

# In-sensor image memorization and encoding via optical neurons for bio-stimulus domain reduction towards visual cognitive processing

**Doeon Lee**

Yonsei University

**Minseong Park**

University of Virginia

**Yongmin Baek**

University of Virginia <https://orcid.org/0000-0002-4387-3042>

**Byungjoon Bae**

University of Virginia

**Junseok Heo**

Ajou University <https://orcid.org/0000-0002-8125-6141>

**Kyusang Lee** (✉ [kyusang@virginia.edu](mailto:kyusang@virginia.edu))

University of Virginia <https://orcid.org/0000-0001-7123-3395>

---

## Article

**Keywords:** In-sensor computing, neuromorphic computing, image classification, edge-computing, artificial neural network

**Posted Date:** April 6th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1507544/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Communications on September 5th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-32790-3>.

# Abstract

As machine vision technology generates large amounts of data from sensors, it requires efficient computational systems for visual cognitive processing. However, the spatially separated sensor/processor architecture suffers from data transportation bottleneck issues due to repeated and redundant data exchange between them. Recently, in-sensor computing systems have emerged as a potential solution for reducing unnecessary data transfer and realizing fast and energy-efficient visual cognitive processing. However, they still lack the capability to process stored images directly within the sensor. Here, we demonstrate an advanced machine vision system for in-sensor visual cognitive processing, emulating a mammalian image encoding process to extract features from the input images, to achieve a domain reduction from the sensor. Heterogeneous integration of a compound semiconductor-based 1-photodiode and oxide-based 1-memristor array enables to memorize images in pixels, and directly conducts a multiply–accumulate operation within the sensor. Since trained weight values are applied as an input to the image-saved crossbar array instead of storing the weight value in it as using the vectorized image as an input for multiply–accumulate operations in our system. Therefore, the input data towards the artificial neural network (ANN) is significantly reduced, and vectorization of the image is no longer needed, thereby minimizing data transportation bottlenecks. We believe the heterogeneously integrated in-sensor computing platform provides an advanced architecture for real-time and data-intensive machine-vision applications via bio-stimulus domain reduction.

# Introduction

Machine vision technology provides the capability to inspect and analyze the surrounding environment using image sensors integrated with processing units.<sup>1-4</sup> Recent advances in image sensing and artificial intelligence (AI) technology have enabled the efficient generation of digital images from the physical world and interpretation of the acquired images, respectively. A key feature of machine vision is real-time object recognition and/or classification via machine learning using an artificial neural network (ANN), which is inspired by the visual reasoning process of mammals.<sup>5</sup> Therefore, machine vision has been widely employed in applications that require in-situ sensing, environment analysis, and interpretation of visual images, such as autonomous vehicles, robotics, and smart production systems.<sup>1,6-8</sup>

In conventional machine vision systems, image sensors are externally connected to memory and processing units by adapting the von Neumann computing architectures. Sensing occurs in the analog domain, which leads to the generation of large amounts of raw data in the image sensor. The large amounts of redundant analog data, including unnecessary background information, must be converted to digital data via an analogue-digital-converter (ADC), then transferred to processing units or cloud-based computing systems<sup>6,9</sup>, in which image processing and analysis are conducted for machine vision applications. Thus, this sensory data processing requires energy-hungry long-distance data communication from the sensors to the memories and processing units with a limited data transfer rate.<sup>1,6,10</sup> This massive amounts of data conversion/transportation causes significant issues in terms of

energy consumption, delayed response time, and the communication bandwidth, all of which are important for machine vision applications with strict delay and power-consumption requirements. Therefore, this data transportation bottleneck issue in integrated sensor and processor systems should be mitigated for fast and efficient machine vision processes.

In contrast, the human vision system outperforms conventional imager and machine vision systems, particularly with regard to unstructured image classification and recognition, low latency, and energy efficiency<sup>6</sup>. The human retina detects images with rod and cone photoreceptors and conveys them through the optic nerve to the brain, which then processes the collected information for visual perception. Interestingly, many of the imaging processes begin in a neural circuit within the retina<sup>11-13</sup>. The retina compresses the large volume of visual signals detected by approximately 100 million photoreceptor cells, and then transmits the encoded data to the brain via the optic nerve, without latency or significant energy consumption. This in-retina computation mechanism can be effectively applied to machine vision systems, alleviating the data transportation bottleneck issues at the sensor/processor interface due to the big data acquisition at the sensor nodes, as well as the heavy computation burden and reliance on the post-processor.

Analogous to the neural circuit and photoreceptor cells in the retina, photodetectors in image sensors can be directly integrated with artificial synapses (e.g., oxide memristors, phase change memories), constructing an artificial neural circuit.<sup>14-16</sup> Integrated analog computing units can store image information as resistance states, while also performing computational tasks to implement an ANN for a cognitive algorithm. In this way, so-called “in-sensor computing,” the image information can be processed within the sensors, significantly reducing data movement by processing it at the edge of the system in a similar way to mammalian vision.<sup>1,6</sup> There have been efforts to demonstrate in-sensor computing systems for analog machine vision using optoelectronic memory devices. However, previously demonstrated in-sensor machine vision systems mostly conducted image memorization and pre-processing (e.g., image enhancement) within the sensors.<sup>2,3,17-25</sup> Thus, all the image data still needs to be transferred to a back-end post-processor for high-level image processing, such as feature extraction, image encoding, and image classification. However, minimizing unnecessary data transportation between the sensor, memory, and computation unit is the key feature and ultimate goal of in-sensor computing to achieve an energy-efficient and latency-free sensor/processor system. As such, advancements in-sensor computing systems require the capability to memorize images and directly perform high-level image processing within the sensor.

In this study, we developed an advanced in-sensor computing system with neuromorphic image memorization and encoding capabilities within the pixel for visual cognitive processing, emulating the biological visual processing system of the mammalian retina. This in-pixel computing system efficiently computes and conveys visual information to minimize data-transportation bottlenecks. In a single sensor, a photodiode is directly integrated with resistive random access memory (ReRAM) to construct 1-photodiode 1-resistive random access memory (1P-1R) pixels, where HfO<sub>2</sub>-based ReRAMs are fabricated

on InGaAs-based p-i-n photodiodes. First, we fabricated a 1P-1R single pixel with an InGaAs photodetector and HfO<sub>2</sub>-based ReRAM and characterized the electrical and optoelectrical properties for the image memorization and data processing of visual information within the sensor. Subsequently, a 16×16 1P-1R crossbar array with an InGaAs photodiode and a HfO<sub>2</sub>-based ReRAM was fabricated and characterized. Subsequently, the imaging of the MNIST handwritten digits was performed, where the visual stimuli of the images were effectively stored in pixels. Subsequently, using the fabricated 1P-1R array, the biomimicking image encoding process was performed by in-memory vector-matrix-multiplication without data transfer. In contrast to typical in-memory computing methods, input image data were stored in ReRAMs and weights were applied to the crossbar array as input voltages, in which 2D-1D vectorization was no longer needed and the size of the crossbar array was significantly reduced. The encoded images were then conveyed to a central processor for image classification. Our process dramatically reduces redundant data movement between the sensor, memory, and processor by performing in-pixel encoding, possibly alleviating data transportation bottlenecks and energy overconsumption.

## Results

Figure 1a shows a schematic illustration of the image encoding and classification processes in the mammalian retinal and brain system. The outputs of the rod and cone photoreceptors are decomposed into approximately 12 parallel information streams, which are then connected to the retinal ganglion cells. Bipolar and amacrine cell activity are combined in a ganglion cell to create diverse encodings of features extracted from the visual world, such as edges, direction, and color; the retina then transmits these pre-processed data to the brain<sup>11-13</sup>. By reducing redundant information, the retina can effectively convey image data to the central processor with a minimal transport delay. In the visual cortex, higher-level visual cognitive processes are conducted using encoded images from the retina<sup>11-13</sup>,

In this study, we designed and demonstrated an in-sensor neuromorphic machine vision system with functionalities of image memorization and processing, by mimicking the above-mentioned neural circuit and visual classification system in the human eye, as shown in Fig. 1b. The image sensor consists of a crossbar array of photodetectors and resistive memory cells, which correspond to photoreceptors and ganglion cells in the retina, respectively. In the retina, the ganglion cells operate as a pre-computing processor unit, whereas the ReRAM in our system serves both as a memory and computation unit, depending on the polarity and magnitude of the applied bias to each pixel. When a reverse bias with respect to the photodiode is applied to the 1P-1R pixels, the sensor operates in a memorization mode, in which incident light stimuli are converted to electrical signals in the photodetectors, and the photocurrents are subsequently stored in the memory cells by changing the conductance of the memory. Under a forward-biased voltage with respect to the photodiode (lower than the threshold voltage for the erase operation), the sensor operates in the computing mode to process the stored image at the pixels via analog in-memory computing for vector-matrix multiplication. Because vector-matrix multiplication is a key operation in the ANN algorithm, we utilized the 1P-1R crossbar array to execute in-sensor image

encoding, which extracts critical features from the original image to alleviate the data transfer burden at the sensor and processor interface, paralleling biological processes in the human retina. Finally, image classification was conducted in the post-processing unit with the encoded images delivered through an ANN. While the encoded images possess compressed information compared to the original images, the ANN successfully classifies the object with less computational load.

Figure 1c shows block diagrams of the image-processing sequence in conventional in-sensor processing systems and our neuromorphic in-pixel computing system.<sup>6,21,23,24</sup> Most previously reported conventional in-sensor processing systems only perform image memorization and pre-processing (low-level processing) within the sensors, such as image contrast enhancement and noise reduction. Meanwhile, the size of the pre-processed images from the sensors ( $N \times N$ ) was still the same as the size of the original images ( $N \times N$ ), and high-level image processing took place in the post-processor. Therefore, conventional systems barely reduce the data traffic load at the sensor/processor interface, as well as the computational burden in the post-processor. In contrast, the fabricated neuromorphic in-pixel computing system in this work memorizes images in each pixel and subsequently conducts image encoding by analog in-memory multiply–accumulate operation (1st high-level processing) by combining the sensing and computing functions. Therefore, the size of the output data from the sensor ( $N$ ) is effectively reduced to the square root of the  $N \times N$  original image by the in-sensor image encoding process, thereby minimizing the transportation of redundant data and reducing the computation load in the post-processor. Therefore, this advance in-sensor computing architecture can significantly reduce the inefficient energy usage and high data latency of smart imaging systems.

First, a  $16 \times 16$  1P-1R in-pixel computing chip was fabricated using p-i-n InGaAs photodiodes and HfO<sub>2</sub>-based ReRAM (see Methods). The optical images of the fabricated chip and optical microscope images of the 1P-1R crossbar array are shown in Fig. 1d and Figs. 1e,f, respectively. Each pixel consisted of an InGaAs photodiode and ReRAM. The row lines share the top electrodes of the ReRAMs (Ta/Pt electrodes) and the column lines share the top electrode of the InGaAs photodiodes (p + electrodes). Prior to operating the 1P-1R array, we studied the optical and electrical characteristics of a single 1P-1R pixel. Figure 2a shows a schematic illustration of a single 1P-1R pixel, and its equivalent circuit diagram is shown in Fig. 2b, where  $V_P$  and  $V_R$  are the applied voltages of the photodiode and ReRAM, respectively, and  $V_{\text{total}} = V_P + V_R$ . The 1P-1R pixel is composed of an InGaAs p-i-n photodiode and HfO<sub>2</sub>-based ReRAM, which converts incoming optical signals to electrical signals and memorizes the optical information as its resistance. Additionally, the ReRAM is utilized as an in-memory computation unit when it operates in the computation mode. The three primary operations of the 1P-1R device, depending on the applied bias voltage ( $V_{\text{Total}}$ ), are depicted in Fig. 2b: i) memorization, ii) computation, and iii) erasing operations. When  $V_{\text{Total}} > 2.5$  V under light illumination, the photodiode is reverse-biased such that an incident optical signal generates a photocurrent to modulate the resistance of the ReRAM by forming a conductive filament (memorization operation). Thus, the optical signal can be stored in the form of resistance in a synaptic device. The stored image data in the ReRAM can then be directly used for high-level in-sensor processing (computation operation). When  $-1.3$  V  $< V_{\text{Total}} < -0.5$  V, the photodiode operates under the ohmic regime

with relatively low resistance ( $< 50 \Omega$ ) compared to a resistance range of the ReRAM ( $> 1 \cdot 10^3 \Omega$ ); in this way, the 1P-1R circuit can be approximated to a single ReRAM circuit. Thus, the 1P-1R crossbar array can be used for synaptic in-memory computing based on Ohm's and Kirchhoff's laws.<sup>6,10,26</sup> Therefore, ReRAM serves as a cross-functional device for both the memory unit and processing unit for high-level in-sensory image processing. When a high negative bias voltage over the RESET threshold voltage ( $V_{\text{Total}} < -1.3 \text{ V}$ ) is applied across the 1P-1R device, the memorized data in the ReRAM are erased (Erase operation). These three operations are key functions for realizing neuromorphic in-pixel image processing with a 1P-1R crossbar array.

Figure 2c shows the current-voltage (I-V) characteristics of the fabricated InGaAs photodiode in the single 1P-1R device under light illumination ( $\lambda = 532 \text{ nm}$ ) with various light intensities, where  $V = V_p$ , and the bottom electrode on the n-InP layer is grounded. The photocurrents generated from the photodiode under reverse bias modulated the resistance states of the connected ReRAM depending on the incident light intensity. The fabricated  $\text{HfO}_2$ -based ReRAM in the single 1P-1R device was also characterized by applying repetitive positive and negative voltage sweeps for the SET and RESET processes, respectively, whereas the bottom electrode (Ti/Pt) of the ReRAM was grounded (Fig. 2d). Initially, to form a conductive filament, a positive d.c. voltage sweep was employed by increasing the voltage from 0 V to 5 V and the voltage was decreased from 5 V to 0 V with 2  $\mu\text{A}$  of compliance current. After the initial channel forming process, repeated SET/RESET operations are performed by applying the positive and negative sweeps of 3 V (SET) and - 2.5 V (RESET), respectively, along the sweep paths indicated in the graph to switch the state of the ReRAM between the high-resistance state (HRS) and low-resistance state (LRS). The I-V curves show stable bipolar switching behavior with abrupt SET and gradual RESET, which is ideal for binary data storage.

The optoelectronic switching behavior of the integrated 1P-1R pixel was then characterized by an I-V measurement under light illumination ( $\lambda = 532 \text{ nm}$  and  $P = 67 \text{ W/cm}^2$ ), as shown in Fig. 2e. Voltage is applied to the top electrode of the ReRAM (Ta/Pt contact) while the top electrode of the photodiode (p-InGaAs contact) is grounded, as shown in the equivalent circuit diagram in Fig. 2b. The ranges of the three corresponding operations, which are explained in Fig. 2b, depending on the applied bias voltage, are indicated by colored areas in the graph. Unlike the electrical-field-driven switching process of a single ReRAM device, the conductive filament channel of the ReRAM in the 1P-1R pixel is grown by photogenerated current from the connected photodiode (memorization operation), whereas RESET switching is still performed by the application of an electrical field. The conductive filament was formed by the first positive voltage sweep under light illumination, corresponding to the forming loop shown in Fig. 2e. After the forming process, the ReRAM is switched to the OFF state by applying a negative voltage sweep, where the light illumination has no effect on the erase operation (Fig. S1). Subsequently, a positive voltage sweep was conducted on the 1P-1R device under light illumination, switching the ReRAM to the ON state and memorizing the light information (see memorization loop in Fig. 2d). Under dark conditions, the ReRAM in the 1P-1R device cannot be switched ON via a positive voltage sweep owing to the lack of sufficient driving current to build the conductive filament, as shown in Fig. S2, because the

current flow is limited by the reverse-biased dark current of the photodiode. This result clearly demonstrates the capability of the 1P-1R device as binary optoelectronic memory.

Moreover, we demonstrate the functionality of the 1P-1R cell as a multistate optoelectronic memory. The memory effect of the 1P-1R unit was predominantly determined by the characteristics of the ReRAM in the pixel. Hence, the multistate capability of the ReRAM enables the 1P-1R pixel to function as a multistate optoelectronic memory device for in-sensor computing applications. Multiple resistance states in the  $\text{HfO}_2$ -based ReRAM are usually achieved by the application of various compliance currents during SET operation.<sup>27,28</sup> Figure S3 presents the I-V curves of the multiple resistance states in the ReRAM by controlling the compliance currents during the SET operations. When a higher compliance current is applied to the ReRAM during the SET operation, a lower resistance state is formed owing to the continuous growth of the oxygen vacancy ( $V_o$ )-based filament with a higher charge injection into the oxide layer.<sup>27,28</sup> The multiple conductance states achieved in Fig. S3 are plotted in Fig. S4, depending on the applied compliance. This behavior is consistent with the characterization results from previously reported oxide-based ReRAMs.<sup>27-29</sup> Inspired by the above-described resistance modulation method in  $\text{HfO}_2$ -based ReRAMs, the photogenerated currents from the photodiode in the 1P-1R system were used as the driving currents to perform a SET operation on the ReRAM during the memorization process. Therefore, multiple resistance states in the 1P-1R system can be enabled by light illumination with different intensities on the photodiode, generating diverse magnitudes of the driving photocurrent during the memorization process. Figure 3a shows multiple memorization and erase processes to demonstrate the analog resistance states in the 1P-1R device by employing double-voltage sweep I-V measurements under illumination with 532 nm wavelength light. Positive voltage sweeps were applied across the photodiode-memristor ( $V_{RP}$ ) to perform a light-driven SET operation, whereas negative voltage sweeps were applied only to the memristor ( $V_R$ ) for a RESET operation. The I-V curves in Fig. 3a clearly display the multiple resistance behaviors depending on the incident light intensity, where higher intensity light results in the formation of a filament channel with higher conductance. Figure 3b shows the conductance of multiple states depending on the incident light power density in the memorization process, which is nearly identical to the multistate characteristic of a single ReRAM.

The endurance characteristics of the 1P-1R optoelectronic memory were measured after the resistance of the ReRAM was set by illuminating lights with seven different intensities (Fig. S5). The measurement results confirm that the 1P-1R device has a stable and reliable endurance property over  $10^4$  s for memory functionality. To validate the continuous image detection and memorization capability using the 1P-1R system, sequential memorization/erase operations are shown in Fig. 3c. Voltage pulses ( $V_{RP}$ ) for memorization (2 V), erase (-3 V), and read (-0.6 V) operations were applied to the 1P-1R device under dark or light conditions, with a power density of  $50 \text{ mW/cm}^2$ . Without light illumination, no data were stored in the memory even with SET pulses, and only leakage currents were observed. However, the imaging data were stored in the memory once the light illuminated the device with the application of a memorization voltage pulse, and the memory was maintained in the form of an LRS until the application of an erasing voltage pulse. When the erasing voltage pulse was applied to the ReRAM, the device was switched back

to the HRS. Repeated memorization/erase processes were successfully completed using the 1P-1R system. This transient switching characteristic can be utilized to perform continuous image memorization/erase processes by applying a voltage pulse train to the 1P-1R focal plane array without data transfer to the external memory.

Using the fabricated 1P-1R array, we demonstrate in-sensor image storage, encoding, and classification. Prior to image encoding and classification, an image-storage operation was demonstrated. Figures 3d-g show the schematic illustrations of the image memorization, read, and erasing processes, respectively, with the  $16 \times 16$  1P-1R focal plane array. A circuit diagram of the pixels is depicted in Fig. 3d, where a ReRAM in a yellow (dark blue)-colored pixel is in the LRS (HRS) state. To control the 1P-1R array, memorization, read, and erase operations (Fig. 2b) were utilized. For the image memorization process, voltage pulses of +5 V were applied across the individual 1P-1R pixels, where the photodiodes were reverse-biased, to store incident image information in the ReRAMs, as shown in Fig. 3e (see Methods for more experimental details). The stored image is then read by applying voltage pulses of -1 V to each 1P-1R pixel, where the photodiode is forward-biased, to read the resistance states of the ReRAMs (Fig. 3f). To erase the saved image in the sensor, voltage pulses of -5 V were applied to each 1P-1R pixel to switch all pixels to the HRS state, enabling the sensor to be ready to capture the next images (Fig. 3g).

In this study, we characterized and operated a fabricated  $16 \times 16$  1P-1R crossbar array for image memorization under light illumination with a wavelength of 532 nm. First, the I-V characteristics of an individual pixel in the array were measured under 532 nm light illumination (see Fig. S6). Forming voltage pulses of +6 V were applied to all 256 pixels under global light illumination ( $P = 70 \text{ mJ/cm}^2$ ) to form the conductive filament channels in the active medium of ReRAMs. Subsequently, READ voltage pulses of -1 V were applied to each pixel to read the conductance state of each ReRAM in the pixels, followed by an application of RESET voltage pulses of -5 V to switch all pixels to the HRS state for the next image memorization process. Figures 3h and i show the  $12 \times 12$  conductance maps of the InGaAs 1P-1R array before and after the forming process, respectively, on a logarithmic scale (see Figs. S7a and b for details). After the forming process, all pixels were effectively switched from the initial HRS state (Fig. 3h) to the LRS state (Fig. 3i). With the operation-ready 1P-1R array, the memorization function was demonstrated by imaging the handwritten digit images of '4' and '8' from the MNIST dataset.<sup>31</sup> First, the '4' handwritten digit image is illuminated on the 1P-1R array, and +5 V (SET) voltage pulses are applied to each pixel to memorize the exposed image in the sensor as shown in Fig. 3e (see Methods for detailed experimental methods). After the image memorization process, -1 V (READ) voltage pulses were applied to all pixels to read the saved image from the sensor. Figure 3j shows the corresponding current map of the 1P-1R array after the memorization process with the digit '4' image, indicating that the captured image is successfully memorized in the sensor. The saved image in the sensor array is then erased by applying voltage pulses of -5 V (RESET) to all pixels, and second image memorization and readout processes were performed under the image exposure of the MNIST handwritten digit of '8' using the identical procedure described above (Fig. 3k).

The 1P-1R crossbar array can be approximated as a 1R crossbar array under a forward bias condition for the photodiodes because the resistance of the forward-biased p-i-n InGaAs photodiode is relatively low compared to that of the ReRAM. Thus, analog neuromorphic computing can be directly performed in the 1P-1R crossbar array using the stored image data in the ReRAMs in the same way as the 1R-based crossbar arrays.<sup>10,16,26,30</sup> Therefore, image processing and encoding based on ANNs can be conducted within the sensor by directly implementing vector-matrix multiplication. This in-sensor vector-matrix multiplication enables an efficient higher-level computation without data transport between the sensor, memory, and processor, reducing significant amounts of energy consumption and processing time.<sup>6,10</sup>

The fabricated 1P-1R crossbar focal plane array fuses sensing, learning, and computing capabilities similar to those of biological retinas. To realize a neuromorphic vision system, we stored the vision information in each 1P-1R cell as a matrix geometry and simultaneously harnessed the data using emulated vision encoding. Previously demonstrated conventional crossbar geometries of neuromorphic in-memory computing systems for image processing are associated with pre-trained weight values in the ANN matrices, and input image data are applied to the crossbar column as a vectorized electrical signal (Fig. 4a)<sup>6,10,14</sup>. Because the format of image data is usually a 2-dimensional (2D)  $N \times N$  array, 2D-to-1D conversion (vectorization) must be applied as a vector input, which is an  $N^2 \times 1$  vector, to the column of the ReRAM crossbar array. In this case, extra complex circuit components (e.g., ADCs, DACs, and multiplexers) must be added to a peripheral circuit to control a large number of input signals, increasing energy consumption and operational complexity.<sup>1,6,10</sup> However, our in-pixel image processing system transposes image data to the weights of the ANN, in which the input image is applied and stored in the crossbar array in a weight vector matrix form, as shown in Fig. 4b. Therefore, the 2D-to-1D conversion of the image data is no longer necessary for this configuration, significantly reducing the circuit complexity and improving the operational efficiency. Moreover, data transportation from image memorization to the image encoding process is significantly diminished because the image information is directly processed in the pixels without any data transfer.

Figure 4c shows the in-pixel computing process using the fabricated 1P-1R array. The  $12 \times 12$  image of '8' is optically mapped onto the 1P-1R array (sensing) and preserved as the conductance of the ReRAMs (learning). Meanwhile, the front-ANN and post-ANN is pre-trained with 10,000 datasets of the MNIST handwritten numbers in the post-processor to extract the optimum weight vector.<sup>31</sup> The pre-trained 1D weight vector is then converted to electrical signals and applied to the 1P-1R array, enabling the physical matrix multiplication for the in-pixel ANN computation via Ohm's and Kirchhoff's law (computing). The output current signals from the voltage-conductance multiplication thus represent the encoded vector of the image '8,' achieved without data transportation by emulating the biological encoding capability. Figure 4d shows the 10 encoded vectors (vector size:  $1 \times 12$ ) for the input digits from '0' to '9,' each digit exhibiting a distinguishable encoded vector. The encoded vector is then fed to the next hidden layers to classify the image in the post-ANN (for more details, see Methods). Figures 4e and f show the classification results from the measured and memorized '4' and '8' digit images. Before training, the activation level of each digit is randomly distributed. However, the activation level of the ANN output

neurons of the ANN is concentrated on a single digit after training. A digit with the highest activation level was adopted as the classified 'answer'. Figures 4g and h show the results of the image classification before and after 100 training epochs for the full precision of 10,000 test digit images, indicating that the classification performance of the ANN was significantly improved after training the ANN. Although the proposed device has an  $N$  times smaller number of weight values compared to conventional in-memory computing methods (Fig. 4a), the classification accuracy is up to 82% with 100 training epochs (Fig. 4i). Further accuracy improvement could be achieved by employing a dual encoding neural layer in the ANN, conserving both row- and column-wise features of images, which can be realized by employing bi-directional peripheral circuitry.<sup>32</sup>

We demonstrated a neuromorphic machine vision system with an in-sensor encoding process inspired by mammalian vision. The focal plane array is based on an InGaAs photodiode directly integrated with HfO<sub>2</sub> ReRAM, constructing the 1P-1R optoelectronic memory and computing pixels. The optoelectronic and memory functionality of the fabricated 1P-1R pixel under light illumination showed reliable digital and multibit memory operation and endurance performance. Furthermore, a 16 × 16 1P-1R crossbar array with an InGaAs photodiode and HfO<sub>2</sub>-based ReRAM was used to perform edge computing of the handwritten numbers. Finally, we demonstrated biological image encoding with the developed 1P-1R crossbar array, utilizing direct image memorization and in-memory vector matrix multiplication. The encoded images were conveyed to the ANN for image classification, which revealed an accuracy of 82% with 100 training epochs. This slightly low classification accuracy is attributed to the structure of the encoding neural network, which consists of twelve 12 × 1 fully connected layers. The architecture of the neural network is inevitably determined by the hardware circuit structure of the 1P-1R crossbar array. The classification accuracy of our sensor system can be further improved using a dual-encoding neural layer in the ANN. The in-sensor computing concept introduced in this study is a novel method for storing and processing image information directly within pixels, and is seamlessly scalable with conventional semiconductor fabrication technology.

## Methods

**Device fabrication.** InGaAs p-i-n layers were grown on a InP substrate by general molecular beam epitaxy (MBE).<sup>33</sup> The 1P-1R crossbar array fabrication starts with a mesa etching of p-InGaAs/InGaAs layers. The mesa areas were protected with a bilayer photoresist (PR; LOR3A/AZ5214) by photolithography, and the unprotected InGaAs area was etched using inductively coupled plasma-reactive ion etching (ICP-RIE; BCl<sub>3</sub> 20 sccm, 600 W ICP power, 150 W forward power, 7 mTorr, 20°C stage temperature for 6 min), followed by wet etching for 1 min in a solution of H<sub>3</sub>PO<sub>4</sub>:H<sub>2</sub>O<sub>2</sub>:H<sub>2</sub>O = 3:1:25, which stopped at the n-InP layer. The PR mask was then removed in the Remover PG (Kayaku Advanced Materials) at 60°C. With a single PR (AZ5214) patterning, n-InP mesa for the bottom metal electrodes was defined, followed by a wet etch process for the n-InP/InP buffer layer using a solution of HCl:H<sub>3</sub>PO<sub>4</sub> = 3:1 (30 s). Next, a dielectric insulator layer of 150 nm Al<sub>2</sub>O<sub>3</sub> was deposited by plasma-enhanced atomic layer deposition (PE-ALD). The via holes were etched with a bilayer PR mask using ICP-RIE (BCl<sub>3</sub> 20 sccm, 50 W ICP power,

200 W forward power, 5 mTorr, 20°C stage temperature for 6 min). The top and bottom electrodes of the photodiodes were simultaneously deposited by photolithography with bilayer PR and e-beam evaporation of Ti/Pt/Au (5/10/50 nm), which was lifted off in the Remover PG at 60°C. Another dielectric insulator of 150 nm Al<sub>2</sub>O<sub>3</sub> was deposited by PE-ALD, and through holes were opened on the bottom electrodes of the photodiodes. Subsequently, the bottom electrodes of the ReRAMs, which are connected to the bottom electrodes of the photodiodes, were deposited using bilayer photolithography, e-beam evaporation of Ti/Pt (5/25 nm), and a lift-off process with the Remover PG. 5.5 nm of a HfO<sub>2</sub> layer was deposited by PE-ALD, followed by metal deposition of Ta/Pt (50/25 nm) on the top electrodes of the ReRAMs using DC magnetron sputtering of Ta (25 W RF power, 5 mTorr, Ar 20 sccm, room temperature for 18 min) and e-beam evaporation of Pt. Finally, the HfO<sub>2</sub> mesa areas were defined by dry etching with ICP-RIE with a bilayer PR mask.

1P-1R single device characterization. All electrical measurements were performed using a semiconductor analyzer (B1500A, Keysight). The devices were illuminated using a diode laser (DJ532, Thorlabs) with a wavelength of 532 nm, where the incident power was controlled by a neutral density filter.

Image memorization. MNIST handwritten digit images of '4' and '8' are printed out on photomasks using a direct write lithography tool (MicroWriter ML3, Durham Magneto Optics Ltd), and the diode laser light with a wavelength of 532 nm ( $P = 50 \text{ mJ/cm}^2$ ) is illuminated on the 1P-1R device by passing through the printed digit images. The projected image ('4' or '8') on the sensor is memorized by applying a +5 V voltage pulse to each pixel using a semiconductor analyzer. After the memorization process, -1 V pulse is applied to each pixel to readout the saved image. Then, the stored image is erased by applying -5 V pulse to each pixel.

In-pixel image encoding and classification. The encoder and classifier models were implemented using Python. We combined a matrix-to-vector encoder (12×12–12) and a fully connected layer classifier with two hidden layers (12-20-16-10) on the Modified National Institute of Standards and Technology (MNIST) dataset. Each original MNIST image was resized to 12×12 pixels, and we trained and tested 10,000 images (with 64 batch sizes for 100 epochs) and 100 MNIST images. For the backpropagation learning process, we employed an RMSprop optimizer, rectifier (softmax for the last output) nonlinearity activations, and an initial time-decaying learning rate (0.001).

## Declarations

### Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

### Supplementary information

The online version contains supplementary material available at <https://doi.org/xxxxx>

## Acknowledgements

D.L., M.P., Y.B., B.B., and K.L. were supported by the U.S. National Science Foundation (NSF) under the grant ECCS-1942868.

## Author contributions

K.L. and D.L. conceived the idea for the devices and experiments. D.L. fabricated the devices and performed the experiments. M.P. performed simulations of image encoding and classification. D.L., M.P., J.H., and K.L. wrote the manuscript with input from other authors. All the authors discussed the results and commented on the manuscript.

## Competing financial interests

The authors declare no competing financial interests.

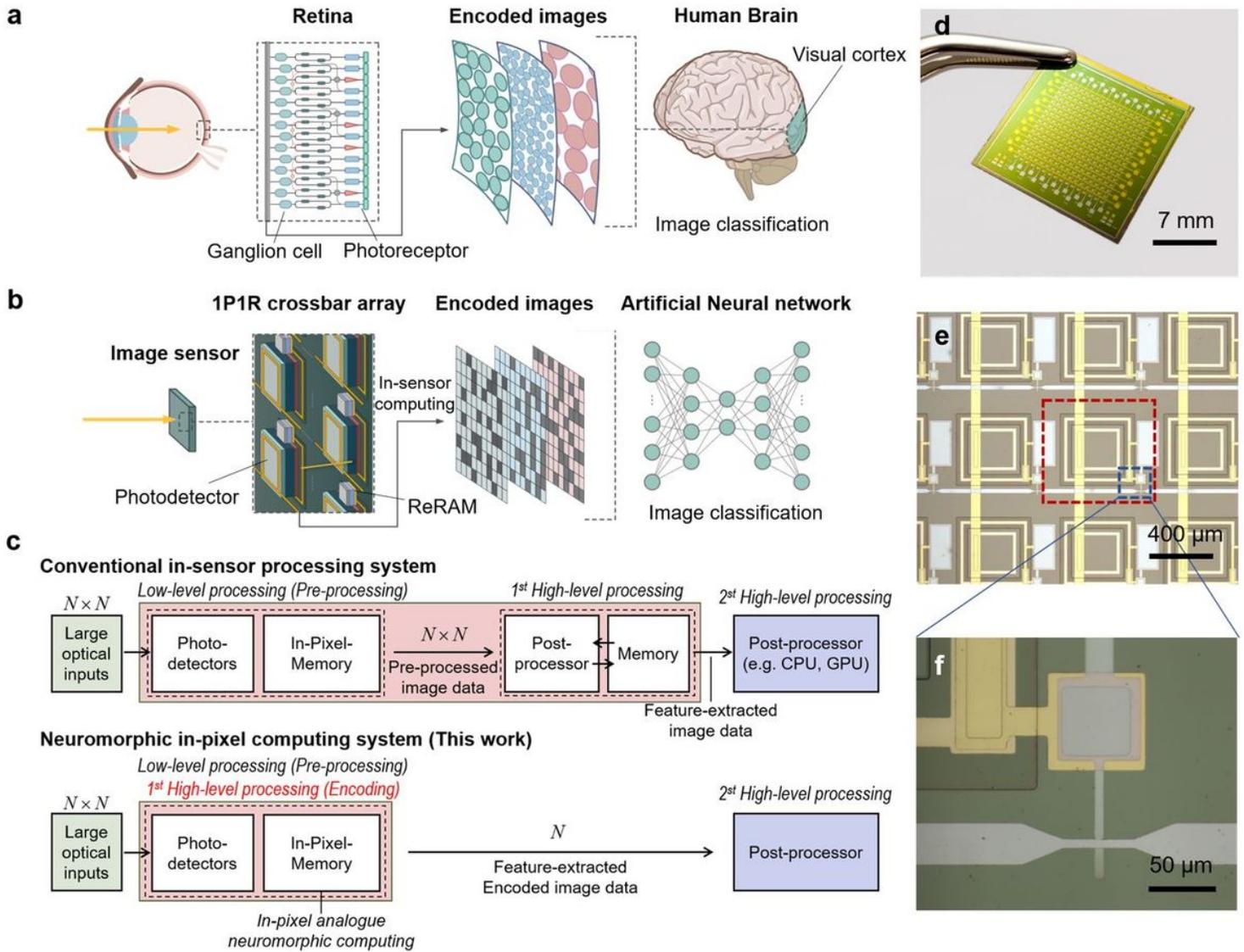
## References

1. Chai, Y. In-sensor computing for machine vision. *Nature* **579**, 32–33 (2020).
2. Jang, H., Liu, C., Hinton, H., Lee, M., Kim, H., Seol, M., Shin, H., Park, S. & Ham, D. An Atomically Thin Optoelectronic Machine Vision Processor. *Adv. Mater.* **32**, (2020).
3. Molina-mendoza, A. J. & Mueller, T. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–65 (2020).
4. Pérez, L., Rodríguez, Í., Rodríguez, N. & Usamentiaga, R. Robot Guidance Using Machine Vision Techniques in Industrial Environments: A Comparative Review. *sensors* **16**, 335 (2016).
5. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
6. Zhou, F. & Chai, Y. Near-sensor and in-sensor computing. *Nat. Electron.* **3**, 664–671 (2020).
7. Zhao, J., Liang, B. & Chen, Q. The key technology toward the self-driving car. *Int. J. Intell. Unmanned Syst.* **6**, 2–20 (2018).
8. Lins, R. G., Ricardo, P., Araujo, M. De & Corazzim, M. In-process machine vision monitoring of tool wear for Cyber-Physical Production Systems. *Robot. Comput. Integr. Manuf.* **61**, 101859 (2020).
9. Yang, Y. Multi-tier computing networks for intelligent IoT. *Nat. Electron.* **2**, 4–5 (2019).
10. Sebastian, A., Gallo, M. Le, Khaddam-aljameh, R. & Eleftheriou, E. Memory devices and applications for in-memory computing. *Nat. Nanotechnol.* **15**, 529–544 (2020).
11. Turner, M. H., Gonzalo, L., Giraldo, S., Schwartz, O. & Rieke, F. Stimulus- and goal-oriented frameworks for understanding natural vision. *Nat. Neurosci.* **22**, 15–24 (2019).
12. Liu, B., Hong, A., Rieke, F. & Manookin, M. B. Predictive encoding of motion begins in the primate retina. *Nat. Neurosci.* **24**, 1280–1291 (2021).
13. Masland, R. H. The Neuronal Organization of the Retina. *Neuron* **76**, 266–280 (2012).

14. Prezioso, M., Hoskins, B. D., Adam, G. C., Likharev, K. K. & Strukov, D. B. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).
15. Wang, Z., Joshi, S., Savel'ev, S., Song, W., Midya, R., Li, Y., Rao, M., Yan, P., Asapu, S., Zhuo, Y., Jiang, H., Lin, P., Li, C., Yoon, J. H., Upadhyay, N. K., Zhang, J., Hu, M., Strachan, J. P., Barnell, M., Wu, Q., Wu, H., Williams, R. S., Xia, Q. & Yang, J. J. Fully memristive neural networks for pattern classification with unsupervised learning. *Nat. Electron.* **1**, 137–145 (2018).
16. Yao, P., Wu, H., Gao, B., Tang, J., Zhang, Q. & Zhang, W. Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–647 (2020).
17. Zhu, Q., Li, B., Yang, D., Liu, C., Tian, Y., Su, X., Wang, X., Qiu, S., Cheng, H. & Sun, D. A flexible ultrasensitive optoelectronic sensor array for neuromorphic vision systems. *Nat. Commun.* **12**, 1–7 (2021).
18. Wang, H., Zhao, Q., Ni, Z., Li, Q., Liu, H. & Yang, Y. A Ferroelectric/Electrochemical Modulated Organic Synapse for Ultraflexible, Artificial Visual-Perception System. *Adv. Mater.* **30**, (2018).
19. Chen, S., Lou, Z., Chen, D. & Shen, G. An Artificial Flexible Visual Memory System Based on an UV-Motivated Memristor. *Adv. Mater.* **30**, (2018).
20. Gao, S., Liu, G., Yang, H., Hu, C., Chen, Q., Gong, G., Xue, W., Yi, X., Shang, J. & Li, R. An Oxide Schottky Junction Artificial Optoelectronic Synapse. *ACS Nano* **13**, 2634–2642 (2019).
21. Choi, C., Leem, J., Kim, M. S., Taqieddin, A., Cho, C., Cho, K. W., Lee, G. J., Seung, H., Bae, H. J., Song, Y. M., Hyeon, T., Aluru, N. R., Nam, S. & Kim, D. Curved neuromorphic image sensor array using a MoS<sub>2</sub>-organic heterostructure inspired by the human visual recognition system. *Nat. Commun.* **11**, 1–9 (2020).
22. Xue, J., Zhu, Z., Xu, X., Gu, Y., Wang, S., Xu, L., Zou, Y., Song, J., Zeng, H. & Chen, Q. Narrowband Perovskite Photodetector-Based Image Array for Potential Application in Artificial Vision. *Nano Lett.* **18**, (2018).
23. Zhou, F., Zhou, Z., Chen, J., Choy, T. H., Wang, J., Zhang, N., Lin, Z., Yu, S., Kang, J., Wong, H. P. & Chai, Y. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* **14**, 776–782 (2019).
24. Wang, T., Meng, J., Li, Q., He, Z., Zhu, H., Ji, L., Sun, Q., Chen, L. & Wei, D. Reconfigurable optoelectronic memristor for in-sensor computing applications. *Nano Energy* **89**, (2021).
25. Hong, S., Cho, H., Kang, B. H., Park, K., Akinwande, D., Kim, H. J. & Kim, S. Neuromorphic Active Pixel Image Sensor Array for Visual Memory. *ACS Nano* **15**, 15362–15370 (2021).
26. Li, C., Hu, M., Li, Y., Jiang, H., Ge, N., Montgomery, E., Zhang, J., Song, W., Dávila, N., Graves, C. E., Li, Z., Strachan, J. P., Lin, P., Wang, Z., Barnell, M., Wu, Q., Williams, R. S., Yang, J. J. & Xia, Q. Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **1**, 52–59 (2018).
27. Jiang, H., Han, L., Lin, P., Wang, Z., Jang, M. H., Wu, Q., Barnell, M., Yang, J. J., Xin, H. L. & Xia, Q. Sub-10 nm Ta Channel Responsible for Superior Performance of a HfO<sub>2</sub> Memristor. *Sci. Rep.* **6**, 1–8 (2016).

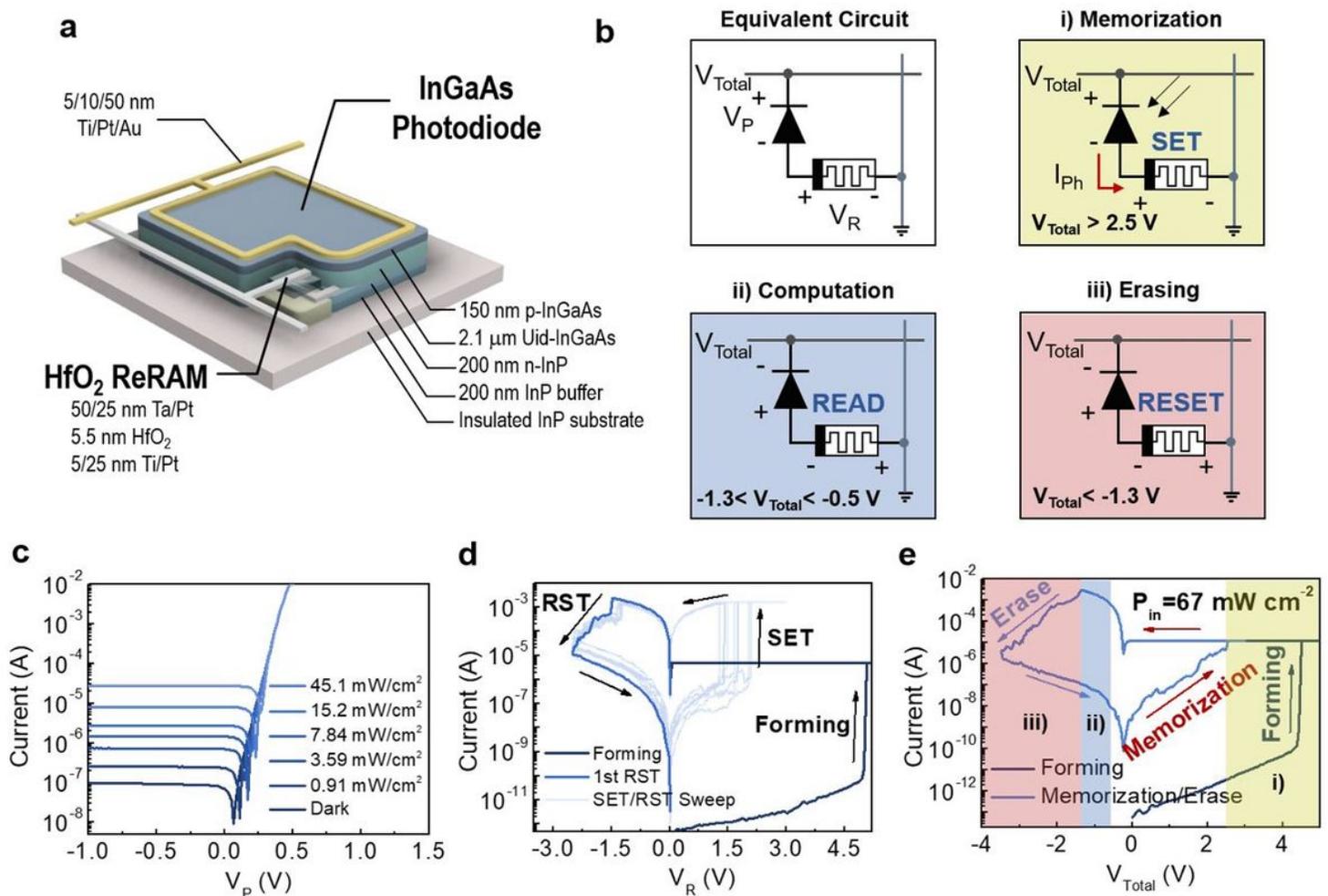
28. Miao, F., Yi, W., Goldfarb, I., Yang, J. J., Zhang, M. X., Pickett, M. D., Strachan, J. P., Medeiros-Ribeiro, G. & Williams, R. S. Continuous electrical tuning of the chemical composition of TaOx-based memristors. *ACS Nano* **6**, 2312–2318 (2012).
29. He, W., Sun, H., Zhou, Y., Lu, K., Xue, K. & Miao, X. Customized binary and multi-level HfO<sub>2</sub> – x-based memristors tuned by oxidation conditions.pdf. *Sci. Rep.* **7**, 1–9 (2017).
30. Lee, H. S., Baek, Y., Lin, Q., Chen, J. M., Park, M., Lee, D., Kim, S. & Lee, K. Efficient Defect Identification via Oxide Memristive Crossbar Array Based Morphological Image Processing. *Adv. Intell. Syst.* **3**, 2000202 (2021).
31. Deng, L. The MNIST Database of Handwritten Digit Images for Machine Learning Research. *IEEE Signal Process. Mag.* **29**, 141–142 (2012).
32. Cai, F., Correll, J., Hwan, L. S., Lim, Y., Bothra, V., Zhang, Z., Flynn, M. & Lu, W. A fully integrated reprogrammable memristor–CMOS system for efficient multiply–accumulate operations. *Nat. Electron.* **2**, 290–299 (2018).
33. Fan, D., Lee, K. & Forrest, S. R. Flexible Thin-Film InGaAs Photodiode Focal Plane Array. *ACS Photonics* **3**, 670–676 (2016).

## Figures



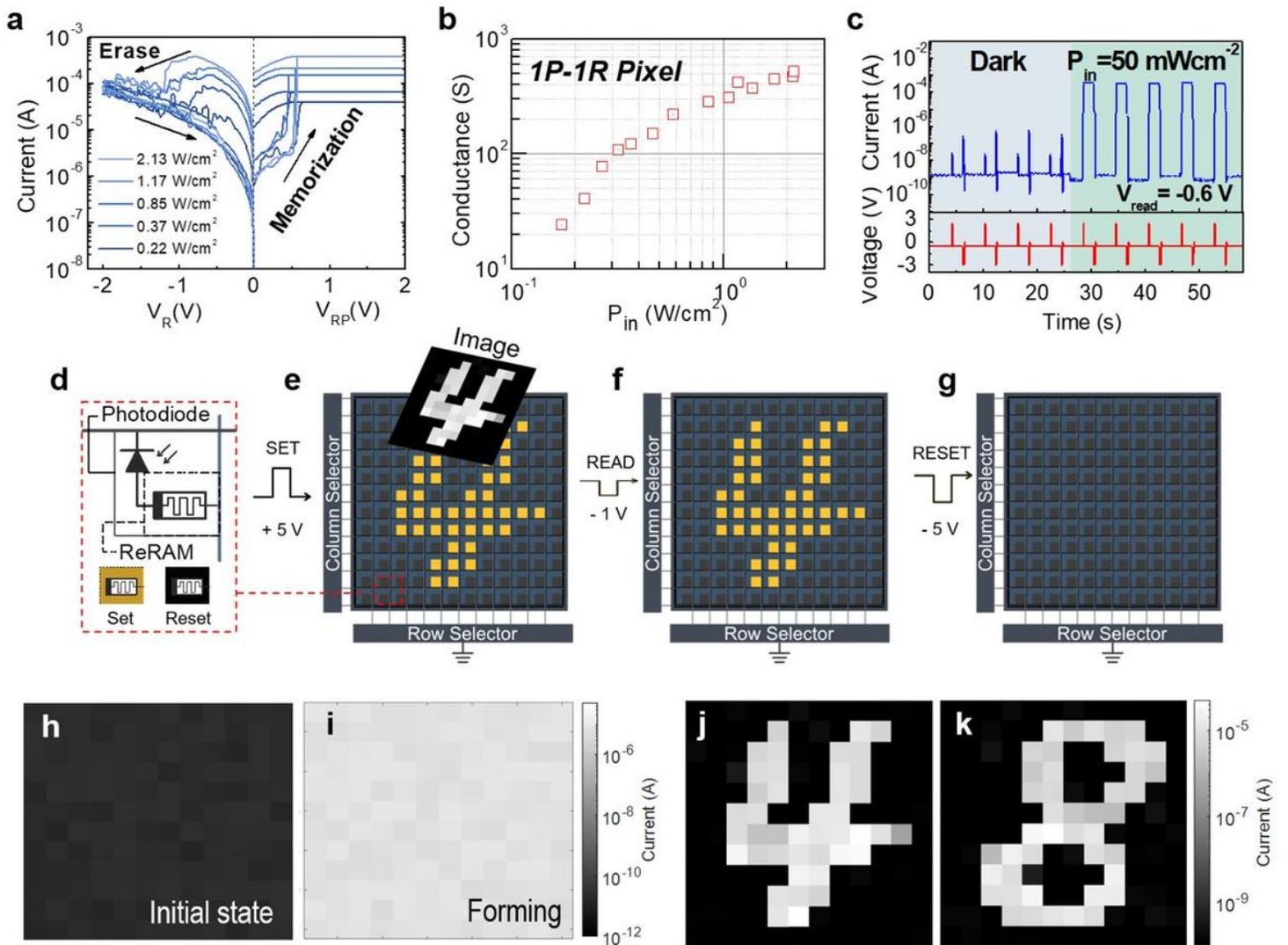
**Figure 1**

**In-sensor computing system with an in-pixel direct computing functionality emulating human vision system.** **a** Schematic illustration of human vision system from the retina to the brain, and its visual cognitive processing with in-retina image encoding process. **b** Schematic illustration of a process flow of the 1P-1R in-sensor computing system, which emulates the human vision system. **c** Block diagrams of image process flows with  $N \times N$  input images in conventional in-sensor processing systems and 1P-1R in-pixel computing system. **d** Optical image of the fabricated 1P-1R focal plane array. Scale bar: 7 mm. **e** A optical microscope image of the fabricated 1P-1R array. Dashed red (blue) box in shows a 1P-1R pixel (ReRAM) in the array. Scale bar: 400  $\mu\text{m}$ . **f** An enlarged optical microscope image of the dashed blue box area in (e). Scale bar: 50  $\mu\text{m}$ .



**Figure 2**

**Design and characterizations of single InGaAs 1P-1R integrated device.** **a** Schematic illustration of the fabricated InGaAs 1P-1R device. **b** Equivalent circuit diagram of the 1P-1R structure, and three key operations in the 1P-1R device depending on applied bias voltage: i) memorization, ii) computation, and iii) erasing operations. Each background color matches with the corresponding operation range colored in **(e)**. **c** I-V curves of the single InGaAs photodiode in the 1P-1R unit under dark and various light illuminations with a wavelength of 532 nm. **d** I-V characteristic of single ReRAM in the 1P-1R unit, sweeping the applied voltage along with the indicated loops in the graph. **e** I-V characteristics of the single 1P-1R integrated device under light illumination with an incident power density of 67 mW/cm<sup>2</sup> and wavelength of 532 nm. The colored regions indicate operation ranges for three main processes shown in **(b)**.



**Figure 3**

**Multi-state optoelectronic memory operation in the single 1P-1R device and characterization and image memorization process of the fabricated 1P-1R crossbar array.** **a** Multi-state memorization process in the single 1P-1R device under light illuminations with various intensities. Each memorization process is followed by an erase process. **b** Conductance of the ReRAM in the 1P-1R unit as a function of the incident light power density during the memorization process. **c** Sequential memorization process in the single 1P-1R device with voltage pulse inputs. Light with  $P_{in}=50$  mW/cm<sup>2</sup> is turned on at 26 s. Input voltage pulses of 2 V and -3 V were used for memorization and erase operation, respectively. Otherwise, a voltage of -0.6 V was applied to the device to readout the resistance state of the ReRAM over time. **d** Circuit diagram for a pixel in the 16×16 1P-1R crossbar array. The pixels are colored in yellow (blue) when the ReRAMs are in a SET (RESET) state in (**e-g**). **e-g** Schematic illustration of image memorization (**e**), readout (**f**), and erasing (**g**) processes in the 1P-1R crossbar array. **h,i** Read current maps before (**h**) and after (**i**) a forming process in the fabricated 1P-1R array. The read voltage is -1 V. **j,k** Readout image maps after image memorization of MNIST handwritten number '4' (**j**) and '8' (**k**). The read voltage is -1 V.

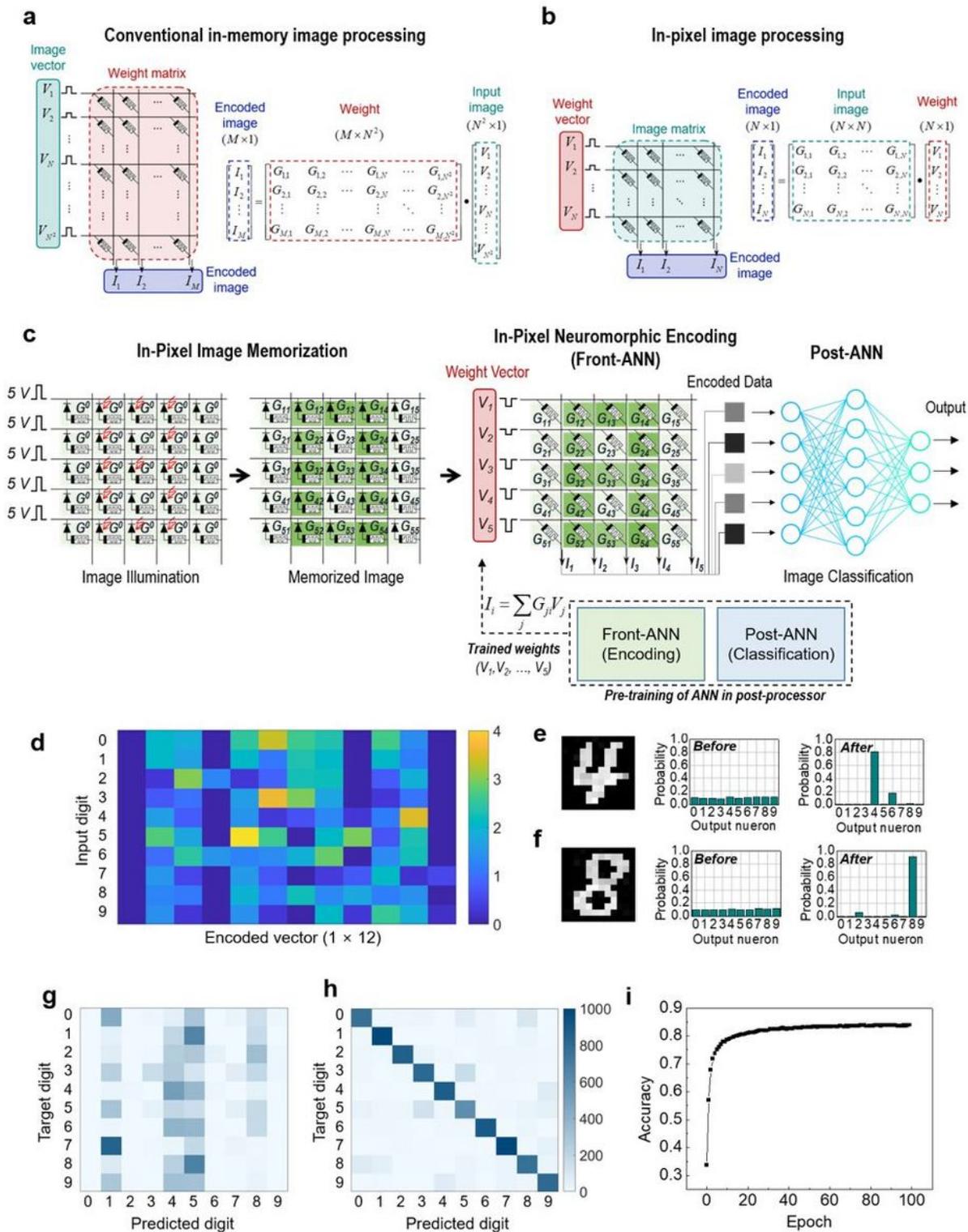


Figure 4

**Image memorization, encoding and classification via an in-pixel neuromorphic computing.** **a,b** Schematic illustration of multiply-accumulate operation for an image encoding process in conventional ReRAM crossbar array (**a**) and 1P-1R crossbar array (**b**). Corresponding matrix-vector multiplication is depicted with parameters of the input voltage ( $V_j$ ), conductance of ReRAM ( $G_{ij}$ ), and output current ( $I_j$ ). **c** Schematic illustration of an example of in-pixel image memorization, encoding, and classification process with a

5×5 1P-1R array. At the initial state, conductance of ReRAMs is  $G^0$ . After image memorization, conductance of each ReRAM is indicated as  $G_{ij}$ . Once an image is memorized in the sensor, pre-trained weight voltages ( $-1.3 \text{ V} < V_i < -0.5 \text{ V}$ ) are applied to the rows of the crossbar to perform a multiply–accumulate operation in the sensor for the encoding process. The encoded data are transferred to a post-ANN to classify the image. **d** Examples of encoded images of MNIST handwritten numbers from 0 to 9. **e,f** Classification results from the memorized ‘4’ (**e**) and ‘8’ (**f**) digit images before and after training the ANN. **g,h** Confusion matrices for a classification result from the 10,000 MNIST handwritten digit images before (**g**) and after (**h**) 100 training epochs. **i** Classification accuracy as a function of the number of training epochs. The classification accuracy is up to 82% with the 100 training epochs.

## Supplementary Files

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

- [Insensorcomputingsupplementaryinformation.pdf](#)