

# In-sensor compressing via programmable optoelectronic sensors based on van der Waals heterostructures for intelligent machine vision

Received: 14 October 2024

Accepted: 11 April 2025

Published online: 24 April 2025



Haoxin Huang<sup>1,13</sup>, Shuhui Shi<sup>2,3,13</sup>, Jiajia Zha<sup>1,13</sup>, Yunpeng Xia<sup>1</sup>, Huide Wang<sup>4</sup>, Peng Yang<sup>5</sup>, Long Zheng<sup>6</sup>, Songcen Xu<sup>7</sup>, Wei Wang<sup>8</sup>, Yi Ren<sup>9</sup>, Yongji Wang<sup>9</sup>, Ye Chen<sup>6</sup>, Hau Ping Chan<sup>1</sup>, Johnny C. Ho<sup>8</sup>, Yang Chai<sup>10</sup> ✉, Zhongrui Wang<sup>3</sup> ✉ & Chaoliang Tan<sup>1,11,12</sup> ✉

Efficiently capturing multidimensional signals containing spectral and temporal information is crucial for intelligent machine vision. Although in-sensor computing shows promise for efficient visual processing by reducing data transfer, its capability to compress temporal/spectral data is rarely reported. Here we demonstrate a programmable two-dimensional (2D) heterostructure-based optoelectronic sensor integrating sensing, memory, and computation for in-sensor data compression. Our 2D sensor captured and memorized/encoded optical signals, leading to in-device snapshot compression of dynamic videos and three-dimensional spectral data with a compression ratio of 8:1. The reconstruction quality, indicated by a peak signal-to-noise ratio value of 15.81 dB, is comparable to the 16.21 dB achieved through software. Meanwhile, the compressed action videos (in the form of 2D images) preserve all semantic information and can be accurately classified using in-sensor convolution without decompression, achieving accuracy on par with uncompressed videos (93.18% vs 83.43%). Our 2D optoelectronic sensors promote the development of efficient intelligent vision systems at the edge.

The rapid advancement of multidimensional visual information processing, encompassing spatial distribution, temporal evolution, phase and wavelength, amid the surge of Internet of Things (IoT) and artificial intelligence (AI), has led to a growing demand on high-

throughput data acquisition. This demand poses significant challenges to conventional signal processing, which is constrained by limited bandwidth<sup>1–3</sup>. Efficiently handling multidimensional data is crucial for edge AI to minimize energy consumption and system

<sup>1</sup>Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China. <sup>2</sup>Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong SAR, China. <sup>3</sup>School of Microelectronics, Southern University of Science and Technology, Shenzhen 518055, China. <sup>4</sup>College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China. <sup>5</sup>College of Integrated Circuits and Optoelectronic Chips, Shenzhen Technology University, Shenzhen 518118, China. <sup>6</sup>Department of Chemistry, The Chinese University of Hong Kong, Hong Kong SAR, China. <sup>7</sup>Department of Electronic & Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China. <sup>8</sup>Department of Materials Science and Engineering, City University of Hong Kong, Hong Kong SAR, China. <sup>9</sup>Department of Chemistry, City University of Hong Kong, Hong Kong SAR, China.

<sup>10</sup>Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong SAR, China. <sup>11</sup>Department of Biomedical Engineering, City University of Hong Kong, Hong Kong SAR, China. <sup>12</sup>Hong Kong Branch of National Precious Metals Material Engineering Research Center (NPMR), City University of Hong Kong, Kowloon, Hong Kong SAR 999077, China. <sup>13</sup>These authors contributed equally: Haoxin Huang, Shuhui Shi, Jiajia Zha. ✉ e-mail: [ychai@polyu.edu.hk](mailto:ychai@polyu.edu.hk); [wangzr@sustech.edu.cn](mailto:wangzr@sustech.edu.cn); [chaoltan@cityu.edu.hk](mailto:chaoltan@cityu.edu.hk)

complexity without compromising performance. However, directly recording multidimensional visual information (e.g., high-speed video and hyperspectral data) using conventional full-frame sensors requires complicated modules, large storage capacity and substantial transmission bandwidth<sup>4,5</sup>. Addressing these limitations to simultaneously capture comprehensive photon information has become a significant challenge in multidimensional high-definition (HD) imaging<sup>6,7</sup>.

With the evolution of compressive sensing (CS) theory<sup>8,9</sup>, the snapshot compressive imaging (SCI) technology has emerged as an effective solution to tackle the abovementioned dilemma<sup>10–13</sup>. In SCI hardware implementations, continuous optical data (video frames or spectral channels) can be encoded first via an optical modulator and then compressed/integrated directly into a single snapshot, which can be reconstructed into the original scene using advanced algorithms. In this case, SCI allows for capturing multidimensional data ( $\geq$  three-dimensional (3D)) via traditional two-dimensional (2D) image sensors without spatiotemporal resolution<sup>10,11</sup>. Unlike full-frame sensors that generate millions of pixel elements, SCI compresses multiple mapped optical information into the same pixel through multi-channel parallel acquisition, thereby enhancing photon utilization and sampling rates. This results in the development of efficient imaging system with low storage and bandwidth requirements, as well as high optical throughput, making SCI highly promising for processing temporal, spectral-temporal, and spatial-temporal data<sup>1,4,12</sup>. Although event-based vision sensors (EVS) can also mitigate the issues of high bandwidth and storage demands through their sparse event data capture<sup>14</sup>, their performance is inherently constrained by a decrease in compression efficiency under static conditions and the inability to generate interpretable images directly without relying on computationally intensive iterative reconstruction algorithms<sup>15,16</sup>. In contrast, SCI utilizes spatiotemporal optical encoding to enable non-iterative reconstruction of full-scene information. This approach not only matches EVS in terms of bandwidth efficiency in high-dynamic scenarios but also significantly enhances the effective information density in static scenes (Supplementary Table 1)<sup>1,12</sup>. The encoding process of SCI is the precondition for compressed imaging. Although various optical modulators have been developed, they are usually expensive and come with limitations. In video SCI, for example, the digital mirror device (DMD) encoder relies on mechanical deflection and that limit its footprint and applicability in diverse scenarios<sup>17</sup>. While liquid crystal encoder has better resolution, it suffers from pixel crosstalk and limited switching speed<sup>18</sup>. Moreover, the encoder and integrator of SCI are physically separated, that poses challenges for precise alignment<sup>19</sup>, and the sequential processing requires independent computing resources, leading to data inconsistency, transmission latency, and increased system complexity<sup>19–22</sup>. These issues present significant obstacles, especially as cutting-edge intelligent applications increasingly prioritize portability and efficiency.

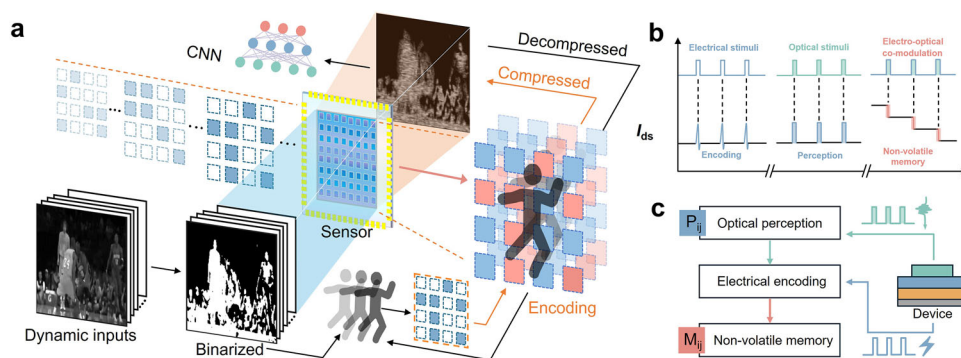
Given the rising prominence of visual sensors based on 2D van der Waals (vdW) materials and their effectiveness across a range of electronic and optoelectronic applications<sup>20,23–26</sup>, constructing functionalized 2D heterojunctions is expected to further advance SCI technology. The ultra-thin nature of 2D materials enables rapid response to optoelectronic modulation, making them potential candidates for use as encoders. Additionally, 2D heterojunctions can integrate sensing, storage and computing functions within a single device to eliminate data transfer between different modules, leading to reduced latency and power consumption<sup>20,27–33</sup>. Currently, reported 2D visual sensors have been utilized to reduce processing loads by removing redundant information while retaining feature information<sup>21,22,34,35</sup>. However, lossless compression of visual information has not been realized. Achieving this could be particularly advantageous in scenarios that demand high acquisition efficiency, reconstruction accuracy, and processing precision, such as industrial inspection, security monitoring, and autonomous driving<sup>1,5,12</sup>. Therefore, integrating sensing, snapshot compression, and computing into a single sensor is a promising approach for optimizing integration functionality. This strategy could enhance multidimensional data acquisition, simplify system complexity, and broaden application scenarios.

Here we report a 2D vdW programmable photoinduced memory sensor (PPMS) that offers simultaneous sensing and processing capabilities for highly compact SCI system, particularly the video and spectral SCI with 3D signals. Leveraging its rich optoelectronic dynamics, our developed PPMS can not only simulate the perception of dynamic motions, but also encode, compress, and store optical signals using its programmable non-volatile conductance under electro-optical co-modulation. This sensor can compress 3D grayscale videos and hyperspectral images into 2D encodings with a compression ratio of 8:1 (from original size to compressed size), then being reconstructed in high fidelity with distinguishable and reproducible characteristics compared to the original dataset. For video compression, the peak signal-to-noise ratio (PSNR) value achieved by the sensor (15.82 dB) is close to that achieved by software (16.21 dB). In addition, we demonstrate that the compressed data retains all semantic information, which can be accurately classified by a convolutional neural network (CNN). This work could deliver an inspiration for constructing compact and efficient optoelectronic sensors, promoting the development of edge intelligent vision systems in the future.

## Results

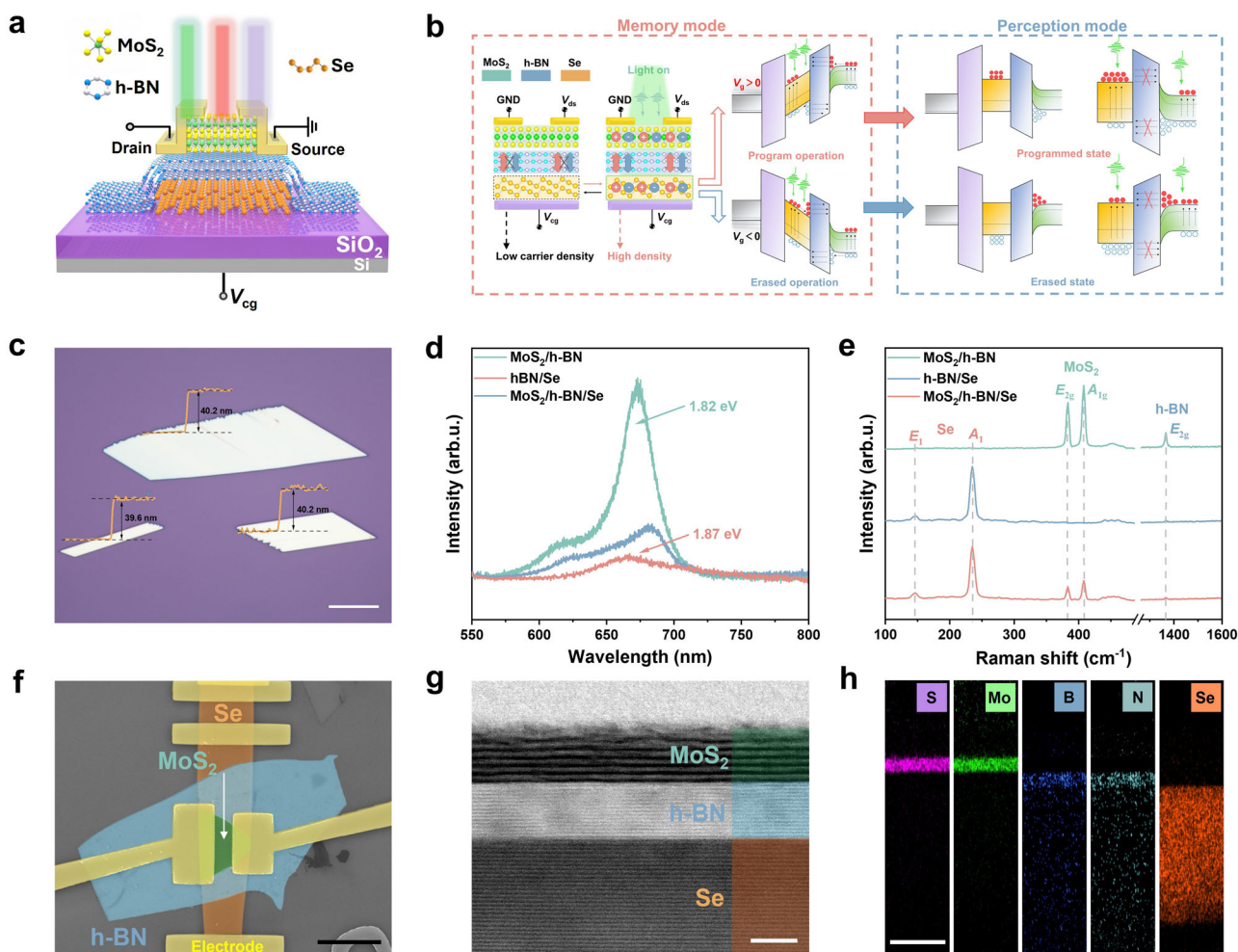
### In-sensor dynamic compressing and computing

Figure 1a illustrates the implementation of in-sensor compressing and downstream computing based on our 2D PPMS. The PPMS integrated incoming 2D frames modulated by electrical masks, effectively compressing 3D videos or hyperspectral data into 2D via SCI. The



**Fig. 1 | In-sensor compressing and computing.** **a** Schematic illustration of compressing dynamic binarized images by using PPMS (Programmable Photoinduced Memory Sensor) and for in-sensor computing, where dynamic Kobe motions

(reproduced with permission from ref. 50) are pre-binarized as the inputs. **b** The relationship between input pulses (both electrical and optical stimuli) and output responses. **c** The dynamic functions enabled by the sensor.



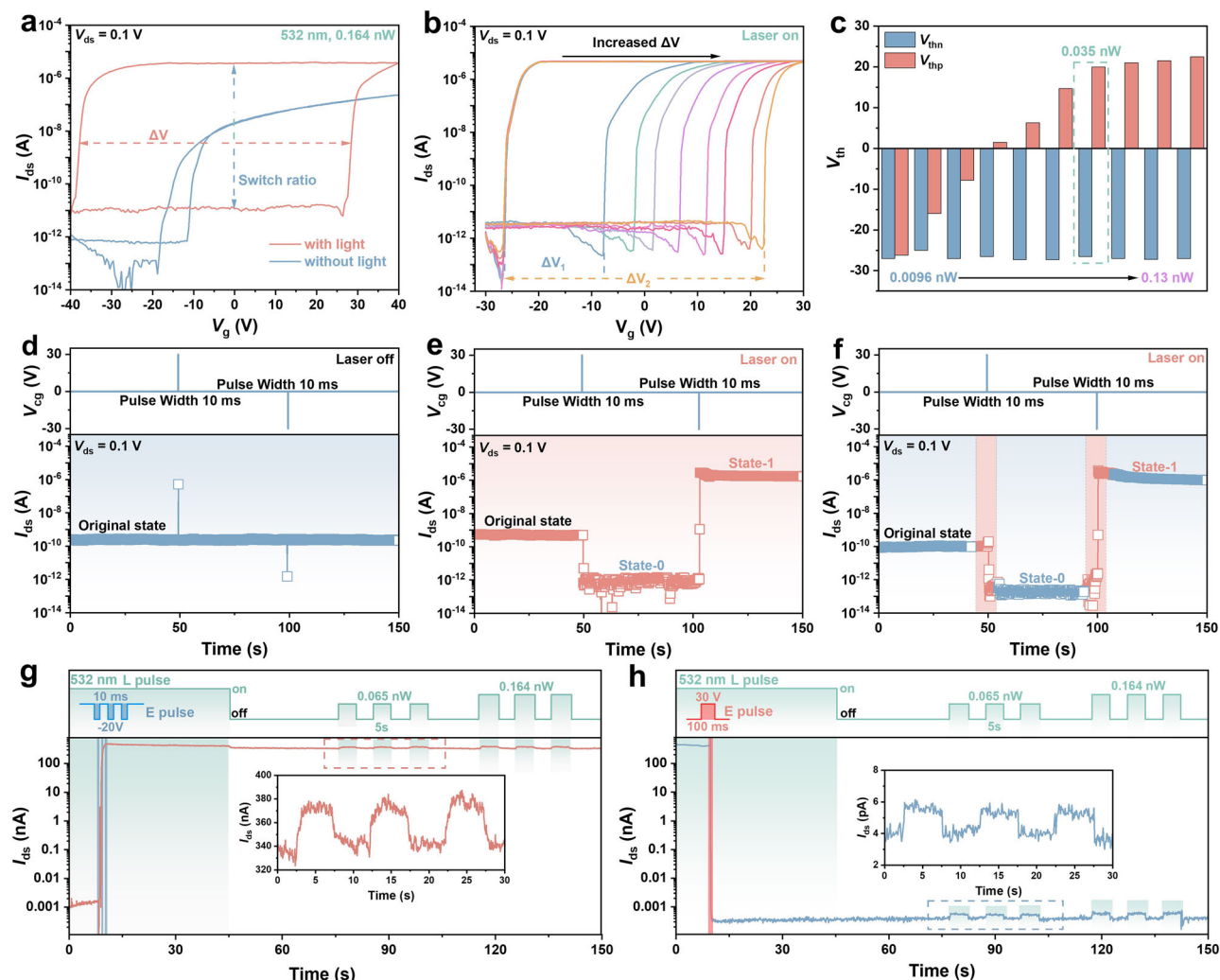
**Fig. 2 | Sensor structure and performance characteristics.** **a** Schematic illustration of the device architecture of PPMS employing three terminals: Drain (D), Source (S), and Control Gate (CG), where  $V_{cg}$  modulates the channel conductivity. **b** Schematic demonstration of memory mode and perception mode of PPMS. Red/blue spheres denote electrons ( $e^-$ )/holes ( $h^+$ ). Under illumination, positive/negative gate bias induces downward/upward band bending of MoS<sub>2</sub>, enabling carrier tunneling (indicated by black arrows) between the charge trapping layer (Se) and channel (MoS<sub>2</sub>) for memory mode. In the absence of gate bias, rapid electron-hole recombination dominates under optical excitation without carrier tunneling, defining the perception mode. **c** Optical image of Se nanosheets with AFM (Atomic Force Microscopy) measured thicknesses (40.2 nm and 39.6 nm) marked by orange

lines (scale bar: 10  $\mu$ m). **d** PL (Photoluminescence) spectra and **(e)** Raman spectra of Se/h-BN/MoS<sub>2</sub> heterostructure, where Raman peaks centered at  $\sim 237$  and  $140$   $\text{cm}^{-1}$  are attributed to the typical  $E_1/A_1$  and  $E_2$  modes of  $t$ -Se<sup>37</sup>, peaks located at 383 and  $408$   $\text{cm}^{-1}$  correspond to the in-plane ( $E_{2g}$ ) and out-of-plane ( $A_{1g}$ ) vibration mode of MoS<sub>2</sub><sup>60</sup>, and the isolated peak at  $1365$   $\text{cm}^{-1}$  is from the high-energy phonon ( $E_{2g}$ ) of h-BN (Hexagonal Boron Nitride)<sup>61</sup>. **f** False-color SEM (Scanning Electron Microscopy) image (scale bar: 10  $\mu$ m) **g** Cross-sectional TEM (Transmission Electron Microscopy) image (scale bar: 5 nm) and corresponding **(h)** Elemental EDS (Energy Dispersive Spectroscopy) mapping of the Se/h-BN/MoS<sub>2</sub> heterostructure (scale bar: 20 nm).

compressed pixel can either be reconstructed into original 3D signal or be classified using a CNN without the need of decompression. Notably, the use of a binarization strategy for input data is primarily aimed at simplifying the representation of complex real-world optical field scenarios. This approach will be discussed in further detail later. The realization of the in-sensor compression depends on the unique photoinduced memory characteristics of our PPMS. As depicted in Fig. 1b, either optical or electrical stimuli result in instant and volatile channel conductance change of the sensor, referring to the perception of input signals. These changes can be non-volatile and further integrated under electro-optical co-modulation, yielding non-volatile multi-level channel conductance. This feature endows our sensor with two working modes, which are labeled as  $P_{ij}$  and  $M_{ij}$ , respectively. As shown in Fig. 1c,  $P_{ij}$  represents the perception of optical signal by pixel( $i,j$ ), while  $M_{ij}$  represents the perception and memorization of the sensed optical signal under electrical encoding. These unique optoelectronic behaviors lay the foundation for dynamic encoding and compressing of 3D visual information.

The device architecture of our specifically designed PPMS for this purpose is illustrated in Fig. 2a, f, which consists of the Se/h-BN/MoS<sub>2</sub> vdW heterostructure. The single-crystalline trigonal selenium ( $t$ -Se) nanosheet is prepared via a modified physical vapor deposition (PVD) method with leaf-like morphology (Fig. 2c)<sup>36,37</sup>, and its crystal structure is corroborated by the high-resolution transmission electron microscopy (HRTEM), where Se atoms covalently bond with each other in molecular chains along the  $c$ -axis ([0001] direction) and these chains stack via vdW forces to constitute the Se nanosheet. (See Supplementary Figs. 1–3 and Note 1 for details). The optical bandgaps for MoS<sub>2</sub> and Se nanosheets are characterized to be 1.82 eV and 1.87 eV, respectively (Fig. 2d). Figure 2e shows the Raman spectra of the vdW heterostructure, in which the typical characteristic peaks match well with previous reports<sup>36,37</sup>. The high quality of the prepared vdW heterostructure is demonstrated by the cross-sectional transmission electron microscope (TEM) image (Fig. 2g), where the clear and sharp boundaries between the layered materials indicate the clean and flat interfaces. The composition is





**Fig. 3 | Working mechanism of the PPMS.** **a** Comparison of the transfer curves measured in the dark and under illumination. Red and blue dashed arrows indicate the threshold voltage shift ( $\Delta V$ ) and switch ratio of sensor during electrical characterization. **b** Transfer curves measured in the sensor under the illumination of different effective laser power (from 0.0175 nW to 0.125 nW). **c** Extracted negative and positive threshold voltages ( $V_{thn}$  and  $V_{thp}$ ) under different laser power. The  $\Delta V$  approaches saturation under an optical power of  $\sim 0.035$  nW, which is highlighted by the green dashed box. Current evolution triggered by positive and negative electrical pulses (**d**) without and (**e**) with illumination ( $P_{532\text{ nm}} = 0.175$  nW). **f** Current

evolution triggered by polarity-dependent electrical pulses assisted under short-term illumination ( $P_{532\text{ nm}} = 0.175$  nW). The orange-shaded region represents the electro-optical co-modulation region, where electrical and optical stimuli collaboratively modulate the sensor response. The dynamic optical responses of the sensor are characterized when it is pre-configured to the (**g**) low-resistance state (state 1) and (**h**) high-resistance state (state 0) with the assistance of illumination (indicated by green shaded areas). The magnified views in Figures **g** and **h** exhibit characteristic photodetection responses, demonstrating that the sensor operates in perception mode under optical excitation.

further evidenced by elemental energy-dispersive X-ray spectroscopy (EDS) mapping (Fig. 2h).

In stark contrast with traditional optoelectronic sensors with fixed charge storage density<sup>22,38–40</sup>, our 2D PPMS exhibits unique photo-induced memory behaviors, enabling two distinct operation modes as shown in Fig. 2b: (1) When the sensor is triggered by electrical pulses under optical illumination, it demonstrates photoinduced nonvolatile memory, and multi-level states can be achieved through consecutive electro-optical co-modulation. (2) When the sensor is exposed solely to optical or electrical stimuli, its conductance change remains volatile and is directly proportional to the stimuli intensity, owing to the limited carrier density in the Se layer. This configuration allows the sensor to detect stimuli but does not retain memory of them. Based on these features, the PPMS can be applied to perform dynamic encoding, allowing in-sensor compression of pixel values over time series for the realization of highly compact SCL.

### Demonstration of working principle

The two-mode operation principle is explained in detail by sweeping the control gate ( $V_{cg}$ ) of the sensor under illumination and dark conditions, respectively (see Supplementary Fig. 4 for the morphology characterization of a typical device). The large memory window recorded under illumination shown in Fig. 3a corroborates the photoinduced memory in our device, which can be ascribed to the increased charge density within Se layer (see Supplementary Fig. 5a, b for detailed characterization)<sup>37,41</sup>. The memory window's dependence on  $V_{cg}$  under illumination is further characterized in Supplementary Fig. 5c, where the linear dependence of  $\Delta V$  on maximum control gate voltage ( $V_{cg, \text{max}}$ ) demonstrates that the stored charge quantity can be effectively tuned by the control gate. It's worth mentioning that  $\Delta V$  reaches 75 V at  $V_{cg, \text{max}} = 40$  V, that equal to a stored charge density of  $5.62 \times 10^{12} \text{ cm}^{-2}$  and is comparable to conventional flash memory device<sup>33</sup> (Supplementary Note 2).

The optical power-dependent storage capacity in PPMS is further explored (Supplementary Fig. 5d–f). Notably, the threshold voltages obtained from positive and negative branches (i.e.,  $V_{\text{thp}}$  and  $V_{\text{thn}}$ ) quickly shifts to negative side under illumination even with an effective optical power as low as 0.0096 nW (i.e.,  $V_{\text{th-Light}} - V_{\text{th-Dark}} < 0$ , as shown in Supplementary Fig. 5d), indicating the activation of hole trapping within Se layer, while large optical power leads to large memory windows (Supplementary Fig. 5e, f). Figure 3b presents the variation of illumination-induced storage capacity, with  $V_{\text{thp}}$  moving from left to right with the increase of laser power (i.e.,  $V_{\text{thp-Light}} - V_{\text{thp-Dark}} > 0$ ), corresponding to the increase of tunneling charge stored in the charge trapping layer. The variation in both  $V_{\text{thn}}$  and  $V_{\text{thp}}$  is further summarized in Fig. 3c, with a transition effective power reaching  $\sim 0.035$  nW. The unique  $V_{\text{th}}$  shifting phenomenon can be attributed to the reduction of the electron tunneling barrier on the Se side caused by illumination ( $\phi_{\text{e-Se}}$ ) and the photosensitive feature of Se (Supplementary Note 3). In short, the finely adjustable charge storage capability grounds the visual input compression. Besides, the conductance of the sensor can be tuned to distinguished states stimulated by varied  $V_{\text{cg}}$  pulses (Supplementary Fig. 6a), proving its potential for analog storage. The unique optoelectronic behaviors in our sensor are further illustrated in  $I_{\text{ds}}-t$  test (Fig. 3d–f). Without illumination, the channel conductance remains unchanged after electrically programming (+30 V/10 ms)/erasing (−30 V/10 ms) operations, while the introduction of illumination induces the distinct and non-volatile OFF state (high-resistance state, denoted as State-0) and ON state (low-resistance state, denoted as State-1) with an extinction ratio over  $10^6$ . Notably, the retention states (state 0 or 1) can be retained after removing illumination (Fig. 3f), demonstrating the non-volatile memory mode. Decent photoinduced endurance and retention characteristics of our sensor are also verified (Supplementary Fig. 6b, c), where no observable performance degradation after >3000 periodic P/E (Programming: +30 V/10 ms; Erasing: −30 V/10 ms) operations and the extinction ratio of  $>10^5$  triggered by  $\pm 40$  V/10 ms pulses could be maintained for more than 3000 s under illumination. In addition, the switching speed of the sensor can reach the time scale of hundreds of microseconds under illumination (Supplementary Fig. 6d, e).

The unique photoinduced memory function is highly related to the photosensitive Se nanoflake. In the dark, Se holds negligible intrinsic carrier density, whereas it increases significantly under illumination (on/off photocurrent ratio  $>10^2$ , Supplementary Fig. 7c, d)<sup>37,41</sup>. The mechanism behind is further revealed by the Fowler–Nordheim (FN) tunneling model (See Supplementary Figs. 8–9 and the Supplementary Note 3 for details)<sup>42,43</sup>. In the dark, the calculated barrier height for electron tunneling ( $\phi_{\text{e-Se}}$ ) from Se to Au layer in Se/h-BN (11 nm)/Au heterostructure is  $\sim 2.40$  eV, while an extra lower barrier close to 0.51 eV appears when the device is under the illumination with a low effective power of  $P_{532\text{ nm}} = 0.042$  nW (Supplementary Fig. 8e). In this case, large amounts of electrons from valence band (VB) or the deep trapping states in the band gap could be excited to the CB with the absorption of incident photons<sup>38</sup>, resulting in lower potential barrier and higher tunneling probability, leading to the obvious photoinduced memory in the sensor.

Besides, the PPMS can also be configured to the perception mode. Figure 3g, h present the current evolution triggered by optical pulses when the device has been electrically tuned to low-/high-resistance states, where only instant photo-response rather than non-volatile positive or negative photoconductivity behaviors (PPC or NPC) appears<sup>44</sup>. In perception mode, PPMS can serve as a photodetector. It's worth pointing out that the sensor presents ultra-high sensitivity towards visible light in the perception mode. As shown in Supplementary Fig. 10a–c, distinct photocurrent values and responsivities can be obtained at erased state, and an admiring photocurrent of  $I_{\text{ph}} = 495$  nA and responsivity of  $R = 7583$  A/W are observed at state 1–6 under the illumination of a low power laser ( $P_{532\text{ nm}} = 0.065$  nW)

(Supplementary Fig. 10b, c). Se is a *p*-type semiconductor with longer hole lifetimes, which has also been confirmed in Supplementary Fig. 7a, b. The ultra-high photoresponsivity can be attributed to the extra electrostatic control of the channel induced by the hole trapping and the strengthened photogating effect at MoS<sub>2</sub>/h-BN surface<sup>45,46</sup>. Moreover, the increased hole density within trapping layer will further mitigate the Schottky barrier between the channel and source/drain electrodes, leading to a higher carrier extraction efficiency<sup>47</sup>.

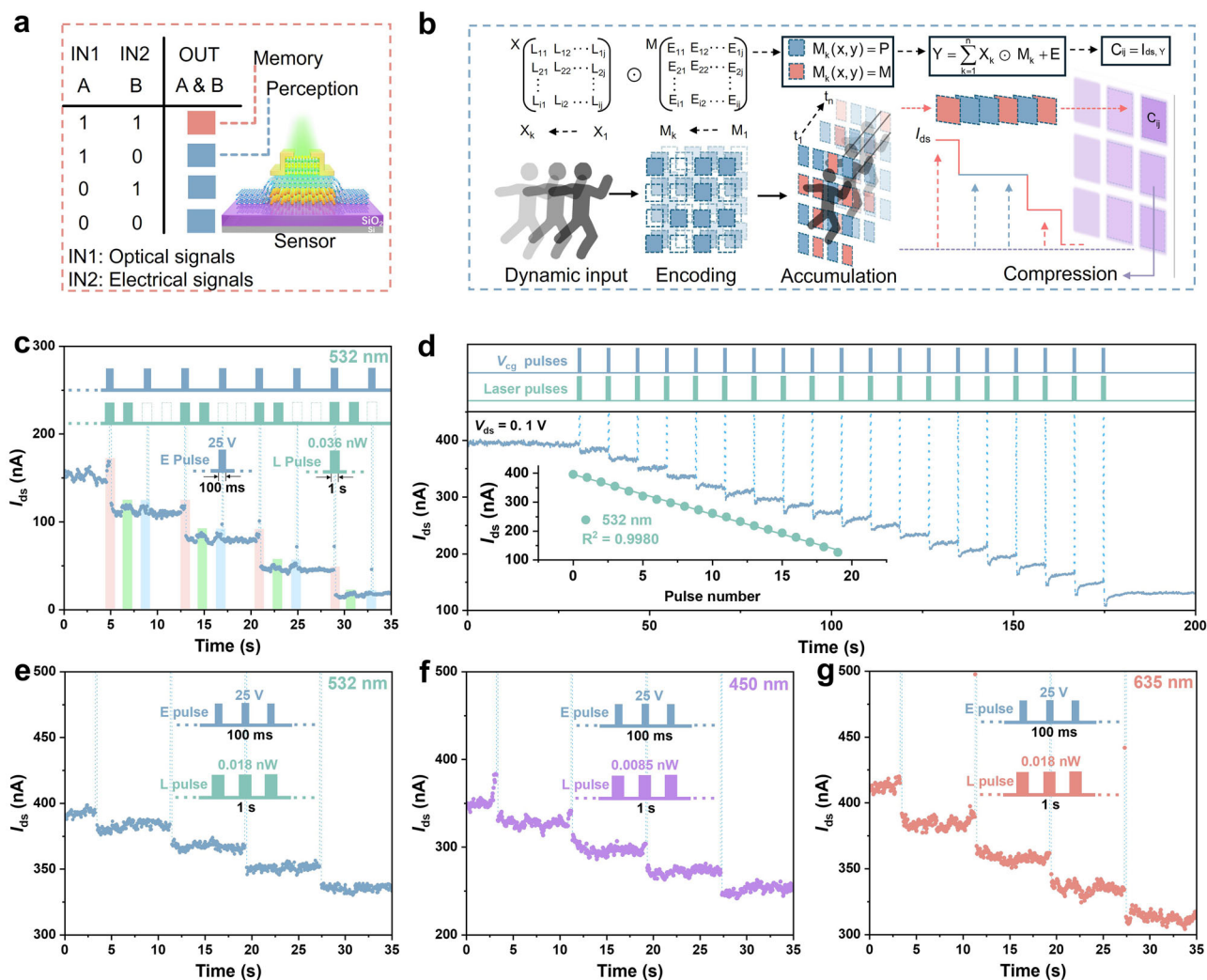
The mechanism behind the perception mode depends on the decreased contact potential difference ( $V_{\text{CPD}}$ ) between Se and MoS<sub>2</sub> under illumination<sup>48</sup>, as revealed in the Kelvin probe force microscopy (KPFM) characterization (Supplementary Fig. 11). Although the trapped carriers create energy band bending of MoS<sub>2</sub>, the tunneling probability of the photogenerated carriers across the tunneling barrier remains low, that inhibits the memory behaviors and leaves along the perception mode (The band diagrams of the sensor working under the two modes has also been depicted in Fig. 2b). The above results fully demonstrate the unique characteristics of our PPMS with reconfigurable perception and memory modes. Noted that a strong enough electric field can still provoke carrier tunneling in the dark, we map the extinction ratio change in the sensor with different P/E pulses in Supplementary Fig. 6f, where two modes can be well distinguished when the sensor works under  $|V_{\text{cg}}| \leq 30$  V and pulse width  $\leq 100$  ms, accordingly. The application of this sensor for subsequent SCI implementation is based on these experimental conditions.

### SCI implementation based on dynamic electro-optical modulation

In conventional SCI system, 3D data (e.g., dynamic visual signals) are first encoded through a temporal-variant mask frame-by-frame (e.g., DMD and spatial light modulator (SLM)) before being integrated by conventional image sensors (e.g., complementary metal oxide semiconductor (CMOS) and charge coupled device (CCD))<sup>4,12,19</sup>. In this way, the frames across temporal or spectral dimension can be compressed into a single snapshot measurement, which can then be used for lossless reconstruction or classified by machine learning models without decompression. However, the physical separation between DMD/SLM mask, CMOS/CCD sensors, and integration/memory unit results in low energy efficiency, large latency, and high system complexity. In our demonstration, the photoinduced memory behaviors enable the integration of perception, encoding, and compressing of dynamic scenes within a single device, minimizing power, latency, and hardware footprint.

The core function in our strategy depends on the optoelectronic AND gate (Fig. 4a), consisting of optical pulse input (IN1) and electrical pulse input (IN2), and the output corresponds to the change of the sensor conductance states. The dynamic compressing principle is designed as follows (Fig. 4a): Only when optical sensing (IN1) and pixel encoding (IN2) functions simultaneously can result in pixel value perception and integration, in which case, the conductance of the optoelectronic sensor experiences non-volatile changes based on the photoinduced memory effect, and such changes accumulate over time. Otherwise, the conductance state will remain unchanged. In this way, the 2D PPMS physically implements the SCI.

To prove this concept, we carried out the dynamic modulation (including sole optical/electrical and electro-optical co-stimulus) within the sensor that had been pre-tuned to low-resistance state. Figure 4c shows the conductance states will not be changed permanently when being triggered by sole electrical pulse trains (E: 25 V/100 ms, 0.25 Hz) or optical pulse trains (L: 532 nm, 0.036 nW/1 s, 0.5 Hz), while the synergic electro-optical pulses can effectively tune the sensor conductance state, manifesting the feasibility of the reconfiguration between photo-detecting and encoding-and-integration. More details and evidence supporting the integrated functions are provided in Supplementary Fig. 12 and Supplementary



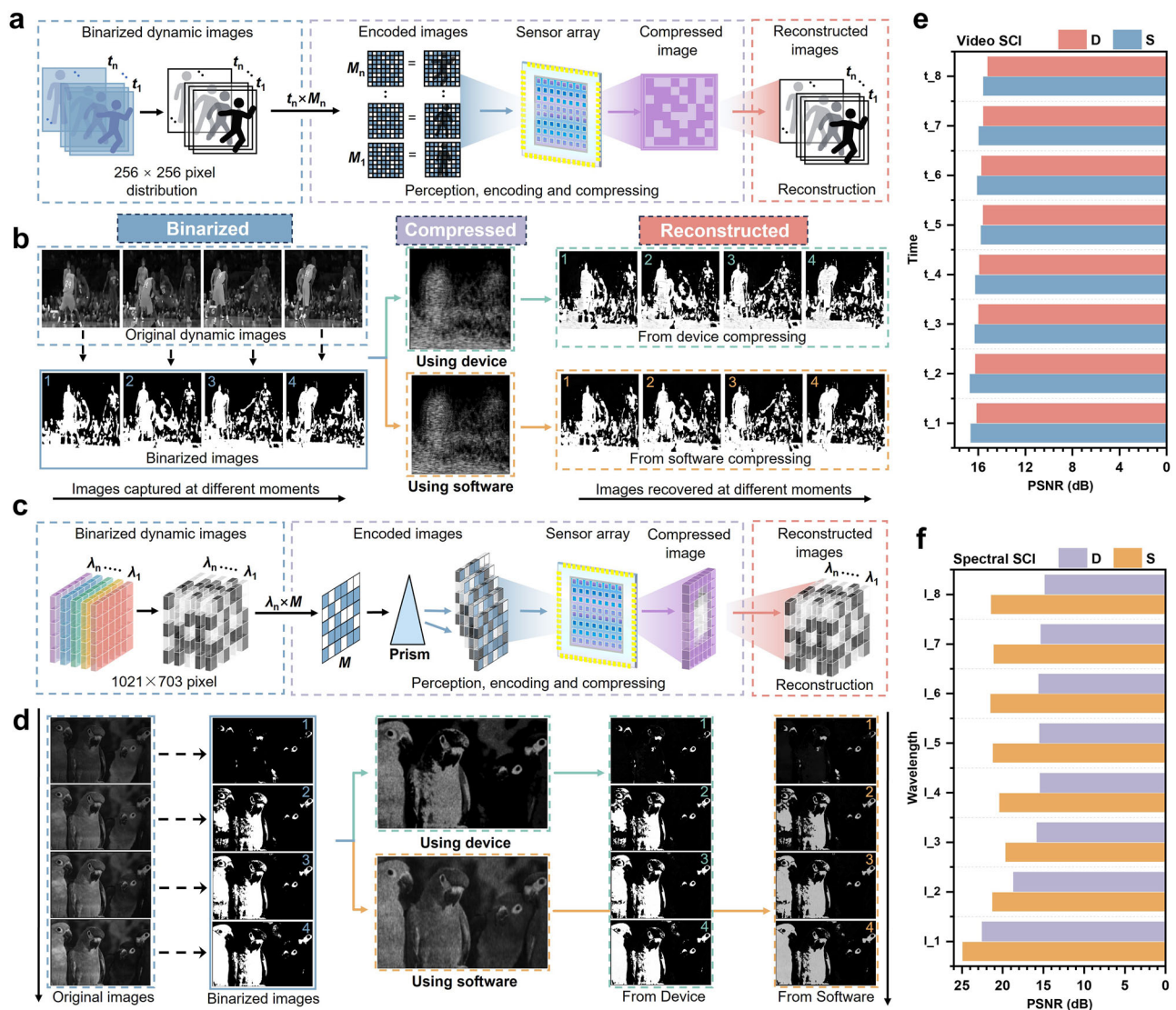
**Fig. 4 | Working principle of snapshot compressive imaging (SCI) enabled by PPMS. a** Demonstration of the optoelectronic AND logic gate for sensing and compressing. **b** An algorithm flow chart for encoding and compressing dynamic images by using the sensor array.  $L_{ij}$  and  $E_{ij}$  denote the optical signal and electrical signal applied to the sensor, respectively.  $X_k$  and  $M_k$  correspond to the pixel array of  $k$ th frame ( $t_k$ ) and  $k$ th mask, respectively.  $C_{ij}$  corresponds to the compressed value at pixel  $(i, j)$ , mapped to the drain-source current of the channel after 2D measurement with the final output matrix denoted as  $Y(I_{ds}, Y)$ . **c** Current evolution in the

sensor under different stimulation modes: Electrical-only (blue bar), optical-only (green bar), and electro-optical co-stimulus (red bar). **d** The sensor can be finely programmed with electro-optical co-stimulus, showing linear decrease with 19 levels. Inset: linear fit between the channel current in each conductance state and the pulse number. The enlarged electro-optical response of the sensor under different lasers: **(e)** 532 nm (0.018 nW/1 s), **(f)** 450 nm (0.0085 nW/1 s), and **(g)** 635 nm (0.018 nW/1 s). The blue dashed lines indicate the transient current variation triggered by electrical pulses.

**Note 4.** Notably, the multiple conductance states modulated by electrical pulses under illumination (Supplementary Fig. 12d) present high linearity and that is crucial for encoding and compressing process. To further verify the possibility of in-sensor multi-frame compression, we shortened the width of the optical pulses to close to real-world applications. As shown in Fig. 4d and e, multi-state conductance decrements over 19 current stages with remarkable uniformity can be observed (E: 25 V/100 ms, L: 0.018 nW/1 s, 0.125 Hz). According to the enlarged figure in Fig. 4d, the non-volatile channel current variation with pulse number (N) can be fitted into a linear relationship with a decent  $R^2 = 0.9980$ . In this case, for the same optical intensity appearing in different frames, the uniform change in conductance guarantees consistency in transmitting pixel values (similar to the linear relationship between induced charge and light intensity in CCD)<sup>49</sup>. This ensures high-quality implementation of the SCI operation, providing a foundation for fidelity in subsequent decompression. Additionally, we demonstrated synergic electro-optical modulation of the

conductance state of the sensor under varying optical pulse intensities, while maintaining a constant electrical pulse intensity (25 V/100 ms). As shown in Supplementary Fig. 13a–c, the absolute value of the average current change ( $\Delta I_{ds, Avg}$ ) incrementally increases with the rise in optical pulse intensity (i.e., 0.014 nW/1 s, 0.030 nW/1 s, and 0.045 nW/1 s). This increment confirms the controllable modulation of tunneling charges between Se and MoS<sub>2</sub>, as the variation in channel current relative to pulse numbers maintains robust linearity (Supplementary Fig. 13d–f). We then plotted the relationship between the average current change and the effective power (Supplementary Fig. 14), where the strong linear correlation between  $\Delta I_{ds, Avg}$  and effective power ( $R^2 = 0.9868$ ) demonstrates our device's capability to resolve continuous analog signals with high fidelity. This suggests its potential for precise analog signal compression, where input power levels can be directly mapped to quantized current states. The charge retention capability of PPMS during the dynamic encoding process is also investigated (Supplementary Fig. 15), revealing the distinct non-volatile memory





**Fig. 5 | Video and spectral SCI enabled by PPMS. a** Schematic illustration of video SCI process. **b** Comparison between the compression and the PnP-enabled reconstruction of dynamic frames from pre-binarized Kobe motion scenes (reproduced with permission from ref. 50) by the device and software, respectively. **c** Schematic illustration of spectral SCI process. **d** Comparison between the

compression and PnP-enabled reconstruction of spectral frames from pre-binarized Bird scenes (reproduced with permission from ref. 54) by the device and software, respectively. Histogram summarizing the peak signal-to-noise ratio (PSNR) values in (e) video and (f) spectral SCI based on device (D) and software (S), respectively.

behavior (retention time > 500 s for each electro-optical co-modulation) with varied optical intensities. This prevents the loss of inter-frame information, laying the foundation for high-quality image decompression.

Besides green light (532 nm), our PPMS can also respond to blue (450 nm, 0.0085 nW/1 s) and red light (635 nm, 0.018 nW/1 s) in the same way. The results are provided in Fig. 4f, g and Supplementary Figs. 16 and 17, which also show decent electro-optical co-modulation with the fitted  $R^2$  equals to 0.9905 and 0.9979 for 450-nm and 635-nm lasers, respectively. The broadband spectrum co-modulation with our sensor is the basis for 3D hyperspectral data compression using SCI. We also supplemented the comparison of PPMS and typical SCI encoders in terms of structure and performance. As shown in Supplementary Table 2 and Supplementary Fig. 18, the frame rate and compression ratio of PPMS can reach 1000 fps and 40:1, respectively, matching the reported encoding strategies<sup>50–52</sup>. Although PPMS shows a limited dynamic range, its potential can be further explored by improving measurement accuracy. For practical applications, the noise level of the device should be improved to achieve high-fidelity

pixel value transmission and compression (the linearity of  $R^2$  value closer to 1).

In addition, a  $3 \times 3$  sensor array is fabricated (See Supplementary Fig. 19a–e and “Methods”). All the nine devices manifest distinct photoinduced memory behaviors with considerable uniformity in  $\Delta V$  (Supplementary Fig. 20). The device-to-device variation in the on/off current ratio ( $10^3$ – $10^5$ , as shown in Supplementary Fig. 19f, g) arises from the inevitable damage introduced by the transfer process and the variation remains within an acceptable range. The successful array fabrication provides a reference for future device scalability.

### Video and spectral SCI based on the optoelectronic sensor

The schematic processes for video and spectral SCI using our sensor are depicted in Fig. 5a and c, respectively. The flowchart outlining the compression and decompression process is shown in Supplementary Fig. 21. Additionally, the detailed steps for the compression and reconstruction of video and spectral data are illustrated in Supplementary Figs. 22 and 23, respectively. A unified

SCI algorithm using element-wise multiplication can be mathematically written as<sup>1</sup>:

$$Y = \sum_{k=1}^n X_k \odot M_k + E \quad (1)$$

Where  $X$  denotes the 3D data cube (e.g., video or hyperspectral data) and is moderated by random 3D mask  $M$  (e.g., electric co-modulation signal).  $k = 1, \dots, n$  signifies the  $k$ th frame ( $t$ ) or spectral band ( $\lambda$ ) of the data cube,  $M_k$  corresponds to  $k$ th mask, and  $E$  is the measurement noise.  $Y$  is the 2D measurement realized by our sensor, or the compression result. Different from traditional SCI systems where the compressed pixel value is derived from the induced charges ( $Q_{\text{Induced}}$ ) within the sensor (e.g., CCD) through photoelectric conversion and analog-to-digital conversion, the pixel value in our PPMS system is derived from the tunneling charge ( $Q_{\text{Tunnel}}$ ) via electro-optical co-modulation (see Supplementary Note 5 for details) and can be fitted according to FNT model<sup>42,43</sup>:

$$I_{\text{FNT}}(V) = \frac{A_{\text{eff}} q^3 m (\alpha V_{\text{cg}})^2}{8\pi h \phi_B d^2 m^*} \exp\left(\frac{-8\pi \sqrt{2m^*} \phi_B^{\frac{3}{2}} d}{3hqV_{\text{ds}}}\right) \quad (2)$$

$$Q_{\text{Tunnel}}(t) = \int_0^T (I_{\text{FNT}}) dt = \frac{\Delta V_T \times C_{\text{CG-FG}}}{q} \quad (3)$$

Where  $V_{\text{cg}}$  and  $T$  denote the intensity and width of the electrical pulse applied to the control gate or back gate, respectively. The proportionality constant  $\alpha$  is used to describe the voltage drop across the floating gate (FG) or charge trapping layer. The total tunneling charge is derived using the integral formula Eq. (3) and is converted into a change in the threshold voltage ( $\Delta V_T$ ). Note that the tunneling current ( $I_{\text{FNT}}$ ) induced by the control gate of PPMS within the working range is negligible in the absence of optical illumination. Therefore, we can precisely control the quantity of charges tunneling each time when  $T$  is held constant:

$$Q_{\text{Tunnel}}(t) = 0 (\text{sole electrical modulation}) \quad (4)$$

$$Q_{\text{Tunnel}}(t) = \text{Constant} (\text{specific electro - optical co - modulation}) \quad (5)$$

For a metal-oxide-semiconductor (MOS) transistor, the  $I$ - $V$  characteristic in the linear region can be described as follows (low  $V_{\text{ds}}$ )<sup>53</sup>:

$$I_{\text{ds}} = \mu C_{\text{ox}} \frac{W}{L} \left( (V_{\text{GS}} - V_T) V_{\text{ds}} - \frac{V_{\text{ds}}^2}{2} \right) \approx \mu C_{\text{ox}} \frac{W}{L} (V_{\text{GS}} - V_T) V_{\text{ds}} \quad (6)$$

Where  $\mu$  is the carrier mobility,  $C_{\text{ox}}$  is the oxide capacitance per unit area,  $W$  is the channel width,  $L$  is the channel length, and  $V_{\text{GS}}$  is the gate-source voltage. Therefore, the  $V_{\text{GS}}$  is to be modified to effective gate voltage ( $V_{\text{GS, eff}}$ ), which is expressed as:

$$V_{\text{GS, eff}} = V_{\text{GS}} - \Delta V_T \quad (7)$$

Then Eq. (6) can be rewritten as:

$$I_{\text{ds}} = \mu C_{\text{ox}} \frac{W}{L} (V_{\text{GS, eff}} - V_T) V_{\text{ds}} \quad (8)$$

Therefore, PPMS can achieve linear regulation of  $I_{\text{ds}}$  through precise adjustment of charges within Se layer via electro-optical co-modulation, which aligns with our experimental observations. The threshold voltage variation can be transformed into grayscale values

through analog-to-digital converter (ADC) conversion. We have also demonstrated the PPMS's ability to sense different light intensities that enable the compression of grayscale information maps (Supplementary Figs. 13 and 14). To simplify the computational load, we here only consider the compression scenario under sole light intensity and pre-binarized the input image. Specifically, the pixel values of the original image, which range from 0 to 255, are first normalized to a range of 0 to 1. A threshold value of 0.2 is then selected for binarization.

Figures 4b and 5a illustrate the in-sensor video SCI which integrates sensing, encoding, and integration/compression of the binarized video, and the masks used are shown in Supplementary Fig. 24. During the encoding process, when both the pixel values ( $X_k(x, y, t)$ ) of the binary high-speed scene and the corresponding mask ( $M_k(x, y, t)$ ) are "1", then the electrical (mask) and optical (data cube) co-modulate the 2D sensor conductance. The encoded pixels, in the form of non-volatile conductance change, will accumulate over time (from frame  $t_1$  to  $t_n$ ). The resulting final conductance represents the integration/compression pixel value ( $C$ ), is proportional to the number of simultaneous optical and electrical pulses. Referring to Fig. 4d, the "fingerprint" current value associated with the number of synergetic electrical and optical pulses can be read and converted into the pixel value of the compressed 2D image ( $C = I_{\text{ds}}, \gamma$ ).

Based on the optoelectronic AND logic gate enabled by our PPMS, we have extended the functionality of a single device across a sensor array (Supplementary Fig. 27a) to facilitate simulated compression processing of dynamic signals across multiple pixels. To evaluate the device performance, we first reconstruct the 2D images from the device compression to produce hardware-compressed reconstructed images. We simultaneously apply the same reconstruction algorithm to the 2D images after software compression to generate software-compressed reconstructed images. Additionally, we binarize multiple frames from the original video, using these binarized images as a reference for comparison. We then calculate the PSNR values for each frame of the reconstructed images from both device and software, comparing them against the corresponding binarized original images (i.e., PSNR(DE\_BI) and PSNR(SW\_BI)) to evaluate the compression effectiveness. Figure 5b shows the compressed videos (below the purple box) and reconstructed 3D video (under the red box). We use a set of eight video frames about Kobe dribbling, each with a pixel resolution of  $256 \times 256$ , for SCI simulation. The left panel of Fig. 5b presents 4 selected frames before and after binarization (detailed enlarged images are shown in Supplementary Fig. 25), with the post-binarization frames still clearly depicting Kobe's shooting dynamics. These frames are then encoded and compressed into a single snapshot 2D measurement by our simulated optoelectronic sensor, as detailed in Supplementary Fig. 22, aligning with the software compression illustrated in the middle panel of Fig. 5b (see Supplementary Fig. 26 for further details). The compressed 2D snapshots are then used to reconstruct the original videos through Plug-and-Play (PnP) algorithm<sup>50,54</sup> (see Supplementary Note 5 for reconstruction process). As shown in the right panel of Fig. 5b, the denoised and reconstructed video frames are highly consistent with the binarized frames with preserved details. Quantitatively, our in-sensor SCI shows a decent average PSNR(DE\_BI) value of 15.81 dB with respect to the PSNR(SW\_BI) of 16.21 dB (Fig. 5e), which demonstrates that our in-sensor SCI can parallel the software SCI performance.

As the 2D PPMS features broadband spectrum response, it could compress 3D hyperspectral data using SCI. The schematic illustration of spectral SCI is shown in Fig. 5c. The spectral scenes ( $X_k(x, y, \lambda)$ ) are collected through the objective lens, followed by the spatial coding and compressing ( $X_k \odot M_k$ ) of the binarized data cube. Similar to conventional single-disperser coded aperture compressive spectral imager (CASSI) system<sup>1,55</sup>, the coded scene here first undergoes spectral dispersion through a dispersive element, such as a prism or a grating (see Supplementary Fig. 23 for



detailed process). Different from the video compression, the 2D PPMS accumulates the encoded frames across different wavelengths, where the current value corresponding to the number of pulses can then be read referring to the linearly changed conductance shown in Fig. 4d.

Figure 5d shows the simulated compression and reconstruction of hyperspectral data, employing the Bird data comprising 24 spectral bands with each sized at  $1021 \times 703$  pixels as the inputs<sup>54</sup>. For quantitative analysis, image data (400–500 nm) with ten continuous wavelengths (step 10 nm) were selected and binarized, which is shown in the left panel of Fig. 5d. Then the spectral data is encoded and compressed using both the PPMS and software. The reconstruction images by PnP algorithm<sup>50,54</sup> is illustrated by the middle (compress) and right (reconstruction) panel of Fig. 5d, and the spectral wavelength interval of 10 nm remains consistent before and after compression. As shown in Fig. 5f and Supplementary Table 3, the average PSNR for images reconstructed using software compression (21.49 dB) is higher than that for images reconstructed using device compression (16.73 dB). This indicates that software compression achieves greater similarity to the original image compared to device compression. It is also noteworthy that the average Structural Similarity Index (SSIM) for images reconstructed from device enabled compressed image (0.69) is close to the average SSIM obtained from software-enabled compressed image (0.79). This suggests that, from a human visual perspective, the quality of reconstruction is comparable between the two methods. Thus, the successful application of video and spectral SCI fully demonstrates the great potential of the 2D optoelectronic sensor in intelligent imaging systems. While it should be noted that under practical situations, SCI is much more complex and requires a proper balance between the decompression algorithm and compression process to achieve high reconstruction quality.

### CNN classification of compressed images

Video classification is one of the most important tasks in machine vision. Usually, videos need to undergo compression before being stored or transmitted. These data packets require decompression to restore explicit semantic information before downstream video classification<sup>56,57</sup>. Here we employ compressed 2D images for video classification. This methodology significantly reduces the additional expenditure of time and energy that is typically associated with the process of decompression<sup>58</sup>.

To perform convolution right at where the data are compressed, the sensor's channel current ( $I_{ds}$ ) should be linear with respect to the applied bias voltage ( $V_{ds}$ ), the latter represents kernels. We examined the relation between channel current of the sensor and  $V_{ds}$  (from  $\pm 1$  mV to  $\pm 100$  mV in step of 1 mV) at non-volatile low-resistance state. As revealed in Supplementary Fig. 28, the calculated conductance with negligible variation perfectly meets our requirements.

The 2D optoelectronic sensor array, as depicted within the dotted box in Supplementary Fig. 27b, performs in-sensor convolution. The array initially encodes and integrates the video data using the array conductance  $G_{ml}$  ( $G_{ml}$ ,  $m = 1, 2, 3 \dots n$ ;  $l = 1, 2, 3 \dots n$ ), which serves as the input feature map (Supplementary Fig. 27a). Then we map different  $3 \times 3$  convolution kernels into bias voltage matrices ( $V_{11}, V_{12}, \dots, V_{33}$ ). Each voltage matrix is split into columns. At each time step, a voltage column vector (e.g.,  $V_{11}, V_{12}, V_{13}$ ) bias the 2D optoelectronic sensor array for vector-vector inner product. The output current vector is  $I_{11}, I_{12}, I_{13}$ . This is repeated with the rest two voltage column vectors, producing  $I_{21}, I_{22}, I_{23}$ ; and  $I_{31}, I_{32}, I_{33}$ , respectively. The currents are then summed over time. According to Ohm's law and Kirchhoff's law<sup>20,59</sup>, the summed currents, reading voltage (kernel), and conductance physically implements vector-

matrix multiplication<sup>49</sup>, or a stride of the convolution:

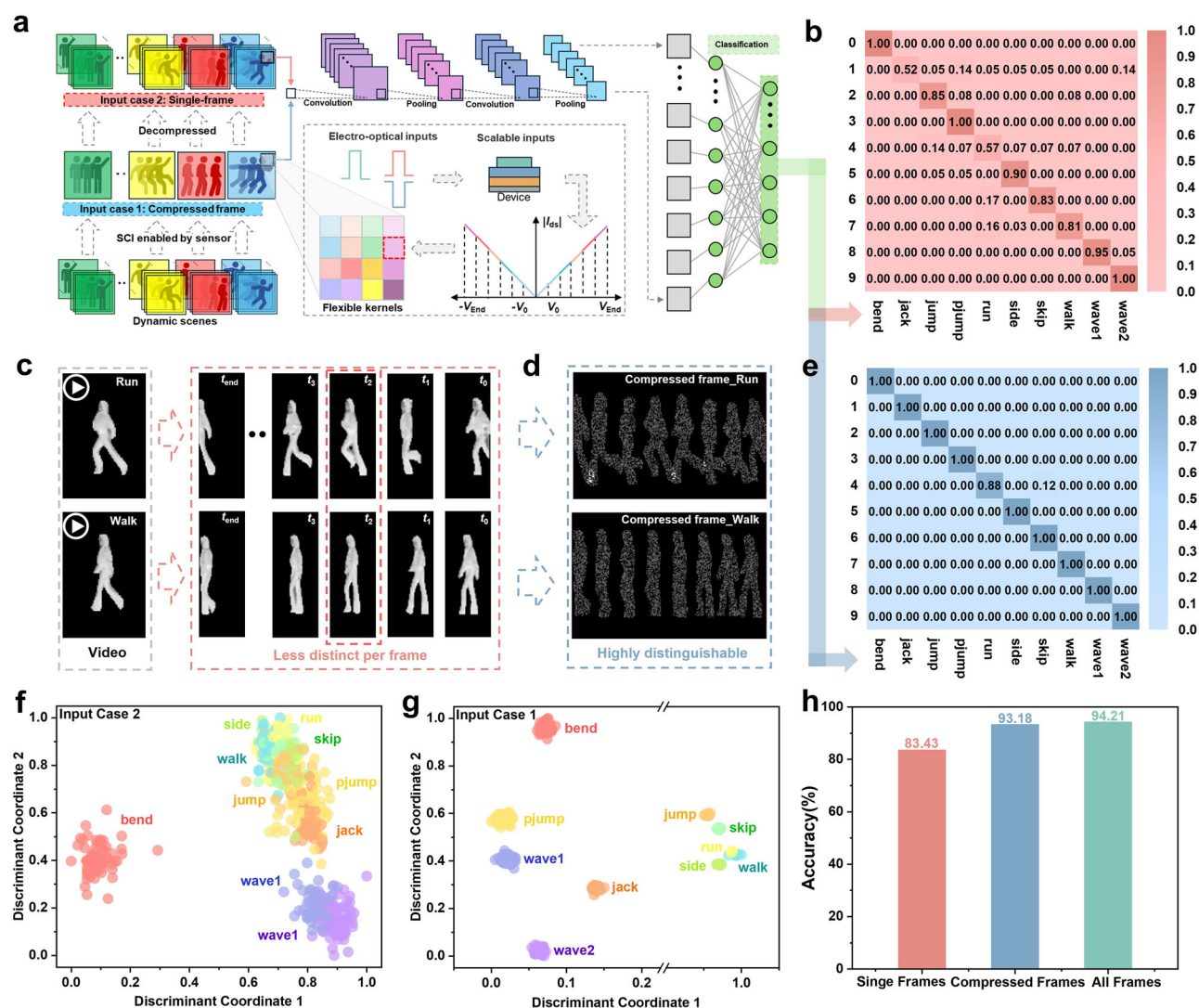
$$I_{kl} = \sum_{m=1}^3 V_{km} G_{ml} \quad (9)$$

Figure 6a illustrates the simulated classification of SCI compression recorded on optoelectronic sensor array using in-device convolution. Here, we selected the Weizmann Human Action Dataset, which are videos about ten different human motion postures (i.e., bend, jack, jump, etc.) for the classification experiment. First, all the videos were subject to background removal. Then the videos are compressed using in-sensor video SCI, resulting in a single compressed frame containing the semantic information of the entire video (input case 1). For comparison, single-frame images selected from each video (input case 2) are classified. Figure 6c, d show compression of two human action movies, i.e., "Run" and "Walk", and the resultant compressed images (for the dynamic movies, please refer to Supplementary Movie 1 and 2), respectively. It can be observed that the compressed images include information from all video frames (Fig. 6d), instead of single frame information in the control experiment (Fig. 6c). This feature is further evidenced in the 2D distribution maps (Fig. 6f, g) and confusion matrices (Fig. 6b, e), which clarify the classification results for both input cases 1 and 2.

Figure 6f, g depict the distribution features of different motions using the two input cases, employing dimensionality reduction through linear discriminant analysis (LDA). In these maps, balls of the same color represent the same motion category, and the quantity of balls of each color indicates the sample size. The spatial separation between different categories highlights their distinguishability, with greater distance suggesting easier differentiation. A comparison of Fig. 6f, g reveals that compressed images (Fig. 6g) encapsulate richer feature information, as evidenced by tighter clustering within the same motion category and clearer separation between different categories, thereby enhancing their distinguishability. Additionally, the confusion matrices in Fig. 6b, e correspond to Fig. 6f, g, respectively, illustrating the classification accuracy across different categories. For example, Fig. 6b shows that the accuracy for correctly classifying 'wave1' is 95%, with a 5% misclassification rate where 'wave1' is confused with 'wave2'. This overlap is visually represented in Fig. 6f, where the blue balls signifying 'wave1' significantly overlap with the purple balls for 'wave2'. In contrast, Fig. 6e demonstrates a 100% classification accuracy for the as-mentioned categories, with Fig. 6g showing no overlap between the two colors of balls.

It is important to note that these two representation methods primarily visualize the classification results for a single motion and do not provide a comprehensive evaluation of classification accuracy. Therefore, we also present the overall classification accuracy based on the two input cases to evaluate the efficacy of the strategies more thoroughly. For a more objective evaluation, we employed a strategy that uses all frames as input for classification. As shown in Fig. 6h, the overall classification accuracies for single-frame, compressed-frame, and all-frame input strategies are 83.43%, 93.18%, and 94.21%, respectively. The results demonstrate that compressed-frame inputs significantly improve classification accuracy (93.18% vs. 83.43%) over single-frame approaches, effectively mitigating information loss inherent to sparse sampling. Remarkably, the compressed-frame strategy attains accuracy comparable to all-frame classification while reducing convolutional parameters by 69% (see "Method" for details). Thus, our compression paradigm eliminates computationally intensive reconstruction steps, enabling direct feature extraction from compressed data—a critical advantage for real-time processing on resource-constrained edge devices.

We have also summarized the recent advances in machine vision based on dynamic behavior of devices fabricated from 2D heterostructures (Supplementary Table 4). Different from existing imaging



**Fig. 6 | In-sensor computing based on 2D PPMS. a** The schematic demonstration of a CNN diagram includes two convolutional layers and one fully connected layer. Two input cases are employed for image classification, namely Case 1: compressed frame and Case 2: single frame. **(b, e)** Confusion matrices of the classification results using the single-frame **(b)** and the compressed-frame **(e)** strategies with in-sensor convolution. The display of **(c)** dynamic frames and **(d)** compressed frames, each contains two human motions (“Run” and “Walk”) that are difficult in

recognition for machine vision. **f, g** The 2D distribution maps of motion features under input Case 1 **(g)** and Case 2 **(f)** derived from linear discriminant analysis (LDA), highlighting the separability of distinct motion categories in reduced-dimensional space. **h** The classification accuracy of the three input strategies. The pre-compression classification strategy achieves superior accuracy (93.18%) compared to single-frame input (83.43%) and attains performance parity with all-frames classification (94.21%).

approaches that primarily focus on post-processing or feature extraction, often compromising image quality and real-time performance, our work explores the potential of 2D heterostructures in achieving in-sensor compression. We have developed a compact compressed imaging prototype device that not only addresses the current research gap in high-quality image acquisition but also provides a versatile platform for future intelligent vision systems.

## Discussion

In summary, we developed an optoelectronic sensor with programmable storage capacity and functionalities. The sensor can work in either optical perception or memory modes, attributed to the photosensitive properties of *t*-Se. Leveraging this property, we designed a specialized optoelectronic AND gate, which enables simultaneous perception, encoding, compression, and storage of dynamic information. The compressed 2D images can be further reconstructed into 3D data using the PnP- algorithm. Our results demonstrate that this sensor provides a more compact and efficient

2D hardware encoder for video and spectra SCI technology. Furthermore, these compressed images keep all relevant semantic information of the video and enable direct in-sensor classification using CNN without decompression. This leads to reduced energy consumption, shorter processing time, and lower hardware costs. Overall, our work unveils the potential of 2D vdW heterostructure for compact and energy-efficient hardware for future intelligent visual systems.

## Methods

### The sensor and array fabrication

*t*-Se with the thickness of 30–50 nm was prepared via a modified PVD method (see Supplementary Note 1 for details), and the obtained products were transferred using a polydimethylsiloxane (PDMS) stamp onto a heavily doped Si covered by 300-nm SiO<sub>2</sub> (Silicon Valley Microelectronics, Inc.). Multilayer h-BN and MoS<sub>2</sub> nanosheets were mechanically exfoliated from their bulk crystals (2D Semiconductor Inc.) first, then transferred onto Se nanoflake step-by-step via a dry

transfer method (Supplementary Fig. 2). The source/drain electrodes were defined using the standard electron beam lithography (EBL) (TESCAN, VEGA3), followed by the electron beam evaporation (EBE) and thermal evaporation of contact metals (Cr/Au, 8/50 nm). The  $3 \times 3$  sensor array was also fabricated via the dry-transfer method: Firstly, three Se nanosheets with lateral size of  $\sim 10 \mu\text{m} \times 50 \mu\text{m}$  were transferred and aligned parallelly. The nanosheets were then patterned into an array of  $3 \times 3$  pixels with each pixel size of  $8 \mu\text{m} \times 8 \mu\text{m}$  (Supplementary Fig. 19) via reactive ion etching (RIE). Then a large h-BN nanosheet ( $> 50 \mu\text{m} \times 50 \mu\text{m}$ ) was exfoliated onto PDMS and transferred onto the  $3 \times 3$  pixels. Finally, a large exfoliated  $\text{MoS}_2$  nanosheet ( $> 50 \mu\text{m} \times 50 \mu\text{m}$ ) was transferred onto the Se/h-BN to finalize the Se/h-BN/ $\text{MoS}_2$  heterostructure arrays followed by the RIE patterning. The electrodes were also defined via EBL followed by deposition of Cr/Au (8/50 nm) as electrodes.

### Material and device characterizations

The material and device morphologies were captured by an optical microscope (Nikon, EBLIPSE LV100ND). The high-resolution transmission electron microscope (HRTEM) and TEM images of crystalline Se were obtained from Tecnai F20 TEM (200 kV), while the selected area electron diffraction (SAED) pattern was captured by Tecnai Spirit TEM (120 kV). Raman spectra of Se and Se/h-BN/ $\text{MoS}_2$  vdW heterostructure were obtained using Renishaw Raman Microscope (wavelength of 532 nm). Atomic force microscope (AFM) (Bruker, Dimension Icon with Scan Asyst) was employed to characterize the thickness of each layer in vdW heterostructure. Electrical/optical measurements of the device were conducted on a probe station equipped with the semiconductor analyzer Agilent 4155C and the source meter Keysight B2902B. The optical pulses were generated via the home-made platform where different optical sources (monochrome lasers of 450, 532, and 635 nm) combined with a mechanical chopper to generate the pulse trains.

### In sensor CNN demonstration

Utilizing light pulse input as a representation of image signals and electric pulse input to signify the mask, we compress binarized multi-frame video signals. Subsequently, a CNN is employed to directly classify the compressed 2D images. The pixel values within the 2D images are substituted by the output current values from the sensor depicted in Fig. 4d. For single frames/compressed frames as inputs, the image classification network architecture comprises two convolutional layers (with output channel counts of 8 and 16, respectively) and a single fully connected layer (encompassing 10 output nodes). For all frames as inputs, the CNN consists of four convolutional layers (with output channels of 8, 8, 16, and 16, respectively) and one fully connected layer to guarantee the capacity-task matching. Following the convolution output, a sigmoid activation function is applied ( $\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$ ). The cross-entropy loss serves as the loss function and is optimized via the Adam optimizer, with an initial learning rate set at 0.001. The network undergoes training for 100 epochs. The dataset, consisting of 990 samples, is partitioned into five subsets, with four subsets allocated for training and the remaining subset reserved for testing. The outcome is derived from the average of five experiments.

### Data availability

The Source Data underlying the figures of this study are available with the paper. All raw data generated during the current study are available from the corresponding authors upon request. Source data are provided with this paper.

### Code availability

The source codes used for simulation and data plotting are available from the corresponding authors upon request.

## References

1. Yuan, X., Brady, D. J. & Katsaggelos, A. K. Snapshot compressive imaging: theory, algorithms, and applications. *IEEE Signal Process. Mag.* **38**, 65–88 (2021).
2. Liu, B., Chen, Y., Liu, S. & Kim, H.-S. Deep learning in latent space for video prediction and compression. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 701–710 (IEEE, Nashville, TN, USA, 2021).
3. Chakraborty, I., Roy, D., Garg, I., Ankit, A. & Roy, K. Constructing energy-efficient mixed-precision neural networks through principal component analysis for edge intelligence. *Nat. Mach. Intell.* **2**, 43–55 (2020).
4. Xu, Y., Lu, L., Saragadam, V. & Kelly, K. F. A compressive hyper-spectral video imaging system using a single-pixel detector. *Nat. Commun.* **15**, 1456 (2024).
5. Park, J., Feng, X., Liang, R. & Gao, L. Snapshot multidimensional photography through active optical mapping. *Nat. Commun.* **11**, 5602 (2020).
6. Wang, Z., Wan, T., Ma, S. & Chai, Y. Multidimensional vision sensors for information processing. *Nat. Nanotechnol.* **19**, 919–930 (2024).
7. Wang, X. et al. Integrated photonic encoder for low power and high-speed image processing. *Nat. Commun.* **15**, 4510 (2024).
8. Candes, E. J. & Wakin, M. B. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
9. Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
10. Gehm, M. E., John, R., Brady, D. J., Willett, R. M. & Schulz, T. J. Single-shot compressive spectral imaging with a dual-disperser architecture. *Opt. Express* **15**, 14013–14027 (2007).
11. Gao, L., Liang, J., Li, C. & Wang, L. V. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature* **516**, 74–77 (2014).
12. Wang, L., Cao, M., Zhong, Y. & Yuan, X. Spatial-temporal transformer for video snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 9072–9089 (2023).
13. Hu, X. et al. HDNet: high-resolution dual-domain learning for spectral compressive imaging. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17521–17530 (IEEE, New Orleans, LA, USA, 2022).
14. Gallego, G. et al. Event-based vision: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 154–180 (2022).
15. Rebecq, H., Ranftl, R., Koltun, V. & Scaramuzza, D. Events-to-video: bringing modern computer vision to event cameras. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3852–3861 (IEEE, Long Beach, CA, USA, 2019).
16. Scheerlinck, C. et al. Fast image reconstruction with an event camera. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 156–163 (IEEE, Snowmass, CO, USA, 2020).
17. Takashima, Y. & Hellman, B. Review paper: imaging lidar by digital micromirror device. *Opt. Rev.* **27**, 400–408 (2020).
18. Ke, X. et al. Liquid-crystal spatial light modulators and their applications. In *Handbook of Optical Wireless Communication*, 1015–1063 (Springer Nature Singapore Pte Ltd., 2024).
19. Martel, J. N. P. et al. Neural sensors: learning pixel exposures for HDR imaging and video compressive sensing with programmable sensors. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 1642–1653 (2020).
20. Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
21. Zhou, Y. et al. Computational event-driven vision sensors for in-sensor spiking neural networks. *Nat. Electron.* **6**, 870–878 (2023).
22. Zhang, Z. et al. All-in-one two-dimensional retinomorphic hardware device for motion detection and recognition. *Nat. Nanotechnol.* **17**, 27–32 (2022).
23. Jiang, Y. et al. A scalable integration process for ultrafast two-dimensional flash memory. *Nat. Electron.* **7**, 868–875 (2024).



24. Dodda, A. et al. Active pixel sensor matrix based on monolayer MoS<sub>2</sub> phototransistor array. *Nat. Mater.* **21**, 1379–1387 (2022).
25. Liu, S. et al. Nanopatterning technologies of 2D materials for integrated electronic and optoelectronic devices. *Adv. Mater.* **34**, 2200734 (2022).
26. Polyushkin, D. K. et al. Analogue two-dimensional semiconductor electronics. *Nat. Electron.* **3**, 486–491 (2020).
27. Wu, G. et al. Ferroelectric-defined reconfigurable homojunctions for in-memory sensing and computing. *Nat. Mater.* **22**, 1499–1506 (2023).
28. Sun, X. et al. Reconfigurable logic-in-memory architectures based on a two-dimensional van der Waals heterostructure device. *Nat. Electron.* **5**, 752–760 (2022).
29. Kumar, D. et al. Artificial visual perception neural system using a solution-processable MoS<sub>2</sub>-based in-memory light sensor. *Light.: Sci. Appl.* **12**, 109 (2023).
30. Marega, G. M. et al. A large-scale integrated vector-matrix multiplication processor based on monolayer molybdenum disulfide memories. *Nat. Electron.* **6**, 991–998 (2023).
31. Yang, Y. et al. In-sensor dynamic computing for intelligent machine vision. *Nat. Electron.* **7**, 225–233 (2024).
32. Zhu, Y. et al. Non-volatile 2D MoS<sub>2</sub>/black phosphorus heterojunction photodiodes in the near- to mid-infrared region. *Nat. Commun.* **15**, 6015 (2024).
33. Zhang, X. et al. Inorganic halide perovskite nanowires/conjugated polymer heterojunction-based optoelectronic synaptic transistors for dynamic machine vision. *Nano Lett.* **24**, 4132–4140 (2024).
34. Wang, F. et al. A two-dimensional mid-infrared optoelectronic retina enabling simultaneous perception and encoding. *Nat. Commun.* **14**, 1938 (2023).
35. Huang, P.-Y. et al. Neuro-inspired optical sensor array for high-accuracy static image recognition and dynamic trace extraction. *Nat. Commun.* **14**, 6736 (2023).
36. Wang, Q. et al. Fabrication and growth mechanism of selenium and tellurium nanobelts through a vacuum vapor deposition route. *J. Phys. Chem. C* **111**, 12926–12932 (2007).
37. Qin, J. et al. Controlled growth of a large-size 2D selenium monosheet and its electronic and optoelectronic applications. *ACS Nano* **11**, 10222–10229 (2017).
38. Lee, S., Peng, R., Wu, C. & Li, M. Programmable black phosphorus image sensor for broadband optoelectronic edge computing. *Nat. Commun.* **13**, 1485 (2022).
39. Wu, L. et al. Atomically sharp interface enabled ultrahigh-speed non-volatile memory devices. *Nat. Nanotechnol.* **16**, 1–6 (2021).
40. Vu, Q. A. et al. Two-terminal floating-gate memory with van der Waals heterostructures for ultrahigh on/off ratio. *Nat. Commun.* **7**, 12725 (2016).
41. Luo, L.-B. et al. Transparent and flexible selenium nanobelt-based visible light photodetector. *CrystEngComm* **14**, 1942–1947 (2012).
42. Vu, Q. A. et al. A high-on/off-ratio floating-gate memristor array on a flexible substrate via CVD-grown large-area 2D layer stacking. *Adv. Mater.* **29**, 1703363 (2017).
43. Britnell, L. et al. Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett.* **12**, 1707–1710 (2012).
44. Wang, Y. et al. Negative photoconductance in van der Waals heterostructure-based floating gate phototransistor. *ACS Nano* **12**, 9513–9520 (2018).
45. Liu, M. et al. Photogating-assisted tunneling boosts the responsivity and speed of heterogeneous WSe<sub>2</sub>/Ta<sub>2</sub>NiSe<sub>5</sub> photodetectors. *Nat. Commun.* **15**, 141 (2024).
46. Yin, Z. et al. Single-layer MoS<sub>2</sub> phototransistors. *ACS Nano* **6**, 74–80 (2012).
47. Prakash, A. et al. Understanding contact gating in Schottky barrier transistors from 2D channels. *Sci. Rep.* **7**, 12596 (2017).
48. Lai, H. et al. Photoinduced multi-bit nonvolatile memory based on a van der Waals heterostructure with a 2D-perovskite floating gate. *Adv. Mater.* **34**, 2110278 (2022).
49. Janesick, J., Elliott, T., Collins, S., Blouke, M. & Freeman, J. Scientific charge-coupled devices. *Opt. Eng.* **26**, 268692 (1987).
50. Yuan, X., Liu, Y., Suo, J., Durand, F. & Dai, Q. Plug-and-Play algorithms for large-scale snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 7093–7111 (2022).
51. Chen, W. et al. Snapshot compressive imaging based digital image correlation: temporally super-resolved full-resolution deformation measurement. *Opt. Express* **30**, 33554–33573 (2022).
52. Zhang, B., Suo, J. & Dai, Q. Event-enhanced snapshot compressive videography at 10K FPS. *IEEE Trans. Pattern Anal. Mach. Intell.* **47**, 1266–1278 (2025).
53. Razavi, B. et al. *Fundamentals of microelectronics*. (John Wiley & Sons, 2021).
54. Qiu, H., Wang, Y. & Meng, D. Effective snapshot compressive-spectral imaging via deep denoising and total variation priors. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9123–9132 (IEEE, Nashville, TN, USA, 2021).
55. Arce, G. R. et al. Compressive coded aperture spectral imaging: an introduction. *IEEE Signal Process. Mag.* **31**, 105–115 (2014).
56. Pranav, S. V., Suraj, P., Pothineni, N., Kadambari, K. V. & Balasubramanyam, P. Optimizing video compression & decompression using deep learning networks. *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, 1–6 (IEEE, Kuala Lumpur, Malaysia, 2021).
57. Tzagkarakis, G., Tsagkatakis, G., Starck, J.-L. & Tsakalides, P. Compressive video classification in a low-dimensional manifold with learned distance metric. *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 155–159 (IEEE, Bucharest, Romania, 2012).
58. Goswami, D., Chakraborty, S. Active learning for video classification with frame level queries. *2023 International Joint Conference on Neural Networks (IJCNN)*, 1–9 (IEEE, Gold Coast, Australia, 2023).
59. Zhou, G. et al. Full hardware implementation of neuromorphic visual system based on multimodal optoelectronic resistive memory arrays for versatile image processing. *Nat. Commun.* **14**, 8489 (2023).
60. Li, H. et al. From bulk to monolayer MoS<sub>2</sub>: evolution of Raman scattering. *Adv. Funct. Mater.* **22**, 1385–1390 (2012).
61. Zhang, Z.-C. et al. An ultrafast nonvolatile memory with low operation voltage for high-speed and low-power applications. *Adv. Funct. Mater.* **31**, 2102571 (2021).

## Acknowledgements

This research is supported by the National Key R&D Program of China (Grant No. 2023YFB2806300). C.T. thanks the funding support from the National Natural Science Foundation of China – Excellent Young Scientists Fund (Hong Kong and Macau) (No. 52122002), the Start-Up Grant (Project No. 9610710) from City University of Hong Kong, ECS scheme (21201821), General Research Fund (11200122) and the Collaborative Research Fund (RGC; no. C2001-23Y and C5001-24Y) from the Research Grant Council of Hong Kong and ITC via Hong Kong Branch of National Precious Metals Material Engineering Research Center (NPMR). Z.W. thanked the support from the National Natural Science Foundation of China (Grant Nos. 62122004, 62374181), Beijing Natural Science Foundation (Grant No. Z210006), Hong Kong Research Grant Council (Grant Nos. 27206321, 17205922, 17212923). This research is also partially supported by ACCESS – AI Chip Center for Emerging Smart Systems, sponsored by Innovation and Technology Fund (ITF), Hong Kong SAR. Y.C. Thanks the funding support from the General Research Fund (15301621) from the Research Grant Council of Hong Kong. P.Y. thanks the funding support from the National Natural Science Foundation of China (Grant No. 62404138) and Shenzhen Science and Technology Program (Grant No. 20231128102926002).

## Author contributions

Z.W. and C.T. proposed the research direction. H.H., C.T. and Z.W. conceived the idea for the project. Y.Chai, Z.W. and C.T. supervised the project. H.H. performed materials synthesis, fabrication and characterization of devices. S.S. conducted simulation of devices. J.Z. contributed to optoelectronic performance measurements and data analysis. H.P.C. provided guidance on theoretical analysis. Y.X., H.W., and P.Y. assisted in materials synthesis. J.H., W.W., Y.R. and Y.W. supported experimental testing. L.Z. and Y.Chen carried out transmission electron microscopy characterization. S.X. performed atomic force microscopy characterization. H.H., S.S., J.Z., Y.Chai, Z.W., and C.T. co-wrote the manuscript. All authors participated in manuscript discussion and revision.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-59104-7>.

**Correspondence** and requests for materials should be addressed to Yang Chai, Zhongrui Wang or Chaoliang Tan.

**Peer review information** *Nature Communications* thanks Vishwanath Saragadam, Yang Wang, and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

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

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025