Ultrafast machine vision with 2D material neural network image sensors

https://doi.org/10.1038/s41586-020-2038-x

Received: 28 August 2019

Accepted: 17 January 2020

Published online: 4 March 2020

Check for updates

Lukas Mennel[1]✉, Joanna Symonowicz[1], Stefan Wachter[1], Dmitry K. Polyushkin[1], Aday J. Molina-Mendoza[1] & Thomas Mueller[1]✉

Machine vision technology has taken huge leaps in recent years, and is now becoming an integral part of various intelligent systems, including autonomous vehicles and robotics. Usually, visual information is captured by a frame-based camera, converted into a digital format and processed afterwards using a machine-learning algorithm such as an artificial neural network (ANN)[1]. The large amount of (mostly redundant) data passed through the entire signal chain, however, results in low frame rates and high power consumption. Various visual data preprocessing techniques have thus been developed[2–7] to increase the efficiency of the subsequent signal processing in an ANN. Here we demonstrate that an image sensor can itself constitute an ANN that can simultaneously sense and process optical images without latency. Our device is based on a reconfigurable two-dimensional (2D) semiconductor[8,9] photodiode[10–12] array, and the synaptic weights of the network are stored in a continuously tunable photoresponsivity matrix. We demonstrate both supervised and unsupervised learning and train the sensor to classify and encode images that are optically projected onto the chip with a throughput of 20 million bins per second.

ANNs have achieved huge success as machine-learning algorithms in a wide variety of fields[1]. The computational resources required to perform machine-learning tasks are very demanding. Accordingly, dedicated hardware solutions that provide better performance and energy efficiency than conventional computer architectures have become a major research focus. However, although much progress has been made in efficient neuromorphic processing of electrical[13–16] or optical[17–20] signals, the conversion of optical images into the electrical domain remains a bottleneck, particularly in time-critical applications. Imaging systems that mimic neuro-biological architectures may allow us to overcome these disadvantages. Much work has therefore been devoted to develop systems that emulate certain functions of the human eye[21], including hemispherically shaped image sensors[22,23] and preprocessing of visual data[2–7], for example, for image-contrast enhancement, noise reduction or event-driven data acquisition.

Here, we present a photodiode array that itself constitutes an ANN that simultaneously senses and processes images projected onto the chip. The sensor performs a real-time multiplication of the projected image with a photoresponsivity matrix. Training of the network requires setting the photoresponsivity value of each pixel individually. Conventional photodiodes that are based, for example, on silicon exhibit a fixed responsivity that is defined by the inner structure (chemical doping profile) of the device, and are thus not suitable for the proposed application. Other technologies such as photonic mixing[24] and metal–semiconductor–metal detectors[25] may, in principle, be suitable, but these device concepts bear additional challenges, such as nonlinear tunability of the photoresponse and bias-dependent (and hence weight-dependent) dark current. We have therefore chosen WSe_2—a 2D semiconductor—as the photoactive material. 2D semiconductors not only show strong light–matter interaction and excellent optoelectronic properties[8,9] but also offer the possibility of external tunability of the potential profile in a device—and hence its photosensitivity—by electrostatic doping using multi-gate electrodes[10–12]. In addition, 2D materials technology has by now achieved a sufficiently high level of maturity to be employed in complex systems[26] and provides ease of integration with silicon readout/control electronics[27].

Figure 1a schematically illustrates the basic layout of the image sensor. It consists of N photoactive pixels arranged in a 2D array, with each pixel divided into M subpixels. Each subpixel is composed of a photodiode, which is operated under short-circuit conditions and under optical illumination delivers a photocurrent of $I_{mn} = R_{mn}E_nA = R_{mn}P_n$, where R_{mn} is the photoresponsivity of the subpixel, E_n and P_n denote the local irradiance and optical power at the nth pixel, respectively, and A is the detector area. $n = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, M$ denote the pixel and subpixel indices, correspondingly. An integrated neural network and imaging array can now be formed by interconnecting the subpixels. Summing all photocurrents produced by the mth detector element of each pixel

$$I_m = \sum_{n=1}^{N} I_{mn} = \sum_{n=1}^{N} R_{mn}P_n \qquad (1)$$

performs the matrix–vector product operation $\mathbf{I} = R\mathbf{P}$, with $R = (R_{mn})$ being the photoresponsivity matrix, $\mathbf{P} = (P_1, P_2, \ldots, P_N)^{\mathrm{T}}$ being a vector that represents the optical image projected onto the chip and $\mathbf{I} = (I_1, I_2, \ldots, I_M)^{\mathrm{T}}$ being the output vector. Provided that the R_{mn} value of each detector element can be set to a specific positive or negative value, various types of ANNs for image processing can be implemented (see Fig. 1c, d), with the synaptic weights being encoded in the photoresponsivity matrix. The expression 'negative photoresponsivity' is to be understood in this context as referring to the sign of the photocurrent.

[1] Institute of Photonics, Vienna University of Technology, Vienna, Austria. ✉e-mail: lukas.mennel@tuwien.ac.at; thomas.mueller@tuwien.ac.at

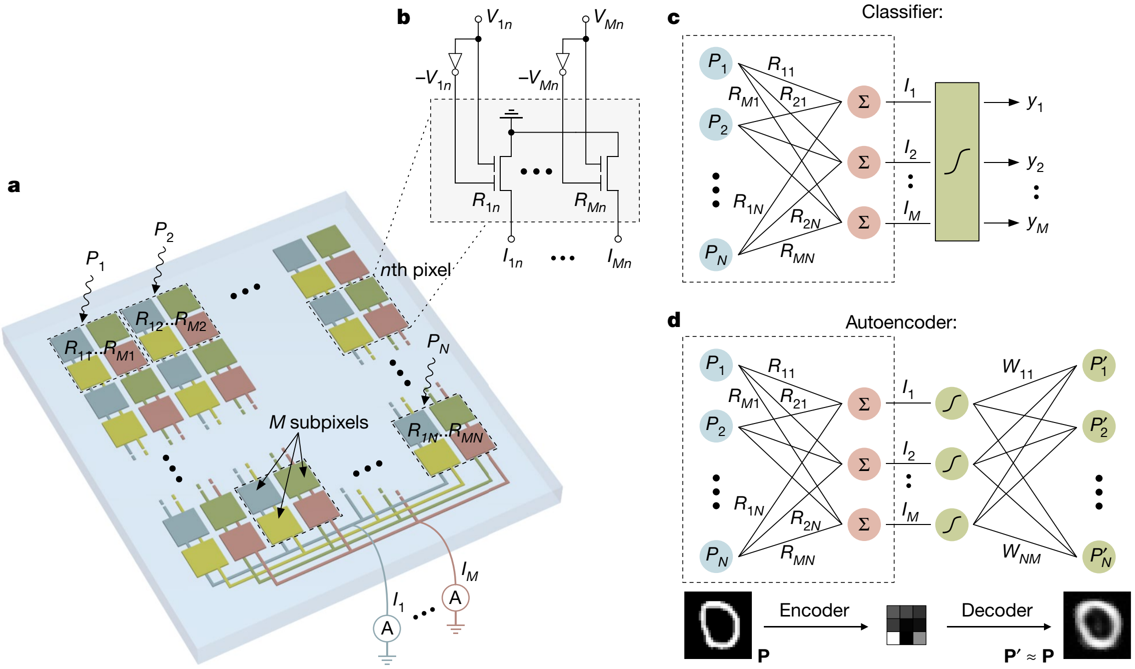


Fig. 1 | Imaging ANN photodiode array. a, Illustration of the ANN photodiode array. All subpixels with the same colour are connected in parallel to generate M output currents. **b**, Circuit diagram of a single pixel in the photodiode array. **c, d**, Schematics of the classifier (**c**) and the autoencoder (**d**). Below the illustration of the autoencoder, shown is an example of encoding/decoding of a 28 × 28 pixel letter from the MNIST handwritten digit database. The original image is encoded to 9 code-layer neurons and then decoded back into an image.

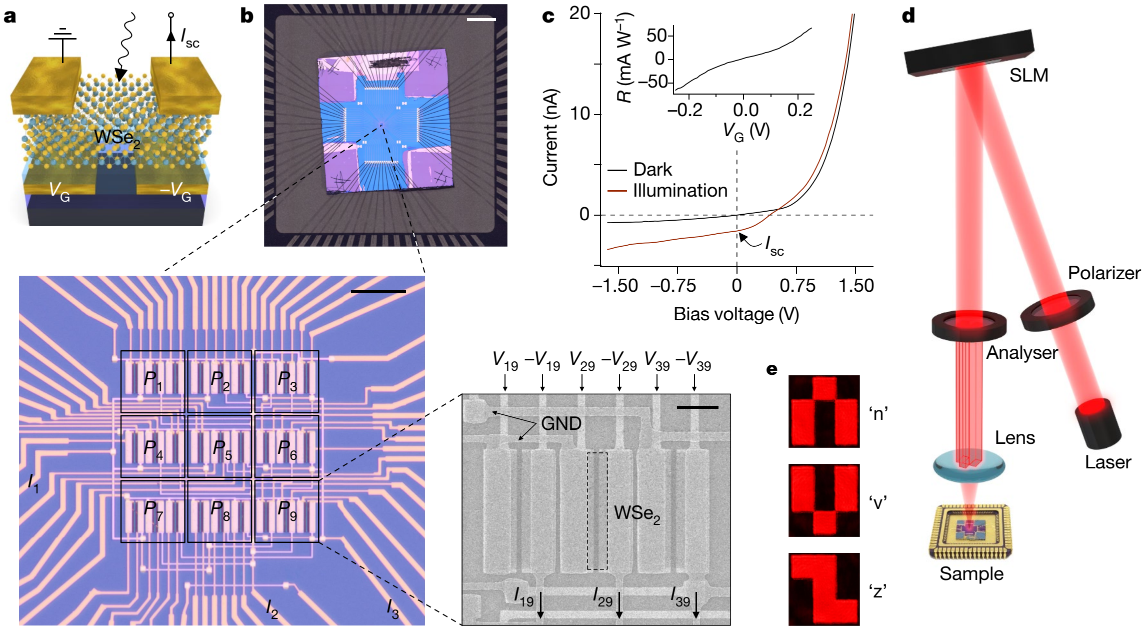


Fig. 2 | Implementation of the ANN photodiode array. a, Schematic of a single WSe_2 photodiode. The device is operated under short-circuit conditions and the photoresponsivity is set by supplying a voltage pair $V_G/-V_G$ to the bottom-gate electrodes. **b**, Macroscopic image of the bonded chip on the chip carrier. Scale bar, 2 mm. First magnification: microscope image of the photodiode array, which consists of 3 × 3 pixels. Scale bar, 15 μm. Second magnification: scanning electron microscopy image of one of the pixels. Each pixel consists of three WSe_2 photodiodes/subpixels with responsivities set by the gate voltages. Scale bar, 3 μm. GND, ground electrode. **c**, Current–voltage characteristic curve of one of the photodetectors in the dark (blue line) and under optical illumination (red line). See also Extended Data Fig. 2a. The inset shows the gate-voltage tunability of the photoresponsivity. **d**, Schematic illustration of the optical setup. Laser light is linearly polarized by a wire-grid polarizer and reflected by a spatial light modulator (SLM). The reflected light is then filtered by an analyser (intensity modulation) and the resulting image is projected onto the photodiode array. **e**, Microscope images of the 3 × 3 pixel letters used for training/operation of the network.

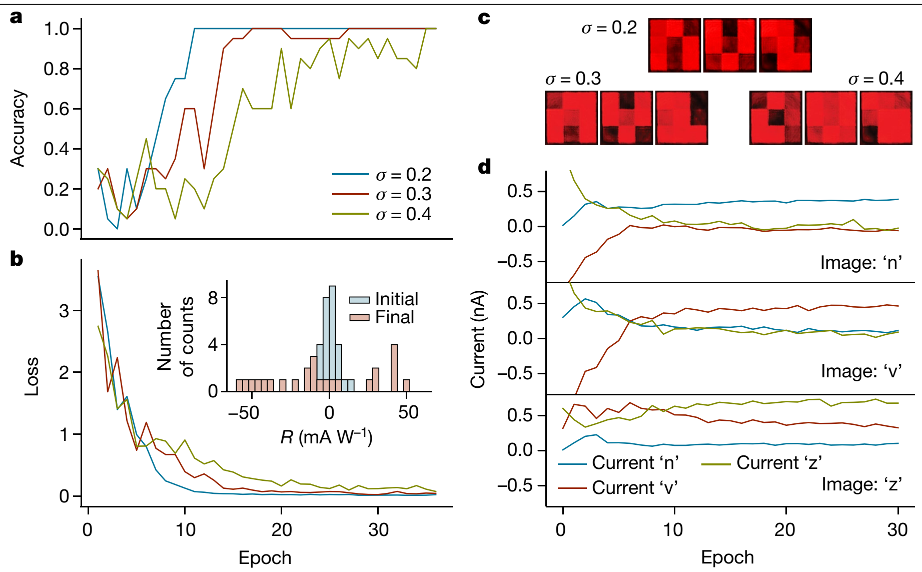


Fig. 3 | Device operation as a classifier. a, Accuracy of the classifier during training for varying artificial noise levels. An image is accurately predicted when the correct neuron has the largest activation. **b**, Loss function for different noise levels during training. The inset shows the initial and final responsivity distributions for $\sigma = 0.2$. **c**, Microscope images of the projected letters with different random noise levels. The complete dataset obtained over 30 epochs of training is shown in Extended Data Fig. 7. **d**, Average currents for each epoch for each projected letter, measured during training with a noise level of $\sigma = 0.2$. Each graph shows the results of a separate experiment, in which the letters 'n' (top), 'v' (middle) and 'z' (bottom) are projected onto the chip, and three currents—corresponding to 'n' (blue), 'v' (red) and 'z' (green)—are measured. In the top graph, for example, the output that corresponds to the letter 'n' (current 'n') is the highest, so the ANN determines that the projected letter is 'n'.

We implemented two types of ANNs: a classifier and an autoencoder. Figure 1c shows a schematic of the classifier. Here, the array is operated as a single-layer perceptron, together with nonlinear activation functions that are implemented off-chip. This type of ANN represents a supervised learning algorithm that is capable of classifying images **P** into different categories **y**. An autoencoder (Fig. 1d) is an ANN that can learn, in an unsupervised training process, an efficient representation (encoding) for a set of images **P**. Along with the encoder, a decoder is trained to attempt to reproduce at its output the original image, $\mathbf{P}' \approx \mathbf{P}$, from the compressed data. Here the encoder is formed by the photodiode array itself and the decoder by external electronics.

Having presented the operational concept of our network, we now come to an actual device implementation. We used a few-layer WSe_2 crystal with a thickness of about 4 nm to form lateral p–n junction photodiodes, using split-gate electrodes (with a ~300-nm-wide gap) that couple to two different regions of the 2D semiconductor channel (Fig. 2a)[10–12]. WSe_2 was chosen because of its ambipolar conduction behaviour and excellent optoelectronic properties. Biasing one gate electrode at V_G and the other at $-V_G$ enables adjustable (trainable) responsivities between −60 and +60 mA W^{-1}, as shown in Fig. 2c. This technology was then used to fabricate the photodiode array shown in Fig. 2b, which consists of 27 detectors with good uniformity, tunability and linearity (see Extended Data Figs. 1, 2b). The devices were arranged to form a 3 × 3 imaging array ($N = 9$) with a pixel size of about 17 × 17 μm^2 and with three detectors per pixel ($M = 3$). The short-circuit photocurrents I_{sc} produced by the individual devices under optical illumination were summed according to Kirchhoff's law by hard-wiring the devices in parallel, as depicted in Fig. 1b. The sample fabrication is explained in Methods, and a schematic of the entire circuit is provided in Extended Data Fig. 3. Each device was supplied with a pair of gate voltages, V_G and $-V_G$, to set its responsivity individually. For training and testing of the chip, optical images were projected using the setup shown in Fig. 2d (for details, see Methods). Unless otherwise stated, all measurements were performed using light with a wavelength of 650 nm and with a maximum irradiance of about 0.1 W cm^{-2}. Despite its small size, such a network is sufficient for the proof-of-principle demonstration of several machine-learning algorithms. In particular, we performed classification, encoding, and denoising of the stylized letters 'n', 'v' and 'z' depicted in Fig. 2e. Scaling the network to larger dimensions is conceptually straightforward and remains a mainly technological task.

To test the functionality of the photodiode array, we first operated it as a classifier (Fig. 1c) to recognize the letters 'n', 'v' and 'z'. During each training epoch we optically projected a set of $S = 20$ randomly chosen letters. Gaussian noise (with standard deviation of $\sigma = 0.2$, 0.3 and 0.4; Fig. 3c) was added to augment the input data[28]. In this supervised learning example, we chose one-hot encoding, in which each of the three letters activates a single output node/neuron. As activation function (the nonlinear functional mapping between the inputs and the output of a node) for the M photocurrents we chose the softmax function $\phi_m(I) = e^{I_m\xi}/\sum_{k=1}^{M} e^{I_k\xi}$ (a common choice for one-hot encoding), where $\xi = 10^{10}$ A^{-1} is a scaling factor that ensures that the full value range of the activation function is accessible during training. As a loss/cost function (the function to be minimized during training) we used the cross-entropy $\mathcal{L} = -\frac{1}{M}\sum_{m=1}^{M} y_m \log\left[\phi_m(I)\right]$, where y_m is the label and $M = 3$ is the number of classes. The activations of the output neurons represent the probabilities for each of the letters. The initial values of the responsivities were randomly chosen from a Gaussian distribution, as suggested in ref. [29], and were different for the supervised- and unsupervised-learning demonstrations. The responsivities were updated after every epoch by backpropagation[30] of the gradient of the loss function

$$R_{mn} \rightarrow R_{mn} - \frac{\eta}{S}\sum_{\mathbf{P}} \nabla_{R_{mn}}\mathcal{L} \tag{2}$$

with learning rate $\eta = 0.1$. A detailed flow chart of the training algorithm is presented in Extended Data Fig. 4d.

In Fig. 3a, b the accuracy and loss are plotted over 35 training epochs. The loss is decreasing quickly for all noise levels and reaches a minimum after 15, 20 and 35 epochs for $\sigma = 0.2$, $\sigma = 0.3$ and $\sigma = 0.4$, respectively. The accuracy reaches 100% for all noise levels, with faster convergence for less noise. In Fig. 3d we show the mean currents for each of the three

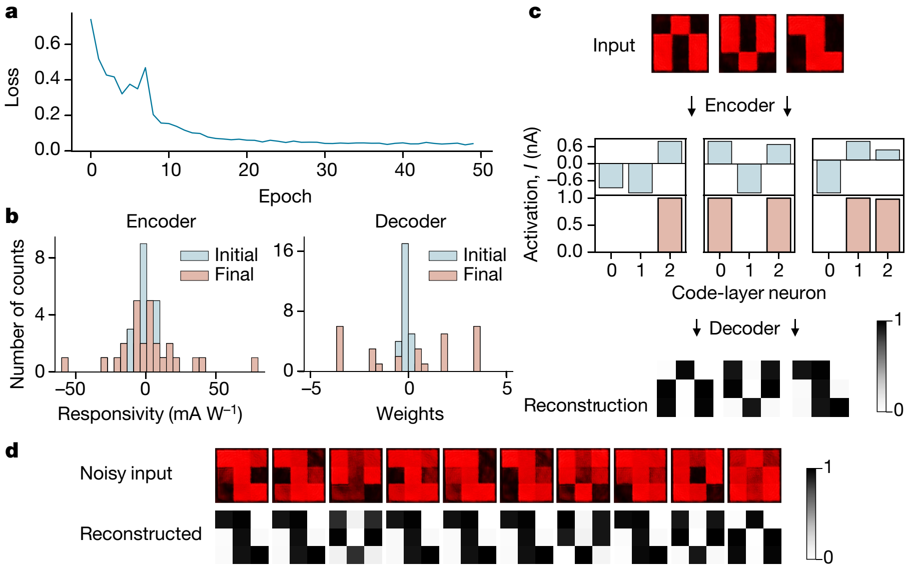


Fig. 4 | Device operation as an autoencoder. a, Loss of the autoencoder during training. The complete dataset of 30 epochs of training is given in Extended Data Fig. 7. **b**, Responsivity and weight distributions before (initial) and after (final) training. **c**, Autoencoding of noise-free letters. The encoder translates the projected images into a current code, which is converted by the nonlinearity into a binary activation code and finally reconstructed into an image by the decoder. **d**, Randomly chosen noisy inputs ($\sigma = 0.3$) and the corresponding reconstructions after autoencoding.

letters during each epoch for $\sigma = 0.2$ (see Extended Data Fig. 5c, d for the other cases). The currents become well separated after about 10 epochs, with the highest current corresponding to the label of the projected letter. The inset in Fig. 3b shows histograms for the (randomly chosen) initial and final responsivity values for $\sigma = 0.2$ (see also Extended Data Fig. 5a, b). The robustness and reliability of the classification results of the analogue vision sensor were verified by comparison of the accuracy and loss with computer simulations of a digital system with the same architecture and learning scheme (Extended Data Fig. 6).

Next, we demonstrate encoding of image patterns with our device operating as an autoencoder (Fig. 1d). We chose logistic (sigmoid) activation functions for the code neurons $\phi_m(I_m) = (1 + e^{-I_m \xi})^{-1}$, again with $\xi = 10^{10}$ A^{-1} as a scaling factor, as well as for the output neurons $P'_n = \phi_n(z_n) = (1 + e^{-z_n})^{-1}$, where $z_n = \sum_{n=1}^{N} W_{nm}\phi_m(I_m)$ and W_{nm} denotes the weight matrix of the decoder. We used the mean-square loss function $\mathcal{L} = \frac{1}{2}||\mathbf{P} - \mathbf{P}'||^2$, which depends on the difference between the original and reconstructed images. The responsivities were again trained by backpropagation of the loss according to equation (2), with a noise level of $\sigma = 0.15$. Along with the encoder responsivities, the weights of the decoder W_{nm} were trained. As shown in Fig. 4a, the loss steeply decreases within the first ~10 training epochs and then slowly converges to a final value after about 30 epochs. The initial and final responsivities/weights of the encoder/decoder are shown in Fig. 4b and Extended Data Fig. 8, and the coded representations for each letter are depicted in Fig. 4c. Each projected letter delivers a unique signal pattern at the output. A projected 'n' delivers negative currents to code-layer neurons 1 and 2 and a positive current to code-layer neuron 3. After the sigmoid function, this causes only code-layer neuron 3 to deliver a sizeable signal. The letters 'v' and 'z' activate two code-layer neurons: 'v', code-layer neurons 0 and 2; 'z', code-layer neurons 1 and 2. The decoder transforms the coded signal back into an output that correctly represents the input. To test the fault tolerance of the autoencoder, we projected twice as noisy ($\sigma = 0.3$) images. Not only did the autoencoder interpret the inputs correctly, but the reconstructions were considerably less noisy (Fig. 4d).

As image sensing and processing are both performed in the analogue domain, the operation speed of the system is limited only by physical processes involved in the photocurrent generation[31]. As a result, image recognition and encoding occur in real time with a rate that is orders of magnitude higher than what can be achieved conventionally. To demonstrate the high-speed capabilities of the sensor, we performed measurements with a 40-ns pulsed laser source (522 nm, ~10 W cm^{-2}). The photodiode array was operated as a classifier and trained beforehand, as discussed above. We subsequently projected two letters ('v' and 'n') and measured the time-resolved currents of the two corresponding channels. In Fig. 5 we plot the electric output pulses, which demonstrate correct pattern classification within ~50 ns. The system is thus capable of processing images with a throughput of 20 million bins per second. This value is limited only by the 20-MHz bandwidth of the used amplifiers, and substantially higher rates are possible. Such a network may hence provide new opportunities for ultrafast machine vision. It may also be employed in ultrafast spectroscopy for the detection and classification of spectral events. We also note that the operation of the vision sensor is self-powered (photovoltaic device) and electrical energy is consumed only during training.

Let us now comment on the prospects for scalability. In our present implementation the weights of the ANN are stored in an external memory and supplied to each detector via cabling. Scaling will require storing the weights locally. This could be achieved, for example, by using ferroelectric gate dielectrics or by employing floating gate devices[32–34]. To demonstrate the feasibility of the latter approach, we present in Extended Data Fig. 9 a floating split-gate photodetector. Once set, this detector 'remembers' its responsivity value and delivers a photocurrent of adjustable sign/magnitude. During training, each detector could then be addressed by its column and row, using the standard infrastructure of active pixel cameras.

Another important question is the number of required subpixels M. As shown in the example in Fig. 1d, a segmentation of each pixel into 3 × 3 subpixels may be adequate for some applications. Given the exponential increase of network complexity with M, increasing the segmentation to 6 × 6 subpixels would already result in a very powerful ANN with a manageable number of 36 analogue outputs. We propose that such a network may also be trained as a binary-hashing[35] autoencoder, eliminating the need for analogue-to-digital conversion. Binary hashing encodes each feature into a binary code of the output signal, which means that a 36-bit digital output allows as many as $2^{36} - 1 \approx 7 \times 10^{10}$

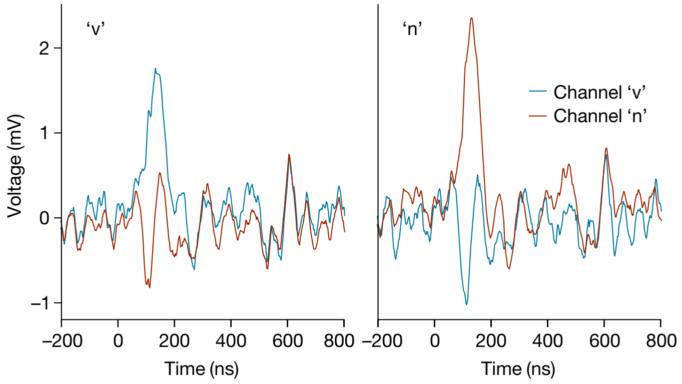


Fig. 5 | Ultrafast image recognition. Projection of two different letters, 'v' and 'n', with a duration of 40 ns, leads to distinct output voltages of the labelled channels.

encodable features. The implementation of an analogue deep-learning network becomes feasible by converting the photocurrents into voltages that are then fed into a memristor crossbar. We finally remark that besides on-chip training, demonstrated here, the network can also be trained off-line using computer simulations, and the predetermined photoresponsivity matrix is then transferred to the device.

In conclusion, we have presented an ANN vision sensor for ultrafast recognition and encoding of optical images. The device concept is easily scalable and provides various training possibilities for ultrafast machine vision applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2038-x.

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Mead, C. A. & Mahowald, M. A. A silicon model of early visual processing. *Neural Netw*. **1**, 91–97 (1988).
3. Lichtsteiner, P., Posch, C. & Delbruck, T. A. 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE J. Solid-State Circuits* **43**, 566–576 (2008).
4. Cottini, N., Gottardi, M., Massari, N., Passerone, R. & Smilansky, Z. A. 33 μW 64×64 pixel vision sensor embedding robust dynamic background subtraction for EVENT detection and scene interpretation. *IEEE J. Solid-State Circuits* **48**, 850–863 (2013).
5. Kyuma, K. et al. Artificial retinas—fast, versatile image processors. *Nature* **372**, 197–198 (1994).
6. Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B. & Delbruck, T. Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. *Proc. IEEE* **102**, 1470–1484 (2014).
7. Zhou, F. et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol*. **14**, 776–782 (2019).
8. Manzeli, S., Ovchinnikov, D., Pasquier, D., Yazyev, O. V. & Kis, A. 2D transition metal dichalcogenides. *Nat. Rev. Mater*. **2**, 17033 (2017).
9. Mueller, T. & Malic, E. Exciton physics and device application of two-dimensional transition metal dichalcogenide semiconductors. *npj 2D Mater. Appl*. **2**, 29 (2018).
10. Pospischil, A., Furchi, M. M. & Mueller, T. Solar-energy conversion and light emission in an atomic monolayer p–n diode. *Nat. Nanotechnol*. **9**, 257–261 (2014).
11. Baugher, B. W., Churchill, H. O., Yang, Y. & Jarillo-Herrero, P. Optoelectronic devices based on electrically tunable p–n diodes in a monolayer dichalcogenide. *Nat. Nanotechnol*. **9**, 262–267 (2014).
12. Ross, J. S. et al. Electrically tunable excitonic light-emitting diodes based on monolayer WSe_2 p–n junctions. *Nat. Nanotechnol*. **9**, 268–272 (2014).
13. Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).
14. Sheridan, P. M. et al. Sparse coding with memristor networks. *Nat. Nanotechnol*. **12**, 784–789 (2017).
15. Li, C. et al. Analogue signal and image processing with large memristor crossbars. *Nat. Electron*. **1**, 52–59 (2018).
16. Kim, K. H. et al. A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications. *Nano Lett*. **12**, 389–395 (2012).
17. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon*. **11**, 441–446 (2017).
18. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
19. Hamerly, R., Bernstein, L., Sludds, A., Soljačić, M. & Englund, D. Large-scale optical neural networks based on photoelectric multiplication. *Phys. Rev. X* **9**, 021032 (2019).
20. Psaltis, D., Brady, D., Gu, X. G. & Lin, S. Holography in artificial neural networks. *Nature* **343**, 325–330 (1990).
21. Kolb, H. How the retina works: much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. *Am. Sci*. **91**, 28–35 (2003).
22. Jeong, K.-H., Kim, J. & Lee, L. P. Biologically inspired artificial compound eyes. *Science* **312**, 557–561 (2006).
23. Choi, C. et al. Human eye-inspired soft optoelectronic device using high-density MoS_2-graphene curved image sensor array. *Nat. Commun*. **8**, 1664 (2017).
24. Schwarte, R. et al. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). In *Proc. SPIE Sensors, Sensor Systems, and Sensor Data Processing* Vol. 3100, 245–253 (SPIE, 1997).
25. Sugeta, T., Urisu, T., Sakata, S. & Mizushima, Y. Metal-semiconductor-metal photodetector for high-speed optoelectronic circuits. *Jpn. J. Appl. Phys*. **19**, 459 (1980).
26. Wachter, S., Polyushkin, D. K., Bethge, O. & Mueller, T. A microprocessor based on a two-dimensional semiconductor. *Nat. Commun*. **8**, 14948 (2017).
27. Goossens, S. et al. Broadband image sensor array based on graphene–CMOS integration. *Nat. Photon*. **11**, 366–371 (2017).
28. Bishop, C. M. Training with noise is equivalent to Tikhonov regularization. *Neural Comput*. **7**, 108–116 (1995).
29. Bengio, Y. in *Neural Networks: Tricks of the Trade* Vol. 7700 (eds Montavon G. et al.) 437–478 (Springer, 2012).
30. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
31. Massicotte, M. et al. Dissociation of two-dimensional excitons in monolayer WSe_2. *Nat. Commun*. **9**, 1633 (2018).
32. Li, D. et al. Two-dimensional non-volatile programmable p–n junctions. *Nat. Nanotechnol*. **12**, 901–906 (2017).
33. Lv, L. et al. Reconfigurable two-dimensional optoelectronic devices enabled by local ferroelectric polarization. *Nat. Commun*. **10**, 3331 (2019).
34. Bertolazzi, S., Krasnozhon, D. & Kis, A. Nonvolatile memory cells based on MoS_2/graphene heterostructures. *ACS Nano* **7**, 3246–3252 (2013).
35. Salakhutdinov, R. & Hinton, G. Semantic hashing. *Int. J. Approx. Reason*. **50**, 969–978 (2009).

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

Methods

Device fabrication

The fabrication of the chip followed the procedure described in ref. [26]. As a substrate we used a silicon wafer, coated with 280-nm-thick SiO_2. First, we prepared a bottom metal layer by writing a design with electron-beam lithography (EBL) and evaporating Ti/Au (3 nm/30 nm). Secondly, we deposited a 30-nm-thick Al_2O_3 gate oxide using atomic layer deposition. Via holes through the Al_2O_3 isolator, which were necessary for the connections between the top and bottom metal layers, were defined by EBL and etched with a 30% solution of KOH in deionized water. Thirdly, we mechanically exfoliated a ~70 × 120 μm^2 WSe_2 flake from a bulk crystal (from HQ Graphene) and transferred it onto the desired position on the sample by an all-dry viscoelastic stamping method[36]. The crystal thickness (about six monolayers, or ~4 nm) was estimated from the contrast under which it appears in an optical microscope. Next, we separated 27 pixels from the previously transferred WSe_2 sheet by defining a mask with EBL and reactive ion etching with Ar/SF_6 plasma. Mild treatment with reactive ion etching oxygen plasma allowed the removal of the crust from the surface of the polymer mask that appeared during the preceding etching step. Then, a top metal layer was added by another EBL process and Ti/Au (3 nm/32 nm) evaporation. We confirmed the continuity and solidity of the electrode structure by scanning electron microscopy and electrical measurements. Finally, the sample was mounted in a 68-pin chip carrier and wire-bonded.

Experimental setup

Schematics of the experimental setup are shown in Fig. 2d and Extended Data Fig. 4a–c. Light from a semiconductor laser (650 nm wavelength) was linearly polarized before it illuminated a spatial light modulator (SLM; Hamamatsu), operated in intensity-modulation mode. On the SLM, the letters were displayed and the polarization of the light was rotated depending on the pixel value. A linear polarizer with its optical axis oriented normal to the polarization direction of the incident laser light functioned as an analyser. The generated optical image was then projected onto the sample using a 20× microscope objective with long working distance (Mitutoyo). Pairs of gate voltages were supplied to each of the detectors individually using a total of 54 digital-to-analogue converters (National Instruments, NI-9264) and the three output currents were measured by source meters (Keithley, 2614B). For time-resolved measurements, a pulsed laser source emitting ~40-ns-long pulses at 522 nm wavelength was used. The output current signals were amplified with high-bandwidth (20 MHz) transimpedance amplifiers (Femto) and the output voltages were recorded with an oscilloscope (Keysight). For the time-resolved measurements, the analyser was removed and the SLM was operated in phase-only mode to achieve higher illumination intensities (~10 W cm^{-2}). The phase-only Fourier transforms of the projected images were calculated using the Gerchberg–Saxton algorithm[37]. For reliable and hysteresis-free operation, the vision sensor was placed in a vacuum chamber (~10^{-6} mbar). Alternatively, a protective dielectric encapsulation layer may be employed to isolate the two-dimensional semiconductor from the environment.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

36. Castellanos-Gomez, A. et al. Deterministic transfer of two-dimensional materials by all-dry viscoelastic stamping. *2D Mater.* **1**, 011002 (2014).

37. Gerchberg, R. W. & Saxton, W. O. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik* **35**, 237–246 (1972).

Acknowledgements We thank B. Limbacher for discussions and A. Kleinl, M. Paur and F. Dona for technical assistance. We acknowledge financial support by the Austrian Science Fund FWF (START Y 539-N16) and the European Union (grant agreement number 785219 Graphene Flagship and Marie Sklodowska-Curie Individual Fellowship OPTOvanderWAALS, grant ID 791536).

Author contributions T.M. conceived the experiment. L.M. designed and built the experimental setup, programmed the machine-learning algorithms, carried out the measurements and analysed the data. J.S. fabricated the ANN vision sensor. S.W. and D.K.P. contributed to the sample fabrication. A.J.M.-M. fabricated and characterized the floating-gate detector. L.M., J.S. and T.M. prepared the manuscript. All authors discussed the results and commented on the manuscript.

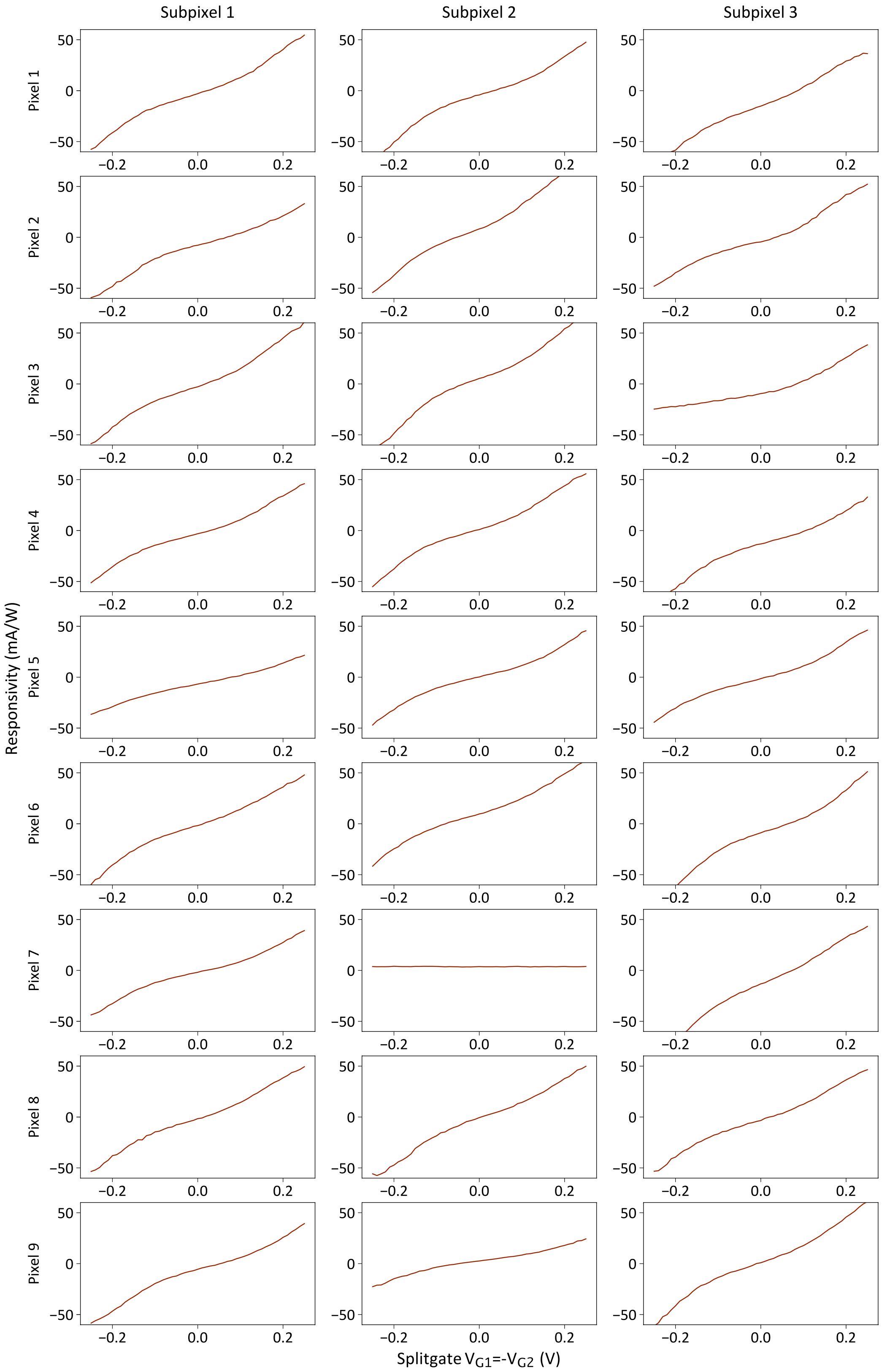
Competing interests The authors declare no competing interests.

Additional information

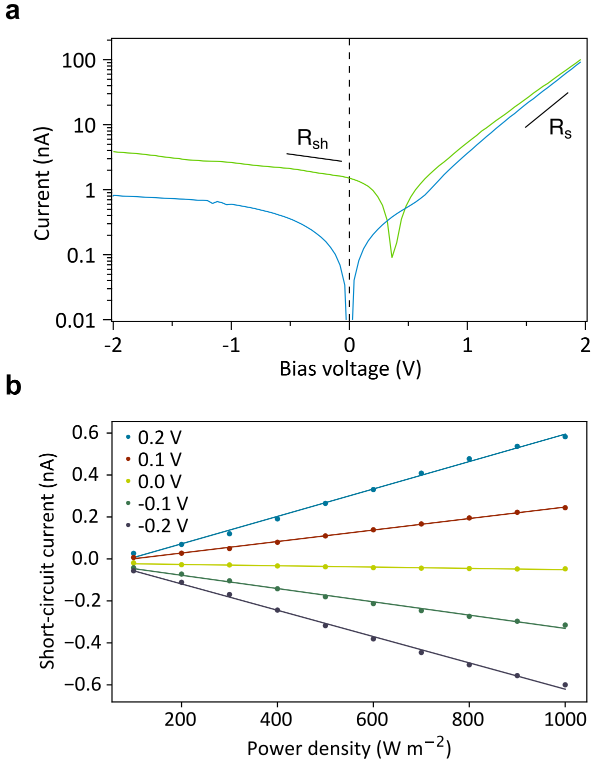
Correspondence and requests for materials should be addressed to L.M. or T.M.

Peer review information *Nature* thanks Yang Chai, Frank Koppens and Sangyoun Lee for their contribution to the peer review of this work.

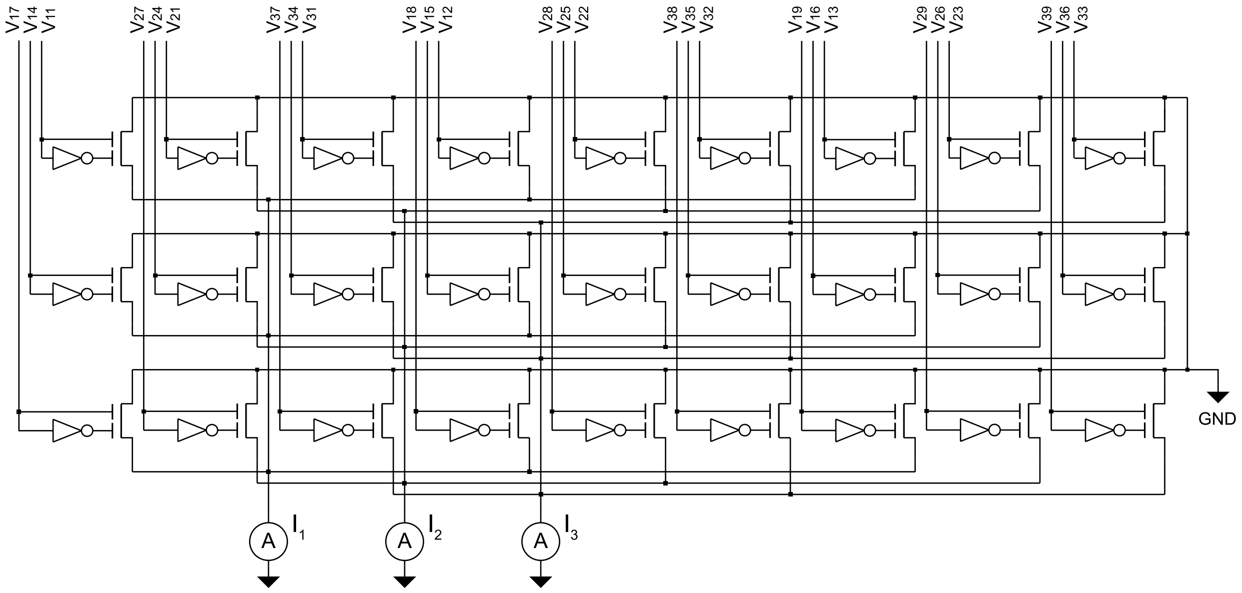
Reprints and permissions information is available at http://www.nature.com/reprints.



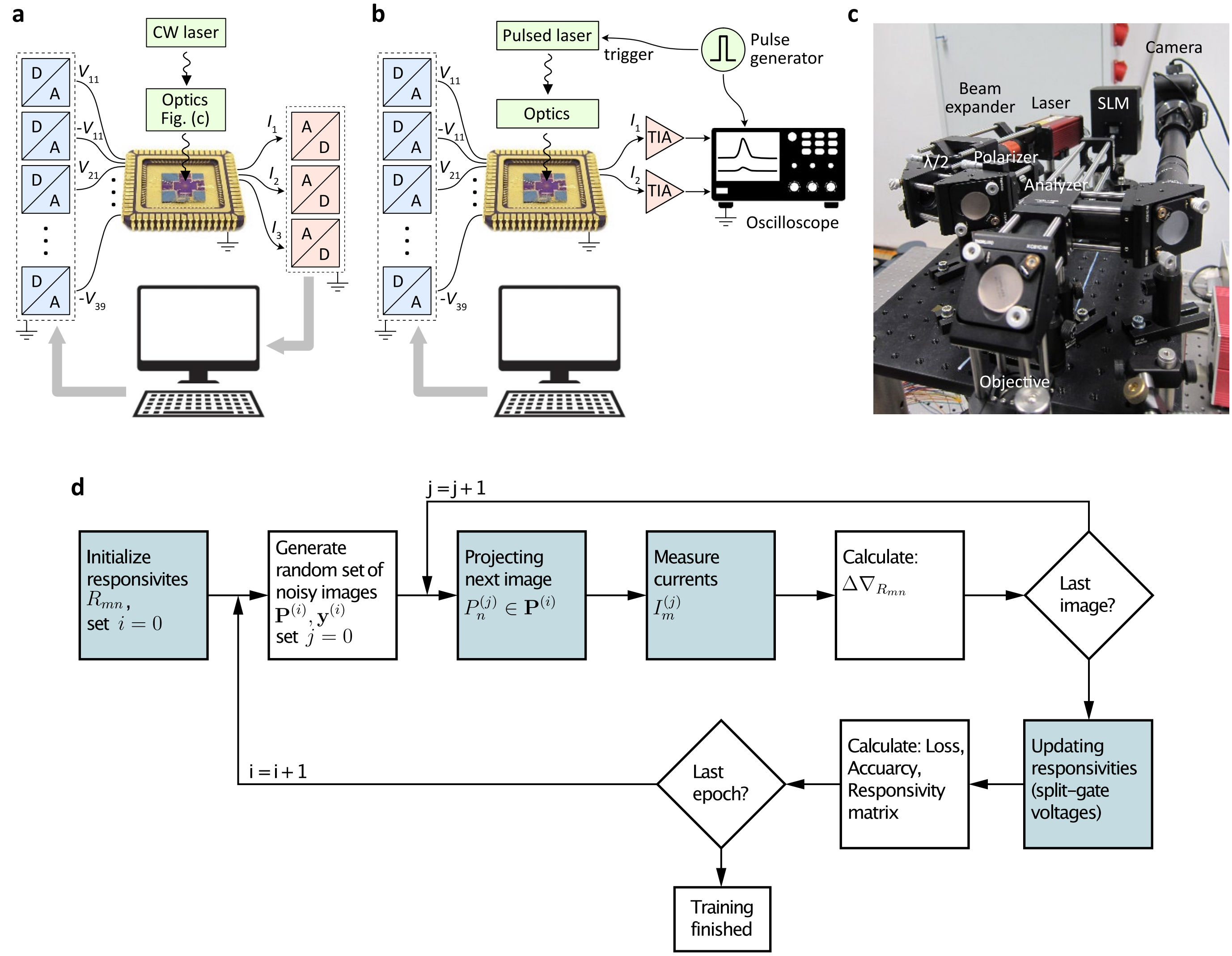
Extended Data Fig. 1 | Photodiode array uniformity. Gate tunability of the responsivities of all 27 photodetectors. One of the detector elements (pixel 7, subpixel 2) did not show any response to light (due to a broken electrical wire), which, however, had no crucial influence on the overall system performance.



Extended Data Fig. 2 | Photodiode characteristics. a, Current–voltage characteristic curve under dark (blue) and illuminated (green) conditions. The series resistance R_s and shunt resistance R_{sh} are ~10^6 Ω and 10^9 Ω, respectively. For zero-bias operation, we estimate a noise-equivalent power of NEP = $I_{th}/R \approx 10^{-13}$ W Hz$^{-1/2}$, where $R \approx 60$ mA W^{-1} is the (maximum) responsivity and $I_{th} = \sqrt{4k_BT\Delta f/R_{sh}}$ the thermal noise, where k_B is the Boltzmann constant, Δf is the bandwidth and T is the temperature. **b**, Dependence of the short-circuit photocurrent on the light intensity for different split-gate voltages. Importantly, the response is linear ($I \propto P$), as assumed in equation (1).

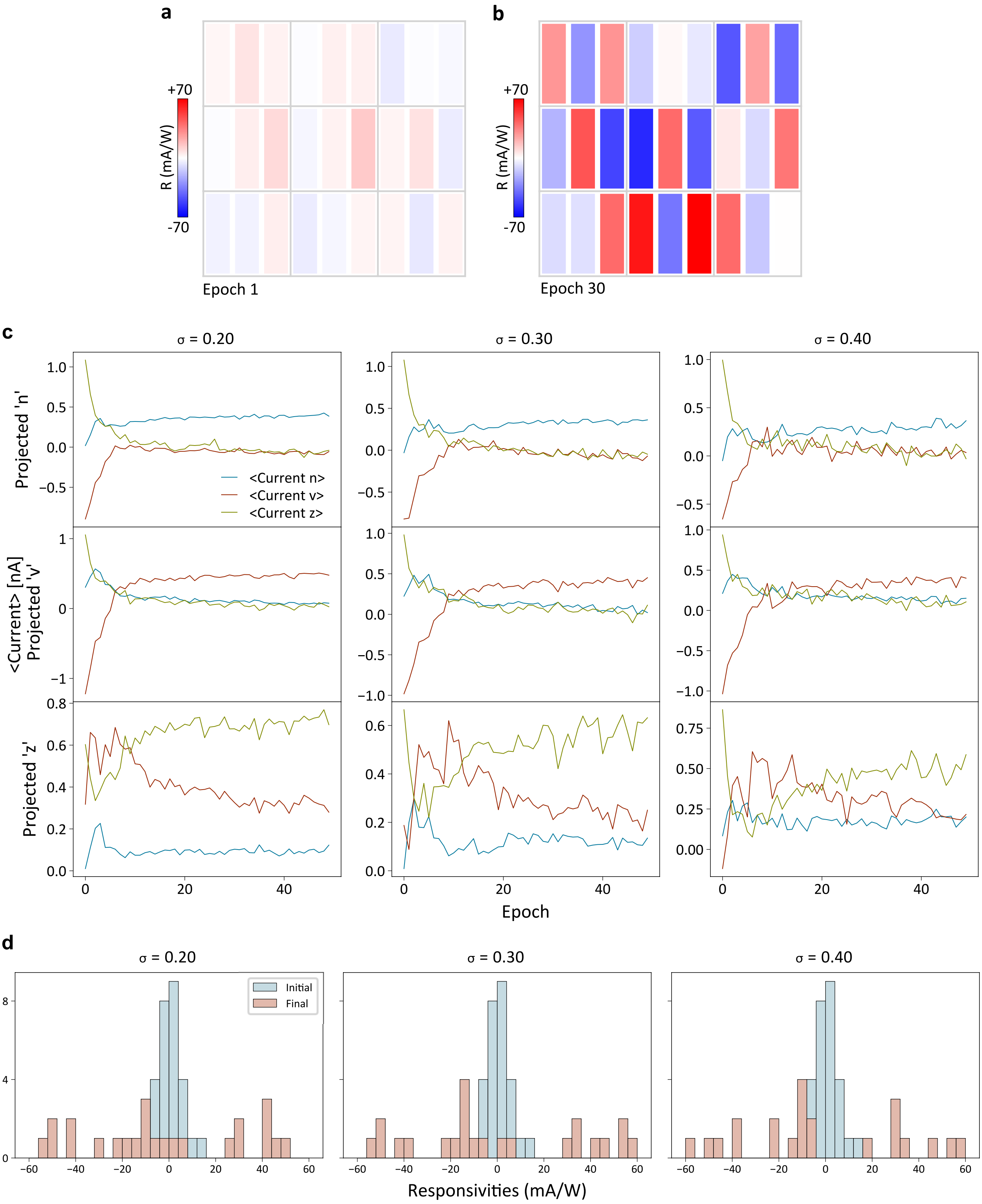


Extended Data Fig. 3 | Circuit of the ANN photodiode array.

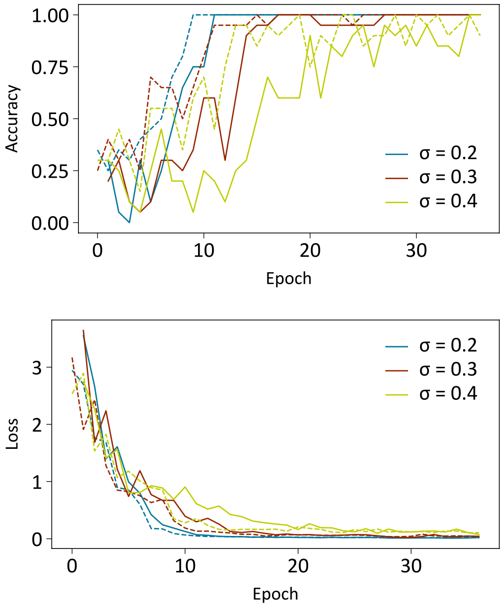


a
CW laser
Optics Fig. (c)
D A
V_{11}
$-V_{11}$
V_{21}
$-V_{39}$
I_1
I_2
I_3
A D
b
Pulsed laser
trigger
Pulse generator
Optics
TIA
Oscilloscope
c
Camera
Beam expander
Laser
SLM
λ/2
Polarizer
Analyzer
Objective
d
Initialize responsivites R_{mn}, set $i = 0$
Generate random set of noisy images $\mathbf{P}^{(i)}, \mathbf{y}^{(i)}$ set $j = 0$
j = j + 1
Projecting next image $P_n^{(j)} \in \mathbf{P}^{(i)}$
Measure currents $I_m^{(j)}$
Calculate: $\Delta\nabla_{R_{mn}}$
Last image?
Updating responsivities (split–gate voltages)
Calculate: Loss, Accuracy, Responsivity matrix
Last epoch?
i = i + 1
Training finished

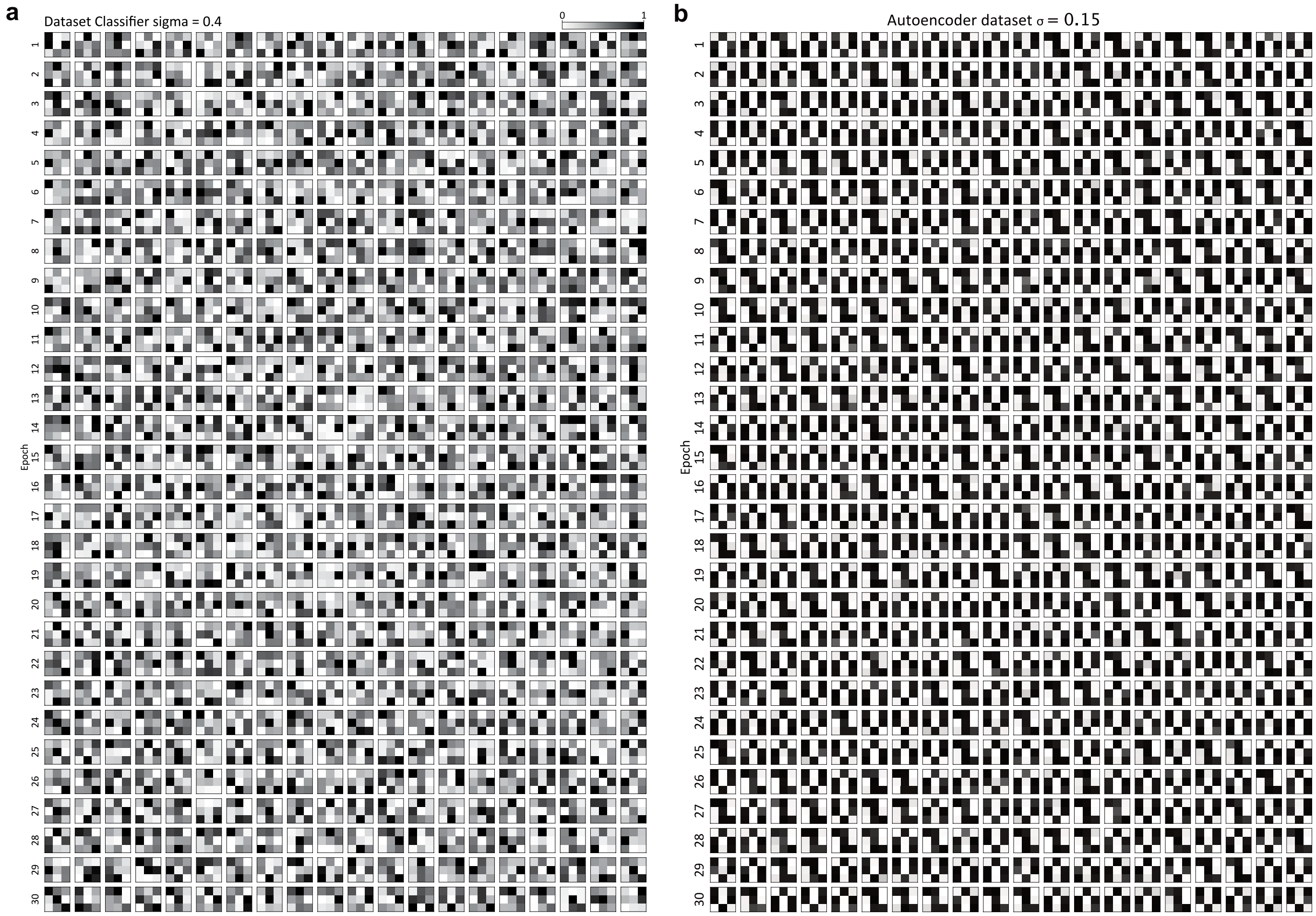
Extended Data Fig. 4 | Experimental setup. a, Experimental setup for training the classifier and the autoencoder. CW, continuous wave. **b**, Experimental setup for time-resolved measurements. TIA, transimpedance amplifier. A pulse generator triggers the pulsed laser as well as the oscilloscope. **c**, Photograph of the optical setup (for schematic see Fig. 2d). **d**, Flow chart of the training algorithm. The blue shaded boxes are interactions with the ANN photodiode array.



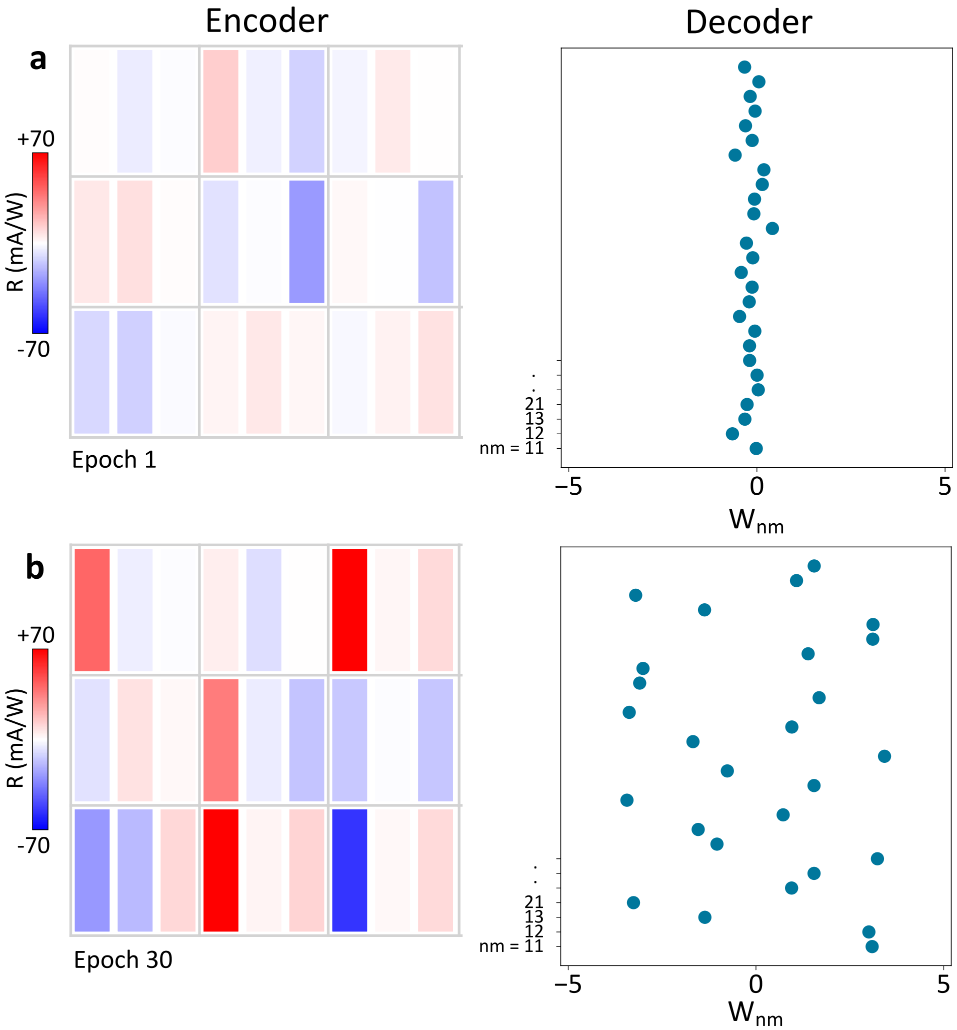
Extended Data Fig. 5 | Classifier training. Photoresponsivity values of all 27 photodetectors with $\sigma = 0.3$ training data. **a**, **b**, Initial (**a**) and epoch 30 (**b**) responsivity values. The weights for the $\sigma = 0.2$ and $\sigma = 0.4$ training data are similar. **c**, Measured currents over all epochs for a specific projected letter and at all three noise levels. **d**, Histogram of the initial and final responsivity values for the three different noise levels.



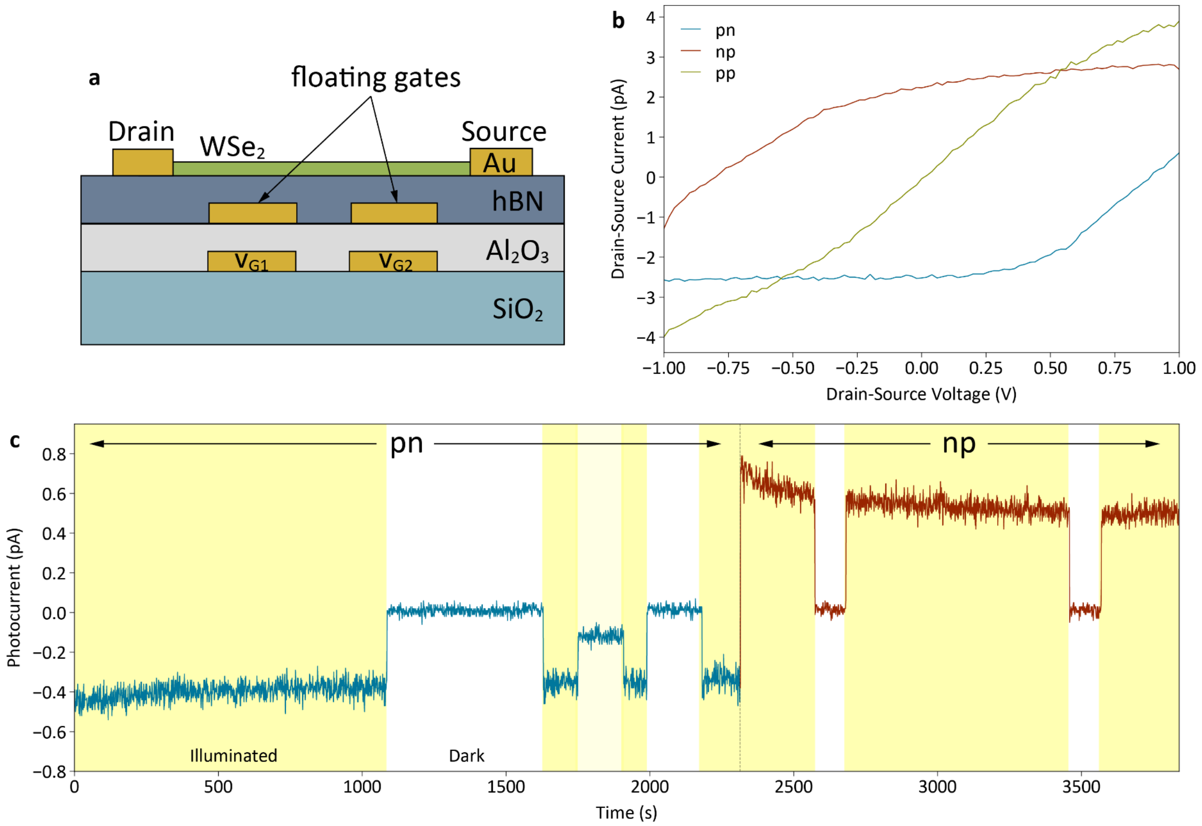
Extended Data Fig. 6 | Comparison with computer simulation. Classifier training of the analogue vision sensor (solid lines) and simulation of the system on a computer (dashed lines) for different data noise levels *σ*. The same ANN architecture, input data, effective learning rate and starting weights have been used. The same accuracy and loss are eventually reached after training. The slightly slower convergence of the analogue implementation compared with the simulation reflects the nonidealities (defective subpixel, device-to-device variations) of the former. Further discussion on the impact of nonidealities is provided in Extended Data Fig. 10.



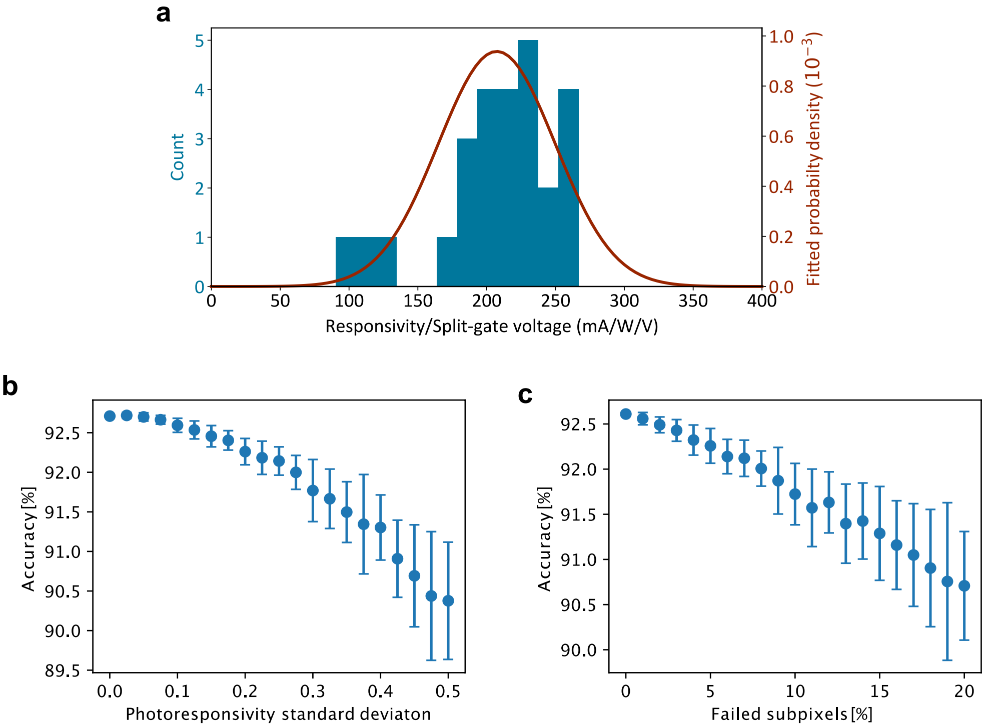
Extended Data Fig. 7 | Training datasets. a, **b**, Dataset of 30 epochs of classifier (**a**) and autoencoder training (**b**) with a test data noise level of $\sigma = 0.4$ and $\sigma = 0.15$ respectively.



Extended Data Fig. 8 | Autoencoder photoresponsivities/weights. a, **b**, Initial (**a**) and epoch 30 (**b**) encoder photoresponsivity values (left) and decoder weights (right).



Extended Data Fig. 9 | Floating-split-gate photodiode with memory. **a**, Schematic of the floating gate photodiode. The addition of 2-nm-thick Au layers, sandwiched between Al$_2$O$_3$ and hexagonal boron nitride (hBN), enables the storage of electric charge when a gate voltage is applied to the device, acting as a floating-gate memory. **b**, Electronic characteristic curves of the photodiode operated in p–n, n–p and p–p configurations. **c**, The ability of the device to 'remember' the previous configuration can be verified from the time-resolved photocurrent measurement. The measurement is performed as follows: the back-gate voltages are set to V_{G1} = +5 V and V_{G2} = −5 V and are then disconnected, that is, there is no longer an applied gate voltage and the only electric field is that generated by the charge stored on the floating electrodes. The short-circuit photocurrent is then measured upon optical illumination. The light is then switched off, at ~1,100 s, with a corresponding drop of the photocurrent to zero. After ~1,600 s, the light is switched on again, causing the current to reach its initial value, and then a smaller value when the intensity of the light is reduced (~1,700 s). After ~2,300 s, the opposite voltage configuration is applied to the back gates (V_{G1} = −5 V and V_{G2} = +5 V), inducing a polarity inversion that also remains permanent. Now, a positive photocurrent (red line) is obtained.



Extended Data Fig. 10 | Robustness of the network. **a**, Detector uniformity, extracted from Extended Data Fig. 1. The fitted Gaussian probability distribution has a standard deviation of $\sigma = 0.205$ (40 mA W^{-1} V^{-1}). **b**, Monte Carlo simulation of a vision sensor with detector responsivities of a given standard deviation. (The photodetectors of the actual device have a measured photoresponsivity standard deviation of 0.205.) Trained on the MNIST database of handwritten digits, the classifier has 784 pixels and 10 subpixels per pixel. For each data point, 50 random photoresponsivity variations were evaluated. **c**, Accuracy dependence on the number of (randomly chosen) defective subpixels. The same ANN and Monte Carlo simulation scheme as in **b** were used. For each data point, 50 random sets of modified photoresponsivities were evaluated.