300. As N increases from 300 to 1,000, both metrics decrease continuously, indicating increased computational error due to the accumulated modal crosstalk. This trend is consistent with the classification accuracies as functions of N: the accuracies decrease slightly as N increases from 1 to 300, but show a more pronounced degradation when N approaches 1,000. Figure 4d compares the classification accuracies of our optoelectronic network with those of the all-electronic baselines. The accuracy improvement varies for different image classification tasks, likely because different tasks have different tolerances to noise[55], which mainly originates from accumulated modal crosstalk as N increases. Nevertheless, accuracy improvements are still observed at large N values. For example, compared with the all-electronic baseline without convolution layers, the simulated optoelectronic network with N = 1,000 achieves accuracy improvements of 2.6% and 1.9% for MNIST and Fashion-MNIST, respectively.

Discussion

Our results establish OAM-preserving vortex-comb multiplexing as a potential route to large-scale parallelism in optoelectronic diffractive neural networks. The key advance is the preservation of modal information through momentum-space matching, which allows spatial and modal channels to be processed simultaneously with relatively low modal crosstalk during optical computing. By extending MDM using OAM beams to vortex combs with arrays of Bessel beams labelled with distinct modal index pairs, this architecture could support large-scale discrete I/O interfaces and multiplexed modes.

By analogy with MDM in optical communications[30,32], a strict definition of MDM in ONNs requires each data-carrying beam to: (i) occupy a distinct mode from an orthogonal modal basis set; and (ii) be effectively (de)multiplexed at the input and output interfaces with little modal crosstalk. Phase and amplitude distortions such as turbulence[42,43] and scattering[44] degrade modal information in MDM, leading to higher modal crosstalk. Therefore, preserving modal information from the input to the output interface is the key to recovering the low-modal-crosstalk computation results without the need for extra post-processing[45,46] or network retraining[41]. A useful distinction in evaluating MDM-based ONNs is that between modal diversity and modal parallelism (Fig. 1i). Modal diversity is defined as the number of modes supported by ONNs, regardless of whether the modal information is preserved, degraded, or completely lost during optical computing. Modal parallelism, in contrast, refers to the number of modes that can be processed simultaneously as independent channels while preserving their modal information throughout ONNs.

An important question is whether orthogonality alone guarantees modal-information preservation in ONNs. In fact, orthogonality is a necessary but not sufficient condition for mode preservation. For example, when

multiple OAM beams with distinct OAM values propagate through a phase-only mask, such as an atmospheric-turbulence phase mask[43] or the diffractive layers in previous OAM-mediated ONNs[38,39], the complex output fields may remain mutually orthogonal because the phase modulation acts as a unitary transformation[31,43]. However, their amplitude and phase profiles can be strongly distorted, and their power can be redistributed over multiple modes in the original modal basis. In such cases, the modal information is not preserved, and additional phase correction[43] or retraining of the optical network[41,46] is required to maintain low-crosstalk MDM-based optical computing. In particular, when optical-network retraining is used[41,46], the training cost typically scales approximately linearly with the number of multiplexed modes N, and additional diffractive layers might be required to suppress modal crosstalk[41,56].

In contrast, our momentum-space matching architecture preserves the modal spectrum to a large extent during optical computing. As a result, the single-diffractive-layer ONN is trained once and could potentially process multiplexed vortex combs in parallel without additional optical redesign or retraining, enabling modal diversity to be converted into available modal parallelism.

Outlook

The I/O interface connecting diffractive ONNs with high-speed optical signal-processing systems remains an important direction for future work. Existing free-space diffractive ONNs usually rely on continuous-image-based interfaces using SLMs, digital micromirror devices, and cameras, which provide massive spatial parallelism but are typically limited to kilohertz modulation or readout speeds[26,27]. This makes them less compatible with high-speed optical systems that transmit data through discrete-channel couplers or fiber arrays[11,15]. By contrast, our ONN uses discrete-port I/O interfaces, where each port is spatially separated from the others and supports high-order modes. Although the present prototype uses low-speed SLMs and a camera, this mode-addressable discrete-port format could potentially be implemented with: (i) multimode couplers[57] or fiber arrays[30,58] to generate and detect vortex combs directly. Moreover, because the vortex-comb multiplexing ONN requires only passive phase modulation (excluding data input and output), it is, in principle, compatible with high-speed free-space optical computing devices[60,61] including mircoLED, VCSEL and photodiode (PD) arrays.

To assess the potential of vortex-comb multiplexing for high-speed optical computing, we estimate the compute density and energy efficiency of ONNs based on SLMs, MDMs and VCSEL arrays (Supplementary Note 10, Table S3,4). For image processing/classification tasks using a 3 × 3 convolution kernel with the same network structure and parameters as shown in Figs. 2a and 3a, a 45-mode vortex-comb multiplexing ONN driven by 1-GHz VCSEL arrays[61] could achieve a compute density of 12.3 TOPS/mm^2 and an energy efficiency of 5.9 TOP/J. Increasing the kernel size from 3 × 3 to 16 × 16 could potentially further improve both metrics by approximately 30-fold. As a reference point, NVIDIA B300[8] reports an FP16 throughput of 2,250 TOPS, corresponding to a compute density of 1.38 TOPS/mm^2 and an energy efficiency of 1.61 TOP/J. We note that this comparison should be regarded as indicative rather than a direct like-for-like benchmark. Nevertheless, the estimates suggest that integrating our ONN with high-speed optical systems could provide a possible route towards high-speed, large-scale optical computing.

Pushing the total parallelism of vortex-comb multiplexing ONNs far beyond the current 35.3k would also be interesting. This could be achieved by further increasing the modal and spatial parallelism of the present architecture and by combining it with multiplexing of other DoFs.

The practical modal parallelism of our ONN is currently limited by accumulated modal crosstalk. As the number of multiplexed vortex combs N increases, the finite numerical aperture and pixel resolution of the optical system[42] constrain the separation between neighbouring Bessel modes, leading to increased modal crosstalk during optical processing of vortex combs such as generation, detection and (de)multiplexing. Moreover, the misalignment of optical input, diffractive, and output layers further increases modal crosstalk[42]. The GS-algorithm-based phase masks for encoding images and kernels result in additional crosstalk[48]. Consequently, the available number of multiplexed modes is set not by the theoretically unbounded mode basis set, but by the crosstalk tolerance of the computing task. Reducing modal crosstalk through optimized modal spacing[41,42], improved hologram design[49,50], complex modulation[51,52] and misalignment resilience[59] would potentially extend the practical modal parallelism towards the simulated 1,000-mode regime.

The spatial parallelism could also be further increased through more advanced modal encoding strategies. In the present architecture, each Bessel beam in a 28 × 28 vortex comb carries one data stream, corresponding to a single image pixel. Previous reports have shown that an entire image can also be encoded onto a single OAM beam[41] or a superposition of multiple OAM beams[39]. It might therefore be possible to encode an image, rather than a single image pixel, onto each Bessel beam within the vortex comb. Such a strategy could, in principle, potentially increase the spatial parallelism by a factor comparable to the number of image pixels encoded on each beam, although this would require careful control of modal crosstalk and encoding/decoding complexity.

Finally, MDM is intrinsically compatible with multiplexing technologies using other DoFs. Combining the vortex-comb multiplexing with polarization-, time-, frequency- and wavelength-division multiplexing could further enhance the total parallelism. However, when multidimensional multiplexing is used, the trade-off between computing capacity and system complexity might need to be further considered.

Materials and Methods

Metrics definition

To quantitatively evaluate the experimentally measured convolution results and the CPU-computed reference (ground truth), we employ three complementary image-quality metrics: PSNR, SSIM structure and MS-SSIM. Throughout, let x denote the experimental image and y the reference (CPU) image, both normalized to the same dynamic range of $[0, 1]$.

PSNR: The PSNR quantifies the pixel-wise fidelity between the two images and is derived from the mean-squared error (MSE). For images of size $M \times N$,

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[x(i,j) - y(i,j)]^2, \quad (1)$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{L^2}{\mathrm{MSE}}\right), \tag{2}$$

where L is the maximum possible pixel value (dynamic range). PSNR is expressed in decibels (dB); higher values indicate smaller pixel-wise deviations from the reference.

SSIM structure: The SSIM assesses perceived image quality by separately comparing three components—luminance, contrast, and structure—between local patches of x and y. For a local window, let μ_x, μ_y be the local means, σ_x^2, σ_y^2 the local variances, and σ_{xy} the local covariance. The three comparison functions are

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},\ c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},\ s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{3}$$

for luminance l, contrast c, and structure s, respectively. The stabilizing constants are $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, and $C_3 = C_2/2$, with $K_1 = 0.01$ and $K_2 = 0.03$.

For the image classification tasks with 28 × 28 images, we evaluate the processed images using only SSIM structure:

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \tag{4}$$

The SSIM structure is insensitive to global brightness and contrast scaling. Therefore, it provides a robust measure of structural fidelity that is not confounded by the intensity normalization[62], which is applied to the optically extracted images in the normalization layer of the electronic classification head. SSIM structure values lie in $[-1, 1]$, with 1 indicating perfect structural correlation, 0 indicating no structural correlation, and negative values indicating local anti-correlation between the two images.

MS-SSIM: For the image processing task using the zebra-walking video with a relatively large frame size of 162 × 162 pixels (compared with the 28 × 28 MNIST images), we use MS-SSIM to evaluate structural similarity across multiple spatial resolutions, which better matches the multiscale nature of human visual perception[62]. Starting from the original images (scale $j = 1$), the images are iteratively low-pass filtered and downsampled by a factor of two, generating a total of M scales. At each scale j, the contrast term $c_j(x,y)$ and structure term $s_j(x,y)$ are computed; the luminance term $l_M(x,y)$ is evaluated only at the coarsest scale M. The MS-SSIM is then defined as

$$\text{MS-SSIM}(x,y) = [l_M(x,y)]^{\alpha_M} \prod_{j=1}^{M} \left[c_j(x,y)\right]^{\beta_j} \left[s_j(x,y)\right]^{\gamma_j}, \tag{5}$$

where the exponents $\alpha_M, \beta_j, \gamma_j$ weight the relative importance of each component and scale. We use $M = 5$ scales with the weights $\beta_j = \gamma_j = \{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ and $\alpha_M = \gamma_M$. MS-SSIM values lie in $[0, 1]$, with higher values denoting greater multiscale structural fidelity.

Datasets details

In the 45-mode vortex-comb multiplexing experiment, we use 10-class MNIST and Fashion-MNIST datasets. For each dataset, we select 9,000 images, of which 8,000 are used for training and 1,000 for testing. The images are evenly divided into N groups, with each group encoded on an individual vortex comb with a distinct modal index pair (ℓ, k_r).

In the 1,000-mode vortex-comb multiplexing simulation, we use seven datasets: MNIST, Fashion-MNIST, Pneumonia-MNIST, Blood-MNIST, OrganaA-MNIST, OrganaB-MNIST and OrganaC-MNIST[54], comprising 10, 10, 2, 8, 10, 10 and 10 classes, respectively. We select 3×10^4, 3×10^4, 2×10^3, 5.5×10^3, 2×10^4, 6×10^3, and 6.5×10^3 images from seven individual datasets. Approximately 8/9 of the images are used for training and 1/9 of the images for testing. The images are divided into N groups, with each group encoded on an individual vortex comb with a distinct modal index pair (ℓ, k_r).

Network structure

The image classification framework consists of a vortex-comb multiplexing ONN, followed by an electronic classification head comprising batch normalization, a ReLU activation, 2×2 max-pooling, and fully connected layers. Input images encoded on each vortex comb are sized at [28, 28]. The optical convolution employs four [3, 3] kernels (a pair of Sobel operators and their negatives), producing a [28, 28, 4] feature tensor that is captured on the camera, digitized, and imported into the electronic network. The [28, 28, 4] tensor is first standardized per channel by batch normalization, passed through a ReLU nonlinearity, and down-sampled by 2×2 max-pooling with a stride of 2 to a [14, 14, 4] tensor. Dropout with a rate of 0.15 is then applied. This [14, 14, 4] tensor is then flattened into a [1, 784] vector and propagated through fully connected layers of sizes [784, C], where C is the number of image classes. The resulting [1, C] vector represents the C-class classification output and is normalized by a Softmax function. The loss is computed by the cross-entropy function against the one-hot image label, and the network is optimized by the Adam algorithm (initial learning rate 1×10^{-3}, mini-batch size 128) for up to 30 epochs, with the best-validation model retained. Backward propagation updates the electronic weights to complete the training, while the optical convolution kernels remain fixed.

References

1. Noy, S. & Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **381**, 187–192 (2023).
2. Kasneci, E. *et al.* ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* **103**, 102274 (2023).
3. Sevilla, J. *et al.* Compute trends across three eras of machine learning. In *2022 International Joint Conference on Neural Networks (IJCNN)* 1–8 (IEEE, 2022).
4. Le Gallo, M. *et al.* A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference. *Nat. Electron.* **6**, 680–693 (2023).
5. Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).

6. Tait, A. N. *et al.* Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
7. Dong, B. *et al.* Partial coherence enhances parallelized photonic computing. *Nature* **632**, 55–62 (2024).
8. NVIDIA. *NVIDIA DGX B300 system: user guide* (2025). **https://docs.nvidia.com/dgx/dgxb300-user-guide/**
9. Ling, Q. *et al.* On-chip optical matrix–vector multiplier based on mode-division multiplexing. *Chip* **2**, 100061 (2023).
10. Sun, S. *et al.* Highly efficient photonic convolver via lossless mode-division fan-in. *Nat. Commun.* **16**, 7513 (2025).
11. Yao, Q. *et al.* 192-dimensional wavelength-/mode-/polarization-multiplexed photonic architecture for highly parallel optical computing. *Natl Sci. Rev.* **13**, nwag350 (2026).
12. Wu, C. *et al.* Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nat. Commun.* **12**, 96 (2021).
13. Yin, R. *et al.* Integrated WDM-compatible optical mode division multiplexing neural network accelerator. *Optica* **10**, 1709–1718 (2023).
14. Yu, X. *et al.* Parallel optical computing capable of 100-wavelength multiplexing. *eLight* **5**, 10 (2025).
15. Feldmann, J. *et al.* Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
16. Zhang, W. *et al.* Optical logic convolutional neural network. *Sci. Adv.* **12**, eaea9278 (2026).
17. Wang, B. *et al.* An all-optical signal processor enabling terabit-per-second real-time equalization. *Science* **392**, eady5344 (2026).
18. Yao, Q. *et al.* Parallel matrix–vector multiplication using a scalable multiport interferometer. *Laser Photon. Rev.* **20**, e01876 (2026).
19. Zheng, H. *et al.* Multichannel meta-imagers for accelerating machine vision. *Nat. Nanotechnol.* **19**, 471–478 (2024).
20. Xu, Z. *et al.* Large-scale photonic chiplet Taichi empowers 160-TOPS/W artificial general intelligence. *Science* **384**, 202–209 (2024).
21. Ahmed, S. R. *et al.* Universal photonic artificial intelligence acceleration. *Nature* **640**, 368–374 (2025).
22. Hua, S. *et al.* An integrated large-scale photonic accelerator with ultralow latency. *Nature* **640**, 361–367 (2025).
23. Dong, B. *et al.* Higher-dimensional processing using a photonic tensor core with continuous-time data. *Nat. Photon.* **17**, 1080–1088 (2023).
24. Petrini, A., Conti, C. & Pierangeli, D. Degree-of-polarization modulation for high-dimensional optical computing. *Nature* **657**, 377–384 (2026).
25. Lin, X. *et al.* All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
26. Luo, M. *et al.* Highly scalable machine vision enabled with meta-optics-based ultra-wide neural network. *eLight* **6**, 10 (2026).

27. Peng, J. *et al.* Optical metasurfaces for general vision processing on the edge. *Nature* **654**, 917–925 (2026).
28. Shastri, B. J. *et al.* Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* **15**, 102–114 (2021)
29. Bente, I. *et al.* The potential of multidimensional photonic computing. *Nat. Rev. Phys.* **7**, 439–450 (2025).
30. Richardson, D. J., Fini, J. M. & Nelson, L. E. Space-division multiplexing in optical fibres. *Nat. Photon.* **7**, 354–362 (2013).
31. Yao, A. M. & Padgett, M. J. Orbital angular momentum: origins, behavior and applications. *Adv. Opt. Photon.* **3**, 161–204 (2011).
32. Wang, J. *et al.* Terabit free-space data transmission employing orbital angular momentum multiplexing. *Nat. Photon.* **6**, 488–496 (2012).
33. Zhao, Z. et al. Dynamic spatiotemporal beams that combine two independent and controllable orbital-angular-momenta using multiple optical-frequency-comb lines. *Nat. Commun.* **11**, 4099 (2020).
34. Céspedes Vicente, O. & Caloz, C. Bessel beams: a unified and extended perspective. *Optica* **8**, 451–457 (2021).
35. Li, X. *et al.* Simultaneous sorting of arbitrary vector structured beams with spin-multiplexed diffractive metasurfaces. *Adv. Photon. Nexus* **3**, 036010 (2024).
36. Q. Wang, J. Liu, D. Lyu, J. Wang, Ultrahigh-fidelity spatial mode quantum gates in high-dimensional space by diffractive deep neural networks. *Light: Sci. Appl.*, **13**, 10 (2024).
37. Zhang, Y. *et al.* Memory-less scattering imaging with ultrafast convolutional optical neural networks. *Sci. Adv.* **10**, eadn2205 (2024).
38. Zhang, K. *et al.* Advanced all-optical classification using orbital-angular-momentum-encoded diffractive networks. *Adv. Photon. Nexus* **2**, 066006 (2023).
39. Fang, X. *et al.* Orbital angular momentum-mediated machine learning for high-accuracy mode-feature encoding. *Light Sci. Appl.* **13**, 49 (2024).
40. Huo, P. *et al.* Broadband and parallel multiple-order optical spatial differentiation enabled by Bessel vortex modulated metalens. *Nat. Commun.* **15**, 9045 (2024).
41. He, C. *et al.* Orbital angular momentum multiplexing diffractive neural networks for high-capacity optical inference. *Nat. Photon.* **20**, 896–904 (2026).
42. Xie, G. *et al.* Performance metrics and design considerations for a free-space optical orbital-angular-momentum-multiplexed communication link. *Optica* **2**, 357–365 (2015).
43. Ren, Y. *et al.* Atmospheric turbulence effects on the performance of a free space optical link employing orbital angular momentum multiplexing. *Opt. Lett.* **38**, 4062–4065 (2013).
44. Zhang, R. *et al.* Coherent optical wireless communication link employing orbital angular momentum multiplexing in a ballistic and diffusive scattering medium. *Opt. Lett.* **44**, 691–694 (2019).
45. Huang, H. *et al.* Crosstalk mitigation in a free-space orbital angular momentum multiplexed communication link using 4 × 4 MIMO equalization. *Opt. Lett.* **39**, 4360–4363 (2014).
46. Xiong, W. *et al.* Convolutional neural network based atmospheric turbulence compensation for optical orbital angular momentum multiplexing. *J. Lightwave Technol.* **38**, 1712–1721 (2020).

47. Cao, S. *et al.* Tailoring perfect optical vortex combs based on diffractive neural networks. *Laser Photon. Rev.* **20**, e02383 (2026).

48. Kim, D. *et al.* Large-scale uniform optical focus array generation with a phase spatial light modulator. *Opt. Lett.* **44**, 3178–3181 (2019).

49. Fan, Z., Qian, C., Jia, Y. *et al.* Holographic multiplexing metasurface with twisted diffractive neural network. *Nat Commun* **15**, 9416 (2024).

50. Shi, Z. *et al.* Super-resolution orbital angular momentum holography. *Nat. Commun.* **14**, 1869 (2023).

51. Ren, H., Fang, X., Jang, J. *et al.* Complex-amplitude metasurface-based orbital angular momentum holography in momentum space. *Nat. Nanotechnol.* **15**, 948–955 (2020).

52. Zhang, Y. *et al.* Direct tensor processing with coherent light. *Nat. Photon.* **20**, 102–108 (2026).

53. Belkin, M., Hsu, D., Ma, S. & Mandal, S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proc. Natl Acad. Sci. USA* **116**, 15849–15854 (2019).

54. Yang, J., Shi, R., Wei, D. *et al.* MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Sci Data* **10**, 41 (2023).

55. Geirhos, R. *et al.* Shortcut learning in deep neural networks. *Nat. Mach. Intell.* **2**, 665–673 (2020).

56. Fontaine, N.K., Ryf, R., Chen, H. *et al.* Laguerre-Gaussian mode sorter. *Nat Commun* **10**, 1865 (2019).

57. Cai, X. *et al.* Integrated compact optical vortex beam emitters. *Science* **338**, 363–366 (2012).

58. Bozinovic, N. *et al.* Terabit-scale orbital angular momentum mode division multiplexing in fibers. *Science* **340**, 1545–1548 (2013).

59. Mengu, D. *et al.* Misalignment resilient diffractive optical networks. *Nanophotonics* **9**, 4207–4219 (2020).

60. Kalinin, K.P., Gladrow, J., Chu, J. et al. Analog optical computer for AI inference and combinatorial optimization. *Nature* **645**, 354–361 (2025).

61. Chen, Z. et al. Deep learning with coherent VCSEL neural networks. *Nat. Photonics* **17**, 723–730 (2023).

62. Wang, Z., Simoncelli, E. P. & Bovik, A. C. Multiscale structural similarity for image quality assessment. In *37th Asilomar Conference on Signals, Systems & Computers* **2**, 1398–1402 (IEEE, 2003).

63. TEXAS INSTRUMENTS. *DLP9500 Data Sheet* (2023). https://www.ti.com/lit/ds/symlink/dlp9500.pdf

Acknowledgement

We thank Jochen Stuhrmann (*Illustrato*) for his assistance with the illustrations. We thank the working group of Prof. Wolfram Pernice for their support and assistance during the research project.

Funding

The research is funded by: European Research Council (grant no. 101200429, PICNIC, WP), European Union's Horizon 2020 research and innovation programme (grant no. 101017237, PHOENICS project, WP, ZZ) and the European Union's Innovation Council Pathfinder programme (grant no. 101259183, NEQIOS project, WP; grant no. 101046878, HYBRAIN project, FBP). Z.Z. acknowledges funding support through a Humboldt Research Fellowship from the Alexander von Humboldt Foundation.

Data and code availability

All data, code, theory details, simulation details that support the findings of this study are available from the corresponding authors on reasonable request.

Competing interests: The authors declare no competing interests.

Author contributions

Conceptualization: Z.Z., W.P.

Methodology: Z.Z., F.P., X.S.M., Z.T., R.X., X.Y.M., I.B.

Investigation: Z.Z., F.P., C.Ü., Z.H., S.J., A.V.

Visualization: Z.Z.,

Funding acquisition: W.P., Z.Z.,

Project administration: W.P.,

Supervision: W.P.,

Writing – original draft: ZZ, W.P.,

Writing – review and editing: All authors

Supplementary Material

Supplementary Text

Figs. S1 to S15

Tables. S1 to S5

Movie S1

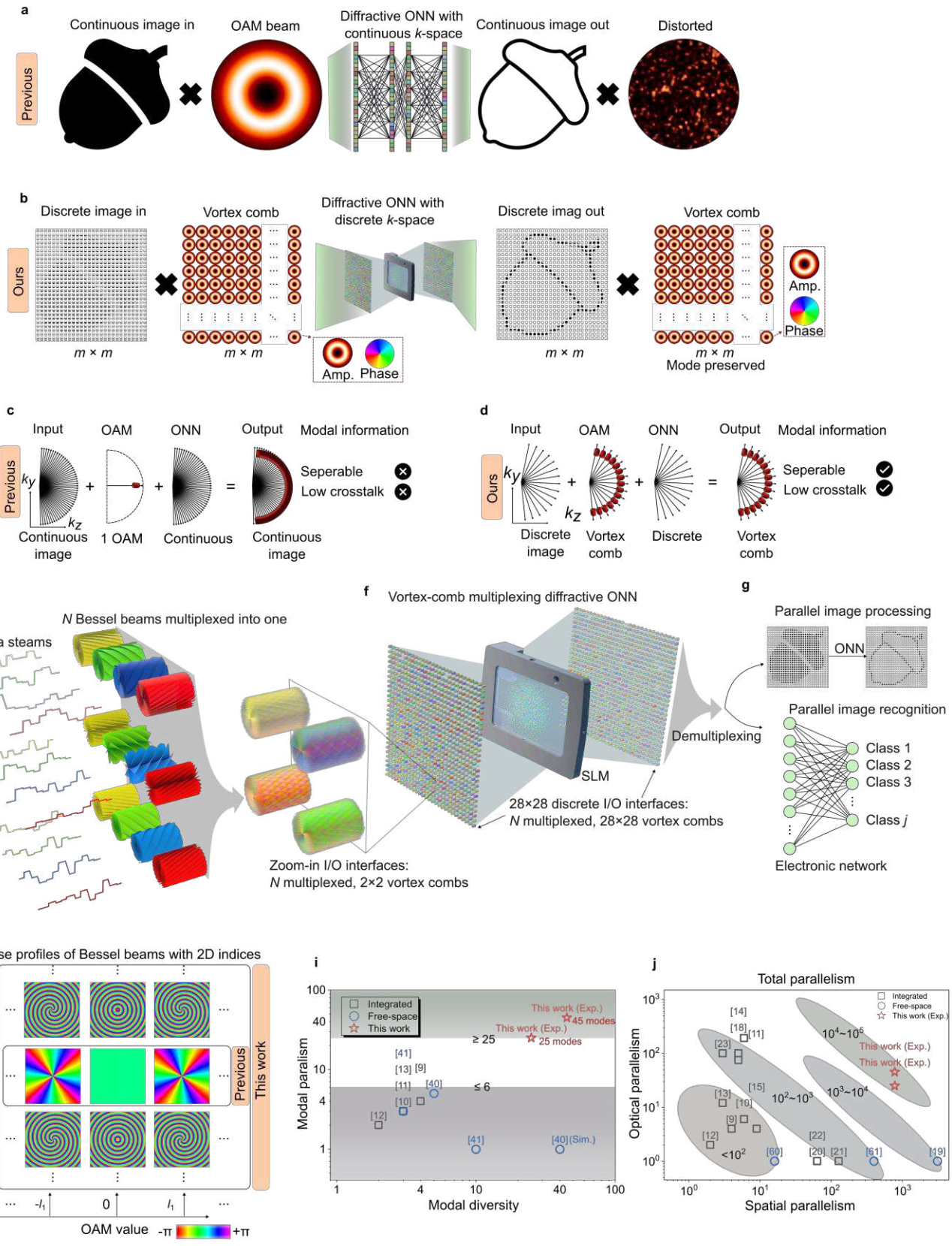
a
Previous
Continuous image in
OAM beam
Diffractive ONN with continuous k-space
Continuous image out
Distorted
b
Ours
Discrete image in
m × m
Vortex comb
m × m
Diffractive ONN with discrete k-space
Amp.
Phase
Discrete imag out
m × m
Vortex comb
Amp.
Phase
m × m
Mode preserved
c
Previous
Input
OAM
ONN
Output
Modal information
Continuous image
1 OAM
Continuous
Continuous image
Seperable
Low crosstalk
d
Ours
Input
OAM
ONN
Output
Modal information
Discrete image
Vortex comb
Discrete
Vortex comb
Seperable
Low crosstalk
e
N Bessel beams multiplexed into one
N data steams
Zoom-in I/O interfaces:
N multiplexed, 2×2 vortex combs
f
Vortex-comb multiplexing diffractive ONN
SLM
28×28 discrete I/O interfaces:
N multiplexed, 28×28 vortex combs
Demultiplexing
g
Parallel image processing
ONN
Parallel image recognition
Class 1
Class 2
Class 3
Class j
Electronic network
h
Phase profiles of Bessel beams with 2D indices
Radial index
OAM value
Previous
This work
i
Integrated
Free-space
This work
This work (Exp.)
45 modes
This work (Exp.)
25 modes
≥ 25
≤ 6
Modal paralism
Modal diversity
j
Total parallelism
Integrated
Free-space
This work (Exp.)
10⁴~10⁶
This work (Exp.)
This work (Exp.)
10³~10⁴
10²~10³
<10²
Optical parallelism
Spatial parallelism

Fig. 1. Concept of the OAM-preserving vortex-comb multiplexing ONN for parallel image processing and recognition. **a,** A continuous image encoded onto an OAM mode propagates through a diffractive ONN with a continuous momentum space, or *k* space. The modal information may not be preserved in the processed output image. **b,** In a momentum-space-matched diffractive ONN, both the input image and the processed output image are encoded onto vortex combs. Each comb consists of an array of Bessel beams with the same amplitude and phase profiles, allowing the modal information to be preserved throughout the ONN. **c, d,** Evolution of wave-vector distributions in momentum space for ONNs without and with momentum-space matching, respectively. Modal information can be preserved when neighbouring OAM components do not overlap in momentum space. **e,** Zoom-in view of N multiplexed vortex combs. Each vortex comb consists of an array of 28 × 28 Bessel beams. At each I/O port, N data streams are encoded onto N Bessel beams with distinct modal index pairs. I/O, input/output **f,** OAM-preserving vortex-comb multiplexing ONN using an SLM as a single diffractive layer for convolutional processing. **g,** Optically extracted features can be directly used for parallel image processing or fed into a lightweight electronic classification head for image recognition. **h,** Example spatial phase profiles of Bessel beams with two-dimensional modal index pairs. **i,** Comparison of modal diversity and modal parallelism of our ONN with those of other ONNs based on mode-division multiplexing (Table S2 for details). The two shaded regions indicate the modal parallelism of ≤ 5 and ≥ 25, respectively. **j,** Comparison of the total parallelism, including both spatial and optical parallelism, of our ONNs with that of other ONNs using discrete-port I/O interfaces. The four shaded regions indicate four ranges of total parallelism, with the corresponding values labelled within the regions. (Table S3 for details).

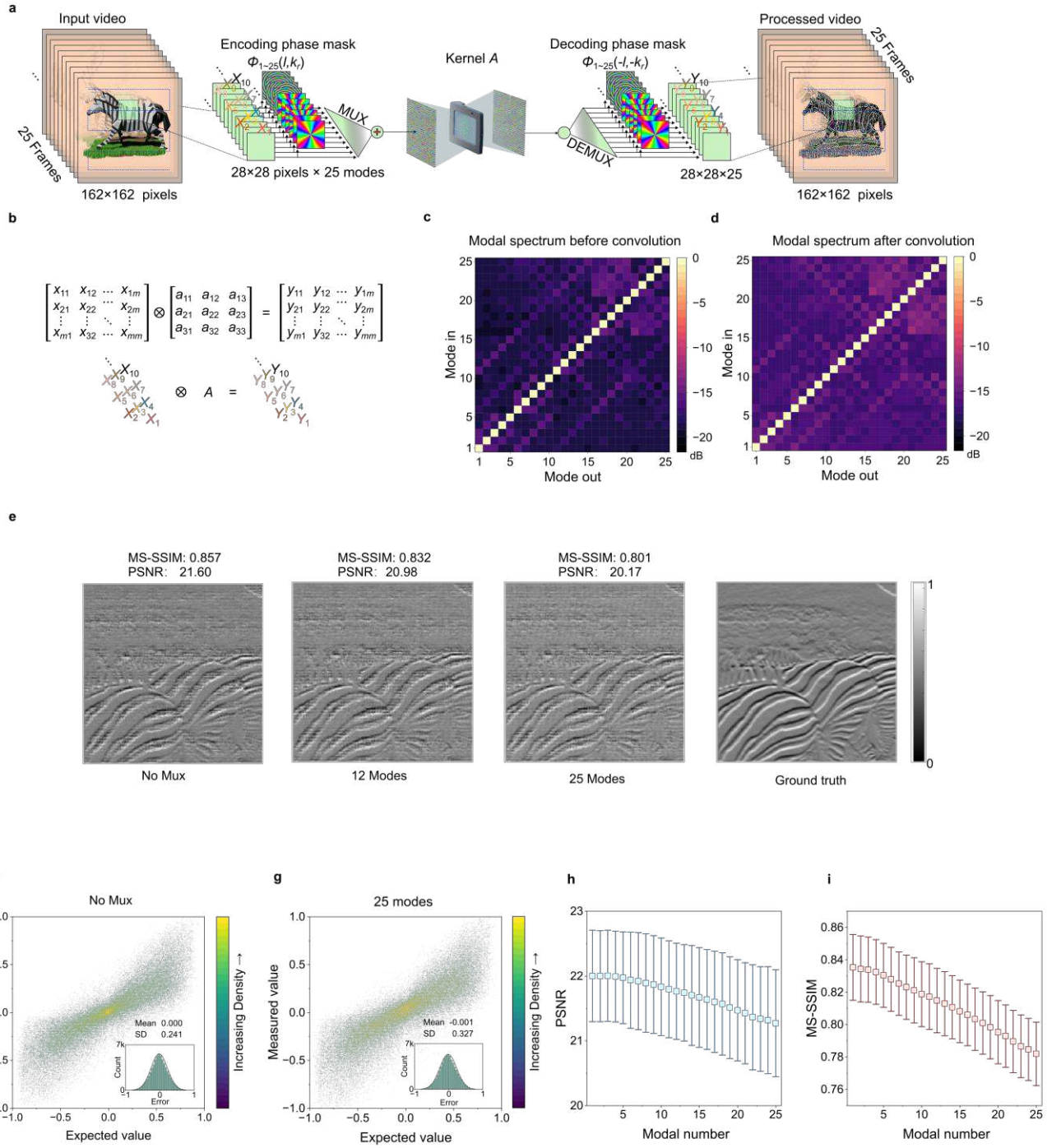

Fig. 2. Experimental demonstration of an OAM-preserving vortex-comb multiplexing ONN for 25-mode parallel image processing. **a,** Workflow of parallel convolutional processing of a zebra-walking video using 25 vortex combs with distinct (ℓ, k_r) index pairs. Each video frame is first divided into a series of patches with 28 × 28 pixels. At each patch position, 25 patches from 25 frames are encoded on independent vortex combs with distinct (ℓ, k_r) index pairs. SLM2 uses a diffractive-layer phase mask that processes the 25 patches in parallel, and converts the encoded vortex combs back into Gaussian combs for camera detection. **b,** Mathematical representation of the convolution operation. Each input patch can be written as a 28 × 28 matrix X. The diffractive layer represents a kernel matrix A (3 × 3 for example). The output matrix Y represents the convolution of X and A. In this image processing task, we send up to 25 input matrices X_i, and

a diffractive layer computes 25 convolution results of Y_i. **c, d,** Modal spectra of vortex combs before and after kernel convolution, respectively, illustrating the preservation of modal information. The two spectra show similar modal distributions, with the average modal crosstalk increasing from approximately -17.97 dB to -16.03 dB. **e,** A representative processed frame after kernel convolution using the ONN without multiplexing and with 12- and 25-mode multiplexing. **f, g,** Measured versus expected convolution results for ONNs without and with vortex-comb multiplexing, respectively. Insets show the corresponding error distributions. SD, standard deviation. **h, i,** Quantitative evaluation of the processed images using the peak signal-to-noise ratio (PSNR) and multiscale structural similarity (MS-SSIM), respectively.

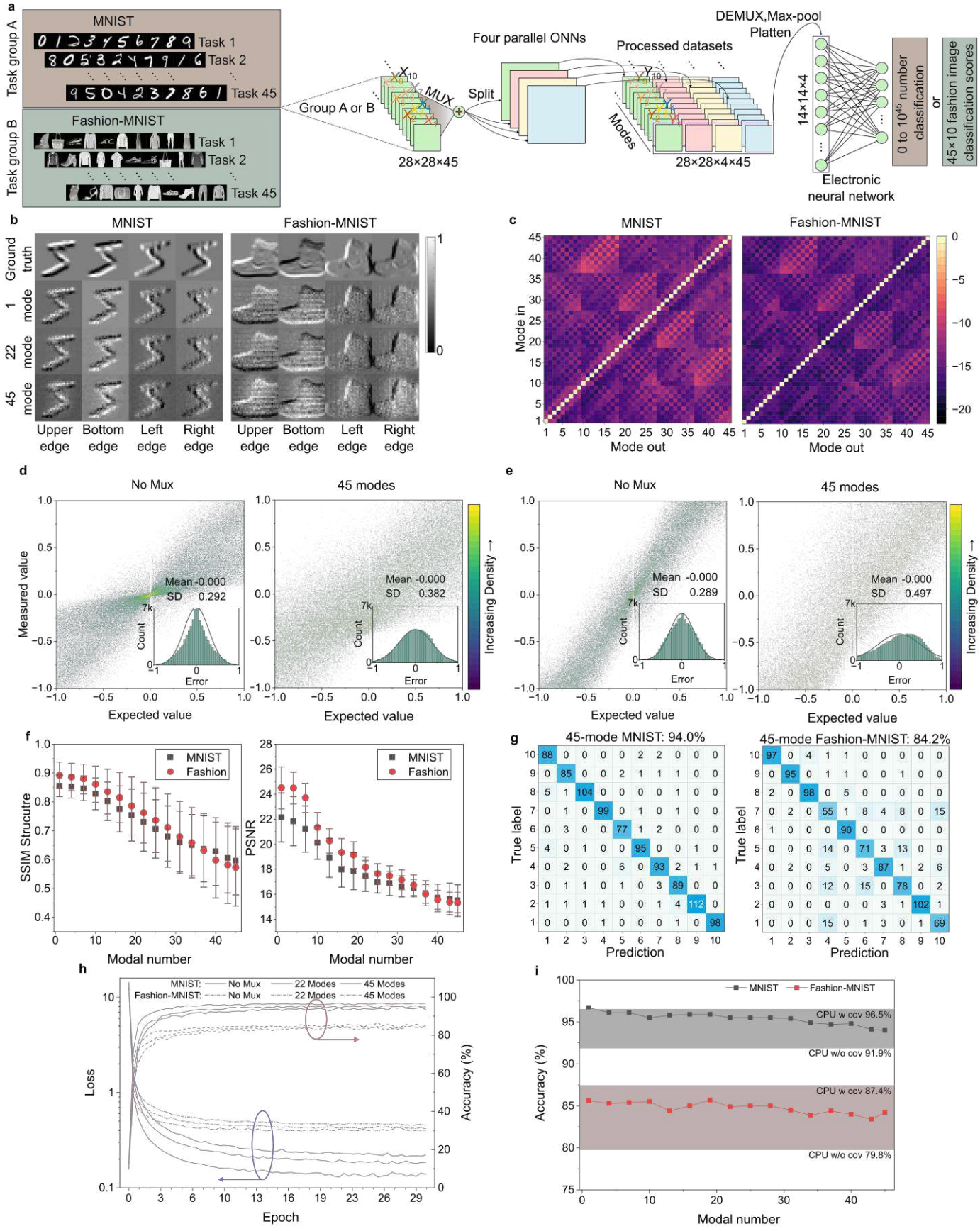
a
Task group A
MNIST
Task 1
Task 2
Task 45
Task group B
Fashion-MNIST
Task 1
Task 2
Task 45
Group A or B
MUX
28×28×45
Split
Four parallel ONNs
Processed datasets
Modes
28×28×4×45
DEMUX,Max-pool
Platten
14×14×4
Electronic
neural network
0 to 10^45 number classification
or
45×10 fashion image classification scores
b
MNIST
Fashion-MNIST
Ground truth
1 mode
22 mode
45 mode
Upper edge
Bottom edge
Left edge
Right edge
c
MNIST
Fashion-MNIST
Mode in
Mode out
d
No Mux
45 modes
Measured value
Expected value
Mean -0.000
SD 0.292
SD 0.382
Count
Error
Increasing Density
e
No Mux
45 modes
Expected value
Mean -0.000
SD 0.289
SD 0.497
f
MNIST
Fashion
SSIM Strucutre
PSNR
Modal number
g
45-mode MNIST: 94.0%
45-mode Fashion-MNIST: 84.2%
True label
Prediction
h
MNIST: No Mux 22 Modes 45 Modes
Fashion-MNIST: No Mux 22 Modes 45 Modes
Loss
Accuracy (%)
Epoch
i
MNIST
Fashion-MNIST
CPU w cov 96.5%
CPU w/o cov 91.9%
CPU w cov 87.4%
CPU w/o cov 79.8%
Accuracy (%)
Modal number

Fig. 3. Experimental demonstration of an OAM-preserving vortex-comb multiplexing ONN for 45-mode parallel image recognition. **a,** Workflow of parallel image classification of MNIST and Fashion-MNIST using 45 vortex combs with distinct (ℓ, k_r) index pairs. A 45-digit/object sequence represented by a 28 × 28 × 45 input tensor is first multiplexed using 45 independent vortex combs with distinct (ℓ, k_r) index pairs. Four parallel ONNs, each with a single SLM-based diffractive layer, process 45 vortex combs in parallel using convolution kernels for upper-, lower-, left- and right-edge detection, respectively. Finally, these optically extracted features are fed into an electronic FC layer for image classification. **b,** Measured convolution results using four detection kernels with the number of multiplexed modes N = 1, 22, and 45. **c,** Modal spectra of vortex combs after kernel convolution. The average modal crosstalk is approximately -14.11 dB and -15.37 dB for MNIST and Fashion-MNIST, respectively. **d, e,** Measured versus expected convolution results for MNIST and Fashion-MNIST, respectively. Insets show the corresponding error distributions. **f,** Quantitative evaluation of the processed images using PSNR and SSIM structure. **g,** Confusion matrices for MNIST and Fashion-MNIST with N = 45. **h,** The loss and classification accuracy as a function of training epoch, with both values calculated using testing sets containing 1,000 images. **i,** Classification accuracy versus the number of multiplexed modes N. The shaded regions indicate the accuracy ranges of all-electronic neural networks without and with a convolutional layer. As N decreases from 45 to 1, the classification accuracies are 2.1-4.8% and 4.4-5.8% higher than those of the all-electronic neural-network baseline without a convolutional layer for MNIST and Fashion-MNIST, respectively. w., with; w/o, without.

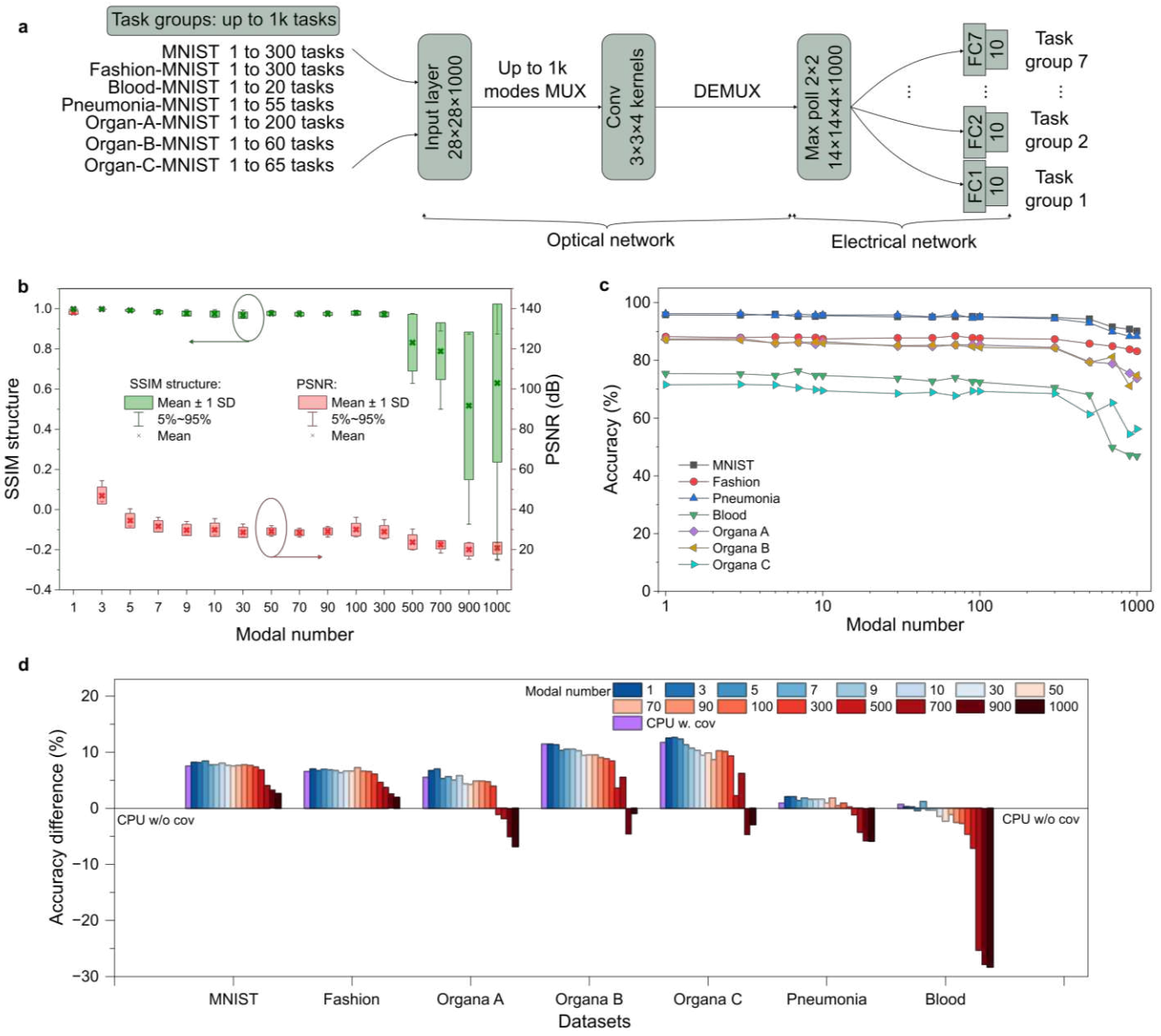

Fig. 4. Simulation of optoelectronic neural networks using OAM-preserving vortex-comb multiplexing for 1,000-mode parallel image recognition across seven datasets. **a,** Workflow of parallel image classification on seven datasets using 1,000 independent vortex combs with distinct (ℓ, k_r) index pairs. **b,** Quantitative evaluation of the processed images using PSNR and SSIM structure, calculated over simulated processed images from all datasets. **c,** Classification accuracies for seven datasets as functions of the number of multiplexed modes N. **d,** Accuracy differences for seven datasets with different N values, compared with those of all-electronic baselines without and with convolutional layers. w., with; w/o, without; cov, convolutional layer.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Movie1.mp4
- OAMcombSupplement.docx