nature communications

Article

https://doi.org/10.1038/s41467-024-53371-6

Ultra-broadband all-optical nonlinear activation function enabled by $MoTe_2$/optical waveguide integrated devices

Received: 12 January 2024

Accepted: 10 October 2024

Published online: 19 October 2024

Check for updates

Chenduan Chen[1,6], Zhan Yang[2,3,6], Tao Wang[1], Yalun Wang[1], Kai Gao[1], Jiajia Wu[1], Jun Wang[2,3], Jianrong Qiu[4] & Dezhi Tan[1,5]

All-optical nonlinear activation functions (NAFs) are crucial for enabling rapid optical neural networks (ONNs). As linear matrix computation advances in integrated ONNs, on-chip all-optical NAFs face challenges such as limited integration, high latency, substantial power consumption, and a high activation threshold. In this work, we develop an integrated nonlinear optical activator based on the butt-coupling integration of two-dimensional (2D) $MoTe_2$ and optical waveguides (OWGs). The activator exhibits an ultra-broadband response from visible to near-infrared wavelength, a low activation threshold of 0.94 μW, a small device size (~50 μm^2), an ultra-fast response rate (2.08 THz), and high-density integration. The excellent nonlinear effects and broadband response of 2D materials have been utilized to create all-optical NAFs. These activators were applied to simulate MNIST handwritten digit recognition, achieving an accuracy of 97.6%. The results underscore the potential application of this approach in ONNs. Moreover, the classification of more intricate CIFAR-10 images demonstrated a generalizable accuracy of 94.6%. The present nonlinear activator promises a general platform for three-dimensional (3D) ultra-broadband ONNs with dense integration and low activation thresholds by integrating a variety of strong nonlinear optical (NLO) materials (e.g., 2D materials) and OWGs in glass.

Artificial intelligence is an important form of technology for human progress, which has a profound impact on the world's economic recovery, social progress, and the life of all people[1–3]. However, the resulting problem is that traditional computers face serious challenges in terms of energy consumption[4], computility[5] and the limitations of Moore's Law are forcing people to find a way to overcome this crisis[6,7]. Photonic integrated devices, on the other hand, have the inherent properties of high speed of photon transport, massive parallelism, and low power consumption, and have therefore received much attention as a promising candidate to realize fast and effective computation[8,9].

Artificial neural networks have been introduced to photonics, i.e., optical neural networks (ONNs), which have been shown to be feasible in photonic integrated circuits and free-space optics[10–18]. These networks utilize both linear and nonlinear operations. Linear operations are implemented in a variety of ways, such as planar light conversion[19], Mach-Zehnder interferometer (MZI) networks[13], wavelength-division multiplexing[20], and networks based on phase-change materials

[1]Zhejiang Lab, Hangzhou, Zhejiang 311121, China. [2]Aerospace Laser Technology and System Department, CAS Key Laboratory of Materials for High-Power Laser, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China. [3]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China. [4]College of Optical Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China. [5]School of Materials Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China. [6]These authors contributed equally: Chenduan Chen, Zhan Yang. e-mail: wctdz@zju.edu.cn

(PCM)[12]. The nonlinear operation involves the implementation of nonlinear activation functions (NAFs) on optical hardware platforms and is crucial for the training and decision-mapping process of the network. Existing solutions are categorized as optoelectronic or all-optical approaches. The response speed of optoelectronic devices limits the NAF based on optoelectronic methods and requires additional power consumption[21–23]. Therefore, an all-optical approach is required to realize large-scale scaling and high-speed applications of neuronal networks.

One of the challenges in achieving an all-optical approach is the insufficient nonlinear response of light in the medium for the photonic device to function as a neuron[24–26]. Although strong nonlinear materials have been discovered and structures that enhance nonlinearity have been designed through nonlinear activation using PCM, the slow recovery time of PCMs (approximately nanoseconds) and the need for additional energy stimulation hinder its suitability for high-speed optical computation[27]. Another form of all-optical NAF involves using two-dimensional (2D) material dispersions in combination with a rib waveguide relying on the evanescent field coupling (EFC)[28]. For example, the third-order nonlinearity of 2D MXene can be used on a silicon-based waveguide to realize the NAF[26]. However, the intrinsic EFC is generally weak along with large loss, which leads to a high activation threshold and high energy consumption, losing one of the most important advantages for optical computing. Furthermore, this approach is only applicable for 2D integration, and the working waveband is limited by the band gap of the waveguides (e.g., silicon). Therefore, there is an urgent requirement to establish a platform with a rapid response rate in a wide spectrum and 3D integration capability for efficient and low energy consumption all-optical NAFs. This will facilitate the development of miniaturized, intelligent, high-speed, and low-power all-optical neural networks.

We have proposed an integrated nonlinear optical activator that is based on butt-coupled 2D materials and OWGs obtained by femtosecond (fs) laser direct writing (FLDW) in glass. FLDW is an advanced rapid prototyping technology that enables the streamlined production of large-scale photonic circuits using a single control program[29–35], which promotes intricate 3D internal machining and yields notable gains in spatial efficiency. The transparent characteristic of glass also enables the device to work in an ultra-broadband. For example, the typical transparent glass usually possesses a very wide bandgap (e.g., $E_g > 3.5$ eV for Eagle glass, $E_g > 8.5$ eV for silica glass) and the OWGs in the glass can work in a range from visible to near-infrared wavelength[35–37]. It is crucial that the waveguide optoelectronic field is tightly coupled to the medium. One possible solution is to utilize a near-surface waveguide or rib waveguide[26,38], but this approach inherits the issues mentioned above. In contrast to the EFC, we realized that butt-coupling can enhance the light-matter interaction with the light directly transmitting through the integrated optical active components for the optical information modulation and decrease the optical loss. Furthermore, the glass matrix with embedded OWGs can act as an excellent substrate to support the 2D materials for integration on the edge side, which is difficult to achieve with other waveguides, such as silicon waveguides. 2D material is considered as a potential functional material for the NAF due to broadband nonlinear optical effect[38,39]. Consequently, an all-optical NAF based on 2D crystals/OWG devices could be realized by butt-integrating layered 2D materials to the end-face of OWGs via a dry transfer technique, which is very favorable for the on-chip integration of nonlinear computing devices.

In this study, we have developed nonlinear activators relying on the butt-coupling by integrating the 2D transition metal dichalcogenides ($MoTe_2$) and OWGs. The activation function is achieved through the utilization of the saturable absorption (SA) and reverse saturable absorption (RSA) effect generated in the integrated system due to the interaction of light with a single-crystal 2D-$MoTe_2$. To validate the principle, we carried out a benchmark machine-learning task. The MNIST handwritten digit recognition achieves remarkable accuracy using the optically measured activation function. Moreover, the same activation function is employed for a more intricate task, i.e., the classification of color images (CIFAR-10), and exhibits exceptional precision.

Results and discussion

Preparation of nonlinear activators

The on-chip device that achieves all-optical NAF consists of OWG and layered 2D material, in which the OWGs were prepared by FLDW. The experimental setup is shown in Fig. 1a, and the detailed preparation parameters are shown in the Methods section. After laser writing, the two ends of the samples are polished to obtain usable OWGs, whose top and side views are shown in Fig. 1b, c. The obvious discreteness of the OWG and the glass can be observed in the optical picture, which is

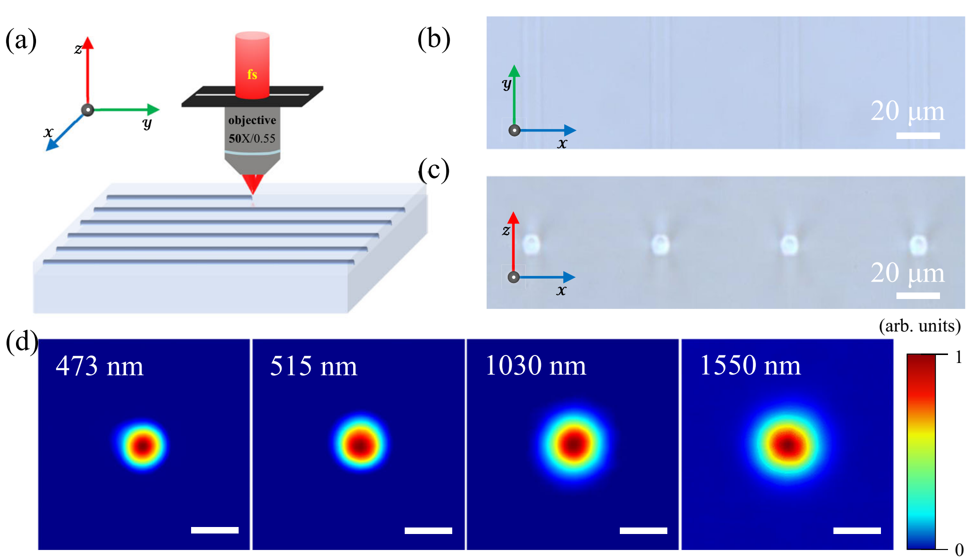

Fig. 1 | Preparation and characterization of optical waveguides (OWGs) in glass. **a** Schematic of femtosecond laser direct writing of OWGs in glass. **b** Top and (**c**) side view of OWGs. **d** The mode field images of OWGs in the range from visible to near-infrared range. Scale bar: 10 µm. The color bar represents the normalized intensity of light passing through the OWGs.

https://doi.org/10.1038/s41467-024-53371-6

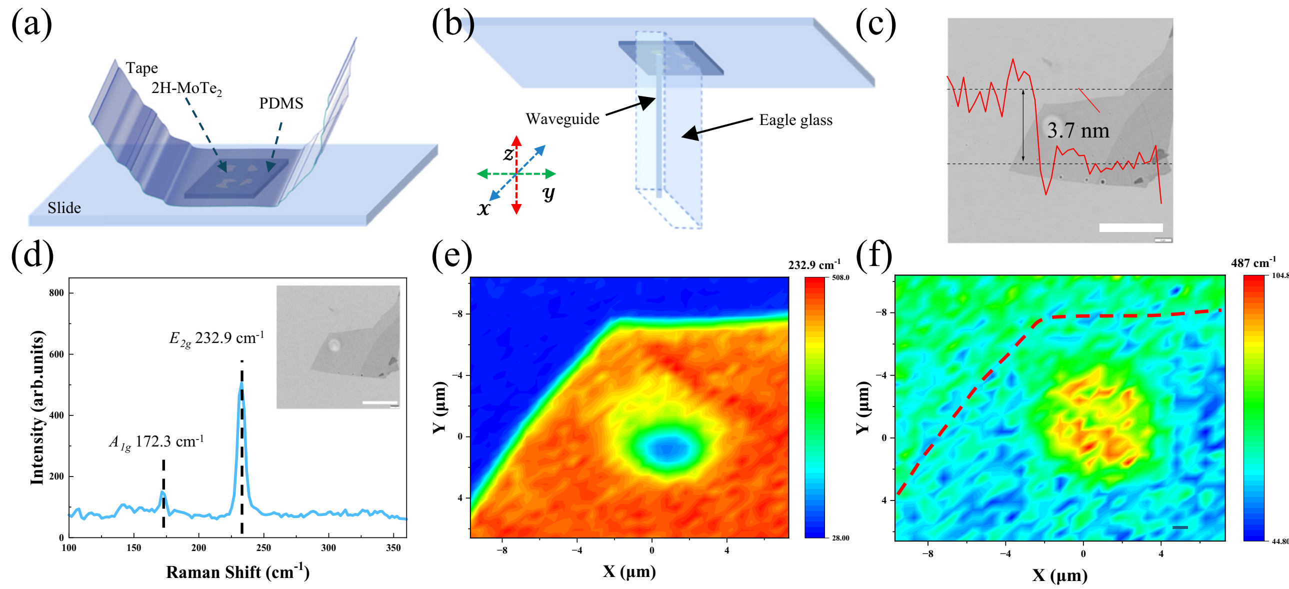

Fig. 2 | Nonlinear activator preparation process. a The process of preparing a few layers of $MoTe_2$, where PDMS stands for polydimethylsiloxane film. **b** Transferring 2D $MoTe_2$ to OWG's end face to form nonlinear activators. **c** Optical picture of a nonlinear activator and information on the thickness of few-layered $MoTe_2$, where the thickness information was obtained by atomic force microscopy. Scale bar: 20 μm. **d** Raman spectra of a few layers of $MoTe_2$. The scale bar in the inste is 20 μm. **e** 232.9 cm^{-1} mapping ($MoTe_2$'s characteristic peak) and (**f**) 487 cm^{-1} mapping (OWG's characteristic peak). The red dashed line is the outline of 2D $MoTe_2$.

mainly caused by the refractive index change due to the laser direct writing in the glass. The typical Raman spectra and mappings are shown in Supplementary Fig. 1, and the increase of the Raman intensity at around 487 cm^{-1} and 604 cm^{-1} confirms the densification of matter in the local region along with an increase of refractive index, which is related to the increase of the four-membered and three-membered rings inside the glass after the FLDW[40], resulting in low-loss light guiding performance[41].

To verify the usability of the OWGs, we investigated its transmission mode field and optical loss by a fiber-butt coupling method (Supplementary Fig. 2a). Initially, we employed an optical fiber to couple light into the OWG, and subsequently collected the light transmitted within the OWG using an objective lens to analyze the impact of light passing through the OWG. The insertion loss of the 1-cm OWG is 0.49 dB and 0.54 dB for 520 and 1550 nm, respectively. The propagation loss was determined to be 0.11 ± 0.02 dB/cm and 0.18 ± 0.02 dB/cm for 520 nm and 1550 nm, respectively (Supplementary Fig. 2b). Consequently, the potential coupling loss between the waveguide and fiber is 0.38 dB/facet. These results confirmed that the written waveguides show low loss performance in a broad range, which provides an excellent platform for photonic integrated devices.

Notably, in the visible to near-infrared region, the optical mode field that traversed through the OWG exhibited excellent symmetry, with a circular symmetry of $R_x/R_y \sim 1$, where R_x and R_y are the radii of the mode field in the X and the Y direction, as shown in Fig. 1d. According to the results, the OWGs with high symmetrical mode field in the visible to near-infrared range were obtained by FLDW technique, which satisfied the application of nonlinear activators in an ultra-broadband wavelength range (> 1000 nm).

For another component of the on-chip nonlinear activator, the 2D material, we selected layered 2H-$MoTe_2$ due to its environmental and operation stability at room temperature[42,43], which ensures stable and efficient use of the device, and potential applications in a broad wavelength range from visible to near-infrared[44,45]. 2D $MoTe_2$ exhibits sub-picosecond carrier lifetimes[46,47], predicting the material's potential for ultrafast computing in the field of optical computing. 2D $MoTe_2$ can significantly enhance the response range and device versatility by adjusting the bias[48,49], which brings inspiration to extend the application range and increase the flexibility of the nonlinear activator. However, the traditional EFC activator can only use a small partial of the light in the waveguides and is not able to make full use of the optical modulation capability of 2D materials. We proposed a butt-coupling activator by integrating the 2D materials on the end surface of the OWGs in the glass. The preparation method of the nonlinear activator is depicted in Fig. 2a–c. Initially, we employed the mechanical exfoliation technique to acquire few-layered $MoTe_2$ from bulk $MoTe_2$, demonstrating its 2H structure through XRD results in Supplementary Fig. 3. Subsequently, employing a 2D material transfer system, we moved the few-layered $MoTe_2$ to the end face of the OWG, creating the nonlinear activator with strong butt-coupling, as shown in Fig. 2c. Simultaneously, to explore the coupling between 2D $MoTe_2$ and the OWG, we performed Raman characterization on the 3.7 nm thick $MoTe_2$ activator, as depicted in Fig. 2d–f. The characteristic peaks of $MoTe_2$, specifically A_{1g} (172.3 cm^{-1}) and E_{2g} (232.9 cm^{-1}), aligned with previous reports[50]. Figure 2e displays the mapping of E_{2g} characteristic peak, distinctly illustrating the morphology of $MoTe_2$ and the outline of the OWG. Worthy, the decrease in the Raman signal from A_{1g} peak at the OWG domain results from the passing of signal through the OWG. Furthermore, in the feature peak mapping results at 487 cm^{-1}, the outline of the OWG is discernible, suggesting effective contact between 2D $MoTe_2$ and the OWG.

Measurement of activation functions

To detect the NLO effect of 2D $MoTe_2$ that was integrated to the end of the OWG, we used the double-balanced detection method as in Fig. 3a. Firstly, the power of the initial light (P_0) is controlled by a tunable optical attenuator. Before being coupled into waveguides, the initial light was divided into two parts by a beam splitter with a splitting ratio of 9:1. The real power of the excitation light was calculated by $90\%P_0$ - P_{loss}, which entered the optical power meter after passing through the nonlinear activator, and P_{loss} was the insertion loss including the propagation loss and coupling loss between the fiber and the OWG without $MoTe_2$ that was determined above. The light with a power of $10\%P_0$ was detected by another optical power meter as a reference. It is noteworthy that this method is able to eliminate the nonlinear effect of

https://doi.org/10.1038/s41467-024-53371-6

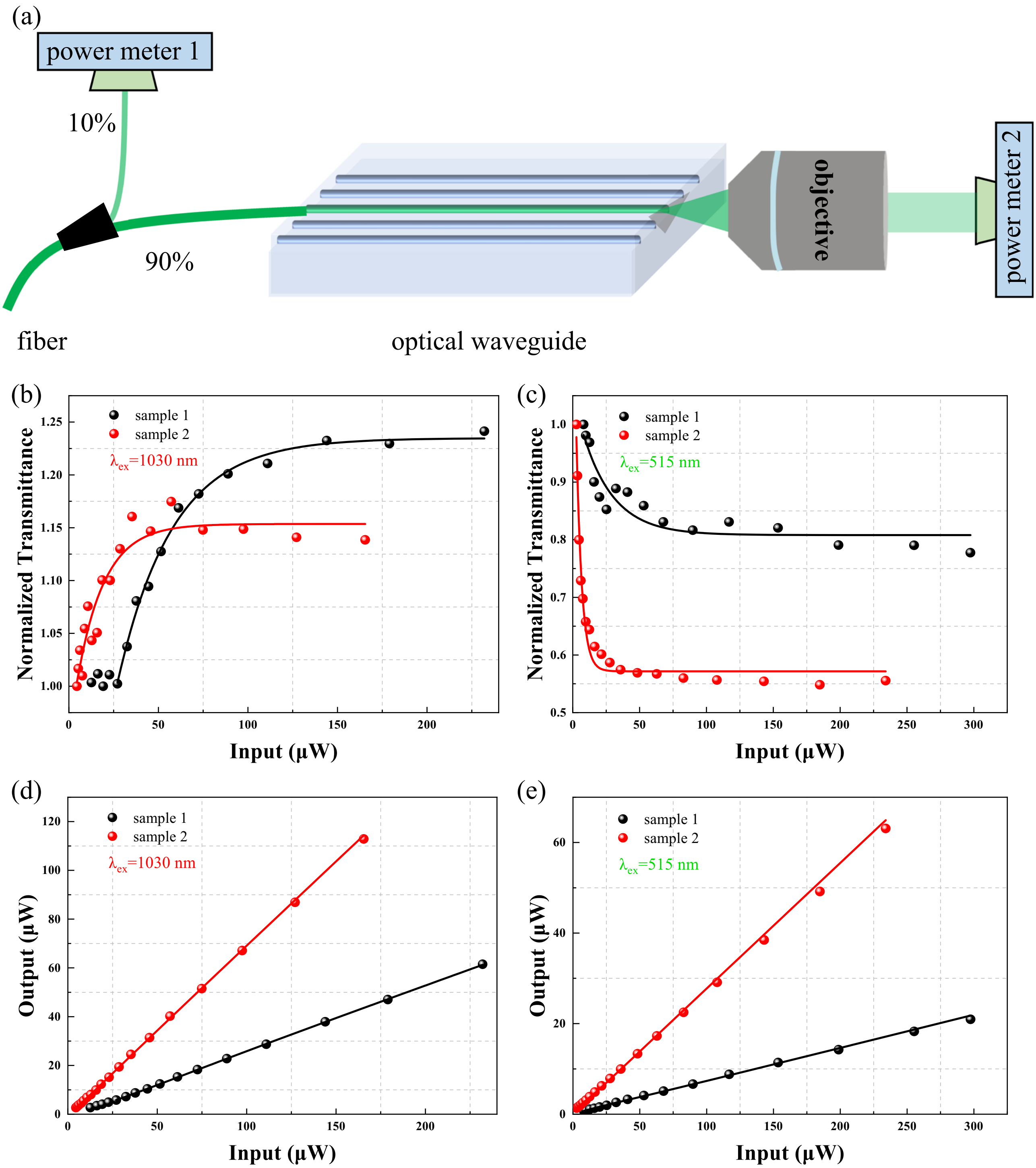

Fig. 3 | Nonlinear optical (NLO) measurements of nonlinear activators. **a** Schematic diagram of the double-balanced detection setup for measuring the NLO properties of integrated devices. Optical response results of few-layer $MoTe_2$ (**b**) under 1030 nm fs excitation and (**c**) under 515 nm fs excitation. The solid line is the fitted data. The nonlinear correlation between optical input and output of the activators (**d**) under 1030 nm fs excitation (**e**) under 515 nm fs excitation. The solid line represents the nonlinear activation function obtained from the result of Fig. 3b, c.

the OWG, ensuring that the resulting nonlinear signal is solely attributed to $MoTe_2$ (see the discussion about Supplementary Fig. 4). Based on the NLO theory, the absorption coefficient of 2D $MoTe_2$ is described as $\alpha = \frac{\alpha_0}{1+I/I_{sat}} + \beta I$, where α is the total absorption, α_O is a linear absorption coefficient, I_{sat} is SA intensity, β is nonlinear absorption coefficient, the first term on the right-hand side denotes SA and the second term denotes RSA.

For analyzing the SA response, the relationship between transmittance T and input light intensity I can be described as[51,52]:

$$T(I) = 1 - \Delta T \times \exp(-I/I_{sat}) - T_{ns} \quad (1)$$

where ΔT is the modulation depth, I is the input intensity, and T_{ns} is the non-saturated loss. For the RSA response coexistence, we could get the following:

$$T(I) = 1 + \Delta T \times \exp(-\beta I) - T_{ns} \quad (2)$$

Figure 3b, c demonstrate the normalized transmittance for two samples (named as S1 and S2) as a function of the incident light intensity under 1030 nm and 515 nm fs laser excitation, respectively. The optical images and the thickness of the few-layered $MoTe_2$ are shown in Supplementary Fig. 5. It is worth mentioning that, due to varying linear transmittance between S1 and S2, as displayed in Supplementary Fig. 6, we have opted for the normalized transmittance with respect to changes in incident light intensity to facilitate the presentation of the results.

The test results in Fig. 3b unequivocally show that, under 1030 nm fs light excitation, the transmittance of few-layered $MoTe_2$ rises with an increase in the incident light intensity, which is attributed to the photobleaching (PB) effect induced by the Pauli exclusion principle, namely, the SA effect. The modulation depth ΔT for S1 and S2 is 0.11 and 0.12, respectively, and the SA intensity I_{sat} is 37.17 μW and 15.17 μW, respectively. Furthermore, we confirm that the nonlinear photoresponse of the integrated devices remains unchanged after storage for one year at room temperature, as shown in Supplementary Fig. 7, which indicates the good stability of 2H-$MoTe_2$. It is interesting that in contrast to the 1030 nm excitation, the optical response under 515 nm fs light excitation is RSA effect, and the transmittance of $MoTe_2$ decreases with the increase of incident light intensity (Fig. 3c). The RSA effect has diverse causes, including the two-photon absorption effect and the excited state absorption effect[53,54], which are elaborated in the Supplementary Fig. 8. The modulation depths ΔT for S1 and S2 are 0.02 and 0.39, respectively. To reveal how the nonlinear properties of the $MoTe_2$/OWG integrated system are affected by the thickness of $MoTe_2$, we have carried out systematical investigations by integrating $MoTe_2$ with varied thicknesses to OWGs and tested their nonlinear photoresponse. All the $MoTe_2$ with a thickness from 10 to 80 nm in the integrated devices exhibited SA behavior under 1030 nm fs light excitation (Supplementary Fig. 9a–h). I_{sat} first decreases and then increases with the increase of thickness, while ΔT changes oppositely, as shown in Supplementary Fig. 9i. The obtained smallest (largest) I_{sat} (ΔT) reaches 2.89 μW (0.177) with the thickness of $MoTe_2$ at approximately 22.3 nm when excited by 1030 nm fs laser. The dependence of the third-order nonlinear susceptibility imaginary part ($Im\chi^{(3)}$) on the thickness may be responsible for the thickness dependence of I_{sat} (see the discussion about Supplementary Fig. 9)[55]. In addition, $MoTe_2$ exhibited SA effects under 800 nm and 1550 nm fs laser excitation, as shown in Supplementary Fig. 10. As a result, the nonlinear response of $MoTe_2$ can operate in an ultra-broad wavelength range with the width larger than 1000 nm, thus supporting the ultra-broadband response of $MoTe_2$/OWG devices.

Near all the 2D materials exhibit similar NLO response, SA, when the energy of the excitation light is bigger than that of the band gap[56,57]. However, the NLO response of $MoTe_2$ is opposite at 1030 and 515 nm. In fact, under 515 nm optical excitation, the nonlinear response changes from RSA to SA as the power density increases, as shown in Supplementary Fig. 11. This is due to the competition between the many-body effect induced absorption enhancement and non-thermalized carriers induced bleaching[53].

From the above nonlinear results, we can get four NAFs in the visible to near-infrared range. The nonlinear threshold was defined as the optical power required to generate a 50% change in the power transmission relative to the transmission with null input[58,59]. The nonlinear thresholds for the four activation functions are, therefore, 47.8 μW (S1 excited by 1030 nm laser), 14.2 μW (S2 excited by 1030 nm laser), 21.7 μW (S1 excited by 515 nm laser) and 5.5 μW (S2 excited by 515 nm laser), respectively. In addition, the activation threshold could be down to 4.9 μW by optimizing the thickness of $MoTe_2$, as shown in Supplementary Fig. 12. Simultaneously, we illustrate the nonlinear correlation between optical input and output as the NAFs in the nonlinear activation unit, depicted in Fig. 3d, e. Therefore, the NAFs can be defined as $f(I) = T(I) \times I$, which play a unique role in the nonlinear operation. In the following chapters, these NAFs are used to recognize and classify pictures in ONNs.

To study the response time of the all-optical device, we have investigated the kinetic process of carrier relaxation in the core material ($MoTe_2$) of the nonlinear activator using transient absorption spectroscopy (TAS) and pump-probe techniques. The relaxation process of photogenerated carriers in the spectral range of 500–900 nm was investigated in detail using 400 nm as the pump light (Fig. 4a–c). The pump-probe results were obtained with 520 nm as the pump light and 1040 nm as the probe light (Fig. 4d). The positive photoinduced absorption (PIA) signal was observed in the range of 500–600 nm. The relaxation process of the photogenerated carrier initially exhibited a negative PB signal in the range of 700–900 nm caused by the state filling effect excited by the pump light (Fig. 4b).

To further investigate the carrier relaxation processes, we fitted and analyzed the TAS (Fig. 4c for 500–900 nm) and pump-probe results (Fig. 4d for 1040 nm). The relaxation kinetic process of 2D $MoTe_2$ can be expressed as:

$$g(t) = A_1 \exp\left(-\frac{t}{\tau_1}\right) + A_2 \exp\left(-\frac{t}{\tau_2}\right) \quad (3)$$

where $g(t)$ is the transient signal at the probe wavelength, A_1, and A_2 are the relative amplitudes, t is the delay time, and τ_1 and τ_2 are the lifetimes of the excited carriers for the fast and slow components of the relaxation process respectively. We attributed the first-lifetime process (τ_1) to the thermal equilibrium of the hot carriers with the lattice. The second-lifetime process (τ_2) is the electron-hole recombination process[60,61]. The time scale of the prompt relaxation process is in the picosecond range from 0.2 to 18.3 ps, which depends on the wavelength and is in agreement with the previously reported values[46,47]. Especially, the fast response time is 0.48 ps and 0.69 ps for 515 nm and 1040 nm, respectively. The fast carrier relaxation process determines the recovering rate of the nonlinear activator to its original state within τ_1. Correspondingly, the response rate of this all-optical nonlinear activator is up to 2.08 THz at 515 nm and 1.45 THz at 1040 nm, which allows for ultrafast information processing.

Principle verification of on-chip NAFs

The schematic of the prepared ONN structure in the glass is shown in Fig. 5a, which includes an input layer, a linear operation, a nonlinear operation and an output layer. The optical interference unit performs the linear operation, and the nonlinear operation is performed by the nonlinear activation unit. The optical interference unit consists of MZIs array, and the nonlinear activation unit consists of 2D-$MoTe_2$ and OWG. The calculation process of the constructed optical neuron is shown in Fig. 5b. The optical signal in the input layer passes (Z_i^{l-1}) through the MZIs array to complete the linear matrix operation ($\sum_i w_{ji}^{(l)} Z_i^{(l-1)}$), that is a_j^l, the input optical signal is weighted and combined. Then, the optical signal is operated through nonlinear operation ($f(a_j^l)$), and finally the output light is obtained (Z_j^l). Therefore, we can take advantage of the NLO properties of 2D $MoTe_2$ to perform nonlinear activation operations.

The Pytorch framework was utilized to perform picture recognition and classification using the experimentally measured SA and RSA NAFs. Specifically, the MNIST dataset and CIFAR-10 dataset were used for the feasibility verification of our all-optical NAFs, as shown in Figs. 6 and 7. The network's performance is evaluated based on its accuracy in classifying images from the MNIST and CIFAR-10 datasets. The MNIST

https://doi.org/10.1038/s41467-024-53371-6

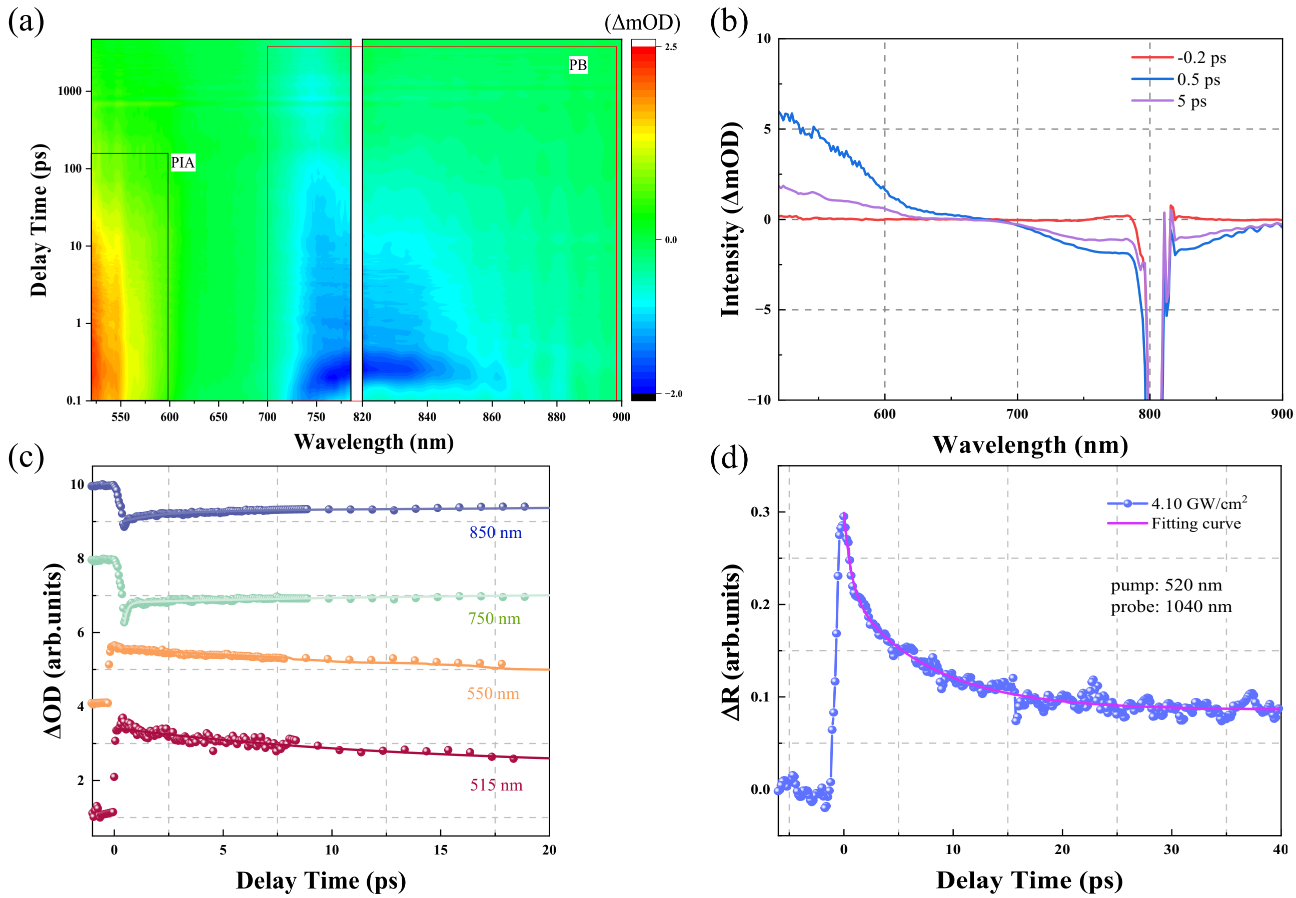

Fig. 4 | Transient photoresponse time of the $MoTe_2$/OWG devices. **a** Transient absorption spectrum of $MoTe_2$ excited by 400 nm at a pulse width of 200 fs. The change in optical absorption is in units of optical density (OD or mOD, where 1 OD = 1×10^3 mOD.) obtained from the probe intensities when the sample is pumped and unpumped. Black box range represents the positive photoinduced absorption (PIA) signal. Red box range is the negative photobleaching (PB) signal. **b** Transient absorption spectra with different pump-probe time delays. **c** Decay profiles of transient differential absorption monitored at different wavelengths. The solid lines represent the corresponding fitting results. **d** Transient reflection dynamics of $MoTe_2$ at 520 nm with a pump intensity of 4.10 GW cm^{-2}, while the solid lines represent the fitting. ΔR is the change in reflection.

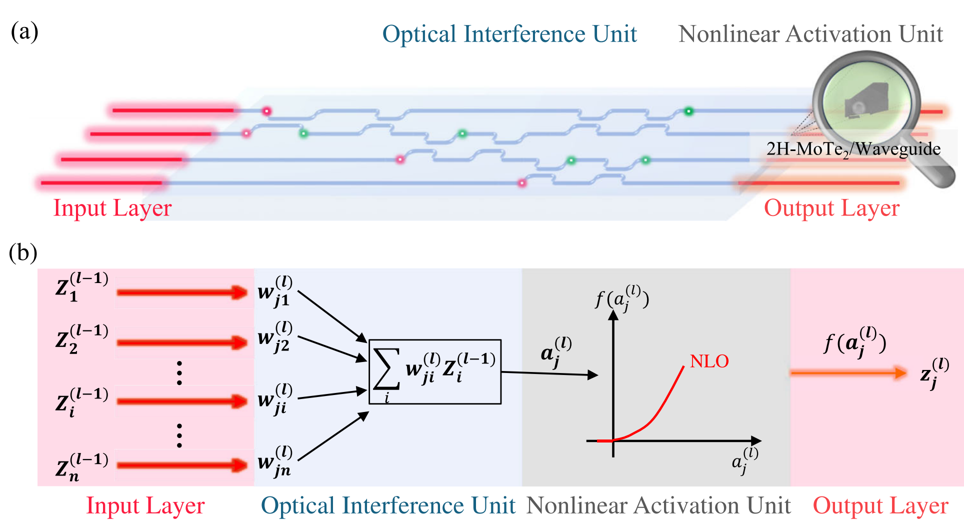

Fig. 5 | Optical neuron computing process. **a** Optical interference and nonlinear activation units comprising optical neural networks (ONNs). **b** Schematic diagram of the computation of artificial neurons. Z_i^{l-1} is the optical signal in the input layer. $w_{ji}^{(l)}$ stands for the linear matrix operation in the optical interference unit. a_j^l is the result of the linear matrix operation. $f(a_j^l)$ stands for the optical signal operated through nonlinear operation. Z_j^l is the final output light.

https://doi.org/10.1038/s41467-024-53371-6

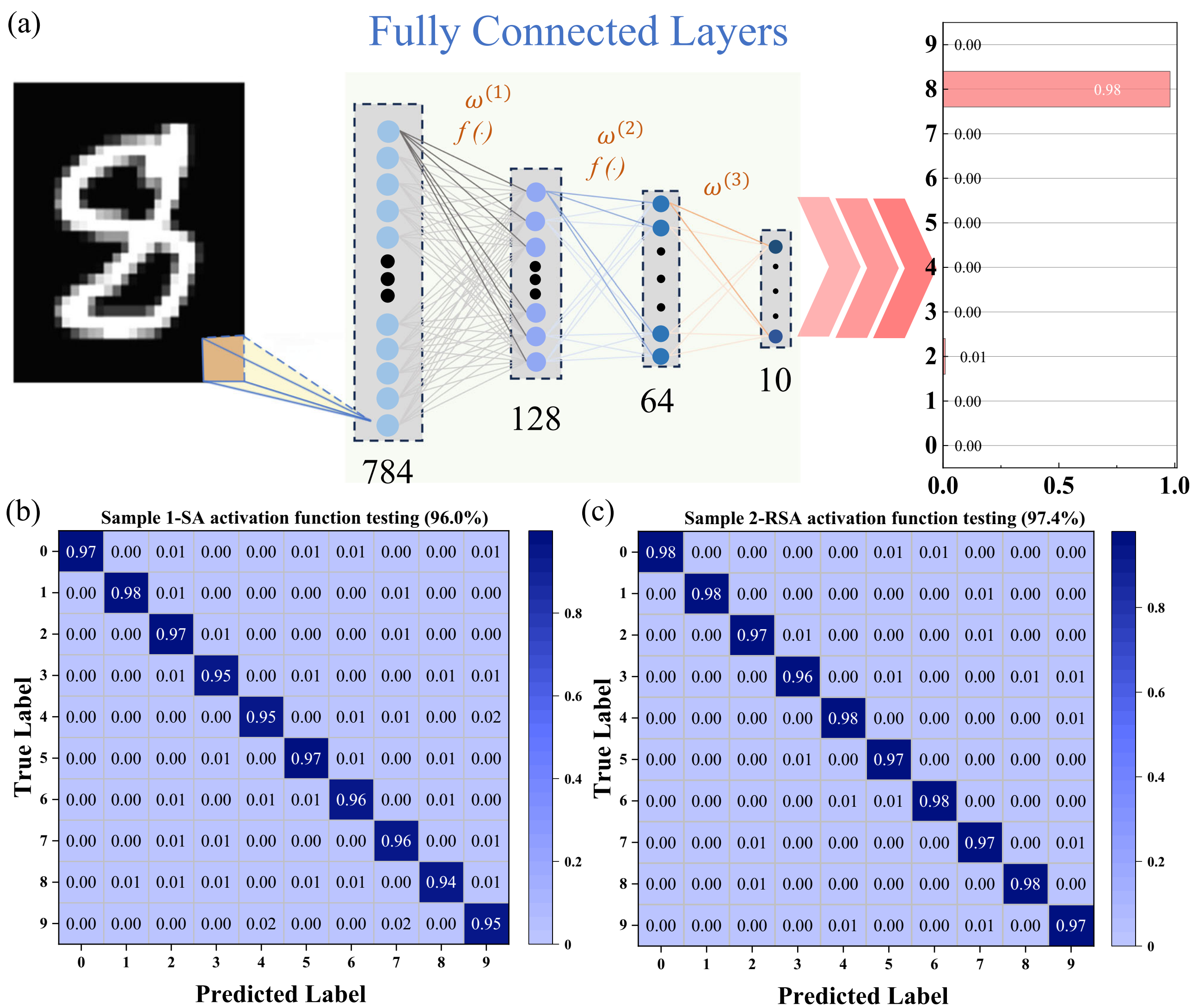

Fig. 6 | Classification performance on MNIST. MNIST is Modified National Institute of Standards and Technology database. **a** Fully connected layer network structure for the MNIST classification task. Construct the confusion matrix of the NAF test dataset from (**b**) the saturable absorption and (**c**) reverse saturable absorption activators.

dataset is comprised of 70,000 grayscale images of handwritten digits, encompassing 60,000 images intended for training and 10,000 images designated for testing purposes. In parallel, the CIFAR-10 dataset is composed of 60,000 color images, with 50,000 images allocated for training purposes and 10,000 images reserved for testing. Images of the MNIST dataset are grayscale with a size of 28 × 28 pixels, and the CIFAR−10 dataset images are RGB-colored with 32 × 32 pixels in size.

For the MNIST dataset, the network consisted of a full connection layer that was three-layered, as shown in Fig. 6a. The initial layer was made up of 128 neurons with 784 × 128 connections. The second and third layers possess 64 and 10 neurons, respectively, equating to 128 × 64 and 64 × 10 connections, correspondingly. The third layer comprises merely ten neurons, representing ten digits ranging from 0 to 9 as its outputs. The initial two networks are activated by the SA or RSA activation function behind each layer, and the last layer utilizes a softmax function to convert input values to probability distributions. The corresponding loss function used is a cross-entropy loss function. For the following hyperparameter, a Stochastic gradient descent optimizer is employed with a learning rate of 0.001. The specific training process is presented in Supplementary Fig. 13. The greyscale values (0–255) of the pictures were normalized to be in the range from 0 to 1, and similarly, the input power of the experimentally measured NAF was normalized to be in the range from 0 to 1[26]. The greyscale information of the pixels in the image was then mapped to the power in order to match with our experimentally measured NAF. Since it is possible for the weights and biases of each neuron to be negative during the simulation, we analogize to the ReLU function by defining $f(x) = 0$ for $x < 0$, i.e., only the part of the NAF that is positive is utilized, which is mainly due to the positive intensity of the optical signal captured by the photodetector.

The four NAFs obtained using the $MoTe_2$ optical nonlinearities can be implemented for the MNIST classification task, as shown in Fig. 6b, c, and Supplementary Fig. 14. The dataset testing accuracy is 96.0% (S1-SA) and 97.4% (S2-RSA), respectively, which is better than or comparable with the previously reported values[62,63]. Moreover, the accuracy reaches 97.6% with a suitable thickness of $MoTe_2$ (17.3 nm) (Supplementary Fig. 15). The results show excellent performance, and our four NAFs are comparable to commonly used electronic neural networks. Across 80 training epochs, we have showcased the accuracy and loss over varying cycle durations for four nonlinear activators, linear activation functions, and ReLU activation functions, as shown in Supplementary Fig. 16. Our findings unequivocally establish the superiority of the nonlinear activation function, which

 https://doi.org/10.1038/s41467-024-53371-6

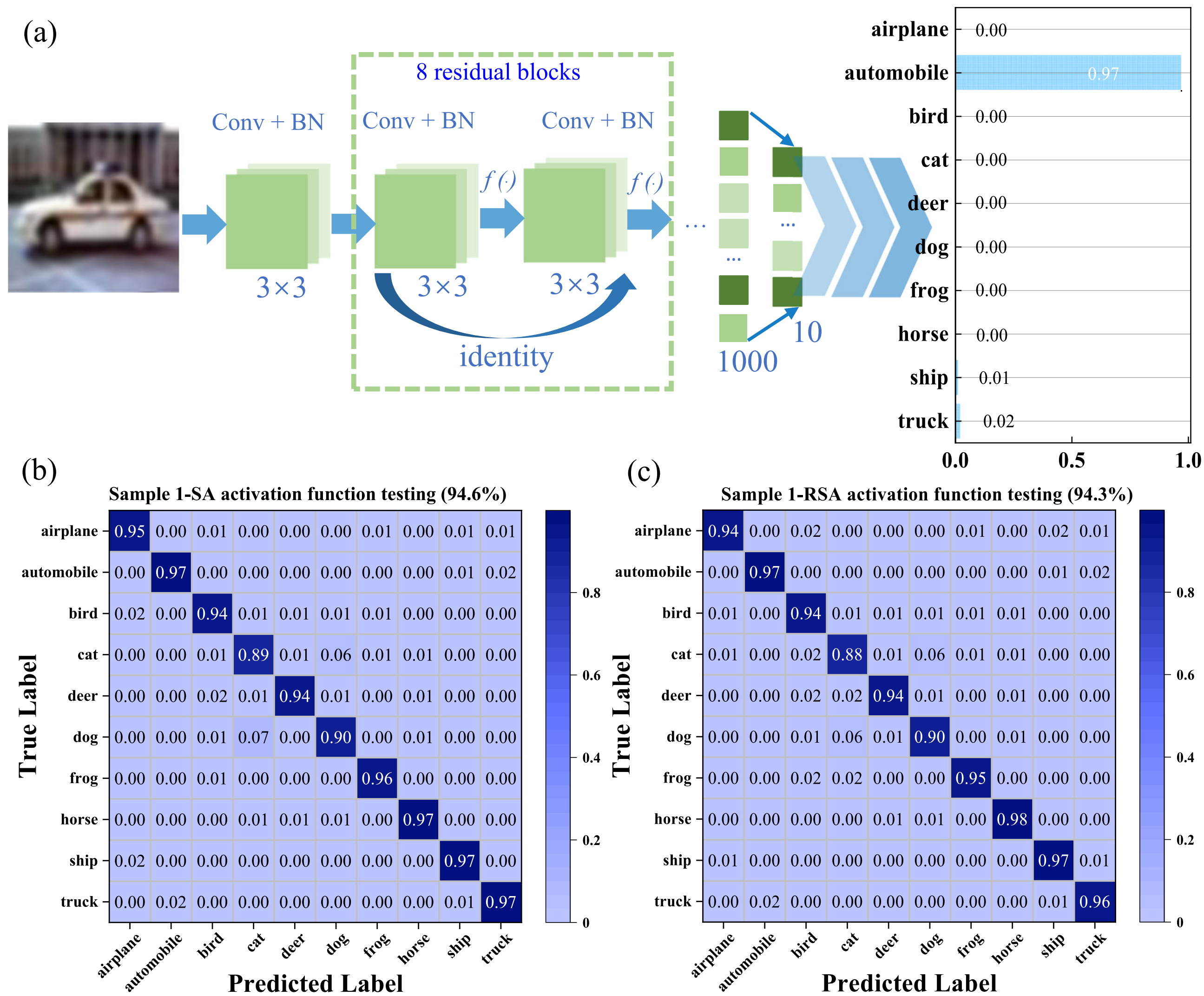

Fig. 7 | Classification performance on CIFAR-10. The CIFAR-10 dataset (Canadian Institute For Advanced Research) is a collection of images that are commonly used to train machine learning and computer vision algorithms. **a** ONN architecture for the CIFAR-10 classification task. **b**, **c** Confusion matrices for the NAF test dataset constructed from the saturable absorption and reverse saturable absorption properties of $MoTe_2$.

is constructed using the optical nonlinearity of $MoTe_2$, in terms of image recognition capability, as it can effectively rival the ReLU function. Furthermore, supplementary materials (Supplementary Fig. 14) present the outcomes of the MNIST handwriting training employing both the linear activation function and the ReLU activation function.

To verify the broader applicability of these SA and RSA NAFs, we trained the activator on the CIFAR−10 dataset. This dataset is more complex in shape and has a richer color palette than the MNIST dataset. Figure 7a demonstrates a closer analysis of the trained network. The network is improved on ResNet-18, and the framework mainly consists of 17 convolutional layers and 2 fully connected layers. The RGB image is subjected to a 3 × 3 downsampling convolutional layer, followed by 4 layers of residual operations corresponding to 8 residual blocks. Each residual block consists of two 3 × 3 convolutional layers with two batch normalization (BN) operations. Each BN operation is followed by a nonlinear activation operation provided by our NAF. The data at the end of the residual operation is output through the fully connected layer. The detailed training procedure is described in the Supplementary Fig. 17. After 250 epochs of training, 10 types of RGB images can be easily identified and classified, as shown in Fig. 7b, c, and Supplementary Fig. 18. The dataset testing accuracies are 94.6% (S1–SA), and 94.3% (S1-RSA), respectively. Furthermore, more detailed testing results and details are shown in Supplementary Figs. 18 and 19. Despite exhibiting a slightly lower test accuracy on the CIFAR-10 dataset compared to the conventional ReLU function (95.5%), it demonstrates a remarkable enhancement in model recognition precision relative to scenarios devoid of any activation function (68.9%). This underscores its inherent advantages and potential applications within the domain of deep learning. Furthermore, the $MoTe_2$ material exhibited a SA effect under 1550 nm excitation, as shown in Supplementary Fig. 20a. The NAF for optical signal input and output (Supplementary Fig. 20b), has an activation threshold of 0.94 μW, which is lower than previously reported threshold[27,64,65]. The activation function can also be used to identify handwritten digits (MNIST) and color photographs (CIFAR−10) (Supplementary Fig. 21). Based on the above results, the optical nonlinearity of $MoTe_2$ on chip can effectively recognize and classify gray and RGB color images.

The high compatibility and integrability enable 2D materials transferred to various substrates, especially to the photonic circuits, such as silicon, silicon nitride, and lithium niobate waveguides, to construct the integrated devices and all-optical nonlinear activators for ONNs. However, most reported all-optical nonlinear activators

Table 1 | Comparison of optical thresholds and sizes of different NAFs

Device	Optical threshold	Test wavelength	Size	Refs.
Ge/Si hybrid waveguide	0.74 mW	~1548 nm	0.5 × 2.58 μm^2	58
add-drop MRR	80 μW	~1552 nm	10^2 μm^2	65
PCM on Si	~2.3 mW	1550 nm	~10^2 μm^2	27
MZI mesh	0.9 mW	-	10^4 μm^2	68
Gra modulator	10 mW	1560 nm	~40 × 10 μm^2	81
Ge-Si modulator	1.1 mW	1550 nm	~30 ×8 μm^2	64
Gra/Si heterojunction	0.5 mW	~2012 nm ~2026 nm	~80 μm^2	62
$MoTe_2$/OWG	4.9 μW	1030 nm	~50 μm^2	This work
$MoTe_2$/OWG	5.5 μW	515 nm	~50 μm^2	This work
$MoTe_2$/OWG	0.94 μW	1550 nm	~50 μm^2	This work

Ge germanium, *Si* represents Silicon, *MRR* stands for micro ring resonator, *Gra* signifies graphene.

based on the 2D materials/OWGs integrated systems rely on EFC, and the intrinsic low efficiency of EFC gives rise to large optical loss and a high activation threshold for ONNs, which prevents full use of the strong optical modulation capability and brings a high energy consumption[62,63,66,67]. In addition, the previous integration is generally limited in two-dimension and the working waveband is limited by the band gap of the waveguide materials. The current butt-coupling integrated system provides a distinctive platform to overcome these issues and enables advanced NAFs. Table 1 shows the different forms of activators. The activators in the form of $MoTe_2$/OWG possess a lower nonlinearity threshold and a wider working waveband compared to the activators consisting of a silicon-based waveguide form. Our activators exhibit ultra-low power consumption down to 0.94 μW, and the single device size is about ~50 μm^2, and both values are one order of magnitude smaller than most reported ones[27,51,64,65,67,68]. Furthermore, FLDW allows for 3D creation of OWGs in glass, which promises 3D integration for high-speed, ultra-broadband and dense integration ONNs, for large-scale photonic computation systems. First, 3D large-scale OWG arrays can be fabricated in glass by FLDW technique[29–35] and integrated $MoTe_2$/OWGs arrays have also been obtained (Supplementary Fig. 22). By using large-scale 2D material transfer techniques[66,69,70], larger-scale and dense integrated devices can be created on one glass chip. Second, the activators formed by the butt-coupling mothed have excellent scalability, especially since it is possible to create multi-layer ONNs within a single substrate (Supplementary Fig. 23a) and multiple substrates (Supplementary Fig. 24). Within a single glass substrate, as shown in Supplementary Fig. 23a, multiple single-layer optical neurons could be prepared on a single glass substrate using FLDW. In each layer, the output of the linear transformation uses the butt-coupling strategy to form activators. Optical fibers can connect each layer of neurons to implement multi-layer ONNs within a single glass substrate. We have demonstrated the coupling of multiple layers of OWGs within a single substrate (Supplementary Fig. 23b). To this end, polymer OWG fabricated by fs laser two-photon polymerization may also be a possible alternative for interconnecting between ONN layers[71–73]. In addition, the NAF structure can be integrated with other layers using other substrates. A multi-layer neural network of optical neurons containing butt-coupled structures can be formed by aligning two $MoTe_2$/OWG devices (Supplementary Fig. 24). Thus, our developed strategy and devices support large-scale, dense integration and multi-layer coupling networks, which implies that parallel and cascaded processing of large-scale data could be realized. Overall, our all-optical activator has the potential for lower optical thresholds and high-density integration and presents a substantial conceptual advance in ONN platforms.

In conclusion, we have successfully created an integrated all-optical nonlinear activator with a low activation threshold, an ultra-fast computational speed, and broadband response, utilizing a butt-coupling 2D $MoTe_2$/OWG platform. The activator functions are derived from the optical nonlinearities of $MoTe_2$, including the SA and RSA effects. The wideband optical nonlinearity significantly enhances the activator's flexibility. The effectiveness of this activation function has received confirmation through the employment of MNIST handwriting dataset. An activation threshold is as low as 0.94 μW, which is one order of magnitude smaller than previously reported values and a high classification accuracy of 97.6% was achieved. The response rate of the all-optical nonlinear activator is up to 2.08 THz, which allows for ultrafast information processing. Additionally, the activator exhibited notable abilities regarding recognition and categorization when utilized on the more intricate and colorful CIFAR-10 dataset. Our approach presents an innovative solution for incorporating optical NAFs, exhibiting substantial potential and feasibility for the future development of 3D all-optical on-chip neural networks. The represented NAFs could also be used to process other data sets, such as video, audio, and text[13,74,75]. Combining the programmability of OWGs written by fs laser[76–78] and the integrability of the light source in glass[29,79,80], the applications of the present NAFs could be extended further.

Methods

Material synthesis

Optical waveguide: A diode-pumped Yb: KGW laser (PHAROS PH2, Light Conversion) provides 169 fs pulses at a central wavelength of 1030 nm with a repetition rate of 1 MHz. The laser beam is focused below the surface of borosilicate glass (Corning Eagle XG) samples. A high-precision XYZ translation stage (Aerotech, ABL1000WB-100) is used to control the 3D movement of the glass sample at the set routine and speed. The pulsed laser was focused through the objective (Nikon LU Plan, 50X, NA = 0.55).

$MoTe_2$: 2H-$MoTe_2$ crystal (product no. 100705) was purchased from Nanjing MKNANO Tech. Co., Ltd. (www.mukenano.com). Few-layered $MoTe_2$ crystals are obtained by mechanical exfoliation method, and $MoTe_2$ is transferred to the end surface of OWG using a self-constructed 2D material transfer system.

Material Characterization: Raman characterization of $MoTe_2$ nanosheets and optical waveguide was carried out using a confocal microscopy system (RENISHAW, Invia) with a 532 nm continuous wave laser. The thickness of the $MoTe_2$ nanosheets was measured using atomic force microscope (Bruker, Dimension ICON). XRD was performed using a Bruker D8 Advance.

Nonlinear optical measurements

The experimental sources of the Intensity-scan on chip were pulses of 169 fs from a diode-pumped Yb: KGW laser (PHAROS PH2, Light Conversion) operating at 1030 nm with a second harmonic at 515 nm and a pulse repetition rate of 1 MHz. The excitation light at other wavelengths (800 nm and 1550 nm) was obtained from a light parametric amplifier (Orpheus-NEO-UP, Light Conversion) with a pulse repetition frequency of 55 KHz.

Ultrafast carrier dynamic measurements

The carrier dynamic properties were measured using time-resolved degenerate pump-probe and transient absorption techniques. In the pump-probe system, the transient reflection was measured using a 380-fs pulse laser with a wavelength of 520 nm and a repetition rate of 10 kHz. The TAS was acquired using a fs transient absorption spectrometer (Light Conversion Harpia). The excitation source was a 200-

fs pulsed laser at 400 nm wavelength with a repetition rate of 1 kHz, and the probe light was a continuous white light spectrum in the range 500–900 nm. The amount of change in optical absorption is in units of optical density (OD or mOD, where 1 OD = 1×10^3 mOD.), which is obtained from the probe intensities when the sample is pumped and unpumped ($\Delta OD = -1000\times\log_{10}I_{pumped}/I_{unpumped}$). When ΔOD > 0, it means that the pumped sample absorbs the probe light, i.e., the photoinduced absorption effect. When ΔOD < 0, it represents an increase in transmission of probe light by the pumped sample, i.e., the photobleaching effect.

Data availability

The Source Data underlying Figs. 2c–f and 4 of this study are available at https://doi.org/10.6084/m9.figshare.27150810. All raw data generated during the current study are available from the corresponding authors upon request.

References

1. Liang, W. et al. Advances, challenges and opportunities in creating data for trustworthy AI. *Nat. Mach. Intell.* **4**, 669–677 (2022).
2. Mosalaganti, S. et al. AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **376**, 1176 (2022).
3. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med.* **28**, 31–38 (2022).
4. Masanet, E., Shehabi, A., Lei, N., Smith, S. & Koomey, J. Recalibrating global data center energy-use estimates. *Science* **367**, 984–986 (2020).
5. Lanza, M. et al. Memristive technologies for data storage, computation, encryption, and radio-frequency communication. *Science* **376**, 1066 (2022).
6. Waldrop, M. M. The chips are down for Moore's law. *Nat. News* **530**, 144 (2016).
7. Wang, S., Liu, X. & Zhou, P. The road for 2D semiconductors in the silicon age. *Adv. Mater.* **34**, 2106886 (2022).
8. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
9. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**, 102–114 (2021).
10. Cheng, Z., Rios, C., Pernice, W. H. P., Wright, C. D. & Bhaskaran, H. On-chip photonic synapse. *Sci. Adv.* **3**, e1700160 (2017).
11. Feldmann J. et al. Calculating with light using a chip-scale all-optical abacus. *Nat. Commun.* **8**, 1256 (2017).
12. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
13. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
14. Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
15. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
16. Qian, C. et al. Performing optical logic operations by a diffractive neural network. *Light Sci. Appl.* **9**, 59–59 (2020).
17. Zhou, T. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
18. Zuo, Y. et al. All-optical neural network with nonlinear activation functions. *Optica* **6**, 1132–1137 (2019).
19. Zuo, Y., Zhao, Y., Chen, Y.-C., Du, S. & Liu, J. Scalability of all-optical neural networks based on spatial light modulators. *Phys. Rev. Appl.* **15**, 054034 (2021).
20. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
21. Tait, A. N. et al. Silicon photonic modulator neuron. *Phys. Rev. Appl.* **11**, 064043 (2019).
22. Amin, R. et al. ITO-based electro-absorption modulator for photonic neural activation function. *APL Mater.* **7**, 081112 (2019).
23. Xu, Z. et al. Reconfigurable nonlinear photonic activation function for photonic neural network based on non-volatile opto-resistive RAM switch. *Light Sci. Appl.* **11**, 288 (2022).
24. Li, H., Wu, B., Tong, W., Dong, J. & Zhang, X. All-optical nonlinear activation function based on germanium silicon hybrid asymmetric coupler. *IEEE J. Sel. Top. Quantum Electron.* **29**, 8300106 (2023).
25. Leuthold, J., Koos, C. & Freude, W. Nonlinear silicon photonics. *Nat. Photonics* **4**, 535–544 (2010).
26. Hazan, A. et al. MXene-nanoflakes-enabled all-optical nonlinear activation function for on-chip photonic deep neural networks. *Adv. Mater.* **35**, 2210216 (2023).
27. 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).
28. Kaichen, Z. et al. The development of integrated circuits based on two-dimensional materials. *Nat. Electron.* **4**, 775–785 (2021).
29. Sun, K. et al. Three-dimensional direct lithography of stable perovskite nanocrystals in glass. *Science* **375**, 307 (2022).
30. Grotevent, M. J. et al. Integrated photodetectors for compact Fourier-transform waveguide spectrometers. *Nat. Photonics* **17**, 59–64 (2022).
31. Chen, X. & Gu, M. Two-beam ultrafast laser scribing of graphene patterns with 90-nm subdiffraction feature size. *Ultrafast Sci.* **2022**, 0001 (2022).
32. Biesenthal, T. et al. Fractal photonic topological insulators. *Science* **376**, 1114–1119 (2022).
33. Zhang, X.-L. et al. Non-Abelian braiding on photonic chips. *Nat. Photonics* **16**, 390–395 (2022).
34. Tan, D., Wang, Z., Xu, B. & Qiu, J. Photonic circuits written by femtosecond laser in glass: improved fabrication and recent progress in photonic devices. *Adv. Photon.* **3**, 024002 (2021).
35. Zhong, L., Wang, Y., Tan, D. & Qiu, J. Toward 3D integration of highly see-through photonic circuits in glass. *Laser Photonics Rev.* **17**, 202200767 (2023).
36. Tan, D. et al. Fabricating low loss waveguides over a large depth in glass by temperature gradient assisted femtosecond laser writing. *Opt. Lett.* **45**, 3941–3944 (2020).
37. Trukhin, A. N. Excitons in SiO_2: a review. *J. Non-Cryst. Solids* **149**, 32–45 (1992).
38. Xia, F., Wang, H., Xiao, D., Dubey, M. & Ramasubramaniam, A. Two-dimensional material nanophotonics. *Nat. Photonics* **8**, 899–907 (2014).
39. Mak, K. F. & Shan, J. Photonics and optoelectronics of 2D semiconductor transition metal dichalcogenides. *Nat. Photonics* **10**, 216–226 (2016).
40. Tan, D., Sun, X. & Qiu, J. Femtosecond laser writing low-loss waveguides in silica glass: highly symmetrical mode field and mechanism of refractive index change. *Opt. Mater. Express* **11**, 848–857 (2021).
41. Tan, D., Sun, X., Li, Z. & Qiu, J. Effectively writing low propagation and bend loss waveguides in the silica glass by using a femtosecond laser. *Opt. Lett.* **47**, 4766–4769 (2022).
42. Huang, J.-H. et al. Polymorphism control of layered $MoTe_2$ through two-dimensional solid-phase crystallization. *Sci. Rep.* **9**, 8810 (2019).
43. Song, S. et al. Room temperature semiconductor–metal transition of $MoTe_2$ thin films engineered by strain. *Nano Lett.* **16**, 188–193 (2015).
44. Faisal, A. et al. Deterministic polymorphic engineering of $MoTe_2$ for photonic and optoelectronic applications. *Adv. Funct. Mater.* **33**, 202302051 (2023).
45. Ding, Y. et al. Broad-band photodetector based on a lateral $MoTe_2$ 1T-2H-1T homojunction. *J. Phys. Chem. C.* **127**, 20072–20081 (2023).

46. Lee, K. et al. Sub-picosecond carrier dynamics induced by efficient charge transfer in $MoTe_2/WTe_2$ van der Waals heterostructures. *ACS Nano* **13**, 9587–9594 (2019).
47. Pan, S. et al. Efficient interlayer electron transfer in a $MoTe_2/WS_2/MoS_2$ trilayer heterostructure. *Appl. Phys. Lett.* **118**, 253106 (2021).
48. Li, C. et al. Waveguide-integrated $MoTe_2$ p–i–n homojunction photodetector. *ACS Nano* **16**, 20946–20955 (2022).
49. Wu, G. et al. $MoTe_2$ p-n homojunctions defined by ferroelectric polarization. *Adv. Mater.* **32**, 201907937 (2020).
50. Reuter, C., Ecke, G. & Strehle, S. Exploring the surface oxidation and environmental instability of 2H-/1T′-$MoTe_2$ using field emission-based scanning probe lithography. *Adv. Mater.* **36**, 202310887 (2023).
51. Yang, Z. et al. MXene-based broadband ultrafast nonlinear activator for optical computing. *Adv. Opt. Mater.* **10**, 202200714 (2022).
52. Chen, Y. et al. Large energy, wavelength widely tunable, topological insulator Q-switched erbium-doped fiber laser. *IEEE J. Sel. Top. Quant.* **20**, 315–322 (2013).
53. Wang, G. et al. Hot-carrier tunable abnormal nonlinear absorption conversion in quasi-2D perovskite. *Nat. Commun.* **13**, 6935 (2022).
54. Li, R. et al. Nonlinear absorption response correlated to embedded Ag nanoparticles in BGO single crystal: from two-photon to three-photon absorption. *Sci. Rep.* **8**, 1977 (2018).
55. Verrone, R.-N. et al. Thickness-dependent optical nonlinearities of nanometer-thick Sb_2Te_3 thin films: implications for mode-locking and super-resolved direct laser writing. *ACS Appl. Nano Mater.* **3**, 7963–7972 (2020).
56. Wang, K. et al. Ultrafast saturable absorption of two-dimensional MoS_2 nanosheets. *ACS Nano* **7**, 9260–9267 (2013).
57. You, J. W., Bongu, S. R., Bao, Q. & Panoiu, N. C. Nonlinear optical properties and applications of 2D materials: theoretical and experimental aspects. *Nanophotonics* **8**, 63–97 (2019).
58. Wu, B., Li, H., Tong, W., Dong, J. & Zhang, X. Low-threshold all-optical nonlinear activation function based on a Ge/Si hybrid structure in a microring resonator. *Opt. Mater. Express* **12**, 970–980 (2022).
59. 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).
60. Breusing, M., Ropers, C. & Elsaesser, T. Ultrafast carrier dynamics in graphite. *Phys. Rev. Lett.* **102**, 086809 (2009).
61. Gao, L. et al. Ultrafast relaxation dynamics and nonlinear response of few-layer niobium carbide MXene. *Small Methods* **4**, 2000250 (2020).
62. Zhong, C. et al. Graphene/silicon heterojunction for reconfigurable phase-relevant activation function in coherent optical neural networks. *Nat. Commun.* **14**, 6939 (2023).
63. Fu, T. et al. Photonic machine learning with on-chip diffractive optics. *Nat. Commun.* **14**, 70 (2023).
64. Shi, Y. et al. Nonlinear germanium-silicon photodiode for activation and monitoring in photonic neuromorphic networks. *Nat. Commun.* **13**, 6048 (2022).
65. Yu, W., Zheng, S., Zhao, Z., Wang, B. & Zhang, W. Reconfigurable low-threshold all-optical nonlinear activation functions based on an add-drop silicon microring resonator. *IEEE Photonics J.* **14**, 5559807 (2022).
66. Meng, Y. et al. Photonic van der Waals integration from 2D materials to 3D nanomembranes. *Nat. Rev. Mater.* **8**, 498–517 (2023).
67. He, T. et al. On-chip optoelectronic logic gates operating in the telecom band. *Nat. Photonics* **18**, 60–67 (2023).
68. Li Q. et al. Optical nonlinear activation functions based on MZI-structure for optical neural networks. In: *2020 Asia Communications and Photonics Conference (ACP) and International Conference on Information Photonics and Optical Communications (IPOC)*), (IEEE, 2020).
69. Nakatani, M. et al. Ready-to-transfer two-dimensional materials using tunable adhesive force tapes. *Nat. Electron.* **7**, 119–130 (2024).
70. Xia, Y. et al. 12-inch growth of uniform MoS_2 monolayer for integrated circuit manufacture. *Nat. Mater.* **22**, 1324–1331 (2023).
71. Gissibl, T., Thiele, S., Herkommer, A. & Giessen, H. Two-photon direct laser writing of ultracompact multi-lens objectives. *Nat. Photonics* **10**, 554–560 (2016).
72. Schumann, M., Bückmann, T., Gruhler, N., Wegener, M. & Pernice, W. Hybrid 2D–3D optical devices for integrated optics by direct laser writing. *Light Sci. Appl.* **3**, e175 (2014).
73. Baghdasaryan, T., Vanmol, K., Thienpont, H., Berghmans, F. & Van Erps, J. Ultracompact 3D splitter for single-core to multi-core optical fiber connections fabricated through direct laser writing in polymer. *Laser Photonics Rev.* **18**, 2400089 (2024).
74. Ji, S., Yang, M. & Yu, K. 3D convolutional neural networks for human action recognition. *IEEE T. Pattern Anal.* **35**, 221–231 (2012).
75. Minaee, S. et al. Deep learning-based text classification: a comprehensive review. *Acm. Comput. Surv.* **54**, 1–40 (2021).
76. Ceccarelli, F. et al. Low power reconfigurability and reduced crosstalk in integrated photonic circuits fabricated by femtosecond laser micromachining. *Laser Photonics Rev.* **14**, 2000024 (2020).
77. Francesco, H. et al. Reconfigurable continuously-coupled 3D photonic circuit for Boson sampling experiments. *npj Quantum Inf.* **8**, 55 (2022).
78. Tan, X. et al. Scalable and programmable three-dimensional photonic processor. *Phys. Rev. Appl.* **20**, 044041 (2023).
79. Gligor, D. et al. Femtosecond laser additive and subtractive micro-processing: enabling a high-channel-density silica interposer for multicore fibre to silicon-photonic packaging. *Int. J. Extrem. Manuf.* **1**, 045002 (2019).
80. van Uden, R. G. H. et al. Ultra-high-density spatial division multiplexing with a few-mode multicore fibre. *Nat. Photonics* **8**, 865–870 (2014).
81. Liao, K. et al. Matrix eigenvalue solver based on reconfigurable photonic neural network. *Nanophotonics* **11**, 4089–4099 (2022).

Acknowledgements

C.D.C. and Z.Y. contributed equally to this work. The authors thank (Dr. R.S.Z. from) Material Experiment Platform at Zhejiang Lab for the assistance in AFM and XRD measurement interpretation. The authors are grateful for financial support from the National Key R&D Program of China (No. 2021YFB2800500, 2021ZD0109904); the National Natural Science Foundation of China (Grant No. 62275233). C.D.C. and K.G. acknowledge the support of China Postdoctoral Science Foundation (No. 2023M743248, 2023M733259).

Author contributions

D.T. conceptualized and supervised the project. C.C. prepared specimens, performed the experiments, and analyzed the data. Z.Y. designed the calculation of optical neural networks. T.W. and Y.W. wrote optical waveguides and carried out Raman characterization. K.G. and J-J.W. schematized the optical neural network. C.C., Z.Y., T.W., Y.W., K.G., J-J.W., J.W., J.Q., and D.T. participated in discussions of the experimental approach and data analysis. D.T. prepared the manuscript with input from all authors. D.T., C.C., and Z.Y. prepared all figures and reviewed the manuscript. All authors contributed critically to the drafts and gave the final approval for publication.

Competing interests

The authors declare no competing interests.

https://doi.org/10.1038/s41467-024-53371-6

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53371-6.

Correspondence and requests for materials should be addressed to Dezhi Tan.

Peer review information *Nature Communications* thanks Chandraman Patil and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.